



Universidade de Brasília (UnB)

Faculdade de Ciência da Informação (FCI)

Graduação em Biblioteconomia

TALLES BRENDÓ CAIXETA RAMOS

**DESENVOLVIMENTO DE FERRAMENTA PARA EXTRAÇÃO DE DADOS DE  
CITAÇÃO E ALTMÉTRICOS VIA APPLICATION PROGRAMMING INTERFACE**

BRASÍLIA

2022

TALLES BRENDON CAIXETA RAMOS

**DESENVOLVIMENTO DE FERRAMENTA PARA EXTRAÇÃO DE DADOS DE  
CITAÇÃO E ALTMÉTRICOS VIA APPLICATION PROGRAMMING INTERFACE**

Monografia apresentada à Faculdade de  
Ciência da Informação da Universidade de  
Brasília, como requisito parcial para  
obtenção do grau de Bacharel em  
Biblioteconomia.

Orientador: Prof. Dr. João de Melo  
Maricato

BRASÍLIA

2022

**FOLHA DE APROVAÇÃO**

**Título:** DESENVOLVIMENTO DE FERRAMENTA PARA EXTRAÇÃO DE DADOS DE CITAÇÃO E ALTMÉTRICOS VIA APPLICATION PROGRAMMING INTERFACE

**Autor(a):** Talles Brendo Caixeta Ramos

Monografia apresentada em **30 de setembro de 2022** à Faculdade de Ciência da Informação da Universidade de Brasília, como parte dos requisitos para obtenção do grau de Bacharel em Biblioteconomia.

Orientador(a) (FCI/UnB): Dr. João de Melo Maricato

Membro Interno (FCI/UnB): Dr. Márcio de Carvalho Victorino

Membro Externo (UEA): Ma. Danielle Pompeu Noronha Pontes

Em 07/05/2021.



Documento assinado eletronicamente por **Danielle Pompeu Noronha Pontes, Usuário Externo**, em 14/10/2022, às 12:20, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Marcio de Carvalho Victorino, Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 17/10/2022, às 06:24, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.



Documento assinado eletronicamente por **Joao de Melo Maricato, Vice-Diretor(a) da Faculdade de Ciência da Informação**, em 19/10/2022, às 17:02, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.

---



Documento assinado eletronicamente por **Talles Brendo Caixeta Ramos, Usuário Externo**, em 24/10/2022, às 09:04, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.

---



A autenticidade deste documento pode ser conferida no site [http://sei.unb.br/sei/controlador\\_externo.php?acao=documento\\_conferir&id\\_orgao\\_acesso\\_externo=0](http://sei.unb.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **8824946** e o código CRC **153A927F**.

---

Para o meu eu do passado, que decidiu se aventurar no mundo maravilhoso da biblioteconomia.

## **AGRADECIMENTOS**

Primeiramente gostaria de agradecer a Universidade de Brasília por ter me proporcionado diversas oportunidades de desenvolvimento intelectual, por ter aberto diversas portas, me permitido conhecer pessoas fascinantes e me proporcionado os melhores anos da minha vida.

Ao meu orientador, Prof. Dr. João de Melo Maricato, pelos conselhos, direcionamentos e oportunidades de aprendizagem concedidas a mim.

Agradeço a Elen Sena Serejo por ter ficado sempre ao meu lado e ter me ajudado nessa aventura que se chama vida. Sem sua ajuda e companhia jamais teria chegado até aqui.

Agradeço a minha família por todo o apoio que me foi fornecido para que eu pudesse ter a chance de estudar e mudar o meu futuro.

Por último e não menos importante, gostaria de agradecer aos membros da banca, pela imensa colaboração.

*“Se cheguei até aqui foi porque me apoiei no ombro dos gigantes.”*

**Isaac Newton**

## RESUMO

Este estudo teve como objetivo o desenvolvimento de uma ferramenta de extração de dados via *Application Programming Interface* (API *Rest*) batizada de *Odisseia metrics* com a finalidade de solucionar dois grandes problemas de uma única vez: a dificuldade que alguns pesquisadores possuem em construir códigos de programação para extração de dados em algumas bases de dados e, promover estudos bibliométricos e altmétricos a partir de grande volume de dados, de maneira que os dados extraídos possam contribuir com análises e mapeamentos mais abrangentes da produção científica brasileira. A ferramenta *Odisseia metrics* foi desenvolvida em linguagem de programação Python, hospedada no Google *Colab* e está na versão 2.0. A sua construção se deu a partir da necessidade de se coletar dados altmétricos e bibliométricos a partir de uma lista de DOIs em um arquivo “.xlsx”. Em um primeiro momento a ferramenta foi construída e o código testado até que se chegasse à total confiabilidade. Posteriormente a ferramenta foi aberta e o código disponibilizado por meio de uma interface gráfica de fácil operação e gratuita. A ferramenta desenvolvida focou-se na extração de dados altmétricos da plataforma Altmetric, de dados de leitores/capturas do Mendeley e dados de citações da base Crossref. A ferramenta foi testada localmente e, posteriormente, por outros pesquisadores e estudantes realizaram coletas de dados. A ferramenta apresenta boa usabilidade e consistência dos dados. Não há necessidade de conhecimento de tecnologias de informação e comunicação avançadas por parte do usuário, possibilitando a coleta de dados para diferentes públicos e tipos de pesquisa. A pesquisa atingiu o seu objetivo, ao disponibilizar uma ferramenta consistente que proporciona, a qualquer interessado, a coleta de grande volume de dados altmétricos e de citação para a realização de estudos a partir de fontes de dados alternativas às comerciais.

**Palavras-Chave:** Mendeley. Crossref. Altmetric. Bibliometria. Altmetria. Application Programming Interface.

## ABSTRACT

This study had the development of an objective research tool (API) called Odisseia Resto de Metrics with an objective of solving two big ones at once: the difficulty that someone in research has in programming codes for data collection and data, promotional in some bibliometric data from the large volume of data, in a selected way it can contribute to more comprehensive data studies and mappings of Brazilian scientific production. The Odisseia metrics tool was developed in Python programming language, hosted on Google Colab and is in version 2.0. Its construction based on the need to select altmetric data from a DOI ".xls" file. At first, the tool was built and the code tested until it arrived with total reliability. Subsequently, the tool was opened and the code made available through an easy-tooperate and free graphical interface. The developed tool focused on the Altmetric altmetric data platform, from Men reader/captures data from the Crossref database. Tool collection was applied locally and later by other researchers and data students. The tool has good usability and data consistency. There is no need for knowledge of advanced information and communication technologies on the part of the user, enabling data collection for different audiences and types of users. From the research, it achieves its objective, by providing a consistent tool that provides, to any interested party, a large volume of altmetric and citation data collection to carry out studies of commercial alternative data sources.

**Keywords:** Mendeley. Crossref. Altmetric. Bibliometrics. Altmetry. Application Programming Interface

## LISTA DE SIGLAS

AA	Acesso Aberto
API	<i>Application Programming Interface</i>
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
DOI	<i>Digital Object Identifier</i>
FI	Fator de Impacto
IDF	<i>International DOI Foundation</i>
ISAN	<i>International Standard Audiovisual Number</i>
ISBN	<i>International Standard Book Number</i>
ISNI	<i>International Standard Name Identifier</i>
ISO	<i>International Standards Organization</i>
ISSN	<i>International Standard Serial Number</i>
JCR	<i>Journal Citations Reports</i>
OA	<i>Open Access</i>
OAI	<i>Open Archives Initiatives</i>
RA	Agência de registro
SCI	<i>Science Citation Index</i>
SciELO	<i>Scientific Electronic Library Online</i>
UnB	Universidade de Brasília
URL	<i>Uniform Resource Locator</i>
WoS	<i>Web of Science</i>

## LISTA DE FIGURAS

Figura 1 - Estrutura de um DOI .....	38
Figura 2 - Concatenação servidor proxy e código DOI resultando em um hiperlink..	40
Figura 3 - Exemplo de DOI com ISSN.....	40
Figura 4 - Exemplo de um código no Notebook Colab .....	48
Figura 5 - Inserindo HTML em um Notebook Colab .....	49
Figura 6 - HTML renderizado em um Notebook Colab .....	49
Figura 7 - Notebook com HTML e Script Python .....	50
Figura 8 - Adicionando botão de upload de arquivos no Notebook Colab .....	55
Figura 9 - Script desenvolvido para escanear tabela.....	56
Figura 10 - Script para coletar os dados da API Altmetric .....	57
Figura 11 - Script para baixar a nova tabela automaticamente.....	57
Figura 12 - Script completo desenvolvido para extração da API Altmetric.....	59
Figura 13 - Script de autenticação da API Mendeley.....	62
Figura 14 - Script extração de dados da API Mendeley.....	63
Figura 15 - Inserindo o sinal hífen caso o DOI falhe.....	63
Figura 16 - Script para baixar nova tabela com resultados extraídos da API Mendeley .....	64
Figura 17 - Código que atualiza o Token do usuário automaticamente .....	64
Figura 18 - Script Completo para extração da API Mendeley .....	65
Figura 19 - Código para extração da API Crossref .....	69
Figura 20 - Inserindo botão Select para escolha da base de Dados desejada .....	70
Figura 21 - Interface gráfica no Google Colab para extração dos dados da Altmetric, Mendeley e Crossreff .....	71
Figura 22 - Selecionando a base Altmetric, Mendeley ou Crossref na ferramenta Odisseia.....	73
Figura 23 - Botão que deve ser executado para a extração de dados da Altmetric, Mendeley ou Crossref .....	73
Figura 24 - Inserindo Ficheiro no formato. xlxs. Para extração de dados da Altmetric, Mendeley, Crossref .....	73
Figura 25 - Exemplo de como a tabela deve ser formatada para a extração dos dados da Altmetric, Mendeley, Crossref.....	74
Figura 26 - Url direcionando para autenticação de conta Mendeley .....	75
Figura 27 - Tela de Login Mendeley que será apresentada ao Usuário .....	75
Figura 28 - Token Mendeley ao final da Url.....	76
Figura 29 - Dados sendo extraídos do Mendeley .....	76
Figura 30 - Exemplo de dados resultantes da extração do Mendeley .....	77
Figura 31 - Exemplo de dados resultantes da extração da Crossref .....	78
Figura 32 - Exemplo de dados resultantes da extração da Altmetric .....	78
Figura 33 - Exemplo de dados resultantes da extração da Altmetric.....	79

## LISTA DE QUADROS

Quadro 1 - Algumas características de um DOI .....	39
Quadro 2 - Dados fornecidos pela API Altmetric .....	53
Quadro 3 - Response Code API Altmetric .....	54
Quadro 4 - Dados extraídos da API Mendeley .....	60
Quadro 5 - Response Code API Mendeley .....	61
Quadro 6 - Response Code API Crossref .....	68

## SUMÁRIO

<b>1. INTRODUÇÃO</b>	13
1.1. JUSTIFICATIVA	16
1.2. OBJETIVOS	19
1.2.1. Objetivo geral	19
1.2.2. Objetivos específicos	19
<b>2. REVISÃO DE LITERATURA</b>	20
2.1. MÉTRICAS DA INFORMAÇÃO	20
2.2. CITAÇÕES E ÍNDICES DE CITAÇÕES	23
2.1.1. Citações e sua utilização para avaliação da ciência	23
2.1.2. Críticas aos indicadores de citações	26
2.3. FONTES DE DADOS: MENDELEY, ALTMETRIC E CROSSREF	27
2.3.1. Mendeley	27
2.3.2. O potencial do Mendeley para estudos alométricos	28
2.3.3. Leitores advindos do Mendeley	29
2.3.4. Cobertura de assuntos do Mendeley	30
2.3.5. Correlação entre citações e indicadores alométricos	31
2.3.6. A plataforma Altmetrics	34
2.3.7. Crossref	36
2.3.8. Identificador Digital de Objetos (DOI)	36
2.4. API'S E RASPAGEM DE DADOS E LINGUAGEM DE PROGRAMAÇÃO	42
<b>3. PROCEDIMENTOS METODOLÓGICOS</b>	45
3.1. PRINCÍPIOS GERAIS DA FERRAMENTA	45
3.2. A ESCOLHA DA LINGUAGEM DE PROGRAMAÇÃO	46
3.3. HOSPEDAGEM E AMBIENTE DE EXECUÇÃO DA FERRAMENTA	47
3.4. FONTES DE DADOS/BASES DE DADOS SELECIONADAS	50
3.5. PASSO PARA O DESENVOLVENDO A FERRAMENTA ODISSEIA	51
<b>4. RESULTADOS: DESENVOLVENDO A FERRAMENTA ODISSEIA</b>	52
4.1. LEVANTAMENTOS NECESSÁRIOS PARA EXTRAIR DADOS DA API – ALTMETRIC	52
4.2. DESENVOLVIMENTO DO CÓDIGO DE EXTRAÇÃO DE DADOS CONFORME API DA BASE ALTMETRIC	54
4.3. LEVANTAMENTOS REALIZADOS PARA EXTRAIR DADOS DA API – MENDELEY	60
4.4. DESENVOLVIMENTO DO CÓDIGO DE EXTRAÇÃO DE DADOS CONFORME API DA BASE MENDELEY	61

4.5.	LEVANTAMENTOS REALIZADOS PARA EXTRAIR DADOS DA API - CROSSREF 67	
4.6.	DESENVOLVIMENTO DO CÓDIGO DE EXTRAÇÃO DE DADOS CONFORME API DA BASE CROSSREF .....	69
4.7.	DISPONIBILIZAÇÃO DOS CÓDIGOS DA ALTMETRIC, MENDELEY E CROSSREF EM UMA MESMA INTERFACE GRÁFICA .....	70
4.8.	Estrutura da Ferramenta Odisseia .....	71
<b>5.</b>	<b>UTILIZAÇÃO E TESTE DA FERRAMENTA .....</b>	<b>72</b>
5.1.	EXTRAINDO DADOS DA ALTMETRIC, MENDELEY E CROSSREF VIA ODISSEIA ....	72
5.1.1.	Autenticação, obtenção e uso do token no Mendeley .....	74
5.1.2.	Extraindo dados da Crossref via odisseia .....	77
5.1.3.	Extraindo dados da Altmetric via odisseia.....	78
5.2.	TESTE DA FERRAMENTA .....	79
<b>6.</b>	<b>CONSIDERAÇÕES FINAIS .....</b>	<b>82</b>
	<b>REFERÊNCIAS .....</b>	<b>85</b>

## 1. INTRODUÇÃO

O ser humano segundo Grogan (1995, p. 51) sempre teve a necessidade e o desejo de conhecer, aprender, obter respostas para suas necessidades informacionais e, com a popularização e a acessibilidade à internet no final do século XX e o avanço contínuo dos sistemas de informações e comunicações, tornou-se possível encontrar respostas para os questionamentos individuais quase que instantaneamente.

O progresso de uma nação está diretamente relacionado ao seu nível de desenvolvimento em ciência e da tecnologia, havendo o potencial desta para auxiliar a solucionar diversos obstáculos sociais essenciais para a vida como a fome, a desigualdade a saúde, etc. Para a tomada de decisões certas é necessário que se tenha acesso a recursos informacionais. Contudo a necessidade de informações precisas e de qualidade não se limitam aos tomadores de decisão de organizações, mas de todas as esferas da sociedade. Em geral, observa-se que estes não conseguem filtrar os documentos e informações úteis em meio a grande quantidade existente.

Uma forma de se obter informações é por meio das publicações científicas. Estas possuem um papel fundamental para o desenvolvimento de novos conhecimentos, técnicas e tecnologias. Os trabalhos científicos preconizam o desenvolvimento intelectual da sociedade, impactando nos mais diversos fatores econômicos e tecnológico e sociais. Monitorar a ciência, seu desenvolvimento e indicadores, é inquestionavelmente uma atividade de extrema relevância na atualidade.

Na era digital, encontrar artigos é mais fácil e rápido comparado ao século passado, contudo, a maior quantidade resulta em uma maior dificuldade de selecionar artigos relevantes. Isto porque existe o crescimento exponencial da produção científica, acarretando na dificuldade em escolher um artigo de boa qualidade dentre as milhares de possibilidades.

Uma forma de identificar artigos relevantes, historicamente, tem sido por meio da análise das citações que determinado documento possui. Entretanto as citações demoram um longo tempo para ocorrer, pois o processo que vai entre a leitura,

produção de novos conhecimentos, avaliação pelos pares e publicação, pode variar entre meses e anos.

Indicadores complementares podem amenizar esse tipo de problema. Estes contam com pontuações advindas de dados altmétricos, como quantidade de downloads, leitores, visualizações, menções em redes sociais, etc. Assim, esses indicadores, apesar de medirem outro tipo de impacto, são capazes de indicar com maior velocidade documentos relevantes, quando comparado aos indicadores de citações. Os impactos altmétricos costumam ocorrer mais rapidamente, quando comparado aos dados de citação, possibilitando a avaliação da ciência de maneira complementar e mais abrangente.

Apesar das diferenças entre os indicadores de citação e altmétricos, a relação entre eles, suas proximidades e diferenças podem ser analisadas, possibilitando a compreensão da sua utilização para a avaliação de pesquisadores e pesquisas e para o processo seleção de artigos relevantes. Diversos estudos são elaborados a fim de averiguar a correlação entre citações e dados altmétricos com o intuito de entender as proximidades, diferenças e previsibilidade de citações a partir de indicadores altmétricos (ROEMER; BORCHARDT, 2015; ZAHEDI; COSTAS; WOUTERS, 2014; THELWALL *et al.*, 2013).

Estes estudos indicam que os indicadores altmétricos devem ser consideradas complementares aos indicadores bibliométricos. Eles contam com potencialidades e limitações, vantagens e desvantagens para a avaliação e monitoramento de uma diversidade de contextos científicos, além de contribuir no processo de seleção de artigos relevantes. Contudo, a obtenção dos dados de citações e altmétricos normalmente exige a utilização de bases de dados de acesso restrito (grandes bases de dados comerciais).

Para a obtenção dos dados de citações e altmétricos, frequentemente são necessárias assinaturas em bases de dados pagas. Algumas das principais fontes de dados de citação utilizados mundialmente são oriundas das bases de dados *Web of Science* e *Scopus*. Tratam-se de duas bases de dados multidisciplinares de acesso pago, com baixa cobertura da ciência desenvolvida por países periféricos.

Com relação a indicadores altmétricos, a base de dados comumente utilizada para a realização de estudos é a plataforma *Altmetric*, da empresa *Digital Science*

que, apesar de ser de assinatura paga, tem disponibilizado o acesso aos dados gratuitamente para o desenvolvimento de pesquisas científicas. Contudo, há organizações que disponibilizam os dados de forma aberta e gratuita. Esses dados crescentemente vêm sendo disponibilizados via API, destacando-se, nesse contexto, as citações abertas, disponibilizadas pela *Crossref* e no de dados altmétricos fornecidos pelo Mendeley.

É reconhecido que artigos científicos brasileiros não são indexados adequadamente nas bases de dados internacionais. Este é obstáculo importante para a disseminação e reconhecimento mundial das publicações brasileiras. Todavia, existe a possibilidade de se medir o impacto da produção científica brasileira por meio de organizações que disponibilizam citações abertas, na qual os artigos publicados por países periféricos possuem uma indexação mais adequada. Além desses obstáculos referentes a medição da ciência brasileira, existe também dificuldades para se coletar dados para estudos altmétricos por diversos pesquisadores.

Estudos altmétricos e bibliométricos utilizando grande volume de dados, que poderiam proporcionar melhor compreensão dos impactos da publicação científica brasileira, deixam de ser realizados devido aos obstáculos que parte dos pesquisadores enfrentam. Entre eles estão: dificuldade em aprender programação, ausência de tempo para se dedicar ao entendimento da programação e tecnologias associadas, limitações relacionadas a computadores com requisitos mínimos de Hardware para extração de uma grande massa de dados, entre outros.

Esses obstáculos materiais, humanos e informacionais dificultam o desenvolvimento de estudos por diversos pesquisadores, pois dados de grande volume normalmente são disponibilizados via *API Rest*, sendo necessários conhecimentos específicos de programação por parte dos pesquisadores para se ter acesso a esses dados.

Este estudo busca amenizar parte desses obstáculos ao propor e disponibilizar uma ferramenta capaz de extrair de dados bibliométricos de citação e altmétricos de algumas bases de dados alternativas às comerciais. Portanto o objetivo desta pesquisa desenvolver uma ferramenta online de fácil acesso, uso e extração de grande volume de dados, para usuários com conhecimento básico de tecnologias da informação, sem a necessidade de conhecimento em programação. Para isso, foi

elaborado metodologias de extração de dados *API Rest* aplicados à plataforma Altmetric, ao Mendeley e às citações da base *Crosreff*. As extrações propostas incluem a implementação em um servidor em nuvem, possibilitando que mesmo usuários com computadores menos potentes possam extrair uma grande quantidade de dados. A pesquisa contempla tanto a apresentação da metodologia e técnicas utilizadas para a construção da ferramenta, quanto passo a passo de como utilizá-la.

### 1.1. JUSTIFICATIVA

Um dos principais fatores que influenciaram o desenvolvimento desse trabalho foi perceber que diversos estudos bibliométricos e altmétricos podem estar deixando de ser executados pelo fato de grande parte dos pesquisadores estar enfrentando alguns obstáculos para a extração de dados, por diversos motivos, como, por exemplo: dificuldade em extrair dados por meio de API; não possuírem conhecimentos de programação; não possuírem computadores com requisitos de hardware mínimos para manipular grandes massas de dados.

O principal fator que justifica o desenvolvimento da ferramenta é a relevância de se analisar o impacto da produção científica brasileira em aspectos ainda pouco explorados, como no caso da análise mediante dados altmétricos e de citação de fontes alternativas às bases comerciais. Considera que as grandes bases internacionais e suas métricas tradicionais de Fator de Impacto (FI), que tem por base a contagem de citações, não conseguem avaliar adequadamente a produção de países periféricos, como é o caso do Brasil.

Os periódicos brasileiros encontram diversos obstáculos para serem incluídos em grandes bases de dados comerciais do exterior. Segundo Mugnaini e Strehl (2008, p. 92), “as publicações dos países em desenvolvimento sempre tiveram dificuldades para atender [...] critérios, principalmente pela falta de apelo internacional e comercial da produção científica dedicada às questões regionais destas nações”. Outro obstáculo apresentado por Mugnaini e Strehl (2008, p. 93) se refere a diferença de contextos na qual cada país se encontra “alcançam os periódicos internacionais aqueles artigos que tratam dos problemas de interesse do “Primeiro Mundo”. Como exemplo, Coura e Willcox (2003) abordam que artigos de Medicina Tropical muitas vezes não são aceitos em revistas estrangeiras pelo fato de não serem consideradas “ciência universal” e sim um “problema local” ou “interesse local”.

Além desses problemas particulares em países em desenvolvimento, há questões mais abrangentes. Conforme apresentado por Maricato e Martins (2017, p. 53) algumas justificativas para criação da área da altmetria se deve “[...] à crise vivenciada pelos filtros “tradicionais” utilizados para determinar a qualidade da informação científica; lentidão no sistema de revisão pelos pares; limitação das análises de citação e do FI; entre outros”. Thelwall e Wilson (2015, p. 1) seguem o mesmo pensamento “[...] as contagens de citações têm uma série de limitações, incluindo o fato de que levam anos para se acumular e, portanto, pode ser muito lento para algumas avaliações”.

Seguindo o mesmo argumento, Thelwall (2018, p. 122) esclarece que “uma desvantagem importante da contagem de citações é que pode levar vários anos para que um artigo seja citado o suficiente para apontar sua probabilidade de impacto a longo prazo”. As citações demoram um longo tempo para ocorrer, pois o processo que vai entre a leitura, produção de novos conhecimentos, avaliação pelos pares e publicação pode variar entre meses e anos.

É perceptível que além da produção científica brasileira encontrar diversos obstáculos para serem reconhecidos e avaliados por periódicos internacionais, há também problemas gerais com os métodos tradicionais. Assim, torna-se necessário investigar se as métricas alternativas podem ser utilizadas como substituto e/ou complementar ao método bibliométrico tradicional no contexto brasileiro. Hipoteticamente, esses indicadores altmétricos podem amenizar alguns obstáculos provocados pelos tradicionais. Piwowar e Priem (2013, p. 11) abordam que os dados altmétricos como contagens de downloads, marcadores (*bookmarks*) e *Tweets* se correlacionam com citações, contudo os dados altmétricos acumulam dados muito mais rapidamente, muitas vezes em dias, semanas ou meses (ao invés de anos).

Assim, esses indicadores, apesar de medirem outro tipo de impacto, são capazes de indicar com maior velocidade documentos relevantes quando comparado aos indicadores de citações. Thelwall e Wilson (2015, p. 1) “uma vez que a altmetria não precisa ser atrasada pelo ciclo de publicação acadêmica, eles podem ser disponíveis para artigos que são muito recentes para atrair muitas citações”, sendo assim a altmetria pode ser uma boa alternativa para se perceber o impacto inicial de publicações recentes. É importante salientar que uma métrica não exclui a outra, pois podem ser consideradas complementares, contando com potencialidades e

limitações, vantagens e desvantagens para a avaliação de diversos contextos científicos.

A escolha de se analisar a produção brasileira com a altmetria (assim como as citações na base *Crossref*) tem como base os pontos levantados Mugnaini e Strehl (2008) e Coura e Willcox (2003). Os autores demonstram que as publicações brasileiras tendem a não alcançar periódicos internacionais e que não possuem apelo internacional. Também considerou os pontos levantados por Maricato e Martins (2017), Thelwall e Wilson (2015), Thelwall (2018), que abordam algumas limitações do formato tradicional de avaliação mediante contagem de citações, especialmente por bases de dados comerciais. Levando em conta esses pontos, percebe-se que a produção científica brasileira já possui um grande obstáculo para ser reconhecida internacionalmente e ainda sofre com as (limitações) dos filtros “tradicionais” fazendo, assim, que a ciência brasileira não seja tão reconhecida e adequadamente mensurada.

Um estudo publicado pela Clarivate Analytics (2019) indicou que o Brasil indexou entre 2011 e 2016 mais de 250 mil artigos na base de dados *Web of Science*, ocupando a 13 posição na produção científica mundial. Embora o Brasil ocupe uma posição de destaque na WoS, ainda há uma sub-representação da produção científica nacional nas bases de dados internacionais. Há uma produção científica significativa de escopo regional ou local não indexada nessas bases, além disso o acesso restrito ao conteúdo dessas bases transforma os estudos sobre a produção e citações científicas um sério obstáculo. Todavia, uma alternativa para os estudos de citação emergiu mais recentemente, quando a *Crossref* passou a disponibilizar abertamente os dados de citação dos artigos que possuem um *Digital Object Identifier* (DOI).

Levando em conta os trabalhos semelhantes já desenvolvidos por outros autores, e o contexto na qual a ciência brasileira se encontra representada no âmbito das bases de dados comerciais internacionais, essa pesquisa tem como objetivo desenvolver uma ferramenta capaz de extrair dados altmétricos que irão auxiliar pesquisadores a medir o impacto de publicações brasileiras mais adequadamente, por meio de fontes de dados bibliométricos e altmétricos, alternativas às comumente utilizadas em larga escala para este tipo de mensuração.

## 1.2. OBJETIVOS

Essa monografia possui objetivo geral e objetivos específicos na qual serão descritos a seguir.

### 1.2.1. Objetivo geral

- Propor uma ferramenta online de extração de dados de citação e altmétricos de bases de dados alternativas utilizando *Application Programming Interface*.

### 1.2.2. Objetivos específicos

Com o propósito de detalhar o objetivo geral do trabalho, são objetivos específicos:

- Apresentar as bases de dados selecionadas para extração de dados de citação e altmétricos;
- Desenvolver um código em Python para extração de dados via *API REST* das bases de dados selecionadas;
- Testar a eficiência da ferramenta desenvolvida para a coleta de dados;
- Implementar a ferramenta em um servidor em nuvem.

## 2. REVISÃO DE LITERATURA

A revisão de literatura abordará conceitos e temas importantes para o entendimento deste trabalho. Primeiro serão abordados alguns tipos de métricas da informação, que são utilizados para o desenvolvimento de diversos tipos de estudos métricos, desde de Bibliometria até Altmétrie.

Após o entendimento dos tipos de estudos métricos, será abordada duas das principais bases de dados Altmétricas, que serão utiliza-las para elaboração deste trabalho: Mendeley, *Altmetric*, bem como a base de dados *Crossref*, que fornece dados de citação.

Serão apresentados conceitos importantes para o entendimento de índices de citações, análises de citações e fator de impacto. O entendimento desses conceitos é de grande importância para a compreensão da necessidade de que trabalhos científicos brasileiros (e de outros países periféricos) utilizem novos indicadores e fontes de dados alternativas.

Também serão apresentados assuntos introdutórios relacionados a programação, para que seja possível um entendimento básico do funcionamento da ferramenta de extração de dados desenvolvida e implementada.

### 2.1. MÉTRICAS DA INFORMAÇÃO

Desde o início do século XX existe o interesse em se desenvolver meios de avaliar, monitorar e medir o desempenho da produção científica. Surgem assim as principais técnicas e métodos para análise e avaliação da ciência, denominados Estudos Métricos da Informação, que agrupam, dentre outras, as áreas de bibliometria, cientometria, informetria, patentometria, webometria e mais recentemente a altmetria. Segundo Lima (1986) *apud* (MATTOS, 2019, p. 21), “[...] os estudos métricos têm como finalidade a aplicação de recursos da ciência na análise da própria ciência, bem como técnica, com aplicações no campo da documentação científica”. Já Oliveira e Grácio (2011, p. 19) destacam que “os estudos métricos compreendem o conjunto de estudos relacionados à avaliação da informação produzida, mais especialmente científica, em diferentes suportes, baseados em recursos quantitativos como ferramentas de análise”.

Araújo (2006, p. 22) destaca cada uma das métricas “[...] são subdisciplinas que se assemelham por serem métodos quantitativos, mas se diferenciam quanto ao objeto de estudo, as variáveis, os métodos específicos e os objetivos”. Os estudos métricos da ciência focam tanto nos insumos quanto nos produtos gerados pela própria ciência, mais precisamente da produção acadêmica publicada pela comunidade científica (NORONHA; MARICATO, 2008; LIMA, 1986).

A Bibliometria (do inglês, *bibliometrics*), termo criado em 1934 por Paul Otlet, “(consiste) na aplicação de técnicas estatísticas e matemáticas para descrever aspectos da literatura e de outros meios de comunicação [...]” (ARAÚJO, p. 12). Nesse contexto, pode-se afirmar que a bibliometria busca por meio de métodos quantitativos desenvolver avaliações e análises de produções científicas na qual podem ser utilizadas em tomadas de decisões, além de ser uma ferramenta de grande valia para monitorar o desempenho da ciência e tecnologia como um todo.

O campo da bibliometria é fortemente relacionado à três leis: i) Lei de Lotka (1926): aplicada à produtividade de autores; ii) Lei de Bradford (1934): à dispersão do conhecimento científico; iii) Lei de Zipf (1949): à frequência de palavras. Segundo Araújo (2006) a bibliometria tem como matéria prima de estudo de artigos, livros, documentos, revistas, autores e usuários.

A Cientometria (do inglês, *Scientometrics*) se utiliza das técnicas bibliométricas aplicadas a ciência. Tem por objetivo medir o desenvolvimento da ciência, tendo como objeto de estudo os campos, disciplinas e assuntos da ciência. “Cientometria é o estudo dos aspectos quantitativos da ciência enquanto uma disciplina ou atividade econômica [...] é um segmento da sociologia da ciência, sendo aplicada no desenvolvimento de políticas científicas”. (MACIAS-CHAPULA, 1998, p. 134).

A Informetria segundo Macias-Chapula (1998, p. 135) tem por objeto de estudo as palavras, documentos e bases de dados. “[...] é o estudo dos aspectos quantitativos da informação em qualquer formato, e não apenas registros catalográficos ou bibliografias, referente a qualquer grupo social”. A informetria traz como diferencial a cobertura de documentos e dados digitais e a diversificação do grupo social de estudo, não se limitando apenas aos cientistas.

Esse subcampo inclui temas como a análise estatística do texto (científico) e os sistemas de hipertexto, biblioteca, circulações, medidas de informação em

bibliotecas eletrônicas, bem como processos e aspectos quantitativos de recuperação de informação, com principal ocupação de descrever e analisar os fenômenos e as leis implicadas na própria informação (OLIVEIRA, 2018, p. 36).

A patentometria tem como objeto de estudo as patentes que, por meio de seu monitoramento e análise, é possível identificar as atividades dos países que estão inovando e gerando novas tecnologias a partir das informações contidas nos documentos de patente. Para Macias-Chapula (1998, p. 137), “o número de patentes reflete as tendências das mudanças técnicas ao longo do tempo [e] determinam o grau aproximado da inovação tecnológica de um país e que o número de citações de patentes mede o impacto da tecnologia”.

A Webometria (do inglês, *Webometrics*) consiste na aplicação de métodos infométricos no ambiente Web. Tem como objeto de estudo páginas da internet, sítios na *World Wide Web* e motores de busca. Thelwall, Vaughan e Björneborn (2005) definem webometria como “estudo quantitativo de todos os fenômenos relacionados com a web”. Mattos (2019) e Gouveia (2013) complementam, reconhecendo que a webometria estuda a rede mundial de computadores *World Wide Web* e que o estudo webométrico acontece “[...] a partir de abordagens infométricas e bibliométricas, que tem como foco os “*links*” hipertextuais, o número de páginas on-line ou cadastradas em mecanismos de busca, a busca por termos específicos e o padrão de uso da *web*.”

A Altmetria (do inglês, *Altmetric*), termo atribuído por Jason Priem em 2010, é uma nova métrica que está em seu período inicial. Surgida no contexto da *Web Social* ou *Web 2.0*, tem ganhado cada vez mais espaço no desenvolvimento de métricas a fim de medir o impacto de produções científicas, se tornando um potencial complemento para estudos que utilizem formas tradicionais de medição de impacto como a bibliometria e cientometria.

Gouveia (2013) define altmetria com “[...] o uso de dados webométricos e cibernômicos em estudos cientométricos”. Para Shema, Bar-Ilan e Thelwall (2014) a Altmetria é uma métrica alternativa que mede o impacto de materiais acadêmicos no âmbito das mídias e redes sociais. Uma das vantagens da métrica alternativa (MARICATO; MARTINS, 2017) é que “diferentemente de outros indicadores comumente utilizados para a avaliação da produção científica, a altmetria capta informações de outros públicos e não somente cientistas e pesquisadores”. A

compreensão da variedade de públicos que utilizam os materiais acadêmicos pode auxiliar na compreensão de como são utilizados os materiais acadêmicos e quanto diversificado pode ser a sua utilização pela sociedade. Sendo assim, a altmetria tem a possibilidade de medir de maneira mais ampla as atividades científicas, indo além dos meios tradicionais.

A altmetria surgiu com a finalidade de ser uma métrica alternativa às métricas tradicionais, que vêm sendo criticadas por pesquisadores. Maricato e Martins (2017, p. 53) argumentam algumas justificativas para criação da área se deve à “[...] crise vivenciada pelos filtros “tradicionais” utilizados para determinar a qualidade da informação científica; lentidão no sistema de revisão pelos pares; limitação das análises de citação e do FI; entre outros.”.

São exemplos de tipos de interações em uma rede ou mídia social medidos pela altmetria: quantidade de vezes que um documento foi compartilhado, quantidades de downloads, visualizações, curtidas, favoritos, salvos, compartilhados, comentários, menções, etc (NASCIMENTO, 2016, p. 67).

## 2.2. CITAÇÕES E ÍNDICES DE CITAÇÕES

A citação é o principal indicador bibliométrico. Ao abordar a temática de citação pretende-se elucidar o motivo pela qual a ciência brasileira precisa utilizar fontes de dados e indicadores alternativos, ao invés de índices de citações internacionais/comerciais.

Como o presente trabalho tem o objetivo desenvolver uma ferramenta de extração de dados altmétricos e bibliométricos de fontes alternativas, possibilitando assim que a ciência brasileira não fique refém de métricas internacionais/comerciais e seus índices de citação, discute-se brevemente o conceito de citação e as fontes frequentemente utilizadas para a realização desse tipo de indicador bibliométrico.

### 2.1.1. Citações e sua utilização para avaliação da ciência

Para Macias-Chapula (1998, p.136), citação é o “[...] meio mais comum de atribuir créditos e reconhecimento na ciência” e pode ser definida como “[...] conjunto de uma ou mais referências bibliográficas que, incluídas em uma publicação,

evidenciam elos entre indivíduos, instituições e áreas de pesquisa, visto que mostram o relacionamento de uma publicação com outra”.

Uma citação, a princípio, tem a finalidade de referenciar alguma ideia, pensamento, opinião ou trabalho de terceiros, contudo há outras funções, Weinstock (1971) expôs quinze finalidades da citação:

- i. prestar homenagem aos pioneiros;
- ii. dar crédito para trabalhos relacionados;
- iii. identificar metodologia, equipamento etc.;
- iv. oferecer leitura básica;
- v. retificar o próprio trabalho;
- vi. retificar o trabalho de outros;
- vii. analisar trabalhos anteriores;
- viii. sustentar declarações;
- ix. informar os pesquisadores de trabalhos futuros;
- x. dar destaque a trabalhos pouco disseminados, inadequadamente indexados ou desconhecidos (não citados);
- xi. validar dados e categorias de constantes físicas e de fatos etc.;
- xii. identificar publicações originais nas quais uma ideia ou um conceito são discutidos;
- xiii. identificar publicações originais que descrevam conceitos ou termos epônimos, ou seja, descobertas que receberam o nome do pesquisador responsável, por exemplo, Doença de Chagas;
- xiv. contestar trabalhos ou ideias de outros;
- xv. debater a primazia das declarações de outros.

A análise de citação é uma ferramenta bibliométrica que se utiliza das citações de trabalhos como matéria prima para elaboração de métricas. Segundo Foresti (1989, p. 3) análise de citação significa “[...] a parte da bibliometria que investiga as relações entre os documentos citantes e os documentos citados, considerados como unidades de análise, no todo ou em suas diversas partes: autor, título, origem

geográfica, ano e idioma de publicação, etc”. A análise de citação pode ser utilizada como método de avaliações de instituições, autores, áreas de estudo, etc. Para Moravcsik e Murugesan (1975) *apud* Vanz e Caregnato (2003 p. 251) “[...] as citações podem medir realizações científicas individuais, de um grupo, de uma instituição, de um país e, até mesmo, podem retratar a evolução da ciência em geral ou de um campo específico”.

A análise de citações está intimamente relacionada com o desenvolvimento de índice de citações. Com o crescente desenvolvimento tecnológico e a popularização dos estudos bibliométricos aplicados a citações, é criado em 1963 por Eugene Garfield, o *Science Citation Index* (SCI), o primeiro índice de citações. Este índice “[...] abriu caminhos para todos que buscassem medir a ciência usando métodos quantitativos e objetivos.” (MACIAS-CHAPULA, 1998, p. 137). O SCI possui índices de citações com uma cobertura temática ampla (artes e humanidades, ciências sociais, e ciências), que permite medir a frequência e o impacto dos títulos de periódicos. A base de dados *Web of Science* (WoS), é a responsável por disponibilizar o SCI através do *Journal Citation Reports* (JCR). O SCI é utilizado mundialmente pela comunidade científica para recuperação de informações e realização de análises bibliométricas e cientométricas.

Por meio do *Journal Citation Reports* (JCR), é possível visualizar avaliações de periódicos indexados na *Web of Science* e por meio do JCR é possível consultar diversos índices, como o fator de impacto da publicação (*Journal Impact Factor* - JIF), o Eigenfactor, o número total de citações da publicação com e sem autocitações, o índice de imediatismo que mede o quão rápido um artigo é citado em cada publicação, dentre outros (CAPES, 2020).

O Fator de Impacto (FI), da WoS, é o mais conhecido método de avaliação bibliométrico de periódicos científicos. O FI conta com metodologia que possui a citação como elemento central. O FI tem por objetivo qualificar revistas científicas, e a sua formula é constituída “[...] por meio de um índice calculado a partir do número médio de citações dos artigos publicados [em determinado periódico] durante os dois últimos anos.” (MARICATO; MARTINS 2017, p. 54).

### 2.1.2. Críticas aos indicadores de citações

A contagem de citações e o fator de impacto (FI) são métricas tradicionalmente utilizadas para elaboração de avaliações de produção científica até os dias atuais. Contudo, essas métricas têm recebido inúmeras críticas, dentre elas a de não conseguir avaliar adequadamente a produção científica de países periféricos, como o Brasil. Em uma pesquisa desenvolvida por Melo, Trinca e Maricato (2021) com a finalidade de identificar o nível de cobertura da *Web of Science* em relação à produção bibliográfica brasileira foi utilizado a produção bibliográfica de 4377 programas da pós graduação dentre os anos 2017 a 2018 presentes na Plataforma Sucupira, na qual se teve como resultado que somente cerca de 46,56% dessas produções estão indexadas na base de dados *Web of Science Core Collection*. Percebe-se pelo estudo dos autores que a avaliação da produção científica brasileira tendo por meio a base *Web of Science*, em alguns casos pode não refletir exatamente a realidade e que por isso é necessário maneiras alternativas de medir o impacto da produção brasileira.

Segundo os periódicos brasileiros encontram diversos obstáculos para serem considerados veículos científicos no exterior. Segundo Mugnaini e Strehl (2008, p. 92), “as publicações dos países em desenvolvimento sempre tiveram dificuldades para atender a estes critérios, principalmente pela falta de apelo internacional e comercial da produção científica dedicada às questões regionais destas nações”.

Outro obstáculo apresentado por Mugnaini e Strehl (2008, p. 93) se refere a diferença de contextos na qual cada país se encontra “alcançam os periódicos internacionais aqueles artigos que tratam dos problemas de interesse do “Primeiro Mundo”. Como exemplo, Coura e Willcox (2003) abordam que artigos de Medicina Tropical muitas vezes não são aceitos em revistas estrangeiras pelo fato de não serem consideradas “ciência universal” e sim um “problema local” ou “interesse local”.

Diante dos problemas de indexação da ciência em bases de dados de citações internacionais, como é o caso da *Web of Science* e Scopus, algumas alternativas podem ser vislumbradas. Um exemplo é o movimento de citações abertas, onde a *Crossref* tem desempenhado um papel relevante. O uso de indicadores de citações dessa base pode proporcionar uma análise alternativa para a análise do impacto da ciência desenvolvida por países com baixa cobertura nas bases de dados internacionais, tal como o Brasil. Porém, para uma compreensão mais profunda do

assunto, é necessário que se entenda o que é o movimento de citações abertas e o papel da *Crossref*.

### 2.3. FONTES DE DADOS: MENDELEY, ALTMETRIC E CROSSREF

Diversas fontes e bases de dados são utilizadas no âmbito dos Estudos Métricos da Informação. Nessa etapa será apresentado as fontes e bases de dados que serão utilizadas para o desenvolvimento deste trabalho. São elas: Mendeley, *Altmetric* e *Crossref*. A base de dados Mendeley será utilizada para exportar dados relacionados a quantidade de leitores/capturas, um tipo proeminente de indicador altmétrico. A base *Altmetric* será utilizada para exportar dados altmetricos de outras redes sociais como o Twitter, Youtube, Facebook, entre outros. Por fim, a base *Crossref* será utilizada para exportar dados relacionados a quantidade de citações que os trabalhos obtiveram.

#### 2.3.1. Mendeley

Dentre as fontes de dados altmétricos, o Mendeley, ferramenta de gerenciamento de referências *online* gratuita, tornou-se um recurso de grande valia para a elaboração de estudos altmétricos ao fornecer metadados, como a quantidade de leitores/capturas, status acadêmico (bibliotecários, alunos de doutorado, professores, etc.), país do perfil e área de atuação. Esta plataforma indica a quantidade de leitores/capturas de um determinado documento científico, a partir do salvamento do item na biblioteca pessoal do usuário (ROBINSON-GARCIA *et al.*, 2017; SCHMITT; JÄSCHKE, 2017). Pelas características apresentadas e pela relevância do Mendeley nos estudos altmétricos, essa ferramenta foi selecionada para a extração de dados no âmbito deste estudo.

Para um pesquisador é de grande importância obter um bom referencial teórico no momento de embasar e desenvolver sua pesquisa. Encontrar trabalhos relevantes em sua área permite se ter uma maior noção do que seja realmente importante para se aprofundar no tema e desenvolver melhores pesquisas. Na tentativa de facilitar o trabalho de pesquisadores foi criado o software Mendeley. Trata-se de uma ferramenta de gerenciamento de referências online, um sistema que funciona como um buscador e organizador de documentos disponíveis na *Web*. O Mendeley também funciona como rede social acadêmica, o que permite o relacionamento entre usuários

e o compartilhamento de artigos. Existe um *feed* no qual é disponibilizado a atualização do usuário, sendo que seus contatos adicionados podem ter acesso às informações.

O Mendeley é uma plataforma que auxilia na organização de documentos acadêmicos, tornando possível a gerenciamento de referências, sendo semelhante a softwares já conhecidos com o mesmo propósito, como, por exemplo, Zotero, Citavi e End-Note. Possui, como um de seus objetivos, facilitar o compartilhamento de trabalhos de pesquisa online pelos seus usuários. Seu nome é derivado de nomes de dois grandes cientistas: Dmitri Ivanovich Mendeleev e Gregor Mendel. A plataforma surgiu em 2008 e foi comprada em 2013 pela empresa Elsevier. Possui versões para o navegador, desktop e mobile, sendo possível a sincronização dos arquivos entre as diferentes plataformas. Possui suporte para os principais sistemas operacionais: Linux, Windows, IOS, Android, Mac OSX. O Mendeley em 2013 possuía 2,5 milhões de usuários, porém não se sabe o número atual de usuários já que a plataforma não possui contagem de usuários em tempo real.

O software, além de atuar como um gerenciador de referências, é uma importante ferramenta altimétrica, proporcionando metadados de grande valia para elaboração de estudos científicos. Percebe-se que o Mendeley se trata de uma ferramenta de gerenciamento de referências *online* gratuita (apesar de ser disponibilizada por uma empresa privada), que proporciona metadados de grande valia para elaboração de estudos altmétricos, como a quantidade de leitores, status acadêmico, países e área de atuação. Diante disso, serão apresentadas algumas características da ferramenta e, brevemente, revisados alguns trabalhos que discutem e utilizam o Mendeley em diferentes perspectivas.

### 2.3.2. O potencial do Mendeley para estudos altmétricos

Existem diversos estudos sobre a ferramenta Mendeley que comprovam o seu potencial de uso para estudos altmetricos em comparação com outras ferramentas. Os estudos de Haustein *et al* (2014); Thelwall e Sud (2016); Thelwall (2018) apresentam que há correlações moderadas ou fortes entre citações e leitores Mendeley. A ferramenta também demonstra possuir uma ótima cobertura das publicações, em que foi constatada nos trabalhos de Araújo (2006) na qual o Mendeley apresentou uma cobertura de 92,43%, Haustein e Larivière (2014) com coberturas de

assuntos que varia entre 62,8% - 81,00%; Zahedi, Costas e Wouters (2017) com cobertura de 76,91% - 93,06%.

O Mendeley foi selecionado, na presente pesquisa, por ser uma ferramenta altmétrica no qual foram apresentados resultados interessantes. Alguns estudos mostram que outras ferramentas altmétricas como Tweets, FbWalls, RH, Blogs, Google+, MSM, Reddits, fóruns, Q&A, Pinner, LinkedIn não apresentam resultados positivos em termos de correlação com citação, possuindo correlações baixas ou negativas além de apresentarem uma baixa cobertura, como é apresentado nos estudos de Thelwall *et al* (2013); Haustein *et al* (2014, p.208); Thelwall e Sud (2016); Araújo (2006); Thelwall (2018).

O Mendeley possui características promissoras para estudos Altmétricos, já que as bases de dados de citações comumente utilizadas para avaliação de produção e citação parecem pouco adequadas para estudos do cenário brasileiro, visto que a produção é sub-representação em bases de dados internacionais/comerciais, como a *Web of Science* e Scopus.

Outro motivo na escolha da utilização do Mendeley se vale pela possibilidade de se obter diversos metadados de forma gratuita e ilimitada, metadados esses que serão utilizados em diversos estudos altmétricos, além do fato de que a ferramenta ser largamente utilizada, possuindo uma grande quantidade de usuários espalhados pelo globo.

### 2.3.3. Leitores advindos do Mendeley

A possibilidade de se obter dados referente aos leitores e documentos do Mendeley de forma gratuita é de grande importância, já que muitos periódicos não costumam fornecer tal informação sobre os artigos por fins econômicos ou pelo valor da informação, como é explicado por Haustein, Larivière (2014, p. 2) “[...] com exceção de alguns periódicos de acesso aberto, as estatísticas não são disponíveis, visto que os editores os consideram muito sensíveis para torná-los disponíveis gratuitamente.”

A ferramenta indica a quantidade de possíveis leitores de um determinado documento científico a partir de quando o usuário salva o item em sua biblioteca pessoal. Os leitores são identificados por seus status acadêmicos como Bibliotecários, alunos de doutorado, professores, etc. Segundo Zahedi, Costas, Wouters, (2017, p.

01). “Essas estatísticas são comumente conhecidas como "estatísticas de leitores", embora, na realidade, as métricas não reflitam necessariamente a "atividade de leitura" real por Usuários do Mendeley”.

No mesmo caminho, Mohammadi, Thelwall e Kousha (2015) abordam que isso se deve ao fato de que mesmo que o usuário tenha salvo um documento em sua biblioteca do Mendeley, não significa necessariamente que um leitor o leu ou mesmo que ocorrerá a leitura futuramente. Em contrapartida há usuários que possam ter lido o artigo, contudo não adicionou em sua biblioteca Mendeley (ou salvou) ou nem sequer utilizam o *software* de gerenciamento de referências. Com essas circunstâncias há diversos estudos e debates sobre o significado e impacto das estatísticas de leitores de softwares gerenciadores de referências.

A fim de identificar se as estatísticas de leitores do Mendeley refletem com fidedignidade os leitores dos artigos e por qual motivo as pessoas utilizam a ferramenta, Mohammadi, Thelwall e Kousha (2015) desenvolveram uma pesquisa online por meio de questionário com 860 usuários da ferramenta escolhidos aleatoriamente, com a finalidade de obter a resposta para esses questionamentos. Foi identificado que “[...] (0,4%) dos usuários não leram nenhum de seus registros favoritos e não pretendia lê-los [...], 82% dos usuários do Mendeley tinham lido ou pretendia ler pelo menos metade das publicações em suas bibliotecas pessoais” (MOHAMMADI; THELWALL; KOUSHA, 2015, p. 6). O estudo concluiu que as estatísticas de leitores disponibilizadas pelo Mendeley são úteis para identificar o público leitor e o impacto social, educacional e profissional que os artigos promovem. Haunschild e Bornmann (2016, p.63) chegaram à conclusão de que “[...] as contagens do Mendeley são vistas como uma possibilidade muito promissora para quantificar o tamanho do público leitor de um jornal dentro e fora do ambiente da ciência.”

#### 2.3.4. Cobertura de assuntos do Mendeley

Em um estudo que teve o objetivo de apresentar a cobertura de artigos de revistas de Ciência da Informação brasileiras no Mendeley, publicado por Araújo (2006), foram analisados 125 artigos de periódicos da área classificados pela Capes como Qualis ‘A’, sendo utilizada métricas alternativas. A ferramenta altmetric.com retornou dados de 36% dos artigos enquanto o Mendeley retornou 92,43%, Twitter retornou 7,57% e o Facebook não retornou dado algum. O Mendeley foi a que mais

retornou resultados ao contrário das redes sociais, que retornaram uma baixa quantidade de dados altmétricos, ficando evidenciado a potencialidade dessa ferramenta para estudos dessa natureza.

Outro estudo que apresenta dados sobre a cobertura do Mendeley foi publicado por Haustein e Larivière (2014). Neste, os autores utilizam dados de leitores do Mendeley com uma amostra de 1,2 milhões de artigos das disciplinas “Pesquisa Biomédica”, “Medicina Clínica”, “Saúde” e “Psicologia” dos periódicos PubMed e *Web of Science*. Os autores descobriram que o Mendeley foi capaz de cobrir 81,0% dos artigos de “Psicologia”, seguido de 72,4% da disciplina “Pesquisa Biomédica”, 67,0% da “Saúde” e por fim 62,8% “Medicina Clínica”, ou seja, atestaram que o Mendeley possui uma boa cobertura dos artigos das disciplinas citadas.

Zahedi, Costas e Wouters (2017) em um estudo para constatar se a cobertura do Mendeley sobre publicações WoS é efetiva, chegaram na conclusão que a ferramenta possui uma alta cobertura de publicações científicas “[...] com valores de cobertura superiores a 60% ou até 80% para publicações WoS dependendo do campo. “Os campos e suas respectivas porcentagem de cobertura Mendeley foram: Ciências biomédicas e da saúde (90,80%), Ciências da vida e da terra (93,06%), matemática e ciência da computação (76,91%), Ciências naturais e engenharia (83,70%), Ciências Sociais e Humanas (92,38 %).

#### 2.3.5. Correlação entre citações e indicadores altmétricos

O Mendeley é uma das poucas ferramentas altmétricas que apresentam resultados altos em termos de correlação entre citação e dados altmétricos. Alguns exemplos de pesquisas, do ponto de vista do estudo das correlações entre citações e indicadores altmétricos, podem ser destacados. Thelwall *et al* (2013) utilizou 208.739 artigos do PubMed e comparou as suas citações com as pontuações altmétricas de 11 mídias sociais (Tweets, FbWalls, RH, Blogs, Google+, MSM, Reddits, fóruns, Q&A, Pingers, LinkedIn). As mídias sociais apresentaram baixa cobertura das 208.739 publicações, menos de 20% das publicações foram recuperadas, sendo identificada apontada uma baixa correlação entre as citações dos artigos e as pontuações altmétricas. Por outro lado, o autor mostra que o gerenciador de referências Mendeley apresentou correlações positivas em um estudo com artigos do periódico F1000, tendo como correlação de 0,69.

No trabalho de Haustein *et al* (2014, p. 208), sobre a correlação entre citações de artigos PubMed e Tweets, foi apresentado que uma maior quantidade de tweets sobre um artigo está associado a uma maior quantidade de citações. Contudo, as correlações em sua maioria foram muito baixas ou negativas para alguns campos. “alguns dos artigos da PubMed mais tweetados eram de tópicos relacionado à comunicação acadêmica em geral (por exemplo, Fukushima), ou eram sobre tópicos curiosos (por exemplo, fratura peniana)”. Haustein *et al* (2014, p. 209) como resultado do seu estudo, traz uma comparação de correlações entre citações e leitores Mendeley, e citações e Tweets. “Enquanto os 390.190 jornais de 2011 com pelo menos um leitor de Mendeley correlacionou-se moderadamente com as citações ( $=0,456$ ), a correlação entre tweets e citações para 63.800 artigos foi muito menor ( $=0,157$ )”.

Thelwall e Sud (2016) desenvolveram um estudo em 2014 para observar o comportamento entre a correlação de citações (Scopus) e leitores do Mendeley pelo período de 2004-2014 com artigos de 50 subcampos de cinco categorias amplas (*agriculture, business, decision science, pharmacy, and the social sciences*). Os autores concluíram que existem correlações positivas entre os dados de citações e leitores, no estudo foi apresentado uma variação, dependendo do campo, entre 0,30 a 0,72. Também concluem que “[...] a contagem de leitores de Mendeley se correlaciona mais fortemente com a citações do que todas os outros indicadores altmétricos testados até agora” (THELWALL *et al.*, 2013 *apud* THELWALL; SUD, p. 3, 2016).

Em outro estudo, Fang *et al* (2020) apresentaram uma análise extensa da presença de dados altmétricos de 12,3 milhões de publicações da *Web of Science*. Os autores verificaram que a maioria dos dados altmétricos concentra-se em publicações nas áreas de Ciências Biomédicas e da Saúde, Ciências Sociais e Humanas e Ciências da Terra e da Vida.

Especificamente sobre a cobertura do Mendeley, Zahedi, Costas e Wouters (2017) verificaram que a ferramenta possui uma alta cobertura de publicações científicas da WoS (valores de cobertura superiores a 60% ou até 80% para publicações WoS dependendo do campo). Os campos e suas respectivas porcentagem de cobertura Mendeley foram: Ciências biomédicas e da saúde (90,80%), Ciências da vida e da terra (93,06%), matemática e ciência da computação

(76,91%), Ciências naturais e engenharia (83,70%), Ciências Sociais e Humanas (92,38 %).

Em um estudo desenvolvido por Thelwall e Wilson (2015) com base em uma amostra de 332.975 artigos de 2009 em 45 áreas médicas indexadas na Scopus, os autores buscaram descobrir se havia correlação (Spearman) entre os leitores do Mendeley (dados alométricos) e a contagem de citações Scopus desses artigos, e concluíram que houve uma correlação forte entre os dados (cerca de  $\rho = 0,7$ ). Das 45 áreas médicas, somente uma área “Guias de Medicamentos” obteve uma correlação baixa ( $< 0,5$ ), contudo os autores explicam que isso está relacionado ao seu pequeno tamanho, já que essa área foi representada somente por 83 publicações. Para Thelwall e Wilson (2015, p. 5) “[...] as correlações entre os leitores de Mendeley e as citações são significativamente positivas e fortes para quase todas as áreas, exceto a pequena categoria de guias de drogas [...]”. Dos 332.927 artigos, 78% tiveram pelo menos um leitor em Mendeley, tornando assim a cobertura do Mendeley bastante positiva, já 73% tiveram pelo menos uma citação na Scopus.

Zahedi, Costas e Wouters (2017, p. 1), elaboraram uma análise a fim de averiguar se a quantidade de leitores do Mendeley é capaz de identificar publicações altamente citadas com mais eficácia do que as pontuações de citações, ou indicadores de impacto baseados em periódicos (JCS). A pesquisa teve como base 9.152.360 publicações de cinco grandes campos científicos classificados por Leiden Ranking de 2013 (*Biomedical and health sciences, Life and Earth sciences, Mathematics and computer Science, Natural sciences and engineering, social sciences and humanities*) com *Digital Object Identifiers* (DOI) dos anos 2004-2013.

Os dados de leitores foram extraídos via *API REST* do Mendeley. Os autores concluíram que as pontuações de leitores são mais eficazes do que as pontuações de citações apresentadas nos periódicos (JCS), para identificar publicações altamente citadas em todos os campos das publicações e em todos os anos. Cerca de 86,5% (7.917.494) de todas as publicações possui pelo menos um leitor no Mendeley, mostrando que o Mendeley conseguiu cobrir de forma significativa grande parte das publicações. Os autores consideraram que as pontuações de leitores do Mendeley foram eficazes para identificar publicações altamente citadas. Zahedi, Costas e Wouters (2017, p. 13) declaram que é “[...] possível argumentar que os leitores

Mendeley e as citações são dois processos diferentes, mas conectados que podem capturar um tipo de impacto semelhante.”

Com o intuito de analisar se o número de leitores do Mendeley apresenta uma correlação (Spearman) positiva com citações posteriores, Thelwall (2018) coletou a quantidade de leitores que as publicações obtiveram dentro de um mês de publicação e correlacionou com a quantidade de citações que as mesmas publicações receberam após 20 meses. A quantidade de leitores foi obtida via API do Mendeley tendo como campo identificador o DOI. Foi obtido como resultado que oito de dez áreas retiradas da Scopus, no ano de 2016, possuíam correlações moderadas ou fortes, indicando, assim, que os leitores do Mendeley podem, em alguns casos, representar um impacto futuro advindo das citações.

#### 2.3.6. A plataforma Altmetrics

Uma quantidade expressiva de discussões e debates sobre o conteúdo acadêmico ocorre ininterruptamente na Internet. Essas informações são rastreadas pela plataforma Altmetric. A ferramenta coleta e agrupa essas informações, auxiliando o usuário a rastrear e acompanhar os trabalhos de seu interesse e, também, proporciona que seja monitora o alcance que o trabalho está obtendo por meio de métricas que a plataforma disponibiliza.

A Altmetric foi lançada em 2012, sendo atualmente uma plataforma que tem como principal produto os dados altmétricos de publicações científicas divulgadas no ambiente digital. Com a Altmetrics é possível saber quantas menções, citações e compartilhamentos uma publicação obteve, além disso a plataforma apresenta as regiões geográficas dos usuários que salvaram ou compartilharam os trabalhos científicos em diversas mídias e redes sociais.

Alguns tipos de dados que o Altmetrics oferece são: quantidade de citações feitas por meio de Bloggers, Twitter, páginas do Facebook, referências do Wikipedia, Google +, Reddited. Também pode ser consultado a quantidade de Leitores do Mendeley, Cannotes e CiteULike. Além dos dados a altmétricos a plataforma também proporciona uma gama de ferramentas e serviços, porém, vale ressaltar que grande parte das ferramentas e serviços da plataforma são pagos.

A plataforma Altmetric apresenta uma pontuação chamada de *Altmetric Score*, a qual reflete o impacto que a publicação científica obteve. Esse Score é gerado conforme a quantidade de leitores, menções, citações que os trabalhos científicos alcançam nos sites e blogs.

Há diversos trabalhos que utilizaram a plataforma *Altmetric* e suas ferramentas para estudos científicos. Lopes (2022), desenvolveu um estudo com o objetivo de analisar a correlação entre as citações do Google Scholar e o *Altmetric Score* dos artigos do Centro de Pesquisa, Inovação e Difusão em Neuromatemática, tendo por objetivo averiguar se as citações científicas estão relacionadas ao impacto social de uma pesquisa. No estudo de Lopes (2022) foram analisados dados de 183 artigos mencionados entre 2013 a 2021. O autor chegou às seguintes constatações: i) No Google Scholar os artigos foram citados 3055 vezes e obteve como média 16.69 citações. i i) No Altmetric Score a (pontuação fornecida pelo Altmetrics) se obteve um Score total de 2217.56, com média de 12.117. Vale ressaltar que esse Score foi composto pelos seguintes dados altmétricos: 4858 leitores advindos do Mendley, CiteUlike e Connotea, 1433 diferentes perfis de fonte de dados, 1153 menções do Twitter, 123 menções em sites de notícias, 70 menções no Facebook, 39 menções em blogs, 15 menções no Google +, 7 menções do Reddit, 3 menções em sites de upload de vídeos. Com os resultados do Google Scholar e Altmetric Score, Lopes (2022) obteve uma correlação de Pearson de 0.557, indicando assim uma correlação moderada.

Outro trabalho que utilizou a Altmetric foi desenvolvido por Torres, Robinson, Arroyo (2022), o qual visava averiguar a cobertura altmétrica da produção científica espanhola na plataforma. Teve-se como base publicações espanholas indexadas na *Web of Science* publicadas entre 2016 a 2020. Os autores descobriram que 55 % das 434.827 publicações científicas espanholas se encontram no Altmetric.com. sendo que as publicações atingiram um total de 3.569.553 menções em diferentes plataformas de mídias sociais. Dentre elas, o Twitter representou 89% dessas menções.

Shri, Kaur e Indu (2021) desenvolveram uma análise da correlação entre citações da Scopus com dados altmétricos, extraídos via Altmetric, de artigos sobre a Covid-19 no contexto da Índia. Teve-se como base 13.825 artigos de origem indiana indexados na Scopus entre 2001 até 2021. Dos 13.825 artigos o Altmetric.com cobriu

4.678 artigos que obtiveram um total de 56.713 citações, 359.411 leitores Mendeley, 148.340 menções no Twitter, 8.116 menções no Facebook, 862 menções em blogs, 174 Menções de vídeo, 219 Menções do Reddit e 123 Menções da Wikipédia. Foi constatado que os dados de citação extraídos possuem fortes correlações de Pearson com os dados Altmétricos. Houve forte correlação entre as menções do Blog ( $r=0,411$ ) e leitores do Mendeley ( $r=0,881$ ). “[...] verificou-se uma correlação significativa alta de que os artigos que possuem alto índice de leitores Mendeley possuem altas citações.” (SHRI, KAUR, INDU, 2021, p. 546).

Percebe-se que Altmetric é uma plataforma muito rica em dados alométricos, os quais podem ser utilizados para desenvolver diversos estudos alométricos. A sua ampla utilização por pesquisadores e a possibilidade de extração de dados gratuitamente por meio de sua API, são algumas das justificativas de sua escolha nesta pesquisa.

#### 2.3.7. Crossref

A Crossref é uma das agências de registros DOI, a qual possui uma iniciativa de Acesso aberto aos seus dados de citação. A base de dados disponibiliza dados de citações de trabalhos acadêmicos disponíveis gratuitamente para o público por meio de buscas diretamente na interface do site na Web ou por meio da API Rest. Para o entendimento das atividades desenvolvidas pela *Crossref*, consideramos necessário abordar sobre o identificador DOI, que é um elemento básico para a extração de dados desta e de muitas fontes de dados (como no próprio Mendeley e na plataforma Altmetric), sendo necessário no momento de se planejar e implementar uma solução para extração de dados, objetivo central deste trabalho.

#### 2.3.8. Identificador Digital de Objetos (DOI)

O Identificador Digital de Objetos (DOI) utilizado largamente nas áreas da informação e documentação foi criado e é gerenciado pela *International DOI Foundation* (IDF) desde 1998. O DOI é uma estrutura alfanumérica que tem por finalidade a identificação exclusiva e persistente de objetos no formato digital, físico ou abstrato. “Seus principais recursos incluem persistência, acessibilidade à rede, interoperabilidade com outros identificadores, infraestrutura tolerante a falhas compartilhada e a capacidade de “resolver” os identificadores de várias formas” (DOI,

2020). O identificador obteve a sua padronização pela *International Standards Organization* (ISO) em 2010 sendo identificado como norma ISO 26324 na qual especifica o funcionamento e características referente ao Identificador Digital de Objetos e nomeia a IDF como autoridade da norma e do DOI.

Para se obter um código DOI é necessária a participação de alguma agência de registro (RA) licenciadas pela IDF (*Crossref*, *DataCite*, *Entertainment ID Registry*). Cada uma possui sua sistemática, preços e características diferentes, contudo seguindo as orientações e normas estabelecidas pela IDF. Segundo (DOI ORG, 2020) “[...] a *Crossref* gerencia DOIs para o setor de publicação científica, o *DataCite* fornece DOIs para referência e compartilhamento de conjuntos de dados científicos, e o *Entertainment ID Registry* (EIDR) fornece identificadores e metadados associados que são usados no setor comercial de filmes e vídeos”. A *Crossref* será a agência de registro utilizada na elaboração deste trabalho.

O *Crossref* é uma agência sem fins lucrativos de registro de DOIs para setores de publicações científicas e conteúdos acadêmicos. “O *Crossref* torna os resultados da pesquisa fáceis de encontrar, citar, vincular e avaliar” (CROSSREF, 2020). A catalogação das publicações científicas dessa plataforma tem assegurada a confiabilidade, pois a ênfase está nos dados referenciais dos documentos, destacando-se informações autorais em sua base de dados.

Além do registro de DOIs, a plataforma possui inúmeras funcionalidades, entre elas estão a recuperação de metadados, que é de suma importância para elaboração desse trabalho. Essa função permite aos usuários a coleta de metadados sobre os trabalhos indexados no base, sendo viabilizado o acesso a informações como o tipo de documento, publicador, data em que foi indexado, autor, quantidade de trabalhos que referenciaram o documento em questão, etc.

Por ser um banco de dados aberto, a *Crossref* torna a consulta e a extração de dados uma tarefa bastante simples. Com isso, pesquisadores conseguem acessar metadados de grande valor para desenvolvimento de métricas e estudos científicos de forma relativamente fácil, via *API REST*.

O identificador DOI é uma combinação alfanumérica de caracteres imprimíveis do padrão Unicode, não possui limitação em sua extensão e não possuindo diferenciação de caracteres maiúsculos e minúsculos. Pode ser dividido em 2

elementos: o prefixo e o sufixo separados pelo caractere “/” (barra). O prefixo é um código exclusivo atribuído a um registrante segundo a organização responsável pelo registro do DOI. Geralmente a organização já possui uma sequência exclusiva podendo ser mais de um para mesma organização, onde diferentes objetos podem ter o mesmo prefixo. “Normalmente, um RA pode emitir um prefixo por cliente, mas também, pode ser apropriado emitir um prefixo por marca ou para algum cluster de produtos reconhecível (por exemplo, uma impressão do editor)” (DOI.ORG, 2020). O sufixo é atribuído definido pelo órgão responsável pelo registro, já um DOI deve ser exclusivo e único. Na Figura 1 é apresentado como um DOI é estruturado.

Figura 1 - Estrutura de um DOI



Fonte: De autoria própria (2022).

O prefixo possui 2 elementos separados pelo caractere “.” (ponto): i) indicador de diretório: na qual deve ser o número “10” sendo o responsável por identificar o item como um objeto digital no sistema responsável pela resolução do DOI; ii) o código de registrante é exclusivo. O código pode ser dividido em subelementos separados pelo caractere “.” (ponto) para devida identificação pelo sistema de resolução do DOI.” O código do registrante pode ainda ser dividido em subelementos para conveniência administrativa, se desejado” (DOI.ORG, 2020)

Algumas das características principais dos códigos DOIs podem ser visualizadas no Quadro 1.

### Quadro 1 - Algumas características de um DOI

Características principais:
<ul style="list-style-type: none"><li>• Possui um prefixo e sufixo separados por “/”;</li><li>• Não diferencia maiúsculas ou minúsculas;</li><li>• Formado por caracteres imprimível do Unicode;</li><li>• Sem limitação de extensão;</li><li>• Deve possuir o indicador de diretório “10” seguido de caractere de ponto “.”</li><li>• Pode possuir codificação hexadecimal.</li></ul>

Fonte: De autoria própria (2022).

#### Características principais:

- Possui um prefixo e sufixo separados por “/”;
- Não diferencia maiúsculas ou minúsculas;
- Formado por caracteres imprimível do Unicode;
- Sem limitação de extensão;
- Deve possuir o indicador de diretório “10” seguido de caractere de ponto “.”
- Pode possuir codificação hexadecimal.

O DOI pode ser utilizado como um *Uniform Resource Locator* (URL) facilitando assim a acessibilidade aos documentos. Por possuir uma atribuição permanente, mesmo que o conteúdo passe por atualizações o DOI não sofrerá alterações “[...] mesmo que um URL seja alterado, o nome DOI ainda funciona e redireciona para o novo local.” (DOI.ORG, 2020). Assim, com o DOI de um objeto é possível acessar seus metadados, identificadores e o endereço digital portador do objeto.

Para se utilizar o DOI como um URL uma maneira é utilizar o servidor *proxy* DOI público padrão (<https://doi.org>) anexando o identificador (por exemplo: 10.5195/biblios.2017.358) ao endereço do servidor proxy, tornando assim que o DOI possa ter a sua resolução mediante busca em navegador de internet, sem necessitar de softwares ou extensões. Na Figura 2 é apresentado a concatenação entre o servidor proxy (<https://doi.org>) e o DOI 10.5195/biblios.2017.358, resultando assim em um hiperlink que irá direcionar para o trabalho representado pelo DOI 10.5195/biblios.2017.358.

Figura 2 - Concatenação servidor proxy e código DOI resultando em um hiperlink



Fonte: De autoria própria (2022).

Além de o identificador ser único, persistente e operável no meio digital ele também possui interoperabilidade com outros identificadores, permitindo que seja possível trabalhar o DOI concomitantemente com outros esquemas de identificadores existentes, um exemplo disso pode ser visto na Figura 3. O órgão responsável pelo registro DOI pode inserir no prefixo ou sufixo do identificador, outro formato de identificador, como por exemplo o ISSN, ISBN, ISAN, ISNI e outros identificadores. O DOI não substitui outros identificadores, e permite que outros identificadores possam fazer parte de sua sintaxe com a finalidade de torna-lo um identificador mais eficiente. “Os princípios orientadores para fazer referência a outros esquemas de identificadores dentro do sistema DOI são maximizar a utilidade para usuários potenciais e maximizar sua eficiência de gerenciamento interno.” (DOI.ORG, 2020)

Figura 3 - Exemplo de DOI com ISSN

**10.11606/ISSN.2238-7714.NO.2017.131888**

Fonte: De autoria própria (2022).

Como foi apresentado, o DOI é um identificador persistente utilizado para identificar documentos digitais. A partir desses identificadores é possível extrair metadados via *API REST* sobre os documentos representados, por meio da plataforma Crossref. Para a confecção de um teste, neste trabalho, será coletado os identificadores DOI, dos trabalhos presentes na plataforma Sucupira, que será melhor abordada no próximo parágrafo.

Estudos utilizando a base de dados *Crossref*

O *Crossref* fornece APIs que proporcionam a extração de dados para estudos de bibliometria e cienciometria. Os metadados que eles oferecem podem acompanhar o crescimento e a importância da ciência, bem como as tendências de publicação e comunicação científica. Há diversos estudos que comprovam a viabilidade do uso da base de Dados *Crossref* no desenvolvimento de pesquisas.

Dentre os estudos, Esarey e Bryant (2018) utilizaram dados da *Crossref* com a finalidade de avaliar a diferença de gênero de citação. Desenvolveram um estudo que teve como base 1907 publicações do período entre 2007 a 2016 sobre ciência política, na qual foi obtido uma média de 28,92 citações por artigo. O estudo teve por finalidade identificar correlações entre o gênero feminino e a quantidade de citações, na qual teve-se como base os metadados do *Crossref* obtidos por meio de API. No estudo foi constatado que os artigos escritos por mulheres, em média, não são menos citados.

Outro trabalho desenvolvido utilizando a base de dados *Crossref* foi feito por Schenkel (2018) no qual se utilizou dos dados de citação da base para averiguar a cobertura de citações aberta em ciência da computação. Foi utilizado cerca de 3,2 milhões de publicações com DOIs. Destes somente 70.000 não foram reconhecidos pela API. Os artigos identificados retornaram 16 milhões de citações.

O trabalho produzido pelos autores Akbaritabar e Stahlschmidt (2019) utilizaram os dados do *Crossref* para monitorar publicações de acesso aberto em bancos de dados bibliométricos. Já o trabalho de Matthias *et al* (2019) utilizou a API do *Crossref* para coletar dados afim de descobrir quantos periódicos de acesso aberto mudaram para um modelo de assinatura. Os autores Hicks *et al* (2019) utilizaram dados advindos da base *Crossref* para desenvolverem uma análise em grande escala para avaliar o impacto do financiamento de pesquisas na consolidação da comunidade acadêmica.

A *Crossref* tem interesse e atuação em relação ao acesso aberto, incluindo citações abertas. Reconhece-se que acesso aberto a publicações científicas é de suma importância para possibilitar e incentivar uma ciência acessível a todos. Quanto mais impedimentos houver para acessar artigos e periódicos *online* de forma gratuita e aberta, maior será a dificuldade de países subdesenvolvidos a acompanhar e desenvolver ciência. Os periódicos tradicionais requerem assinaturas que não possuem preços acessíveis, principalmente para países não desenvolvidos, tornando

assim um obstáculo para acesso a produções científicas. Nesse interim e seguindo os mesmos princípios, surge um conceito relativamente novo, que é o de citação aberta.

O movimento de citações abertas surge na premissa de resolver os obstáculos referentes a acesso a dados de citações e dados bibliográficos, que em geral só são disponibilizados por meio de assinaturas em bases de dados. As citações abertas permitem que o público, independente de condições financeiras, possam elaborar estudos e avaliações. Assim, revistas, bases de dados, sites e organizações têm de forma crescente disponibilizado seus dados de citação de forma aberta, havendo destaque para o *Crossref*. Essa agência responsável pelos registros DOI, possui uma iniciativa de acesso aberto ao fornecer seus dados de citações abertamente, gratuitamente ao público, por meio de *API REST*.

Pode-se constatar que a base de dados Crossref tem sido utilizada de forma eficiente por diversos pesquisadores, em diferentes contextos em todo o globo, auxiliando assim no desenvolvimento de diversas pesquisas. Observa-se que a Crossref disponibiliza dados citações gratuitamente, que podem ser obtidos via sua API REST. As pesquisas indicam que a base Crossref consegue cobrir uma parcela significativa das publicações e um volume alto de citações, demonstrando ser uma fonte de dados de citação promissora. Diante desses fatores, decidiu-se utilizar esta fonte de dados no âmbito desta pesquisa.

#### 2.4. API'S E RASPAGEM DE DADOS E LINGUAGEM DE PROGRAMAÇÃO

As APIs são utilizadas para proporcionar a comunicação entre diferentes aplicativos. Uma API é criada quando uma empresa de *software* tem a intenção de que outros criadores de *software* desenvolvam produtos associados ao seu serviço. Com o uso das APIs, um determinado *software* pode utilizar informações de outro *software* para que haja comunicação entre eles. São exemplos *softwares* que utilizam o posicionamento do Google Maps ou sites que permitam o usuário usar sua conta do Facebook para fazer o *login*.

Este Trabalho irá utilizar API s de algumas bases da dados com a finalidade de coletar dados alométricos e de citação, sendo assim, é necessário que essas bases de dados possuam API s abertas e que permitam a coleta de dados. É importante considerar também qual tipo de parâmetro a API s de cada base irá solicitar.

Segundo Haddaway (2016, p. 187) a Raspagem da *Web* consiste no uso de um programa para extrair dados de arquivos HTML na internet. A *Web Scraping* pode ser entendida então como uma técnica que se utiliza de programação ou uso softwares para explorar dados da Internet e extrai-los de forma organizada e padronizada. Segundo Haddaway (2016, p. 187) a extração ocorre por meio de programas que interagem com as páginas da *Web* por meio de APIs. Interface essa que pode ser traduzido como Interface de Programação de Aplicativos.

As aplicações das API's podem ser as mais diversas, a exemplo das apresentadas por Sirisuriya (2015): comparação de preços online, desenvolvimento de pesquisas, extração de preços de mercadorias e produtos, monitoramento de patentes, análise de mercado, monitoramento de produções. Existem infinitas outras possibilidades que irão depender do interesse do usuário, não se limitando a profissionais da tecnologia da informação

Na atualidade torna-se cada mais evidente a importância, muitas vezes decisiva, de aquisição de informações e dados confiáveis e de qualidade. Caso o usuário colete os dados que necessitam manualmente em diversas fontes de informações, esse processo pode se tornar bastante caro, trabalhoso, demorado, chegando ao ponto de ser inviável. Assim, a utilização de API's tem a finalidade de otimizar a coleta de dados, na qual, proporcionará ao interessado maior rapidez e qualidade aos dados, subsidiando melhores análises dos dados e tomadas de decisões.

Um computador é formado tanto por hardwares quanto por *softwares*. Os *hardwares* são as peças físicas concretas, aquilo que é possível de manusear, já os *softwares* são programas executáveis pelo usuário que possibilitam realizar as suas tarefas. Esses *softwares* são construídos através de linguagens de programação, formadas por conjuntos de códigos interpretados por um compilador, que executam e processam os comandos escritos, transformando-os na linguagem utilizada pela máquina. Os compiladores e/ou interpretadores são, ferramentas utilizadas na tradução e interpretação da linguagem de programação, também chamada de código-fonte, para a linguagem executável pela máquina, chamado de código de máquina.

A primeira linguagem de programação data do século XIX. Desde então o avanço científico na área de programação tem sido cada vez mais revolucionário,

existindo os mais variados tipos de linguagens, com diferentes focos e vantagens. As linguagens de programação podem, em geral, ser divididas em baixo e alto nível. Essa nomenclatura não está associada a qualidade ou hierarquia de uma linguagem sobre a outra, mas sim, da sua proximidade com o código da máquina.

Uma linguagem de programação é de baixo nível quando está mais próximo do código de máquina, sendo por vezes até binária. Um exemplo de linguagem baixo nível é a uma linguagem chamada Assembly, criado na década de 50, sendo uma das linguagens pioneiras da computação.

Já as linguagens de alto nível são as de mais fácil legibilidade humana, cujos comandos são mais compreensíveis e os códigos mais intuitivos. Esse tipo de linguagem é estruturado em Scripts que são traduzidos posteriormente para código de máquina. Exemplos dessas linguagens de programação são o Python, Java, JavaScript, C++, C, C#, entre outras. Esse trabalho será desenvolvido utilizando a linguagem de alto nível Python, cuja justificativa de escolha será detalhada nos procedimentos metodológicos.

### 3. PROCEDIMENTOS METODOLÓGICOS

Nesta etapa são apresentados os procedimentos metodológicos adotados para a realização da pesquisa. São detalhadas as atividades e levantamentos realizados que foram necessários para a construção da ferramenta, levando-se em conta todo o conteúdo que foi abordado na revisão de literatura.

A presente pesquisa, que por seu estilo pode ser denominada como sendo de apresentação de produto, pode ser classificada a partir de princípios apresentados por Hernández Sampieri, Fernández Collado e Baptista Lucio (2006) como aplicada (busca a resolução de um problema social, prático e concreto), quantitativa (utiliza lógica de programação para o desenvolvimento da ferramenta), de cunho exploratório e descritivo (apresenta a ferramenta proposta e seus detalhes operacionais).

#### 3.1. PRINCÍPIOS GERAIS DA FERRAMENTA

A ferramenta desenvolvida foi denominada *Odisseia Metrics* em referência ao poema de Homero. A meta principal da ferramenta é que ela fosse capaz de extrair dados alométricos partindo do princípio que a ferramenta deveria ser: i) gratuita; ii) de fácil acesso; iii) de fácil manutenção; iv) utilizada de forma intuitiva. O identificador DOI deve ser usado como parâmetro único nessa ferramenta, medida foi tomada com o objetivo de simplificar o projeto, extração de dados com vistas a garantir a qualidade e consistência dos dados extraídos.

A escolha do público alvo é princípio de suma importância, já que é necessário definir para quem a ferramenta será destinada e moldar a sua construção, para atender as necessidades do público e seus objetivos. A ferramenta vislumbrou ter o objetivo de atender principalmente ao público que não possui domínio com linguagens de programação e que desejam desenvolver estudos alométricos e de citação, permitindo, assim, que a dificuldade ou ausência de conhecimento de programação não sejam um obstáculo para o desenvolvimento de estudos científicos.

A escolha do público se deu pelo fato de existir um grande número de acadêmicos que se interessa pela temática de métricas relacionadas a produção científica, mas que, por não terem conhecimento em programação, podem desistir da temática ou acabam desenvolvendo um trabalho de coleta de dados manual de forma

exaustiva. Levando esses fatos em consideração a ferramenta pretende remover esses obstáculos e promover ainda mais os estudos de métricas.

### 3.2. A ESCOLHA DA LINGUAGEM DE PROGRAMAÇÃO

Diversas linguagens de programação foram pensadas para o desenvolvimento deste trabalho, entre elas: Java, JavaScript, C++, C#, mas, no final foi escolhida a linguagem que mais condizia com a realidade acadêmica e com os objetivos deste trabalho, a linguagem Python. Python é uma linguagem de programação de alto nível lançada em 1991 por Guido van Rossum. Desde então, vem sendo bastante utilizada por desenvolvedores de *softwares*. A linguagem Python é muito utilizada na área de ciências de dados tendo uma ampla quantidade de módulos e *frameworks* que potencializam o seu uso.

A linguagem de programação selecionada para o desenvolvimento da ferramenta foi o Python e alguns dos motivos podem ser sumarizados:

- É uma linguagem que geralmente é ensinada nos cursos de graduação e pós graduação das instituições de ensino superior do país.
- É comumente utilizada para análise de dados.
- Possui uma gama de bibliotecas e frameworks dedicados à análise de dados.
- É uma linguagem fácil e com vasto material na Internet para o seu aprendizado.
- É muito bem documentada e que dificilmente se tornará obsoleta em um curto espaço de tempo.

As grandes possibilidades relacionadas a análises de dados que a linguagem Python traz é um dos principais motivos de sua escolha, já que o objetivo da ferramenta será coletar dados que posteriormente serão analisados por algum pesquisador. Vale levar em conta, também, que a ferramenta poderá receber *upgrades* e ter seu desenvolvimento continuado pela comunidade acadêmica. É uma linguagem acessível e que já consta na grade curricular da grande parcela dos cursos acadêmicos (a exemplo do curso de Biblioteconomia da UnB). Futuramente, caso seja necessário, é possível que haja uma união com outras linguagens de programação e até de marcação.

### 3.3. HOSPEDAGEM E AMBIENTE DE EXECUÇÃO DA FERRAMENTA

Quanto a hospedagem seguiu-se a premissa de que o usuário conseguisse acessar a ferramenta de forma fácil e rápida. Como é um projeto acadêmico sem fins lucrativos, considerou-se essencial que a hospedagem ocorresse de forma gratuita, por não haver previsão de recursos financeiros para o desenvolvimento e manutenção da ferramenta. A princípio foi pensado em apenas disponibilizar o código fonte na Internet para que os interessados pudessem acessá-lo e usar por conta própria. Porém isso excluiria uma gama de usuários que não sabem como configurar um ambiente para execução de códigos de programação, indo contra um dos nossos objetivos, que é promover e facilitar estudos alométricos por parte da comunidade científica e acadêmica.

Tendo isso em pauta, acabamos por encontrar uma excelente solução que resolveria tanto a hospedagem do código quanto o ambiente de execução do código: a utilização do *Google Colab*. O *Google Colab* ou *Google Colaboraty* é um ambiente online que permite a hospedagem e a execução de códigos Python sem necessidade de configuração, sendo de acesso gratuito e de fácil compartilhamento. Percebe-se que o *Google Colab* é uma ferramenta muito potente, completa e gratuita, sendo esses aspectos significantes para a escolha da plataforma para servir de hospedagem da ferramenta Odisseia.

Outro aspecto importante é o fato de a plataforma permitir *Scripts Python*, a qual foi a linguagem selecionada para o desenvolvimento deste trabalho. Vale ressaltar que a execução do *Notebooks Colab* ocorre nos servidores do Google, não necessitando assim que o usuário possua um computador com grande potência de hardware, removendo assim mais um obstáculo para o usuário e facilitando que mesmo os usuários com computadores de baixa performance possam executar estudos com uma grande massa de dados. Diante desses argumentos e das análises feitas o *Google Colab* se mostrou uma excelente opção para construção da *Odisseia Metrics*.

O *Google Colab* permite executar linhas de códigos em seu ambiente chamado, *Notebook Colab*. Esse permite que sejam inseridos diversos códigos na qual podem ser executados individualmente. Um exemplo de código no *Google Colab* pode ser visualizado na Figura 4.

Figura 4 - Exemplo de um código no *Notebook Colab*

```
Código 1

[3] seconds_in_a_day = 24 * 60 * 60
seconds_in_a_day

86400

Código 2

[2] print("hello World")

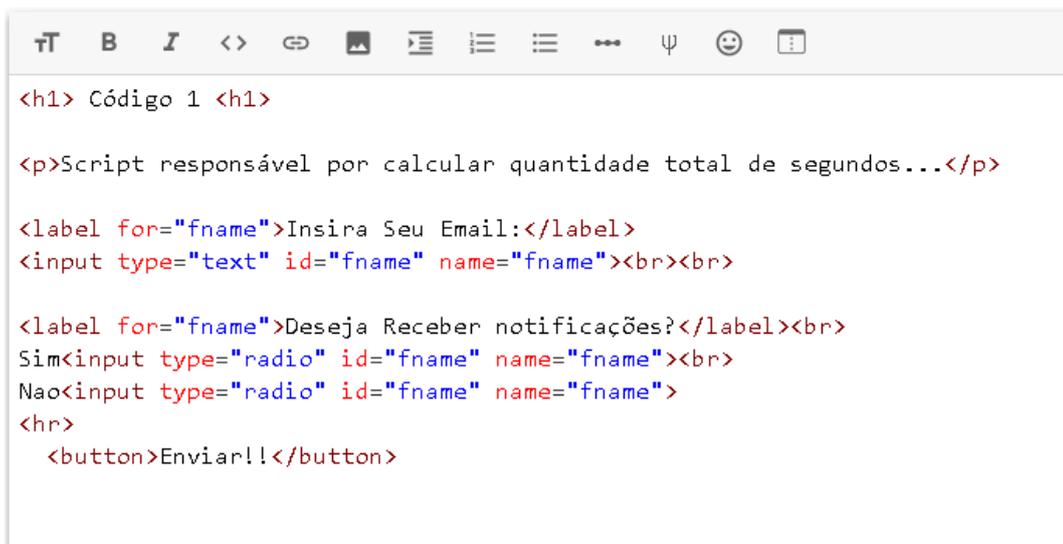
hello World
```

Fonte: De autoria própria (2022).

Observe que na Figura 4, temos dois códigos diferentes que estão em um mesmo *Notebook Colab*. O primeiro código se refere ao cálculo de segundos que um dia possui, e logo em seguida, o *script* imprime a variável retornando o valor total do cálculo. O segundo *script* imprime a frase *hello world* no console. Percebe-se que os *scripts* não possuem relação nenhuma um com o outro, mas ocupam um mesmo ambiente de desenvolvimento (*Notebook Colab*), ou seja, é permitido, em uma mesma página, haver diversos códigos distintos com execuções isoladas.

Outro ponto forte do *Google Colab* é a possibilidade de juntar códigos executáveis com linguagem de marcação ou *rich text*, permitindo assim que o usuário consiga utilizar HTML e inserir imagens no Notebook. Veja o exemplo de código inserido na Figura 5 e respectivo resultado na Figura 6.

Figura 5 - Inserindo HTML em um *Notebook Colab*



```

<h1> Código 1 </h1>

<p>Script responsável por calcular quantidade total de segundos...</p>

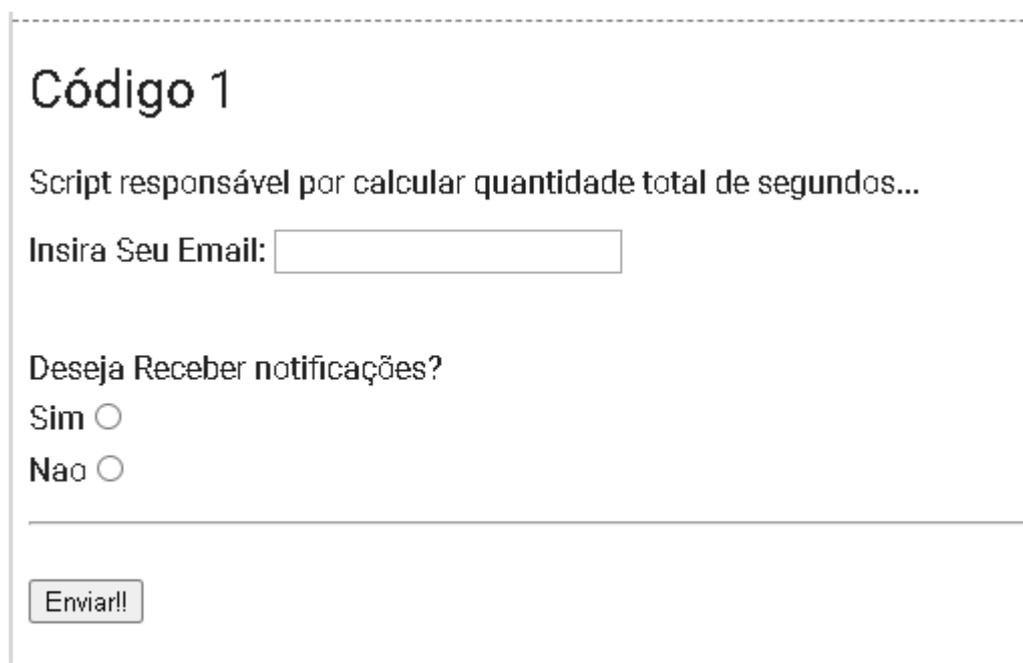
<label for="fname">Insira Seu Email:</label>
<input type="text" id="fname" name="fname"><br><br>

<label for="fname">Deseja Receber notificações?</label><br>
Sim<input type="radio" id="fname" name="fname"><br>
Nao<input type="radio" id="fname" name="fname">
<hr>
<button>Enviar!!</button>

```

Fonte: De autoria própria (2022).

Figura 6 - HTML renderizado em um Notebook Colab



**Código 1**

Script responsável por calcular quantidade total de segundos...

Insira Seu Email:

Deseja Receber notificações?

Sim

Nao

---

Fonte: De autoria própria (2022).

A possibilidade de inserir HTML, Markdown, Latex, Imagens etc. abre muitas possibilidades para instruir o usuário na utilização da ferramenta, já que por meio dessas funcionalidades é possível guiar passo a passo do usuário, inserir instruções, tutoriais, imagens explicativas, redirecionar para vídeos e uma infinidade de outras possibilidades. Um exemplo de inserção de instruções no Google Colab pode ser visualizado na Figura 7.

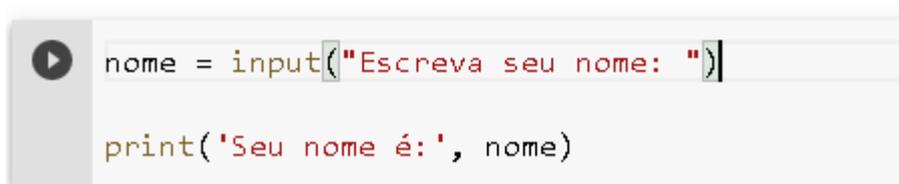
Figura 7 - Notebook com HTML e Script Python

👋 Olá, Jovem Padawan

Não sabe, como começar 😨?

SIMPLES, Siga o passo a passo a seguir:

- Aperte o Botão Play ▶
- Digite seu nome
- Veja a mensagem sendo imprimida com seu Nome



```
nome = input("Escreva seu nome: ")  
  
print('Seu nome é:', nome)
```

Fonte: De autoria própria (2022).

Os Notebooks são armazenados na conta do Google Drive do usuário, indicando assim que o Google *Colab* e o Google drive possam se relacionar. Por meio de configurações é possível utilizar documentos, planilhas e imagens hospedados no Google Drive diretamente no *Notebook Colab*. Considerando que a ferramenta estará hospedada em um *Notebook Colab*, o usuário terá acesso por meio de *Link*, sendo necessário apenas que o se tenha acesso à um dispositivo conectado à *Internet* com navegador internet atualizado.

Para conseguir utilizar a ferramenta é necessário que o usuário possua conta no Google e que esteja logado na hora da execução. Essa autenticação é necessária pelo fato de o Google *Colab* não poder ser utilizado sem que o usuário esteja logado em uma conta Google.

### 3.4. FONTES DE DADOS/BASES DE DADOS SELECIONADAS

A escolha das bases teve como pré-requisitos possuir uma API acessível gratuitamente e ser possível a utilização do DOI como parâmetro. Foram selecionadas três bases de dados para o desenvolvimento da ferramenta: Mendeley, Crossref e Altmetric.

A escolha do Mendeley se deu pelo fato de possuir dados alométricos bastante interessantes que é o caso da quantidade de leitores de um trabalho científico. Além disso a escolha do Mendeley foi motivada devido à gama de trabalhos científicos desenvolvidos no exterior se utilizando dos dados alométricos do Mendeley e a forte correlação com dados de citação.

A escolha do *Crossref* se deu pelo motivo de que o identificador único DOI ter sido escolhido como parâmetro principal para extração de dados pela ferramenta. Devido a relação do *Crossref* com DOI e também pelo fato da API do *Crossref* ser de fácil acesso e bem documentada, a base de dado se tornou uma candidata acessível para inclusão no projeto. Além disso, acredita-se que a *Crossref* é uma base de dados de citação muito promissora, sobretudo para a análise de informações de países que são sub-representados em bases de dados comerciais internacionais.

A Plataforma Altmetric foi adicionada à lista por possuir uma API acessível e por ter uma gama de dados alométricos que agregam: o Twitter, Facebook, Youtube, Google + e outras redes sociais. Esses dados advindos de redes sociais podem ser bastante interessantes para análise de onde há possibilidade de se medir outros impactos as publicações científicas.

### 3.5. PASSO PARA O DESENVOLVENDO A FERRAMENTA ODISSEIA

A construção da ferramenta foi realizada em 7 etapas, sendo elas:

- 1 Levantamentos necessários para extrair dados da API - Altmetric
- 2 Desenvolver código de extração de dados conforme API da base Altmetric.
- 3 Levantamentos necessários para extrair dados da API – Mendeley
- 4 Desenvolver código de extração de dados conforme API da base Mendeley.
- 5 Levantamentos necessários para extrair dados da API – Crossref.
- 6 Desenvolver código de extração de dados conforme API da base Crossref.
- 7 União dos Três códigos desenvolvidos e adicionar interface gráfica básica

Após essa fase da pesquisa, a etapa final foi testar a ferramenta a partir de uma lista de DOIs e, por fim, foi elaborado um manual com o passo a passo para a utilização da ferramenta pelo usuário.

Adicionalmente, foi realizado um teste de veracidade da API com objetivo de verificar se os dados estavam sendo extraídos corretamente. Para este teste, foi

selecionado, aleatoriamente, uma lista de 200 DOI's de artigos disponíveis na Plataforma Sucupira. Esses DOIs foram utilizados para a extração dos dados por meio da ferramenta *Odisseia Metrics*, nas fontes de dados suportadas: *Altmetrics.com*, *Mendeley* e *Crossref*. Posteriormente, desse montante extraído, foram selecionados 30 DOI's aleatoriamente de cada uma das fontes de dados. Estes foram inseridos manualmente um a um na API do *Mendeley*, *Crossref* e *Altmetrics* e os seus dados foram comparados aos resultados disponibilizados pela *Odisseia Metrics*.

#### **4. RESULTADOS: DESENVOLVENDO A FERRAMENTA ODISSEIA**

##### **4.1. LEVANTAMENTOS NECESSÁRIOS PARA EXTRAIR DADOS DA API – ALTMETRIC**

Pretende-se que a ferramenta *Odisseia* seja capaz de extrair diversos dados do *Altmetric* por meio de uma lista de DOIs, na qual retornará diversos dados altmétricos como: quantidade de menções do trabalho no Facebook, Youtube, Reddit, Blogs, quantidade de leitores/capturas no Twitter, Citeulike, Connotea. O outro dado extraído é o "is\_oa", o qual informará se o trabalho científico se enquadrará em Acesso Aberto ou não.

A primeira ação necessária para o desenvolvimento da ferramenta foi a busca na base de dados do *Altmetric*. Consultando a API da base *Altmetric* <https://api.altmetric.com> foi possível ter visão dos dados que ela fornece de forma gratuita e também a possibilidade de analisar quais rotas para a API podem ser utilizadas. A versão da API disponível para uso foi a *Altmetric Details v1*. Esta permite fazer uma requisição do tipo GET tendo como parâmetro o DOI. No caso, optou-se por utilizar a URL [https://api.altmetric.com/ {doi}](https://api.altmetric.com/{doi}) que irá receber como parâmetro o identificador DOI.

Considerando que o DOI foi definido como o parâmetro principal utilizado pela ferramenta, definindo-se a rota e o parâmetro, posteriormente foi necessário definir os dados a serem coletados com a ferramenta. Após a análise da documentação da plataforma *Altmetric*, identificou-se a lista de Dados retornados pela API e que são extraídos pela ferramenta *Odisseia* (Quadro 2).

Quadro 2 - Dados fornecidos pela API Altmetric

Campos	Descrição
0	List of DOIs
title	Title of the publication
journal	Name of publication journal
authors	List of author names
type	Type of publication (e.g article, dataset etc.)
is_oa	Is the article Open Access? Not good data, came from Scopus in 2011 and never updated
cited_by_msm_count	Number of the news sources that have mentioned the publication
cited_by_fbwalls_count	Number of the pages that have shared on Facebook
cited_by_videos_count	Number of the Youtube/Vimeo channels
cited_by_rdts_count:	Number of Reddit threads posted about this publication
cited_by_feeds_count	Number of blogs that have mentioned the publication
cited_by_gplus_count	Number of the accounts that have shared on Google+
cited_by_posts_count	A "post" is any online document that links to one or more research objects (i.e. a post is a mention or a group of mentions). This field contains the number of distinct posts that include one or more mentions of the research object in question.
cited_by_tweeters_count	Number of the Twitter accounts that have tweeted this publication
readers>citeulike	Readers in citeulike
readers>Mendeley	Readers in Mendeley
readers>connotea	Readers in Connotea
readers_count	Total Reader Counts

Fonte: documentação da API Altmetric: <https://www.altmetric.com/products/altmetric-api/>

O Response Code é a maneira padrão de os servidores comunicarem informações sobre o sucesso ou falha de uma solicitação. Quando o usuário faz uma solicitação de dados a um servidor da *Web*, um código de status de resposta HTTP de três caracteres é retornado. Os números retornados no cabeçalho de resposta HTTP possuem significados que indicam o sucesso ou falha da operação. O Response Code utilizado pela API Altmetric é apresentado no Quadro 3.

Quadro 3 - Response Code API Altmetric

Cód. de status HTTP	Descrição
200	Sucesso. O corpo da resposta deve conter os dados solicitados.
403	Você não está autorizado para esta chamada. Algumas chamadas e tipos de consulta só podem ser feitos por detentores de uma chave de API e/ou licença para a versão Full Access da API.
404	A Altmetric não tem detalhes sobre o artigo ou conjunto de artigos que você solicitou.
429	Você está sendo taxa limitada. Se você ainda não o fez, solicite uma chave de API.
502	A versão da API Altmetric Details Page que você está usando está atualmente em manutenção.

Fonte: Documentação API Altmetric: <https://www.altmetric.com/products/altmetric-api/>

Caso o Status Code retornado seja 200 (indicando sucesso na requisição de dados a um Servidor Web), a extração de dados será executada, caso contrário será retornado ao usuário uma mensagem “erro”. Caso o Status Code retornado for diferente de 200 indicando assim que houve algum tipo de falha na busca da informação, será inserido na tabela o sinal hífen “-”, ao invés dos dados solicitados.

Para realizar a busca, o usuário deverá colocar a lista de DOIs na qual deseja extrair os metadados em uma tabela no formato .xlsx na qual o título da coluna com os Identificadores DOIs tenha como título “0” (número zero). Caso o usuário deseje pular a linha ele deve inserir o sinal de hífen “-” ao invés do DOIs.

#### 4.2. DESENVOLVIMENTO DO CÓDIGO DE EXTRAÇÃO DE DADOS CONFORME API DA BASE ALTMETRIC

O primeiro passo para desenvolver o código de extração de dados foi a criação um Notebook Colab para inserimos um *Script Python*. Não foi necessário realizar outras configurações, pois utilizamos as configurações já nativas do Google Colab.

Com o Notebook já criado, foi iniciado a programação de um Código Python. O primeiro passo da ferramenta após inicializada foi coletar o arquivo .xlsx contendo a lista de DOIs adicionadas pelo usuário, então o primeiro passo foi permitir que seja possível inserir uma tabela na ferramenta, para isso inserimos um botão de *upload*, permitindo que o usuário consiga inserir uma tabela, do seu computador pessoal para o ambiente Colab, de forma fácil. Com isso criamos o botão com comando apresentado na Figura 8.

Figura 8 - Adicionando botão de upload de arquivos no Notebook Colab



Fonte: De autoria própria (2022).

Acionando o botão *upload*, torna-se possível que o usuário insira a tabela desejada no ambiente, contudo é necessário que os dados que estão sendo inseridos tenham sido checados, tratados e validados previamente. O sistema irá tentar coletar os dados da tabela inserida e, caso seja identificado algum erro, será retornado uma mensagem de erro ao usuário. Nesse caso o usuário deverá verificar se a tabela estará conforme foi pré-estabelecido para o funcionamento da ferramenta. O *Script* desenvolvido para escanear os dados da tabela inserida com a lista de DOIs é disponibilizado na Figura 9.

Figura 9 - Script desenvolvido para escanear tabela

```
#tentando escanear a tabela inserida pelo usuário
try:
    df = pd.read_excel (fn)
    df[0] = df[0].astype('category')
    quant_inicio =0
    quant_linhas= df[0].count()
    total_resultado = pd.DataFrame(df)
    output.clear()

except:
    !rm token.xlsx
    !rm $fn
    output.clear()
    display(Markdown("# ERRO!\n # Não houve Upload da tabela"))
    sys.exit()
```

Fonte: De autoria própria (2022).

Após a verificação da tabela, a extração dos dados pode ser realizada. A extração é feita por meio de *requests* (requisição) na qual será inserido o DOI por meio de um *Loop For* (execução contínua da requisição de cada DOI). Também é verificado se o *Status Code* terá sucesso e, caso tenha, as informações serão armazenadas. Caso contrário será inserido o sinal hífen “-” em todos os campos respectivos ao DOI buscado. Esse hífen, indica que o DOI solicitado é inexistente ou inconsistente. O Script desenvolvido para a extração dos dados é apresentado na Figura 10.

Figura 10 - Script para coletar os dados da API Altmetric

```

try:
    if total[k]!='-':
        #fazendo requisição da Api utilizando o DOI como parâmetro
        x= requests.get('https://api.altmetric.com/v1/doi/'+total[k])
        if x.status_code == 200:
            x = x.json()
            for contagem in range (0,13):
                if status[contagem] in x:
                    lista[contagem].append(x[status[contagem]])
                else:
                    lista[contagem].append('-')
            lista[13].append(x['readers']['citeulike'])
            lista[14].append(x['readers']['mendeley'])
            lista[15].append(x['readers']['connotea'])
            lista[16].append(x['readers_count'])

        else:
            remover=[lista[x].append('-') for x in range(0,17)]
    else:
        remover=[lista[x].append('-') for x in range(0,17)]
except:
    remover=[lista[x].append('-') for x in range(0,17)]

```

Fonte: De autoria própria (2022).

Após iniciar a extração todos os dados, os dados serão armazenados em uma lista, com os dados organizados em linha, ou seja, para cada linha com DOI haverá colunas com os dados altmétricos das mídias cobertas pela plataforma Altmetric.

Figura 11 - Script para baixar a nova tabela automaticamente

```

for cont in range (0,17):

    total_resultado.insert(loc=(cont+1), column=status[cont],value=lista[cont])

total_resultado.to_excel (r'tabela_altmetrics.xlsx', index = False, header=True)
!rm $fn
files.download('tabela_altmetrics.xlsx')
print("Coleta Completa \nArquivo Baixado\n Legenda dos Metadados: https://docs.goo
-----

```

Fonte: De autoria própria (2022).

Após chegar ao último DOI da tabela inserida pelo usuário, os resultados armazenados, são inseridos automaticamente em um arquivo no formato. xlsx com o nome "tabela\_Altmetric". O *download* dessa lista com os dados altmétricos é feito de forma automática na máquina do usuário, contendo todos os resultados coletados. O

código desenvolvido para a realização deste processo é encontrado na Figura 11. Por fim, o código completo desenvolvido, contendo todas as etapas apresentadas até este momento, que resulta na extração dos dados finais de extração da API Altmetric, pode ser visualizado na Figura 12.

Figura 12 - Script completo desenvolvido para extração da API Altmetric

```

try:
#-----
def altmetric():

    total=[]
    lista = [[] for _ in range(17)]

    # nome dos campos que serão extraídos
    status=['title', 'journal', 'authors', 'type', 'is_oa',
            'cited_by_msm_count', 'cited_by_fbwalls_count',
            'cited_by_videos_count', 'cited_by_rdt_count',
            'cited_by_feeds_count', 'cited_by_gplus_count',
            'cited_by_posts_count', 'cited_by_tweeters_count',
            'readers>citeulike', 'readers>mendeley',
            'readers>connotea', 'readers_count']

    # 'quant_linhas' é o número da linha atual dentro da Array onde está a tabela inteira
    for k in range (0,quant_linhas,desc='Coletando'):

        total.append(df[0].loc[k])

    try:
        if total[k]!='-':
            #fazendo requisição da Api utilizando o DOI como parâmetro
            x= requests.get('https://api.altmetric.com/v1/doi/'+total[k])
            if x.status_code == 200:
                x = x.json()
                for contagem in range (0,13):
                    if status[contagem] in x:
                        lista[contagem].append(x[status[contagem]])
                    else:
                        lista[contagem].append('-')
                lista[13].append(x['readers']['citeulike'])
                lista[14].append(x['readers']['mendeley'])
                lista[15].append(x['readers']['connotea'])
                lista[16].append(x['readers_count'])

            else:
                remover=[lista[x].append('-') for x in range(0,17)]
        else:
            remover=[lista[x].append('-') for x in range(0,17)]
    except:
        remover=[lista[x].append('-') for x in range(0,17)]

    output.clear()

    for cont in range (0,17):

        total_resultado.insert(loc=(cont+1), column=status[cont],value=lista[cont])

    total_resultado.to_excel (r'tabela_altmetrics.xlsx', index = False, header=True)
    !rm $fn
    files.download('tabela_altmetrics.xlsx')
    print("Coleta Completa \nArquivo Baixado\n | legenda dos Metadados: https://docs.google.c
#-----

```

Fonte: De autoria própria (2022).

#### 4.3. LEVANTAMENTOS REALIZADOS PARA EXTRAIR DADOS DA API – MENDELEY

A documentação referente a API do Mendeley pode ser consultada no link <https://api.Mendeley.com/apidocs/docs>. A documentação fornece várias rotas acessíveis para extração de dados, permitindo até executar de testes de chamadas da API no próprio site, desde que o usuário esteja logado com sua conta Mendeley. A rota utilizada pelo Odisseia foi “/catalog” do tipo de requisição GET, tendo como Query Parameter o DOI (String), na qual indica o DOI do documento que deseja ser consultado. Outro parâmetro utilizado foi o campo “*view=status*” que é o responsável por filtrar os dados que se deseja. Ambos os parâmetros são realizados via URL.

São extraídos dados referentes aos “leitores” do Mendeley, retornando em cada coluna um tipo de leitor distinto para cada DOI. Os dados a serem extraídos com a *Odisseia Metrics* são apresentados no Quadro 4.

Quadro 4 - Dados extraídos da API Mendeley

<u>Tipo</u>	<u>Tipos de leitores</u>
<u>Quant Leitores/Mendeley</u>	<u>Total de Leitores</u>
<u>Other</u>	<u>Outros tipos de leitores</u>
<u>Professor</u>	<u>Professor</u>
<u>Professor &gt; Associate Professor</u>	<u>Professor Associado</u>
<u>Researcher</u>	<u>Pesquisadores</u>
<u>Student &gt; Bachelor</u>	<u>Estudante de bacharel</u>
<u>Student &gt; Doctoral Student</u>	<u>Estudante de Doutorado</u>
<u>Student &gt; Master</u>	<u>Estudante de mestrado</u>
<u>Student &gt; Ph. D. Student</u>	<u>Estudante de Ph.d</u>
<u>Student &gt; Postgraduate</u>	<u>Estudante pós graduado</u>
<u>Lecturer</u>	<u>Lecturer</u>
<u>Lecturer &gt; Senior Lecturer</u>	<u>Senior Lecturer</u>
<u>Librarian</u>	<u>Bibliotecário</u>
<u>Unspecified</u>	<u>Não definido</u>

Fonte: Documentação da API do Mendeley: <https://api.Mendeley.com/apidocs/docs>

A rota Mendeley Utilizada pelo Odisseia, poderá retornar os seguintes Status Code, os quais são apresentados no Quadro 5.

Quadro 5 - Response Code API Mendeley

HTTP Status Code	Descrição
200	Documents found
400	Invalid view requested
400	No filter specified
400	Either document id or author profile id should be given, not both
404	Document not found

Fonte: Documentação da API do Mendeley: <https://api.Mendeley.com/apidocs/docs>

Em caso de o Status Code ser 200, significará que houve sucesso na busca dos dados, caso contrário será retornado a mensagem de “erro” ao usuário. Caso o Status Code retornado for diferente de 200 indicando assim que houve algum tipo de falha na busca da informação, será inserido na tabela o sinal hífen “-”, ao invés dos dados referentes aos leitores.

Para a extração dos dados do Mendeley é necessário que o usuário acesse com login e senha a sua conta Mendeley, por meio de um *link* que será disponibilizado para o usuário. Após o usuário acessar sua conta Mendeley, ele deverá copiar uma chave Token que lhe será fornecida e colar na ferramenta Odisseia.

O usuário deverá colocar a lista de DOI's na qual deseja extrair os metadados em uma tabela no formato .xlsx na qual o título da coluna com os Identificadores DOIs tenha o número “0” (zero). Caso o usuário deseje pular a linha ele deve inserir o sinal de hífen “-” ao invés do DOIs.

#### 4.4. DESENVOLVIMENTO DO CÓDIGO DE EXTRAÇÃO DE DADOS CONFORME API DA BASE MENDELEY

O primeiro passo foi realizar a autenticação do usuário, necessária para se acessar sua conta Mendeley para poder coletar o Token. Após inserir o Token, a extração é autorizada. A parte referente a autenticação Mendeley teve como base o *Script* disponibilizado pelo usuário Juan Pablo Alperin em sua conta GitHub

(<https://github.com/jalperin?tab=repositories>). O código utilizado para a realização do processo de autenticação da API do Mendeley é apresentado na Figura 13.

Figura 13 - Script de autenticação da API Mendeley

```

CLIENT_ID = ''
CLIENT_SECRET = ''
REDIRECT_URI = 'http://localhost/'
AUTHORIZE_URL = 'https://api-oauth2.mendeley.com/oauth/authorize'
TOKEN_URL = 'https://api-oauth2.mendeley.com/oauth/token'

def make_authorization_url():
    params = {"client_id": CLIENT_ID,
             "response_type": "code",
             "redirect_uri": REDIRECT_URI,
             "duration": "temporary",
             "scope": "all"}
    url = AUTHORIZE_URL + "?" + urllib.parse.urlencode(params)
    return url

def get_token(code):
    client_auth = requests.auth.HTTPBasicAuth(CLIENT_ID, CLIENT_SECRET)
    post_data = {"grant_type": "authorization_code",
                "code": code,
                "redirect_uri": REDIRECT_URI}
    response = requests.post(TOKEN_URL,
                             auth=client_auth,
                             data=post_data)
    token_json = response.json()
    return token_json["access_token"], token_json["refresh_token"]

def renew_token():
    client_auth = requests.auth.HTTPBasicAuth(CLIENT_ID, CLIENT_SECRET)
    headers = {"Authorization": "bearer " + access_token}
    post_data = {"grant_type": "refresh_token",
                "refresh_token": refresh_token}
    response = requests.post(TOKEN_URL,
                             auth=client_auth,
                             data=post_data)
    token_json = response.json()
    return token_json["access_token"], token_json["refresh_token"]

```

Fonte: De autoria própria (2022).

No momento da autenticação é disponibilizado um *link* para que o usuário possa acessar e, por fim liberar seu acesso, o passo a passo dessa funcionalidade será apresentada mais a frente, no tutorial de como utilizar a ferramenta.

Após a autenticação dá-se o início a extração de dados via request. Pode-se observar que a API do Mendeley usa como parâmetros o DOI, o Token inserido pelo usuário e, ainda, um parâmetro estático “&view=stats” o qual é responsável por identificar os dados referente aos DOIs. O resultado retornado é apresentado no

formato Json, sendo necessário a utilização de alguns filtros para que os dados possam ser utilizados. O código utilizado para a extração dos dados é disponibilizado na Figura 14.

Figura 14 - Script extração de dados da API Mendeley

```

try:

    novo = 'Bearer <'+access_token+'>'
    link = 'https://api.mendeley.com/catalog/?doi='+total[x]+'&view=stats'

    z = requests.get(link,
                    headers = { 'Authorization': novo })
    z = z.json()

    if z!=[]:

        z = str(z).replace("'", '')
        z = z[1:-1]
        z = str(z).replace("True", "True")
        z = str(z).replace("False", "False")

        if "reader_count" in z:
            dect = str(z).find('reader_count')
            dectfinal=str(z).find('reader_count_by_user_role')

            if dect > 0:

                nova = str(z)[((dect)-1):((dectfinal)-3)]
                nova='{'+ nova+'}'

            h = json.loads(nova)

```

Fonte: De autoria própria (2022).

Assim como na busca do Altmetric, caso a busca pelo DOI apresente alguma falha (não seja encontrado) , será adicionado o sinal de hífen “-” em todos os campos referente ao DOI em questão. O código elaborado para tal processo é apresentado na Figura 15.

Figura 15 - Inserindo o sinal hífen caso o DOI falhe

```

except:

    leitores.append('-')
    for contagem in range (0,13):
        lista[contagem].append('-')

```

Fonte: De autoria própria (2022).

Após todos os DOIs serem verificados e armazenados em uma lista, ao final do programa essa lista será convertida, de Jason, para um arquivo “.xlsx” O *Script* que realiza esta ação é apresentado na Figura 16.

Figura 16 - Script para baixar nova tabela com resultados extraídos da API Mendeley

```
output.clear()
total_resultado.insert(loc=1, column='Quant Leitores/Mendeley',value=leitores}]

for contagem in range (0,13):
    total_resultado.insert(loc=(2+contagem), column=status[contagem],value=lista[contagem])

#total_resultado
total_resultado.to_excel (r'tabela_mendeley.xlsx', index = False, header=True)
print("Coleta Completa \nArquivo Baixado")
!rm $fn
files.download('tabela_mendeley.xlsx')
```

Fonte: De autoria própria (2022).

Uma mensagem de “Coleta completa” será apresentada ao usuário e, concomitantemente, um arquivo com o nome “tabela\_Mendeley.xlsx” contendo todos os dados extraídos será baixado automaticamente após as buscas na API serem finalizadas com sucesso.

Vale observar que o Token tem um prazo de tempo para expirar. Para solucionar esse problema, foi adicionado um código para que o *token* seja atualizado a cada 3500 DOIs buscados. Sendo assim o usuário poderá buscar uma grande massa de dados sem se preocupar com a expiração do Token, visto que essa ação foi automatizada com o código apresentado na Figura 17.

Figura 17 - Código que atualiza o Token do usuário automaticamente

```
if contador_de_voltas >=3500:

    access_token, refresh_token = renew_token()

    contador_de_voltas=0
```

Fonte: De autoria própria (2022).

Por fim, podemos observar como ficou o resultado do código completo de extração da API Mendeley (Figura 18).

Figura 18 - Script Completo para extração da API Mendeley

```

def mendeley():

    total=[]
    for x in range(quant_inicio,quant_linhas):

        total.append(df[0].loc[x])

    CLIENT_ID = ''
    CLIENT_SECRET = ''
    REDIRECT_URI = 'http://localhost/'
    AUTHORIZE_URL = 'https://api-oauth2.mendeley.com/oauth/authorize'
    TOKEN_URL = 'https://api-oauth2.mendeley.com/oauth/token'

    def make_authorization_url():
        params = {"client_id": CLIENT_ID,
                  "response_type": "code",
                  "redirect_uri": REDIRECT_URI,
                  "duration": "temporary",
                  "scope": "all"}
        url = AUTHORIZE_URL + "?" + urllib.parse.urlencode(params)
        return url

    def get_token(code):
        client_auth = requests.auth.HTTPBasicAuth(CLIENT_ID, CLIENT_SECRET)
        post_data = {"grant_type": "authorization_code",
                    "code": code,
                    "redirect_uri": REDIRECT_URI}
        response = requests.post(TOKEN_URL,
                                 auth=client_auth,
                                 data=post_data)
        token_json = response.json()
        return token_json["access_token"], token_json["refresh_token"]

    def renew_token():
        client_auth = requests.auth.HTTPBasicAuth(CLIENT_ID, CLIENT_SECRET)
        headers = {"Authorization": "bearer " + access_token}
        post_data = {"grant_type": "refresh_token",
                    "refresh_token": refresh_token}
        response = requests.post(TOKEN_URL,
                                 auth=client_auth,
                                 data=post_data)
        token_json = response.json()
        return token_json["access_token"], token_json["refresh_token"]

    output.clear()

    print ("\n          Acesse o link abaixo e copie o seu código:")

    print (make_authorization_url())

    code = str(input("\nInsira seu Cod: "))
    output.clear()

    if code == '':
        exit

```

```

t=[]
leitores=[]
posgradu=[]
profAssociado=[]
estudantePhd = []

#lista=[]
status= ['Other', 'Professor', 'Professor > Associate Professor', 'Researcher',
         'Student > Bachelor', 'Student > Doctoral Student', 'Student > Master',
         'Student > Ph. D. Student', 'Student > Postgraduate', 'Lecturer',
         'Lecturer > Senior Lecturer', 'Librarian', 'Unspecified']

lista = [[] for _ in range(13)]
lista69=[]
contador_de_voltas=0

for x in trange (quant_inicio,quant_linhas,desc='Coletando'):

    contador_de_voltas+=1

    if contador_de_voltas >=3500:

        access_token, refresh_token = renew_token()

        contador_de_voltas=0

    if total[x]!='-':

        try:

            novo = 'Bearer <'+access_token+'>'
            link = 'https://api.mendeley.com/catalog/?doi='+total[x]+'&view=stats'

            z = requests.get(link,
                             headers = { 'Authorization': novo })
            z = z.json()

            if z!=[]:

                z = str(z).replace("'", "")
                z = z[1:-1]
                z = str(z).replace("True","True")
                z = str(z).replace("False","False")

                if "reader_count" in z:
                    dect = str(z).find('reader_count')
                    dectfinal=str(z).find('reader_count_by_user_role')

                    if dect > 0:

                        nova = str(z)[((dect)-1):((dectfinal)-3)]
                        nova='{'+ nova+'}'

                    h = json.loads(nova)

```

```

h = json.loads(nova)

quantidade_de_leitores=(h['reader_count'])
status_academico =(h['reader_count_by_academic_status'])

if int(quantidade_de_leitores) == 0:

    for contagem in range (0,13):

        if status[contagem] in status_academico:
            lista[contagem].append('0')

        else:

            for contagem in range (0,13):

                if status[contagem] in status_academico:
                    lista[contagem].append(status_academico[status[contagem]])
                else:
                    lista[contagem].append('0')

            leitores.append(quantidade_de_leitores)

        else:
            leitores.append('-')
            for contagem in range (0,13):
                lista[contagem].append('-')

    except:

        leitores.append('-')
        for contagem in range (0,13):
            lista[contagem].append('-')

    else:
        leitores.append('-')

        for contagem in range (0,13):
            lista[contagem].append('-')
            #print(leitores[x])

output.clear()
total_resultado.insert(loc=1, column='Quant Leitores/Mendeley',value=leitores)

for contagem in range (0,13):
    total_resultado.insert(loc=(2+contagem), column=status[contagem],value=lista[contagem])

#total_resultado
total_resultado.to_excel (r'tabela_mendeley.xlsx', index = False, header=True)
print("Coleta Completa \nArquivo Baixado")
!rm $fn
files.download('tabela_mendeley.xlsx')

```

Fonte: De autoria própria (2022).

#### 4.5. LEVANTAMENTOS REALIZADOS PARA EXTRAIR DADOS DA API - CROSSREF

Conforme a Documentação referente a API da Base Crossref, que pode ser acessada pelo endereço eletrônico disponível no Link: [https://api.Crossref.org/swaggerui/index.HTML#/Works/get\\_works\\_\\_doi\\_](https://api.Crossref.org/swaggerui/index.HTML#/Works/get_works__doi_) utilizamos um método GET e o DOI como parâmetro. São extraídos somente o campo “cited\_by” via API Crossref, referente a quantidade de pessoas que citaram o trabalho científico representado pelo DOI em questão.

Para facilitar o acesso à API do Crossref já existe uma biblioteca disponibilizada pela própria Crossref, chamada Habanero na qual por meio do módulo *counts* é possível obter uma contagem de citações de um DOI.

Utilizando essa biblioteca tornou-se possível extrair a quantidade de citações que cada DOI possui. O Response Code da API da Crossref conta com dois códigos para indicar quando o DOI buscado foi ou não identificado. Respective códigos estão apresentados no Quadro 6.

Quadro 6 - Response Code API Crossref

Código	Descrição
200	Doi was identified
404	Doi was not identified

Fonte: Documentação da API do Crossref:

[https://api.Crossref.org/swaggerui/index.HTML#/Works/get\\_works\\_\\_doi\\_](https://api.Crossref.org/swaggerui/index.HTML#/Works/get_works__doi_)

Em caso de o Status Code ser 200, significará que houve sucesso na busca dos dados, caso contrário será retornado a mensagem de “erro” ao usuário.

Caso o Status Code retornado for diferente de 200 indicando assim que houve algum tipo de falha na busca da informação, será inserido na tabela o sinal hífen “-”, ao invés dos dados referentes aos leitores.

Para realizar a busca, o usuário deve fazer o upload de uma lista de DOIs que deseja extrair as citações em uma tabela no formato .xlsx. Nesta, o título da coluna com os Identificadores DOIs deve ter o número “0” (zero). Caso o usuário deseje pular a linha ele deve inserir o sinal de hífen “-” ao invés do DOIs.

#### 4.6. DESENVOLVIMENTO DO CÓDIGO DE EXTRAÇÃO DE DADOS CONFORME API DA BASE CROSSREF

Para extrair os dados da Api do Crossref iremos por meio de um ciclo coletar cada DOI do arquivo .xlsx e enviar como parâmetro para a API Crossref por meio do módulo 'counts.citation' na qual irá retornar o número de contagens de citações para cada DOI, ao final os resultados serão armazenados em uma planilha que ao final do processamento será gerado um arquivo .xlsx com os dados extraídos. Na Figura 19 o código desenvolvido para a extração dos dados pode ser visualizado.

Figura 19 - Código para extração da API Crossref

```
def crossref():

    total=[]
    for x in range(quant_inicio,quant_linhas):

        total.append(df[0].loc[x])

    y=[]
    output.clear()

    for x in trange (quant_inicio,quant_linhas,desc='Coletando' ):

        if total[x]!='-':

            try:
                valor = counts.citation_count(doi =str(total[x]))

                y.append(valor)
            |
            except:
                y.append('-')

        else:
            y.append('-')

    output.clear()
    total_resultado.insert(loc=1, column = 'cited by/crossref',value=y )
    total_resultado.to_excel (r'tabela_crossref.xlsx', index = False, header=True)
    print("Coleta Completa \nArquivo Baixado")
    !rm $fn
    files.download('tabela_crossref.xlsx')
```

Fonte: De autoria própria (2022).

Caso o código obtenha sucesso, será feito download automático dos dados em um documento no formato '.xlsx' com o nome 'tabela\_Crossref.xlsx'.

#### 4.7. DISPONIBILIZAÇÃO DOS CÓDIGOS DA ALTMETRIC, MENDELEY E CROSSREF EM UMA MESMA INTERFACE GRÁFICA

Após o desenvolvimento os três códigos para extração dos dados em cada fonte de dados/base (Altmetric, Mendeley e Crossref), eles foram disponibilizados em uma mesma interface gráfica utilizando-se o Google Colab. Para isso, primeiramente inserimos um botão do tipo Select, para listar o nome das três bases, para que o usuário possa selecionar de qual base deseja extrair os dados. A Figura 20 apresenta o código utilizado para a inserção do botão Select que o usuário deverá acionar no momento de coletar os dados desejados.

Figura 20 - Inserindo botão Select para escolha da base de Dados desejada

```
Base = 'Altmetrics' #@param ["Mendeley","Crossref","Altmetrics" ]

if Base=='Crossref':
    crossref()
elif Base=='Mendeley':
    mendeley()
else:
    altmetric()
```

Fonte: De autoria própria (2022).

Para que o código em Python da Figura 20 ficasse em segundo plano, foi desenvolvido o seguinte Script: #@title #####Selecione a Base de dados. {display-mode: "form"}. Isso proporcionou a exibição de uma interface gráfica mais amigável ao usuário. A Interface gráfica disponibilizada no Google *Colab* pode ser visualizada na Figura 21. Nesta, a ferramenta Odisseia fornecerá acesso as três bases de dados a partir de uma interface amigável única, na qual o usuário poderá selecionar uma base que deseja extrair os dados.

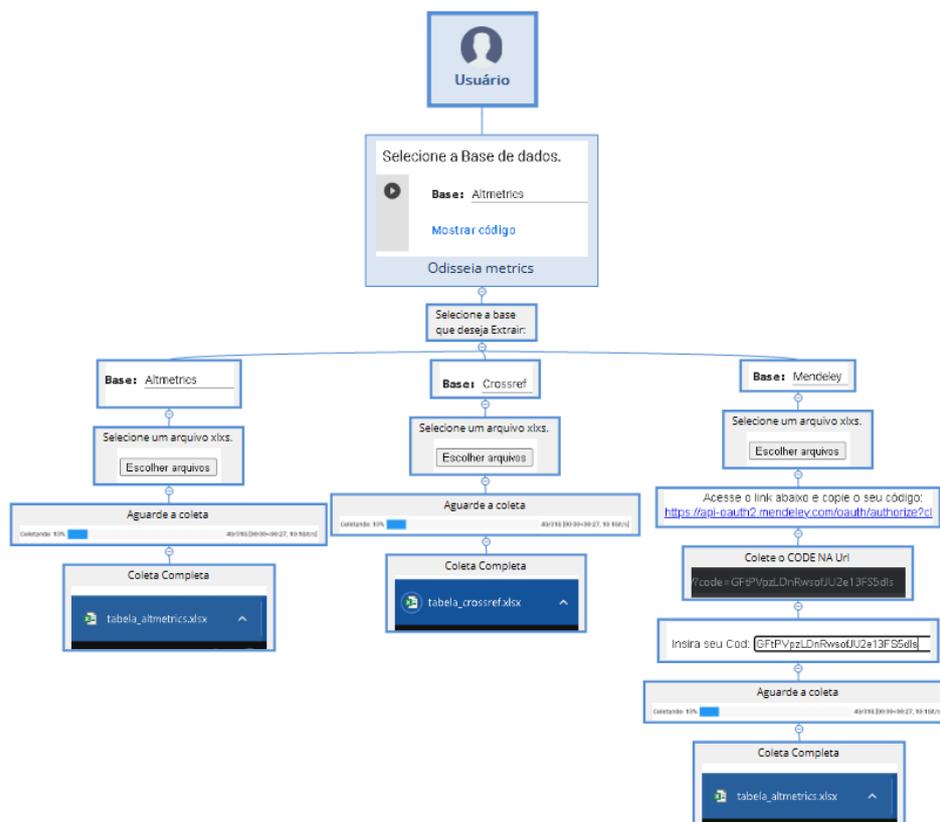
Figura 21 - Interface gráfica no Google Colab para extração dos dados da Altmetric, Mendeley e Crossreff



Fonte: De autoria própria (2022).

#### 4.8. Estrutura da Ferramenta Odisseia

Figura 22 - Estrutura da Ferramenta Altmetrics



Fonte: De autoria própria (2022).

## 5. UTILIZAÇÃO E TESTE DA FERRAMENTA

Primeiramente será explicado e exemplificado o processo necessário para extração de dados de cada uma das bases de dados, começando pelo Mendeley e em seguida Crossref e Altmetric. Sempre que necessário são fornecidos também Links para acessar material de apoio. Posteriormente, serão apresentados resultados de testes mais robustos, com uma lista de DOIs de artigos extraídos da Plataforma Sucupira.

### 5.1. EXTRAINDO DADOS DA ALTMETRIC, MENDELEY E CROSSREF VIA ODISSEIA

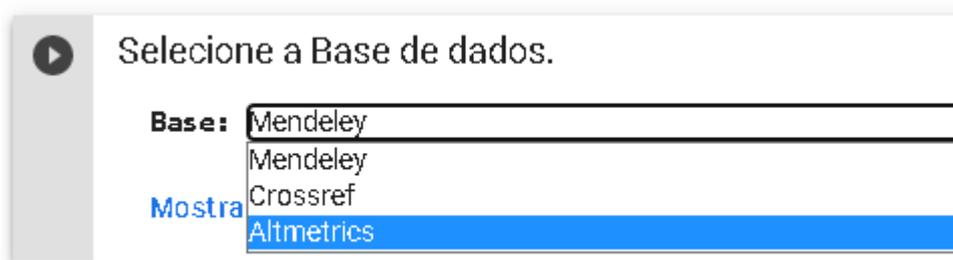
A ferramenta Odisseia permite extrair dados da Altmetric, Mendeley e Crossref a partir de uma lista de DOIs que o usuário deverá inserir na ferramenta. A extração ocorre por meio de API REST utilizando cada DOI individualmente como parâmetro. Os pré-requisitos necessários para se extrair dados da Altmetric, Mendeley e Crossref são: estar logado em uma conta do Google (para poder acessar o Google Colab) e possuir um arquivo .xlsx com uma lista de DOIs que se deseja extrair. O Mendeley é a única fonte de dados que pede informações complementares, ou seja, o usuário deve possuir conta Mendeley (para coletar o *Token* de usuário).

Primeiramente serão apresentados os procedimentos comuns a cada uma das fontes de dados (Altmetric, Mendeley e Crossref) e, posteriormente, são apresentadas características específicas do Mendeley para a extração dos dados. Por último, são apresentados exemplos de planilhas com exemplo dos resultados da extração para cada uma das fontes de dados.

Para todas as fontes de dados, o usuário deverá acessar o seguinte endereço eletrônico do Google *Colab*, link este que dá acesso ao Odisseia *Metrics*: [https://colab.research.google.com/drive/1L75mJbqteVE0kMoTMK1\\_kLVXWFZpQumi](https://colab.research.google.com/drive/1L75mJbqteVE0kMoTMK1_kLVXWFZpQumi)

. Na página será necessário selecionar uma das bases de dados que se deseja extrair os dados, conforme pode ser visualizado na Figura 22.

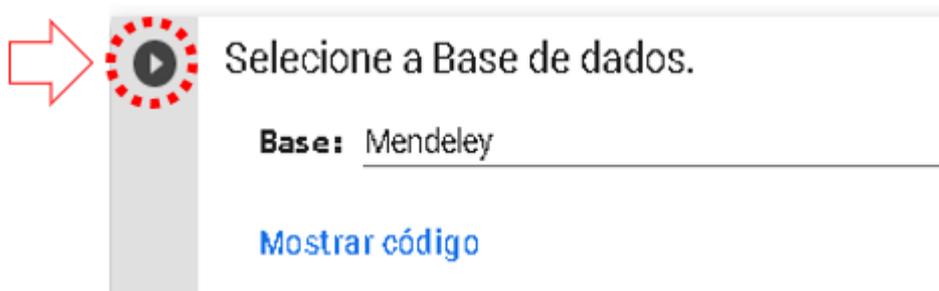
Figura 23 - Selecionando a base Altmetric, Mendeley ou Crossref na ferramenta Odisseia



Fonte: De autoria própria (2022).

Após selecionar o usuário ter selecionado de qual base de dados deseja extrair os dados altmétricos (no exemplo da Figura 23 foi o Mendeley), o usuário deve executar a aplicação, clicando no botão “*play*” (Figura 23).

Figura 24 - Botão que deve ser executado para a extração de dados da Altmetric, Mendeley ou Crossref



Fonte: De autoria própria (2022).

Após a inicialização (Figura 23) será apresentado ao usuário um botão do tipo “Choose File” escrito “Escolher Arquivo” na qual irá permitir que o usuário selecione um arquivo no formato .xlsx (conforme Figura 24).

Figura 25 - Inserindo Arquivo no formato. xlsx. Para extração de dados da Altmetric, Mendeley, Crossref



Fonte: De autoria própria (2022).

Esse arquivo deve possuir uma coluna identificada como “0” e as linhas subsequentes devem possuir DOIs. Caso se deseje que a linha seja pulada ou ignorada, basta inserir um hífen “-” ao invés do DOI (exemplificado na Figura 25). Uma tabela em formato .xlsx, contendo lista de DOIs de exemplo, pode ser baixada em: [https://drive.Google.com/file/d/1Avuup2by2lkhc2BFXOCdLye\\_60FzQOID/view](https://drive.Google.com/file/d/1Avuup2by2lkhc2BFXOCdLye_60FzQOID/view).

Figura 26 - Exemplo de como a tabela deve ser formatada para a extração dos dados da Altmetric, Mendeley, Crossref

	A	B	C	D
1	0			
2	-			
3	-			
4	-			
5	10.18226/23190639.V5N2.02			
6	10.18226/23190639.V5N2.02			
7	10.21714/2178-8030GEP.V19.4687			
8	10.21714/2178-8030GEP.V19.4687			
9	-			
10	10.21171/GES.V13I34.2346			

Fonte: De autoria própria (2022).

Após inserir a tabela, caso a aplicação retorne uma mensagem “Erro!” reinicie a página e tente novamente e verifique se a tabela está no formato adequado. Caso a aplicação ocorra de forma correta, o próximo passo será fazer uma autenticação no Mendeley para que seja possível a extração de dados. No caso da extração de dados da Altmetric e Crossref não é solicitado nenhum código

#### 5.1.1. Autenticação, obtenção e uso do token no Mendeley

Primeiramente o usuário deve realizar a autenticação, que tem a finalidade de identificar a conta de quem está coletando estes dados. Portanto, é necessário que o usuário possua uma conta no Mendeley. Será apresentado um link na qual o usuário será redirecionado para o site do Mendeley (Figura 26). Após a autenticação, será disponibilizado um Token que deverá ser inserido no Odisseia (Figura 26 à frente de onde aparece “Insira seu cod:”).

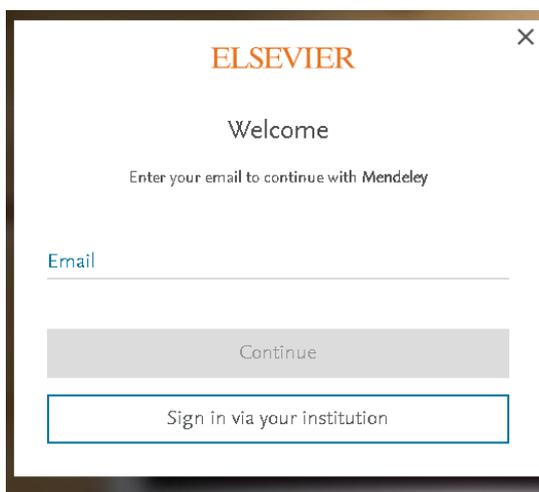
Figura 27 - Url direcionando para autenticação de conta Mendeley

Acesse o link abaixo e copie o seu código:  
[https://api-oauth2.mendeley.com/oauth/authorize?client\\_](https://api-oauth2.mendeley.com/oauth/authorize?client_)  
Insira seu Cod:

Fonte: De autoria própria (2022).

Ao clicar no *link*, após ser direcionado para o site do Mendeley (Figura 27), será solicitado o e-mail e Senha da conta Mendeley do usuário. Vale ressaltar que a ferramenta Odisseia não coleta de dados pessoais do usuário.

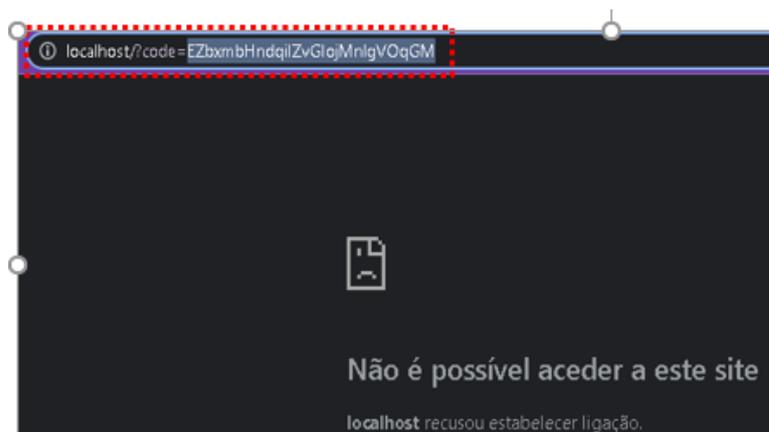
Figura 28 - Tela de Login Mendeley que será apresentada ao Usuário



Fonte: De autoria própria (2022).

Após o usuário preencher com o e-mail e senha (Figura 27) ele será direcionado para uma página do tipo "localhost", semelhante à da Figura 28. Aqui vale observar que o token que precisamos, será disponibilizado ao final do link da página *localhost*, logo após o sinal de igual. "=". (conforme destacado na Figura 28).

Figura 29 - Token Mendeley ao final da Url



Fonte: De autoria própria (2022).

O usuário deverá copiar o *token* ao final da URL e inseri-lo no Odisseia, pois será por meio desse token que o Mendeley irá identificar o usuário e irá autorizar a extração de dados. Esse código deverá ser inserido (colado), na caixa de texto apresentada na Figura 26 (à frente de “Insira seu cod:”).

Após inserir no Odisseia bastará apertar a tecla “Enter” (Figura 26) que a ferramenta irá começar a extrair os dados e também irá permitir que o usuário acompanhe o processo por meio de uma barra “Loading.” (Figura 29). Quando a barra de Loading finalizar, o arquivo com os dados extraídos será baixado automaticamente em formato .xlsx.

Figura 30 - Dados sendo extraídos do Mendeley



Fonte: De autoria própria (2022).

Após a o *download* automático da tabela ocorrer, será possível verificar os dados extraídos. No caso do Mendeley haverá na primeira coluna a lista de DOIs originalmente inserida pelo usuário e nas demais colunas estarão os dados de leitores Mendeley (totais e segmentados por categorias profissionais ou de atuação), conforme Figura 30.

Figura 31 - Exemplo de dados resultantes da extração do Mendeley

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
DOI	Leitores/Me	Other	Professor	Associate	Researcher	Student > Bach	> Doctoral	Student > Mat	> Ph. D. St	> Postgr	Lecturer	> Senior	Librarian	Inspecifi
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10.18226/-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10.18226/-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10.21714/-	39	0	2	0	0	7	1	12	1	2	0	0	1	0
10.21714/-	39	0	2	0	0	7	1	12	1	2	0	0	1	0
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Fonte: De autoria própria (2022).

Quando uma linha que possui DOI tiver como retorno o sinal de hífen “-” , isso pode significar: DOI inválido ou o trabalho científico referente a esse DOI não está indexado no Mendeley. Atualizações futuras estão sendo planejadas para permitir um retorno mais exato nesse quesito, permitindo que o usuário entenda com melhor precisão quando houver cada um dos casos citados acima.

#### 5.1.2. Extraindo dados da Crossref via odisseia

A Ferramenta permite fazer uma requisição do tipo GET na API do Crossref, inserindo o DOI como parâmetro. Diferentemente do Mendeley, para extrair esses dados não será necessário autenticação do usuário. Os pré-requisitos para se extrair dados da Crossref são: estar logado em uma conta Google (para poder acessar o Google Colab) e um arquivo .xlsx com uma lista de DOIs.

Os procedimentos iniciais utilizados para a extração dos dados do Mendeley são idênticos para todas as fontes de dados. Portanto, não serão aqui repetidos. Os passos apresentados nas Figuras 22, 23, 24 e 25, devem ser realizados para cada uma das bases de dados que se tenha interesse em extrair os dados.

Ao serem realizados os procedimentos, os dados de citação da Crossref serem baixados automaticamente. Como os dados de citação são diferentes dos de leituras/capturas do Mendeley um exemplo é apresentado na Figura 31. Nesta Figura, há uma coluna referente ao DOI buscado e outra coluna referente à quantidade de citações que esse DOI teve. Infelizmente a ferramenta não é capaz de extrair os dados de citação por ano.

Figura 32 - Exemplo de dados resultantes da extração da Crossref

0	cited by/crossref
-	-
-	-
-	-
10.18226/:	1
10.18226/:	1
10.21714/:	0
10.21714/:	0
-	-

Fonte: De autoria própria (2022).

### 5.1.3. Extraíndo dados da Altmetric via odisseia

A Ferramenta também permite fazer uma requisição do tipo GET na API da Altmetric, inserindo o DOI como parâmetro. Os pré-requisitos para extração de dados da Altmetric são os mesmos utilizados para a extração e dados da Crossref, ou seja, estar logado em uma conta Google (para poder acessar o Google Colab) e um arquivo .xlsx com uma lista de DOIs.

Os procedimentos de seleção da base de dados e realização do upload do arquivo são os mesmos apresentados nas Figuras 22, 23, 24 e 25. Após a realização dos procedimentos, quando a barra de Loading finalizar, o arquivo com os dados extraídos será baixado automaticamente em formato .xlsx., à exemplo do apresentado na Figura 32. Para cada DOI, os dados altmétricos resultantes serão apresentados nas respectivas colunas, com metadados título, nome da revista, autores, tipo de documento e se o mesmo é *open access* e respectivos indicadores altmétricos.

Figura 33 - Exemplo de dados resultantes da extração da Altmetric

	A	B	C	D	E	F	G	H	
	0	title	journal	authors	type	is_oa	by_msm	by_fbwalls	by_vi
2	10.1590/E:	ALGUNS D	Educação	['Eliana A]	article	FALSO	-	-	-
3	10.1016/J.	Class A β-	Journal of	['Rebeca F	article	FALSO	-	-	-
4	10.15560/:	Freshwater	Check List	['Maria A	article	FALSO	-	1	-
5	10.1007/S:	Interactio	Education	['Marcelo	article	FALSO	-	-	-
6	10.4000/C	Território:	Confins	['Ricardo C	article	FALSO	-	-	-
7	10.1002/B	The dilem	BUSINESS	['Patrícia C	article	FALSO	-	-	-

Fonte: De autoria própria (2022).

Figura 34 - Exemplo de dados resultantes da extração da Altmetric

<b>o</b>	10.1002/BSD2.10	10.4000/CONFINS.13980
<b>title</b>	The dilemma of environme	Territórios em disputas
<b>journal</b>	BUSINESS STRATEGY AND D	Confins
<b>authors</b>	['Patrícia Gonçalves Roque]	['Ricardo Gilson da Cost
<b>type</b>	article	article
<b>is_oa</b>	FALSO	FALSO
<b>cited_by_msm_count</b>	-	-
<b>cited_by_fbwalls_count</b>	-	-
<b>cited_by_videos_count</b>	-	-
<b>cited_by_rdt_count:</b>	-	-
<b>cited_by_feeds_count</b>	-	-
<b>cited_by_gplus_count</b>	-	-
<b>cited_by_posts_count</b>		1 3
<b>cited_by_tweeters_count</b>		1 2
<b>readers&gt;citeulike</b>	0	0
<b>readers&gt;mendeley</b>	22	6
<b>readers&gt;connotea</b>	0	0
<b>readers_count</b>		22 6

Fonte: De autoria própria (2022).

## 5.2. TESTE DA FERRAMENTA

A princípio foram extraídos da Plataforma Sucupira da CAPES dados de cerca de 400 mil artigos científicos publicados por pesquisadores brasileiros associados a programas de pós-graduação stricto sensu no Brasil, abrangendo o período de 2017 à 2018, contudo nem todos os registros possuíam DOI's e muitos possuíam DOI's inoperantes, então se fez necessário filtrar os registros para obtermos uma lista de artigos somente com DOI's funcionais. Após remover os artigos com DOI's ausente, os DOI's restantes foram testados via API Rest no DOI.ORG, na qual se utiliza o DOI como parâmetro e retorna como resultado se ele é funcional ou não. Ao final se obteve 320.749 artigos científicos publicados por pesquisadores brasileiros, extraídos da Plataforma Sucupira da CAPES com DOI's válidos e funcionais. Desses dados foram coletados o identificador DOI de cada um e inseridos em um arquivo xlsx.

Os testes tem o objetivo de extrair a quantidade de leitores do Mendeley, a quantidade de Citações Crossref e por último a quantidade citações no Twitter do Altmetrics utilizando o Odisseia *Metrics*. O primeiro passo foi extrair a quantidade de leitores, logo em seguida o de citações e, por último, a quantidade de citações no Twitter pois não é possível extrair das 3 bases de dados ao mesmo tempo.

A coleta de leitores do Mendeley levou cerca de 8 horas e 34 minutos para ser completada. Ao final foi gerado um arquivo. xlsx possuindo a quantidade de leitores para cada DOI e as categorias de leitores (estudante, bibliotecário, professor, aluno de mestrado, etc.). Os 320.749 artigos científicos resultaram em 837.281 de leitores, sendo que a média foi de 26,10 leitores por DOI.

A coleta de Citações Crossref levou cerca de 6 horas e 12 minutos para ser completada, na qual foi gerado um arquivo. xlsx possuindo a quantidade de citações para cada DOI. Os 320.749 artigos científicos resultaram em 209.931 de citações, sendo que a média foi de 6,55 citações por DOI.

Devido à algumas restrições de limite de extrações na API do Altmetrics (na qual esse limite não é fornecido pela documentação da API) utilizamos uma amostra menor de dados, sorteando 80.000 DOI s dentre os 320.749. A coleta de Citações no Twitter levou cerca de 11 hora e 54 minutos para ser completada, na qual foi gerado um arquivo. xlsx possuindo a quantidade de citações no Twitter para cada DOI. Os 80.000 DOIs resultaram em 8.784 menções no Twitter, sendo que a média foi de 0,11 citações por DOI, não foi identificado erros na busca do montante, todos os DOI's que foram utilizados eram funcionais e por conseguinte retornaram o valor esperado. Caso o limite de coleta da API do Altmetrics seja atingido a planilha gerada ao final irá corresponder somente até o último DOI coletado, os demais DOI's serão substituídos por "-".

Os testes executados mostram que a ferramenta consegue extrair de forma efetiva uma lista de DOI nas três bases de dados, provando assim que é eficiente para o que ela foi proposta, contudo vale ressaltar que cada base tem as suas características e forma de funcionamento.

Alguns trabalhos foram desenvolvidos utilizando o Odisseia Metrics, entre eles o desenvolvido pelos autores Pontes, Maricato e Silva (2022) desenvolveram um estudo para medir a correlações entre indicadores de citação e altmétricos, em relação à países e idiomas. O estudo teve como base, dados de 22.170 artigos extraídos da base Scopus, entre os anos de 2017 a 2021. Por meio da ferramenta Odisseia foi possível a extração dos indicadores de citação (Crossref) e também de metadados advindos do Mendeley. O estudo teve como conclusão que as correlações mais altas são advindas dos artigos em inglês, tendo como exceção as correlações entre

Crossref e Twitter que obtiveram uma correlação mais alta o idioma Português. Os autores não reportaram erros de coleta e processamento de dados ao utilizarem a ferramenta Odisseia.

Outro trabalho que se utilizou da ferramenta Odisseia foi a Dissertação de mestrado da autora Alves (2022) que buscava identificar o impacto bibliométrico e altmétrico em artigos e preprints relacionados à COVID-19. O trabalho da autora teve como base dados de 6.523 preprints e 83.477 artigos de periódicos, dos anos 2020 e 2021, coletados na base de dados Dimensions, utilizando termos relacionados à COVID-19. A ferramenta Odisseia nesse trabalho foi utilizado para consultar dados da API do Mendeley e do Altmetrics para os 90.000 registros, ao qual estes dados foram utilizados para efetuar os cálculos dos indicadores bibliométricos e altmétricos. A ferramenta possibilitou a coleta de quantidade de Citações, Leitores/capturas Mendeley e Menções no Twitter. Os autores também não relataram nenhuma dificuldade ao utilizarem a ferramenta Odisseia para a extração dos dados.

A ferramenta Odisseia também foi utilizada para o desenvolvimento do projeto de Iniciação científica na UnB (RAMOS, 2021), ao qual se teve como base 320.749 artigos publicados por pesquisadores brasileiros, no período entre 2017-2018, disponibilizados na plataforma Sucupira da CAPES. O objetivo do trabalho foi medir a correlação entre citações abertas do Crossref e leitores Mendeley em diferentes campos do conhecimento no contexto da ciência brasileira. A ferramenta foi utilizada para extrair a quantidades de citações e quantidades de leitores Mendeley que esses trabalhos tiveram no decorrer de 2017-2018 e com isso possibilitar medir a correlação entre citações e leitores dos artigos brasileiros.

O trabalho desenvolvido por Maricato e Demachk (2022) analisou 11.955 artigos publicados no Portal de Periódicos da Universidade Federal de Goiás com o objetivo de se obter a cobertura de fontes de dados bibliométricos e altmétricos e as correlações entre indicadores de citação bibliométricos e indicadores altmétricos. A ferramenta Odisseia foi utilizada para extrair dados da plataforma Altmetric, o número de leitores do Mendeley e o número de citações no Crossref dos 11.955 artigos. Não foi mencionado nenhuma dificuldade ou erro na utilização da ferramenta para a extração dos dados.

Conforme descrito nos procedimentos metodológicos, a última etapa da pesquisa foi a realização de um teste de veracidade da API com objetivo de verificar se os dados estavam sendo extraídos corretamente pela Odisseia *Metrics*. Da lista de 200 DOI's aleatórios coletados na Plataforma Sucupira, utilizados para extração dos dados nas três fontes de dados, 30 foram selecionados (aleatoriamente), para análise manual. Estes foram inseridos manualmente um a um na API do Mendeley, Crossref e Altmetric.com e os seus dados foram comparados aos resultados extraídos pela Odisseia *Metrics*. Constatou-se que entre os dados coletados manualmente (dados de 30 DOIs) e os extraídos pela ferramenta Odisseia foram idênticos. Ou seja, da amostra selecionada, chegou-se 100% de veracidade da ferramenta, indicando que a ferramenta é confiável para extração de dados das três fontes.

Observa-se que a ferramenta se mostrou eficiente, confiável e adequada para o desenvolvimento de trabalhos altmétricos e bibliométricos, possibilitando a extração de dados de forma fácil e rápida, auxiliando pesquisadores a desenvolverem seus estudos, sem a necessidade de domínio de linguagens de programação.

## **6. CONSIDERAÇÕES FINAIS**

Percebe-se que a Ferramenta Odisseia permite que pesquisadores e estudantes, que até então precisavam ter um domínio em tecnologia e programação, não necessitem ser especialistas na área de Tecnologia da Informação e, muito menos, que saiba programar para conseguir extrair massas de dados e desenvolver estudos baseados nas fontes e bases de dados Altmetric.com, Mendeley e Crossref. A ferramenta foi testada com um grande volume de dados, demonstrando ser funcional e consistente.

A utilização do Google *Colab* facilitou o desenvolvimento e a implementação da ferramenta, por meio de uma interface amigável e de fácil compreensão. A escolha do Google *Colab* facilita a inserção, a extração e a recuperação desses dados, sendo grande parte desse processo automática e relativamente veloz. A ferramenta também

já trata os dados de saída, necessitando do usuário apenas a adaptação dos dados de acordo com suas necessidades de pesquisa.

Foi possível perceber que a ferramenta cumpre adequadamente com o seu propósito, que é possibilitar que qualquer pessoa consiga extrair dados de forma fácil, mesmo que não tenha conhecimentos avançados de Tecnologias da Informação e hardwares potentes. É necessário que a ferramenta seja utilizada por outros pesquisadores para testá-la constantemente. É possível que as fontes e bases de dados utilizadas para extração de dados modifiquem suas metodologias de disponibilização de dados via suas API, assim, futuramente pode haver necessidade de atualização da ferramenta.

A Odisseia *Metrics*, naturalmente, pode ser aperfeiçoada e ampliada futuramente. Aqui serão abordadas algumas futuras melhorias que poderão ser implementadas paulatinamente na ferramenta. Grande parte das melhorias sugeridas se referem à velocidade de extração e precisão dos resultados.

Uma melhoria importante na ferramenta seria que salvasse os dados no google driver, sem a necessidade de que seja feito o download automático para o computador pessoal. Essa funcionalidade poderia facilitar para o usuário executar a busca da lista de DOI's no Odisseia sem a necessidade de acompanhar a extração, possibilitando assim que ele consiga o resultado da extração posteriormente diretamente pelo Google Driver.

Para ganho de tempo, poderia ser incluída na ferramenta alguma opção que possibilitasse a extração em mais de uma base simultaneamente, ou seja, que a API REST funcionasse de maneira simultânea. Atualmente a ferramenta busca em uma base por vez. Observa-se que a limitação de processamento dos dados está mais relacionada às próprias bases de dados do que no Google *Colab*. Assim, a busca simultânea poderia proporcionar uma busca mais ágil.

Por vezes quando o Odisseia não conseguir executar alguma tarefa, será retornado ao usuário a mensagem “Erro!” impossibilitando que o usuário identifique o motivo do erro e conseqüentemente não auxiliando o usuário a encontrar a solução. Seria importante retornar Informações sobre o erro que o usuário teve e quais caminhos ele deve tomar para que obtenha sucesso na execução da ferramenta.

Por fim, seria relevante que outras bases e fontes de dados alométricas e bibliométricas fossem incorporadas no Odisseia *Metrics*. Em relação à indicadores, observou-se que as fontes de dados não informam dados em relação à data de menção, leitura/captura e citação. Seria muito bom se houvesse alguma atualização das APIs por parte das fontes e/ou que fossem criados mecanismos diretamente na Odisseia para que esta limitação fosse resolvida.

## REFERÊNCIAS

- AKBARITABAR, A.; STAHLSCHMIDT, S. Merits and Limits: Applying open data to monitor open access publications in bibliometric databases. 13 fev. 2019. DOI [10.48550/arXiv.1902.03937](https://doi.org/10.48550/arXiv.1902.03937). Disponível em: <http://arxiv.org/abs/1902.03937>. Acesso em: 9 set. 2022.
- ALMEIDA, C. C.; GRACIO, M. C. C. Produção científica brasileira sobre o indicador “Fator de Impacto”: um estudo nas bases SciELO, Scopus e Web of Science. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 24, n. 54, p. 62–77, 4 jan. 2019. <https://doi.org/10.5007/1518-2924.2019v24n54p62>.
- ALVES, Larissa de Araújo. **Vantagens de citação e altmétricas em preprints e artigos na temática COVID-19**. Orientador: Prof. Dr. João de Melo Maricato. Brasília, 2022. Dissertação (Mestrado em Ciência da Informação) - Faculdade de Ciência da Informação, Universidade de Brasília, Brasília, 2022.
- ARAÚJO, C. A. A. Bibliometria: evolução histórica e questões atuais. **Em Questão**, v. 12, n. 1, p. 11–32, 10 dez. 2006. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/16>. Acesso em: 09 set. 2022.
- BAR-ILAN, J., Shema, H., & Thelwall (2014). Bibliographic References in Web 2.0. *In*: B. Cronin, & C. Sugimoto (eds.), **Beyond Bibliometrics: Harnessing Multi-dimensional Indicators of Performance** (pp. 307–325). Cambridge, MA: MIT Press.
- CAPES. **Plataforma Sucupira**. 2020. Disponível em: <https://www.capes.gov.br/avaliacao/plataforma-sucupira>. Acesso em: 19 ago. 2020.
- COURA, J. R.; WILLCOXII, L. C. B. Impact factor, scientific production and quality of Brazilian medical journals. **Memórias do Instituto Oswaldo Cruz**, v. 98, p. 293–298, abr. 2003. <https://doi.org/10.1590/S0074-02762003000300001>.
- CROSSREF. **You are Crossref – Crossref**. 2020. Disponível em: <https://www.Crossref.org/>. Acesso em: 19 ago. 2020.
- DOI. **Digital Object Identifier System**. 2020. Disponível em: <https://www.doi.org/>. Acesso em: 19 ago. 2020.
- Érika Demachki & João de Melo Maricato (2022): Coverage of Data Sources and Correlations Between Altmetrics and Citation Indicators: The Case of a Brazilian Portal of Open Access Journals, *Serials Review*, DOI: 10.1080/00987913.2022.2066967
- ESAREY, J.; BRYANT, K. Are Papers Written by Women Authors Cited Less Frequently? **Political Analysis**, v. 26, n. 3, p. 331–334, jul. 2018. <https://doi.org/10.1017/pan.2018.24>.
- FERRAZ, R. R. N.; QUONIAM, L.; MACCARI, E. A. Inovação No Planejamento Anual E Trienal Do Preenchimento Da Plataforma Sucupira : Uso da Ferramenta Computacional Scriptlattes. out. 2014. **17º Seminários em Administração [...]**. São Paulo, Brazil: Universidade de São Paulo, Faculdade de Economia, Administração e

Contabilidade, out. 2014. p. 1–19. Disponível em: <https://hal.archives-ouvertes.fr/hal-01894781>. Acesso em: 9 set. 2022.

FORESTI, Nóris. **Estudo da contribuição das revistas brasileiras de biblioteconomia e ciência da informação enquanto fonte de referência para a pesquisa**. 1989. Dissertação (Mestrado) – Departamento de Biblioteconomia da Universidade de Brasília, UnB, Brasília, 1989.

GOUVEIA, F. C. Almetria: métricas de produção científica para além das citações. **Liinc em Revista**, v. 9, n. 1, 19 maio 2013. DOI [10.18617/liinc.v9i1.569](https://doi.org/10.18617/liinc.v9i1.569). Disponível em: <https://revista.ibict.br/liinc/article/view/3434>. Acesso em: 9 set. 2022.

GROGAN, Denis J. **A prática do serviço de referência**. Brasília: Briquet de Lemos, 1995.

HADDAWAY, N. The Use of Web-scraping Software in Searching for Grey Literature. **Grey Journal**, v. 11, p. 186–190, 9 out. 2015. Disponível em: [https://www.researchgate.net/publication/282658358\\_The\\_Use\\_of\\_Web-scraping\\_Software\\_in\\_Searching\\_for\\_Grey\\_Literature](https://www.researchgate.net/publication/282658358_The_Use_of_Web-scraping_Software_in_Searching_for_Grey_Literature). Acesso em: 09 set 2022.

HAUNSCHILD, R.; BORNMANN, L. Normalization of Mendeley reader counts for impact assessment. **Journal of Informetrics**, v. 10, n. 1, p. 62–73, fev. 2016. <https://doi.org/10.1016/j.joi.2015.11.003>.

HAUSTEIN, S.; LARIVIÈRE, V. **Mendeley as the source of global readership by students and postdocs?** [S. l.: s. n.], 2014.

HAUSTEIN, S.; LARIVIÈRE, V.; THELWALL, M.; AMYOT, D.; PETERS, I. Tweets vs. Mendeley readers: How do these two social media metrics differ? **it - Information Technology**, v. 56, n. 5, p. 207–215, 28 out. 2014. <https://doi.org/10.1515/itit-2014-1048>.

HERNÁNDEZ SAMPIERI, R.; FERNÁNDEZ COLLADO, C.; BAPTISTA LUCIO, P. **Metodologia da pesquisa**. São Paulo: McGraw-Hill, 2006.

HICKS, D. J.; COIL, D. A.; STAHLER, C. G.; EISEN, J. A. Network analysis to evaluate the impact of research funding on research community consolidation. **PLOS ONE**, v. 14, n. 6, p. e0218273, 18 jun. 2019. <https://doi.org/10.1371/journal.pone.0218273>.

JALPERIN. **Mendeley\_oauth2\_example.py.GitHub**. Disponível em: <https://conferencias.unb.br/index.php/iniciacaocientifica/26CICUNB17DF/paper/view/34587>. Acesso em: 25 abr. 2022.

LOPES, C. T. O.; PESCHANSKI, J. A.; SUCA, E. G.; DA PAIXÃO, F. **Citações vs Almetric Attention Score: Uma análise dos artigos do CEPID NeuroMat**. preprint. [S. l.]: SocArXiv, 1 fev. 2022. DOI [10.31235/osf.io/dhja3](https://doi.org/10.31235/osf.io/dhja3). Disponível em: <https://osf.io/dhja3>. Acesso em: 9 set. 2022.

MACIAS-CHAPULA, C. A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, v. 27, p. nd-nd, 1998. <https://doi.org/10.1590/S0100-19651998000200005>.

MARICATO, J. M.; MARTINS, D. L. Almetria: complejidades, desafíos y nuevas formas de medición y comprensión de la comunicación científica en la web social.

**Biblios: Journal of Librarianship and Information Science**, n. 68, p. 48–68, 2017. <https://doi.org/10.5195/biblios.2017.358>.

MATTHIAS, L.; JAHN, N.; LAAKSO, M. The Two-Way Street of Open Access Journal Publishing: Flip It and Reverse It. **Publications**, v. 7, n. 2, p. 23, 3 abr. 2019. <https://doi.org/10.3390/publications7020023>.

MATTOS, M. C. C. M. **Estudos Métricos da Informação**. Uniasselvi, Indaial. 2019. Disponível em: <https://www.uniasselvi.com.br/extranet/layout/request/trilha/materiais/livro/livro.php?codigo=38195>. Acesso em: 5 ago. 2020

MELO, J. H. N. de; TRINCA, T. P.; MARICATO, J. M. Limites dos indicadores bibliométricos de bases de dados internacionais para avaliação da Pós-Graduação brasileira: a cobertura da Web of Science nas diferentes áreas do conhecimento. **Transinformação**, v. 33, p. e200071, 2021. <https://doi.org/10.1590/2318-0889202133e200071>.

MOHAMMADI, E.; THELWALL, M.; KOUSHA, K. Can Mendeley bookmarks reflect readership? A survey of user motivations. **Journal of the Association for Information Science and Technology**, v. 67, n. 5, p. 1198–1209, maio 2016. <https://doi.org/10.1002/asi.23477>.

MOUNCE, R. Open Access and Altmetrics: Distinct but Complementary. **Bulletin of the American Society for Information Science and Technology**, v. 39, 1 abr. 2013. <https://doi.org/10.1002/bult.2013.1720390406>.

MUGNAINI, R.; STREHL, L. Recuperação e impacto da produção científica na era google: uma análise comparativa entre o google acadêmico e a web of science 10.5007/1518-2924.2008v13nesp1p92. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 13, n. 1, p. 92–105, 16 maio 2008. <https://doi.org/10.5007/1518-2924.2008v13nesp1p92>.

NASCIMENTO, A. G. do; BARROS, M. **Altmetria Para Bibliotecários: Guia Prático de Métricas Alternativas Para Avaliação da Produção Científica**. 1ª edição. [S. l.]: Scortecci Editora, 2017.

NORONHA, D. P.; MARICATO, J. de M. Estudos métricos da informação: primeiras aproximações. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, p. 116–128, 24 abr. 2008. <https://doi.org/10.5007/1518-2924.2008v13nesp1p116>.

OLIVEIRA, E. F. T. de. **Estudos Métricos da Informação no Brasil: Indicadores de Produção, Colaboração, Impacto e Visibilidade**. [S. l.]: Editora Oficina Universitária, 2021.

OLIVEIRA, E. F. T. de; GRACIO, M. C. C. Indicadores bibliométricos em ciência da informação: análise dos pesquisadores mais produtivos no tema estudos métricos na base Scopus. **Perspectivas em Ciência da Informação**, v. 16, n. 4, p. 16–28, dez. 2011. <https://doi.org/10.1590/S1413-99362011000400003>.

PIWOWAR, H.; PRIEM, J. The power of altmetrics on a CV. **Bulletin of the American Society for Information Science and Technology**, v. 39, n. 4, p. 10–13, 2013. <https://doi.org/10.1002/bult.2013.1720390405>.

- PONTES, D. P. N.; MARICATO, J. M.; SILVA, M. R. da. Correlações entre indicadores de citação e altmétricos: análise de artigos brasileiros com colaboração internacional. **Encontro Brasileiro de Bibliometria e Cientometria**, v. 8, p. 660–668, 20 jul. 2022. Disponível em: <https://ebbc.inf.br/ojs/index.php/ebbc/article/view/59>. Acesso em: 09 set. 2022.
- RAM, S.; KAUR, B.; INDU I, I. Social Media Metrics of Indian Covid-19 Research: An Altmetric Analysis. **Journal of Young Pharmacists**, v. 13, n. 3s, p. s42–s47, 8 dez. 2021. <https://doi.org/10.5530/jyp.2021.13s.69>.
- RAMOS, T. B. C. Correlação entre citações abertas do Crossref e leitores Mendeley em diferentes campos do conhecimento no contexto da ciência brasileira. *In*: 26º CONGRESSO DE INICIAÇÃO CIENTÍFICA DA UNB E 17º DO DF, 20 jan. 2021. **26º CONGRESSO DE INICIAÇÃO CIENTÍFICA DA UNB E 17º DO DF [...]**. [S. l.: s. n.], 20 jan. 2021. Disponível em: <https://conferencias.unb.br/index.php/iniciacaocientifica/26CICUNB17DF/paper/view/34587>. Acesso em: 9 set. 2022.
- SANT'ANNA, H. C.; ALVES, J. C. R. Análise de dados da Plataforma Sucupira sobre a Pós-Graduação em Design no Brasil (2013-2017): uma primeira aproximação. **Revista de Design, Tecnologia e Sociedade**, v. 5, n. 2, p. 1–18, 30 dez. 2018. .
- SCHENKEL, R. Integrating and exploiting public metadata sources in a bibliographic information system Extended Abstract. *In*: BIR 2018 Workshop on Bibliometric-enhanced Information Retrieval. 2018. Disponível em: <http://ceur-ws.org/Vol-2080/paper2.pdf>. Acesso em: 09 set. 2022.
- SCHENKEL, R. Integrating and exploiting public metadata sources in a bibliographic information system., p. 6, 2018. Disponível em: <http://ceur-ws.org/Vol-2080/paper2.pdf>. Acesso em: 09 set. 2022.
- SIRISURIYA. A Comparative Study on Web Scraping. *In*: IEEE 8th International Conference on Industrial and Information System, 2015. Disponível em: <http://ir.kdu.ac.lk/bitstream/handle/345/1051/com-059.pdf?sequence=1&isAllowed=y>. Acesso em: 09 set. 2022.
- THELWALL, M. Early Mendeley readers correlate with later citation counts. **Scientometrics**, v. 115, n. 3, p. 1231–1240, jun. 2018. <https://doi.org/10.1007/s11192-018-2715-9>.
- THELWALL, M. Webometrics. **Annual review of information ...**, 1 jan. 2005. Disponível em: <https://www.academia.edu/709949/Webometrics>. Acesso em: 9 set. 2022.
- THELWALL, M.; HAUSTEIN, S.; LARIVIÈRE, V.; SUGIMOTO, C. R. Do Altmetrics Work? Twitter and Ten Other Social Web Services. **PLoS ONE**, v. 8, n. 5, p. e64841, 28 maio 2013. <https://doi.org/10.1371/journal.pone.0064841>.
- THELWALL, M.; SUD, P. Mendeley readership counts: An investigation of temporal and disciplinary differences. **Journal of the Association for Information Science and Technology**, v. 67, n. 12, p. 3036–3050, 2016. <https://doi.org/10.1002/asi.23559>.

THELWALL, M.; WILSON, P. Mendeley readership altmetrics for medical articles: An analysis of 45 fields. **Journal of the Association for Information Science and Technology**, v. 67, n. 8, p. 1962–1972, ago. 2016. <https://doi.org/10.1002/asi.23501>.

TORRES-SALINAS, D.; ROBINSON-GARCÍA, N.; ARROYO-MACHADO, W. Coverage and distribution of altmetric mentions in Spain: a cross-country comparison in 22 research fields. **Profesional de la información**, v. 31, n. 2, 4 abr. 2022. DOI [10.3145/epi.2022.mar.20](https://doi.org/10.3145/epi.2022.mar.20). Disponível em: <https://revista.profesionaldelainformacion.com/index.php/EPI/article/view/86801>. Acesso em: 9 set. 2022.

USP. Entenda o que é Acesso Aberto. [s. d.]. **ABCD - Agência de Bibliotecas e Coleções Digitais**. Disponível em: <https://www.abcd.usp.br/apoio-pesquisador/aceso-aberto-usp/entenda-o-que-e-aceso-aberto/>. Acesso em: 9 set. 2022.

VANZ, S. A. de S.; CAREGNATO, S. E. Estudos de Citação: uma ferramenta para entender a comunicação científica. **Em Questão**, v. 9, n. 2, p. 295–307, 2003. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/75>. Acesso em: 09 set. 2022.

WEINSTOCK, Melvin. Citation Index. *In*: KENT, Allen; LANCOUR, Harold (Ed.). **Encyclopedia of Library and Information Science**. New York: M.Dekker, 1971. V. 5, p. 19.ia. Disponível em: <http://garfield.library.upenn.edu/essays/V1p207y1962-73.pdf>. Acesso em: 09 set. 2022.

ZAHEDI, Z.; COSTAS, R.; WOUTERS, P. Mendeley readership as a filtering tool to identify highly cited publications. **Journal of the Association for Information Science and Technology**, v. 68, n. 10, p. 2511–2521, 2017. <https://doi.org/10.1002/asi.23883>.