

UNIVERSIDADE FEDERAL DE SÃO CARLOS  
CENTRO DE EDUCAÇÃO E CIÊNCIAS HUMANAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

RODRIGO CÉSAR PERISSINI

**BIBLIOANALYTICS: aplicação web para democratização de  
análises bibliométricas**

SÃO CARLOS - SP  
2022

RODRIGO CÉSAR PERISSINI

**BIBLIOANALYTICS: aplicação web para democratização  
de análises bibliométricas**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade Federal de São Carlos, para obtenção do título de Mestre em Ciência da Informação.

**Área de concentração:** Conhecimento, Tecnologia e Inovação.

**Linha de pesquisa:** Conhecimento e Informação para Inovação.

**Orientador:** Prof. Dr. Leandro Innocentini Lopes de Faria.

SÃO CARLOS - SP  
2022

**Perissini, Rodrigo César**

**Biblioanalytics: aplicação web para democratização de análises bibliométricas / Rodrigo César Perissini -- 2022. 128f.**

**Dissertação (Mestrado) - Universidade Federal de São Carlos, campus São Carlos, São Carlos**

**Orientador (a): Leandro Innocentini Lopes de Faria**

**Banca Examinadora: Roniberto Morato do Amaral, Ronaldo Ferreira de Araújo**

**Bibliografia**

**1. Biblioanalytics. 2. Bibliometria. 3. Aplicação web. I. Perissini, Rodrigo César. II. Título.**

**Ficha catalográfica desenvolvida pela Secretaria Geral de Informática (Sin)**

**DADOS FORNECIDOS PELO AUTOR**

**Bibliotecário responsável: Ronildo Santos Prado - CRB/8 7325**



**UNIVERSIDADE FEDERAL DE SÃO CARLOS**

Centro de Educação e Ciências Humanas  
Programa de Pós-Graduação em Ciência da Informação

---

**Folha de Aprovação**

---

Defesa de Dissertação de Mestrado do candidato Rodrigo Cesar Perissini, realizada em 27/09/2022.

**Comissão Julgadora:**

Prof. Dr. Leandro Innocentini Lopes de Faria (UFSCar)

Prof. Dr. Roniberto Morato do Amaral (UFSCar)

Prof. Dr. Ronaldo Ferreira de Araújo (UFAL)

O Relatório de Defesa assinado pelos membros da Comissão Julgadora encontra-se arquivado junto ao Programa de Pós-Graduação em Ciência da Informação.

## RESUMO

Com a industrialização da ciência, a mesma passou a ser encarada como fonte produtiva e econômica, exponencializando a quantidade de publicações geradas por pesquisas científicas e gerando uma necessidade de mecanismos para facilitar sua recuperação e organização como as fontes de informação *online*; e também de mecanismos para trabalhar com esse gigantesco conjunto de dados, como é o caso das ferramentas bibliométricas. O problema é que as ferramentas disponíveis apresentam limitações que dificultam a democratização de análises bibliométricas, principalmente ao campo científico brasileiro, como por exemplo a utilização somente da língua inglesa; barreiras financeiras e/ou intelectual, dada a necessidade de instalar e configurar a ferramenta, e também por algumas serem pagas; e entre outras. Essa pesquisa tem como objetivo desenvolver uma ferramenta bibliométrica gratuita para a web com foco na democratização de análises bibliométricas. Para a construção do BiblioAnalytics se utilizou da pesquisa-ação para coleta dos requisitos necessários perante ao NIT/MATERIAIS, que é um grupo de pesquisa e desenvolvimento presente na UFSCar, e da comparação com a ferramenta VantagePoint, quando necessário, para validar o correto funcionamento de algumas de suas funcionalidades. Por fim, após a construção do BiblioAnalytics, pôde ser observado que a ferramenta funciona mesmo em um ambiente gratuito na *web*, e consegue obter resultados relevantes e corretos, apesar de limitações (como a geração de matrizes para um conjunto muito grande de dados), fornecendo mecanismos que contribuem com a democratização de análises bibliométricas, o que faz com que a mesma possa ser utilizada por um conjunto muito maior de pesquisadores e alunos, dado que a mesma pode ser acessada até mesmo por um aparelho celular.

**Palavras-chave:** BiblioAnalytics. Bibliometria. Aplicação *web*. Plataforma Lattes. Web of Science.

## ABSTRACT

With the industrialization of science, it came to be seen as a productive and economic source, exponentially increasing the number of publications generated by scientific research and generating a need for mechanisms to facilitate its retrieval and organization as the online information sources; and also a need for mechanisms to work with this gigantic set of data, such as bibliometric tools. The problem is that the available tools have limitations that make it difficult to democratize bibliometric analyzes, especially in the Brazilian scientific field, such as the use of only the English language; financial and/or intellectual barriers, given the need to install and configure the tool, and also because some are paid; and among others. This research aims to develop a free bibliometric web application with a focus on the democratization of bibliometric analysis. For the construction of BiblioAnalytics, action research was used to collect the necessary requirements within NIT/MATERIAIS, which is a group of research and development in UFSCar, and the comparison with the VantagePoint tool, when necessary, to validate the correct functioning of some of its functionalities. Finally, after the construction of BiblioAnalytics, it could be observed that the tool works even in a free environment on the web, and manages to obtain relevant and correct results, despite limitations (such as the generation of matrices for a very large set of data), providing mechanisms that contribute to the democratization of bibliometric analyzes, which means that it can be used by a much larger group of researchers and students, given that it can be accessed even by a cell phone.

**Keywords:** BiblioAnalytics. Bibliometrics. Web Application. Plataforma Lattes. Web of Science.

## LISTA DE FIGURAS

<b>Figura 1</b> – Visualização de um indicador de ligação onde são mostradas as palavras-chave mais utilizadas pelos autores em uma busca sobre bibliometria na Web of Science. ....	23
<b>Figura 2</b> – Uma edição do livro “O Senhor do Anéis” representada nos formatos JSON e XML.....	26
<b>Figura 3</b> – Exemplo de uma publicação científica no formato RIS. ....	28
<b>Figura 4</b> – Um exemplo de registro telefônico no formato JSON.....	30
<b>Figura 5</b> – Duas publicações científicas em um arquivo no formato ISI-CE. ....	33
<b>Figura 6</b> – Duas publicações científicas em um arquivo no formato RIS modificado disponibilizado pelo syncLattes. ....	35
<b>Figura 7</b> – Aplicação de um arquivo de tesouro em um conjunto de três publicações que contemplam nomes de mesmos autores escritos de maneiras distintas.....	38
<b>Figura 8</b> – Comparação entre total de publicações por ano obtidas pelo projeto syncLattes em comparação com dados do Somos UFSCar .....	41
<b>Figura 9</b> – Índices disponíveis na <i>Web of Science Core Collection</i> por meio do acesso Capes.....	44
<b>Figura 10</b> – Índices disponíveis na <i>Web of Science Core Collection</i> para a Universidade de Manchester por meio do acesso completo.....	44
<b>Figura 11</b> – Visualizando um conjunto de publicações na Web of Science .....	45
<b>Figura 12</b> – Tipos de dados aceitos no <i>software</i> bibliométrico VOSviewer .....	47
<b>Figura 13</b> – Escolhendo uma fonte de dados no <i>software</i> bibliométrico VOSviewer. ....	48
<b>Figura 14</b> – Diferentes tipos de análises oferecidas para dados processados no <i>software</i> VOSviewer. ....	48
<b>Figura 15</b> – Os quatro diferentes tipos de visualizações disponíveis no programa VOSviewer. ....	49
<b>Figura 16</b> – Visão geral da arquitetura de microsserviços implementada no syncLattes	51
<b>Figura 17</b> – Exemplos de relatórios disponíveis ao utilizar a ferramenta scriptLattes.....	53
<b>Figura 18</b> – Uma lista de autores e uma matriz de coocorrência de autores gerados na ferramenta VP. ....	54
<b>Figura 19</b> – Rede de autores criada na ferramenta VP. ....	55
<b>Figura 20</b> – Bibliometrix e o fluxo de trabalho adequado para o mapeamento científico.	57
<b>Figura 21</b> – Diagrama de caso de uso compreendendo os requisitos necessários para a ferramenta BiblioAnalytics. ....	65
<b>Figura 22</b> – Diagrama genérico representando a estrutura tecnológica do BiblioAnalytics .....	73
<b>Figura 23</b> – Um Registro Completo obtido da Web of Science no formato Microsoft Excel contendo várias publicações. ....	78
<b>Figura 24</b> – Exemplo de JSON obtido após utilização da biblioteca convert-excel-to-json e feita a seleção dos campos necessários. ....	79
<b>Figura 25</b> – Exemplo de uma publicação em formato JSON após todos os processamentos gerais terem sido realizados. ....	81
<b>Figura 26</b> – JSON obtido após a realização de todos os tratamentos específicos e a adição do usuário que importou e a data de criação.....	83
<b>Figura 27</b> – Publicação contida em um arquivo fornecido pelo <i>software</i> syncLattes no formato JSON.....	88

<b>Figura 28</b> – Publicação no formato JSON após todos os processamentos feitos pelo BiblioAnalytics .....	89
<b>Figura 29</b> – Exemplo de uma matriz de coautoria .....	93
<b>Figura 30</b> – Telas de login, registro e menu principal da ferramenta BiblioAnalytics. ...	102
<b>Figura 31</b> – Telas de importação, análise e aplicação de tesauro da ferramenta BiblioAnalytics .....	103
<b>Figura 32</b> – Telas de deleção e de documentos e manuais .....	104
<b>Figura 33</b> – Visão técnica integral do BiblioAnalytics .....	105
<b>Figura 34</b> – Tesauro criado para padronização da instituição <i>Chinese Academy of Sciences</i> no BiblioAnalytics .....	110
<b>Figura 35</b> – Comparativo das matrizes de “Document Type” ou “Tipo de Documento” geradas por meio das ferramentas BiblioAnalytics e VP.....	112



## LISTA DE QUADROS

<b>Quadro 1</b> – Alguns dos principais indicadores bibliométricos para análise de publicações científicas.....	23
<b>Quadro 2</b> – Ordem de tagueamento de uma linha no formato no formato RIS. ....	28
<b>Quadro 3</b> – Algumas das principais tags disponíveis no formato RIS.....	29
<b>Quadro 4</b> – As diversas tags disponíveis no formato ISI-CE. ....	32
<b>Quadro 5</b> – Mapeamento dos campos do XML do Lattes utilizados para a composição dos campos Dublin Core Qualificado no DSpace. ....	34
<b>Quadro 6</b> – Representação de uma tabela em formato CSV.....	36
<b>Quadro 7</b> – Índices de química da Web of Science Core Collection.....	42
<b>Quadro 8</b> – Índices de citação da Web of Science Core Colection.....	43
<b>Quadro 9</b> – Ferramentas bibliométricas, tipo de acesso, principais usos e publicações ao longo dos anos.....	58
<b>Quadro 10</b> – Conjunto de validações a serem realizadas no BiblioAnalytics e as metodologias empregadas para geração de resultados.....	63
<b>Quadro 11</b> – As motivações que propiciaram a escolha da Plataforma Lattes e Web of Science como fontes de informação para o BiblioAnalytics.....	74
<b>Quadro 12</b> – Aplicação de uma expressão regular para extração dos autores (AU) de uma publicação no formato ISI-CE. ....	76
<b>Quadro 13</b> – Campos extraídos de um Registro Completo na Web of Science, suas colunas e suas contrapartes no BiblioAnalytics.....	78
<b>Quadro 14</b> – Processamentos gerais realizados pelo BiblioAnalytics nos campos extraídos de uma publicação da Web of Science.....	82
<b>Quadro 15</b> – Comparativo entre os arquivos no formato RIS expondo os metadados oferecidos pelos <i>softwares</i> scriptLattes e syncLattes.....	85
<b>Quadro 16</b> – Campos relevantes extraídos do JSON do syncLattes e sua nova nomenclatura no BiblioAnalytics. ....	86
<b>Quadro 17</b> – Processamentos gerais efetuados em publicações do formato JSON do syncLattes durante uma importação do BiblioAnalytics e os campos em que as mesmas são aplicadas.....	87
<b>Quadro 18</b> – Exemplos de JSONs durante a geração de uma matriz na ferramenta BiblioAnalytics e sua matriz resultante. ....	94
<b>Quadro 19</b> – Exemplo de tesouro em arquivo de texto para as ferramentas VP e VOSviewer.....	96
<b>Quadro 20</b> – Comparação da contagem do campo “Publication Type” para a WoS e PL. ....	107
<b>Quadro 21</b> – Os dez elementos com maior coocorrência para os campos “Organizations” e “Authors” no conjunto de publicações da WoS e da PL após a extração.....	108
<b>Quadro 22</b> – Comparativo da contagem dos campos “Organizations” da WoS e “Authors” da PL entre as ferramentas BiblioAnalytics e o VantagePoint após todos os processamentos na geração de uma lista.....	109
<b>Quadro 23</b> – Comparativo de uma lista de “Organizations” no BiblioAnalytics antes e após a aplicação de um tesouro para a <i>Chinese Academy of Sciences</i> .....	111
<b>Quadro 24</b> – Comparativo de tempo entre a geração de uma lista e de uma matriz.....	115
<b>Quadro 25</b> – Listas e matrizes disponíveis para geração durante a utilização do BiblioAnalytics.....	117

## LISTA DE SIGLAS E ABREVIações

API	<i>Application Programming Interface</i>
AWS	<i>Amazon Web Services</i>
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CSS	<i>Cascading Style Sheets</i>
CSV	<i>Comma-separated values</i>
HTML	<i>HyperText Markup Language</i>
IP	<i>Internet Protocol</i>
ISI	<i>Institute for Scientific Information</i>
ISI-CE	<i>ISI Common Export Format</i>
JSON	<i>JavaScript Object Notation</i>
ORCID	<i>Open Researcher and Contributor ID</i>
PL	Plataforma Lattes
RIS	<i>Research Information Systems</i>
RTF	<i>Rich Text Format</i>
SGBD	Sistema de Gerenciamento de Banco de Dados
SPA	<i>Single Page Application</i>
UFSCar	Universidade Federal de São Carlos
VP	VantagePoint
WoS	Web of Science
XML	<i>Extensible Markup Language</i>

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	13
1.1 Problema de Pesquisa.....	14
1.2 Objetivos gerais e específicos .....	14
1.3 Justificativa.....	15
<b>2 REFERENCIAL TEÓRICO</b> .....	18
2.1 Bibliometria: da obtenção dos dados à visualização da produção científica .....	18
2.1.1 A bibliometria e sua importância .....	18
2.1.2 Obtenção e preparação dos dados .....	20
2.1.3 O processo de contagem .....	21
2.1.4 Indicadores para representação da informação .....	22
2.1.5 Exploração da bibliometria por meio de métricas alternativas .....	24
2.2 Metadados e a interoperabilidade em ambientes digitais.....	25
2.2.1 O formato RIS .....	27
2.2.2 O formato JSON.....	30
2.2.3 O formato ISI-CE.....	31
2.2.4 O formato RIS modificado do syncLattes .....	34
2.2.5 O formato CSV.....	36
2.3 Vocabulários controlados e tesouros .....	37
2.4 Fontes de informação .....	39
2.4.1 Plataforma Lattes .....	40
2.4.2 Web of Science .....	42
2.5 Ferramentas para auxílio em análises bibliométricas .....	46
2.5.1 VOSviewer .....	46
2.5.2 syncLattes.....	50
2.5.3 scriptLattes.....	52
2.5.4 VantagePoint .....	53
2.5.5 Bibliometrix .....	56
2.5.6 Outras ferramentas .....	58
2.6 Democratização do acesso e conhecimento e seu contexto em ferramentas bibliométricas .....	59
<b>3 MATERIAIS E MÉTODOS</b> .....	62
3.1 Abordagem e tipologia de pesquisa.....	62
3.2 Requisitos e viabilização .....	64

3.3 Escopo tecnológico para a construção de uma ferramenta bibliométrica.....	66
3.3.1 Aplicações para <i>web</i> .....	66
3.3.2 <i>Single Page Applications</i> .....	68
3.3.3 Servidor de APIs .....	69
3.3.4 Bancos de dados .....	71
3.3.5 Concatenação de tecnologias e a definição da arquitetura de um <i>web app</i> .....	71
3.4 Fontes de dados: seleção e utilização de fontes de informação online .....	73
3.5 O processo de importação e tratamento de metadados.....	75
3.5.1 Web of Science .....	75
3.5.2 Plataforma Lattes .....	84
3.6 Analisando metadados: construção de listas e matrizes .....	89
3.6.1 Construção de listas.....	90
3.6.2 Construção de matrizes .....	92
3.7 Refinamento das análises por meio da utilização de tesouros.....	95
3.8 Possibilitando análises altimétricas por meio de mecanismos de acesso .....	97
3.9 Mecanismos extras para auxiliar na democratização de análises bibliométricas .....	99
<b>4 RESULTADOS</b> .....	101
4.1 A ferramenta BiblioAnalytics e seu funcionamento .....	101
4.2 Análise das contagens oferecidas pelo BiblioAnalytics por meio de listas e matrizes em diferentes cenários .....	106
4.2.1 Contagens após a extração de metadados .....	106
4.2.2 Contagens após todos os processamentos: a geração de uma lista .....	107
4.2.3 Contagens com a aplicação de um tesouro.....	110
4.2.4 Contagens na geração de uma matriz.....	111
4.3 Validação da ferramenta.....	113
4.3.1 Contexto geral.....	113
4.3.2 Contagens.....	113
<b>5 CONSIDERAÇÕES FINAIS</b> .....	118
<b>REFERÊNCIAS</b> .....	120

## 1 INTRODUÇÃO

A industrialização da ciência deu início a uma época onde a ciência começou a ser encarada como força produtiva e econômica, o que fez com que a quantidade de pesquisas e publicações científicas, juntamente com a quantidade de pesquisadores, crescessem a números jamais presenciados na história (SANTOS, 1978). Com isso, nos dias atuais, a pesquisa científica impacta cada vez mais no desenvolvimento social e econômico, uma vez que ela é utilizada para tentar explicar diversos acontecimentos do dia a dia, se tornando cada vez mais importante (DORTA-GONZÁLEZ; DORTA-GONZÁLEZ, 2010).

Com o aumento da quantidade de produções e cientistas, ocorre também a necessidade de mais meios de publicação (BOURDIEAU, 1983) e, conseqüentemente, a necessidade de se organizá-las e recuperá-las. Tentando amenizar as questões, têm surgido diversas fontes de informação no contexto *web*, sendo alguns exemplos o Google Scholar, a Web of Science (WoS), a Scopus e a Plataforma Lattes (PL) (COBO et. al., 2011). Porém, ainda assim, mesmo que o pesquisador tenha acesso às mesmas, é um processo trabalhoso de conseguir criar reflexões em meio a todas essas opções e seus grandes acervos.

Pesquisas de cunho bibliométrico, e também suas similares como a altimetria, vêm ganhando ainda mais relevância graças a todo esse universo informacional em constante expansão e a capacidade das mesmas em analisar grandes conjuntos de dados fornecidos por fontes de informação, predominantemente online nos tempos atuais, como as citadas anteriormente. Todavia, ao se utilizar puramente a bibliometria o processo pode ser custoso e demorado sendo necessário de se utilizar mecanismos que possam auxiliar nesse tipo de estudo, como ferramentas tecnológicas.

As ferramentas, hoje em dia, estão cada vez mais presentes nesse tipo de estudo (PERISSINI; FARIA, 2021), facilitando, viabilizando e até mesmo democratizando (sobre o prisma de tornar algo popular) estudos que poderiam ser considerados inviáveis. As mesmas, assim como as fontes de informação, apresentam suas próprias seus próprios objetivos (mineração textual, visualização e entre outros) e suas próprias características como por exemplo os formatos de metadados de darão suporte, a presença ou não de mecanismos que facilitem a recuperação e representação da informação como vocabulários controlados e tesouros e também possíveis processamentos a serem realizados no conjunto de

informações processado (MORAL-MUÑOZ et al., 2020).

## 1.1 Problema de Pesquisa

O principal problema da contextualização apresentada é que diversas barreiras são enfrentadas por pesquisadores e alunos na realização de estudos métricos, seja na recuperação ou na análise dos dados, como por exemplo o acesso a uma ferramenta ou fonte de informação, sendo válido ressaltar que o acesso pode ser analisado sob diversos prismas como o financeiro (necessidade uma assinatura ou equipamento necessário para utilizar), por uma questão de conhecimento e interpretação (dificuldade em configurar ou utilizar uma ferramenta) ou até mesmo por uma questão linguística (não compreender o idioma utilizado) (ALMEIDA et al., 2012).

Nesse contexto, a existência, se possível, de uma ferramenta que forneça mecanismos para que a mesma seja acessível a um número ainda maior de pesquisadores, ou até mesmo outras pessoas interessadas nos estudos métricos, é extremamente importante, contribuindo com a popularização e viabilização de estudos que talvez não seriam possíveis.

## 1.2 Objetivos gerais e específicos

Dado o contexto apresentado e suas questões contemporâneas, nesta seção é apresentado o objetivo geral e o objetivo específico do trabalho em questão, onde o objetivo geral tem como foco resumir e apresentar a ideia central do trabalho em questão e os objetivos específicos de apresentar de forma mais detalhada os resultados esperados no desenvolvimento desta dissertação.

Iniciando pelo objetivo geral, o foco central deste estudo é desenvolver uma ferramenta bibliométrica gratuita para a *web* com foco na democratização de análises bibliométricas.

Dado o objetivo geral, tem-se como objetivos específicos desta pesquisa:

- Mapear a literatura tecnocientífica de modo a compreender e revistar temas contemporâneos e relevantes para o contexto deste estudo.
- Propor um modelo sistemático de uma ferramenta capaz de contribuir com a democratização de análises bibliométricas.

- Implementar uma prova de conceito da solução, de modo a atender os objetivos, requisitos e especificações definidos.
- Validar a solução e seus respectivos resultados de modo a certificar o correto funcionamento da mesma e verificar suas particularidades.

### 1.3 Justificativa

A contagem ou medição é tópico de estudo e discussão desde os primórdios da ciência dada sua relevância (MILANEZ, 2015, p. 28). Pensadores e pesquisadores de diferentes épocas, como Horácio (65-5 AC) e Heike K. Onnes (1853-19 DC), mesmo separados por centenas de anos, convergem para o mesmo ponto: a importância dos processos de contagem (FARIA, 2015, p. 1).

Em contextos atuais das áreas da Ciência da Informação, Biblioteconomia e outras áreas da informação, o processo de contagem é parte integrante e indissociável da bibliometria e entre outras. A bibliometria é atualmente um tópico de pesquisa de extrema relevância e aplicabilidade (GUEDES, 2012). Conforme Ravelli *et al.* 2009, a importância dos estudos e reflexões bibliométricas é sustentada principalmente pela necessidade de se conhecer e avaliar a produtividade e qualidade da pesquisa de autores e suas instituições, além de também serem cada vez mais requisitados e utilizados para quantificação da produção e informações relevantes nas mesmas.

Dada a importância da bibliometria, do processos de contagens e das inúmeras bases de dados que estão surgindo, diversas ferramentas (*softwares*) estão sendo desenvolvidas buscando auxiliar no processamento, contagem, visualização e acesso de gigantescas quantidades de dados de modo automatizado, sendo alguns exemplos de *softwares* que atuam nessa temática o Bibliometrix<sup>1</sup> (ARIA; CUCCURULLO, 2017), VantagePoint (VP)<sup>2</sup> (PORTER; CUNNINGHAM, 2004), Gephi<sup>3</sup> (BASTIAN; HEYMANN; JACOMY, 2009), VOSviewer<sup>4</sup> (ECK; WALTMAN, 2009), BibExcel<sup>5</sup> (PERSSON; DANELL; SCHNEIDER, 2009), syncLattes<sup>6</sup> (MATIAS, 2015) e entre

---

<sup>1</sup> <<https://www.bibliometrix.org/>>

<sup>2</sup> <<https://www.thevantagepoint.com/>>

<sup>3</sup> <<https://gephi.org/>>

<sup>4</sup> <<https://www.vosviewer.com/>>

<sup>5</sup> <<https://homepage.univie.ac.at/juan.gorraiz/bibexcel/>>

<sup>6</sup> <<https://github.com/nitmateriais/synclattes>>

muitos outros. O problema é que muitas dessas ferramentas buscam apenas atender ao seu objetivo técnico e não atendem a especificidades de outros grupos, como por exemplo o do pesquisador brasileiro, dificultando sua utilização pelos mesmos.

É possível elencar pontos presentes nas mesmas que podem dificultar sua utilização por pesquisadores em geral, e também de pesquisadores de uma região específica, como por exemplo os brasileiros. São alguns desses pontos: (i) infraestrutura: a necessidade de o pesquisador ter que ter um computador de mesa, muitas vezes potente, para instalar a mesma e conseguir realizar seus estudos, sendo exemplos as ferramentas VP, VOSviewer, Gephi e entre outras; (ii) o licenciamento: algumas ferramentas são pagas (VP, HistCite e entre outras); (iii) fontes de informação: em grande parte oferecem suporte apenas a fontes de informação internacionais como as bases de dados WoS, Dimension e Scopus, não oferecendo mecanismos de se trabalhar com a PL, que é bastante relevante para o campo brasileiro; (iv) facilidade (ou dificuldade) de uso e configuração: muitas ferramentas pressupõem um conhecimento prévio de conceitos relacionados a estudos métricos para sua utilização ou até mesmo de computação para a correta instalação e configuração, sendo exemplos a configuração de um ambiente em linguagem R para o Bibliometrix ou a elaboração de filtros com expressões regulares no VP; (v) entre outros.

A democratização do conhecimento, sob o prisma do acesso, é essencial para a ampliação do campo científico e também para sua popularização e compreensão perante à população em geral (AULER; DELIZOICOV, 2001). Conforme exposto sobre as ferramentas bibliométricas, nota-se a ausência de uma ferramenta que contribua com a democratização de análises bibliométricas no campo científico brasileiro, como por exemplo uma ferramenta que: (i) funcione diretamente na *web* (sem a necessidade que o pesquisador tenha um computador ou infraestrutura potente para fazer os processamentos necessários de sua análise bibliométrica); (ii) seja grátis; (iii) forneça mecanismos de se fazer análises bibliométricas tanto com publicações internacionais quanto nacionais; (iv) tenha fácil utilização (sem a necessidade de instalação ou de conhecimento de linguagens de programação para utilização e configuração da mesma); (v) esteja disponível também em língua portuguesa.

Em um mundo onde os aparelhos celulares vêm ganhando cada vez mais espaço perante computadores tradicionais (ENGE, 2021), a existência de uma ferramenta que atenda parte dos requisitos expostos fomentaria ainda mais o campo



científico, uma vez que suas características podem auxiliar na popularização e de análises bibliométricas a um grupo maior de pessoas, independentemente de sua condição financeira e conhecimento, sendo capaz até mesmo de ser acessada por meio de um *smartphone*. Tal ferramenta poderia até mesmo ser utilizada em salas de aula de instituições de ensino por meio dos dispositivos móveis dos discentes para que sejam explorados conceitos e utilizações dos estudos métricos e também por pesquisadores em seus estudos bibliométricos, sendo, portanto, de extrema importância a existência de uma ferramenta com tais características.

## 2 REFERENCIAL TEÓRICO

A exploração do referencial teórico é tema imprescindível na elaboração de estudos científicos dado que o mesmo tem como foco ser uma revisão de pesquisas e discussões feitas por outros autores sobre temas a serem abordados (LAKATOS; MARCONI, 2010).

Esta seção tem como objetivo explorar o referencial teórico de temas relevantes para a elaboração deste estudo científico e construção de uma ferramenta bibliométrica, explorando temas como a bibliometria, metadados, vocabulários controlados e tesouros, fontes de informação, ferramentas bibliométricas e a sua democratização.

### 2.1 Bibliometria: da obtenção dos dados à visualização da produção científica

Os estudos bibliométricos são extremamente importantes no contexto contemporâneo da *Web* 4.0, onde inúmeros repositórios e plataformas estão sendo criados e disponibilizados para atender às necessidades de seus usuários, contendo registros que previamente só se encontravam disponíveis em acervos físicos de bibliotecas e entre outros (CHU; OUYANG; XU, 2009). Dada a importância da bibliometria, é abordado nesta seção a relevância da bibliometria e os diversos processos comumente utilizados para a realização de um estudo bibliométrico.

#### 2.1.1 A bibliometria e sua importância

A bibliometria pode ser definida como um campo de estudo quantitativo voltado para diferentes tipos de contagens e elaboração de indicadores com base em dados de diferentes literaturas científicas, como artigos e livros, e também de patentes (OKUBO, 1997). Diversas outras definições contribuíram para o entendimento desse conceito, como a de Pritchard (1969) que retrata o mesmo como “estudos que tentam quantificar os processos de comunicação escrita” e também a de Macias-Chapula (1988) que definem a bibliometria como “o estudo de aspectos quantitativos da produção, disseminação, socialização e evidência da informação registrada”. Faria (2015, p. 3, apud RAO, 1986) e Zhu *et al.* (1999) ainda completam as definições ressaltando que, dependendo da finalidade do estudo bibliométrico, podem ser

quantificados desde o texto do autor quanto o nome dos autores, título, fonte, idioma, palavras-chave, classificações e citações.

No contexto do campo científico brasileiro, a bibliometria torna-se objeto de reflexão na década de 70 com o surgimento do primeiro curso de pós-graduação em Ciência da Informação do Brasil, sendo vinculado, com base no nome atual, ao conhecido Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT) (ALVARADO, 1984). Conforme exposto por Araújo e Alvarenga (2011), o curso em questão “foi pioneiro na introdução da bibliometria no Brasil, envolvendo estudos de quantificação e descrição do conhecimento registrado”. Desde então, a mesma é essencial para o processo de transparência das instituições brasileiras, onde estudos bibliométricos retratam e qualificam os mais diversos campos científicos brasileiros, suas instituições e suas contribuições.

Quando se fala de bibliometria e seus processos de quantificação da comunicação escrita, é imprescindível comentar sobre suas leis, são elas que guiam os estudos bibliométricos até os dias atuais. Segundo Alvarado (1984), as três principais leis bibliométricas são: lei de Bradford, lei de Lotka e leis de Zipf. A Lei de Bradford é focada na obra e no seu meio de publicação, busca estudar o fenômeno da dispersão que ocorre em publicações científicas de modo a identificar a qualidade dos periódicos. A lei de Lotka tem como foco a produtividade dos autores e se assemelha à regra de Pareto, exemplificando que em geral muitos autores publicam pouco e poucos autores publicam muito. Por fim as leis de Zipf estão diretamente relacionadas com a contagem de palavras em uma determinada obra, evidenciando as maiores ocorrências e a adequação das mesmas nesse contexto, o pode auxiliar tanto na indexação quanto na recuperação da informação (FARIA, 2015; GUEDES; BORSCHIVER, 2005; KROEFF *et. al.*, 2015).

A bibliometria é atualmente um tópico de pesquisa de extrema relevância e aplicabilidade em suas diversas aplicações, seja por meio de uma pesquisa que busca avançar o conhecimento e métodos dos estudos métricos ou uma pesquisa que se utiliza da bibliometria como metodologia para elaboração da mesma (GLANZEL, 2003). Conforme Ravelli *et. al.* 2009, a importância dos estudos e reflexões bibliométricas é sustentada principalmente pela necessidade de se conhecer e avaliar a produtividade e qualidade da pesquisa de autores e suas instituições, além de também serem cada vez mais requisitados e utilizados para quantificação da produção e informações relevantes nas mesmas. Araújo e Alvarenga (2011)

ponderam que a bibliometria, quando analisada como área de estudo da Ciência da Informação, assume um papel extremamente importante na análise do campo científico de um país, por meio da análise bibliométrica da produção científica do mesmo. Guedes (2012, p. 101) também explora e resume a importância da bibliometria:

Em suma, as abordagens teóricas, práticas e descritivas da Bibliometria contribuem sobretudo para a adequada mensuração da informação e produção de indicadores, que funcionam como ferramenta de relevância crucial para a cadeia de tomadas de decisão em sistemas de recuperação da informação, de avaliação e de comunicação científica; otimizando, em última análise, o gerenciamento de recursos e de resultados pretendidos, no âmbito da gestão da informação e do conhecimento científico, entre indivíduos, disciplinas, organizações e países (GUEDES, 2012, p. 101).

### 2.1.2 Obtenção e preparação dos dados

Quando se inicia um estudo bibliométrico, é imprescindível que seja dada a devida atenção a um dos principais elementos do mesmo: o dado, ou objeto, a ser analisado. É por meio da análise do mesmo que o estudo será realizado e que ponderações, reflexões e contribuições serão geradas durante sua elaboração. A obtenção do mesmo é etapa essencial em um estudo bibliométrico sendo que em alguns casos é necessário posterior preparação dos mesmos, principalmente quando se trata de dados obtidos de bases de dados (FARIA, 2001; VAAN RAN, 1997).

A obtenção do objeto pode ser realizada por meio de diversas fontes de informações (tema que será explorado com maiores detalhes posteriormente) como bases de dados presentes em ambientes digitais, por meio do ítem físico, como o caso do livro presente em um acervo, e entre outras (ARAUJO; FACHIN, 2015; BLATMANN, 2015). É importante ressaltar que o “local” de obtenção do dado em questão não é algo absoluto e, portanto, poderá ser alterado dependendo do tipo de estudo, resultado esperado e ferramenta a ser utilizada pelo mesmo de acordo com as necessidades dos autores.

Sobre a preparação dos dados, apesar de não ser uma etapa necessária em todos os estudos bibliométricos, é frequentemente realizada quando se está utilizando dados provenientes de bases de dados ou fornecidos por outras ferramentas, dado

que raramente os dados recuperados podem ser utilizados diretamente por *softwares* de análise bibliométrica automatizada (MOGEE, 1997). Cada ferramenta é desenvolvida de modo a atender um conjunto pré-definido de dados e suas características e, para isso que isso seja possível, são delimitados os formatos de metadados suportados pela mesma, seja o formato CSV, RIS, ISI-CE ou outros (FARIA, 2001). Caso o conjunto de dados a ser analisado por uma ferramenta não seja compatível com o formato de metadados suportado pela mesma, é necessário a realização da preparação dos dados, processo que precisa se ajustar às condições em que o tratamento da informação é realizado (VAAN RAN, 1997).

Faria (2001) ainda elenca outros pontos importantes sobre a preparação dos dados:

- A informação extraída de uma base de dados pode necessitar ser complementada com informações provenientes de outras fontes.
- Ajusta-se às condições em que o tratamento informacional é realizado.
- Estima-se que seu esforço consuma cerca de 60% de todo o esforço empreendido no tratamento automatizado da informação.
- O processo deve ser preferencialmente automatizado para que se evitem gargalos durante o processamento.

### 2.1.3 O processo de contagem

A contagem ou medição é tópico de estudo e discussão desde os primórdios da ciência dada sua relevância (MILANEZ, 2015, p. 28). Pensadores e pesquisadores de diferentes épocas convergem para o mesmo ponto: a importância dos processos de contagem. Isso pode ser evidenciado pela fala de Horácio (65-5 AC) onde há séculos atrás defendia que “há uma medida para tudo” e também contemporaneamente pelo descobridor da supercondutividade, Heike K. Onnes (1853-19 DC), que pregava “medir é conhecer” (FARIA, 2015, p. 2).

Em contextos atuais das áreas da Ciência da Informação, Biblioteconomia e outras áreas da informação, o processo de contagem é parte integrante e indissociável da bibliometria e entre outras. A contagem pode ser feita para diversos elementos de uma obra, a critério do estudo que deseja ser efetuado, como por exemplo a contagem de palavras, referências, conceitos que mais aparecem e entre muitos outros (ALVARADO, 2002, 2008; FARIA, 2015).

Além da definição do que será objeto de contagens, onde o mesmo será obtido e se será necessário a sua preparação, outro ponto essencial é estabelecer o tipo de contagem que será feita. Um dos principais exemplos é sobre a contagem de dados referentes à autoria que, conforme Alvarado (2002), pode ser feita de modo direta, completa ou ajustada/parcial:

- **Contagem direta:** autores principais ou mais relevantes são creditados (contabilizados na contagem) com uma contribuição e os secundários excluídos.
- **Contagem completa:** tantos os autores principais e secundários são contabilizados igualmente em sua contribuição.
- **Contagem ajustada/parcial:** cada autor recebe uma fração igualitária de sua contribuição. Exemplo: quatro autores publicaram o mesmo artigo e, portanto, cada um recebe um quarto de contribuição na contagem.

#### 2.1.4 Indicadores para representação da informação

Conforme exposto pelas palavras de Guedes (2012), um dos principais resultados de estudos bibliométricos é a produção de indicadores com base nos dados bibliométricos analisados, são eles que fornecem uma visualização dos resultados de uma análise para sua melhor compreensão. Esse processo transforma um conjunto de metadados, que muitas vezes é ignorado ou não compreendido pelo pesquisador ou leitor de uma publicação, em algo palpável capaz de gerar conhecimento e possivelmente novas reflexões.

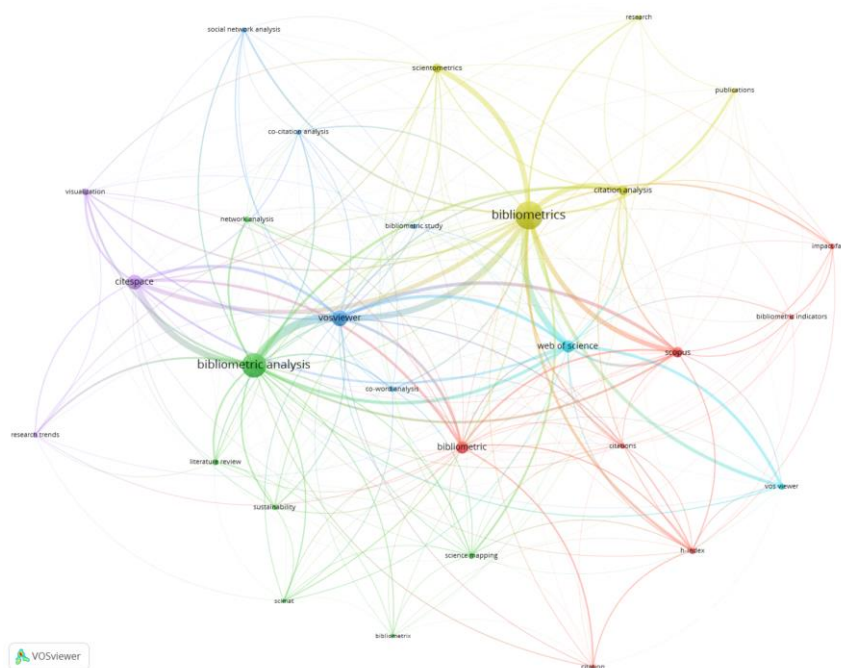
Existe atualmente uma gama de indicadores bibliométricos para focos diversos, como os indicadores bibliométricos de produção científica e indicadores para patentes, sendo alguns dos principais indicadores de produção científica elencados e explicados no Quadro 1 (FARIA, 2015). Outro exemplo seria o Índice H que combina tanto indicadores de produtividade e impacto para avaliar a relevância de um pesquisador (BORNMAN; DANIEL, 2007; DORTA-GONZÁLEZ; DORTA-GONZÁLEZ, 2010; LIMA; VELHO; FARIA, 2012).

**Quadro 1** – Alguns dos principais indicadores bibliométricos para análise de publicações científicas.

Indicadores de atividade	Indicadores de impacto ou fator de impacto	Indicadores de ligação ou relacionais
Utiliza-se da contagem de publicações para geração de listas de frequência e rankings de publicações.	Utiliza-se de dados de citações para medir a importância de autores, instituições, países, periódicos e entre outros.	Utiliza-se de dados de coocorrências de citações, palavras-chaves, publicações para evidenciar a colaboração científica. Um exemplo pode ser visto na Figura 1.

Fonte: Elaborado pelo com base nas teorias de Faria (2015), OKUBO (1997) e CALLON *et al.* (1993).

Ainda sob o prisma dos indicadores bibliométricos, existe um conjunto de indicadores que vem ganhando popularidade no meio científico: os indicadores de colaboração. Os indicadores de colaboração, sub conjunto dos indicadores relacionais (GRÁCIO, 2018), buscam medir a relação de um objeto de análise com seus pares em um determinado contexto, podendo ser esse contexto a coocorrência de palavras-chaves, autores, citações e entre outros (FARIA, 2015). Conforme pode ser visto no exemplo da Figura 1, é apresentada a coocorrência de palavras em um conjunto de publicações, podendo ser analisado quais as temáticas relevantes dentro desse universo (tamanho do círculo), quais os diferentes núcleos temáticos (por meio da cor) e também qual a frequência de coocorrência entre as temáticas (espessura da linha).

**Figura 1** – Visualização de um indicador de ligação onde são mostradas as palavras-chave mais utilizadas pelos autores em uma busca sobre bibliometria na Web of Science.

Fonte: Elaborada pelo autor

### 2.1.5 Exploração da bibliometria por meio de métricas alternativas

A bibliometria não é o único modo de se estudar publicações científicas e seu impacto, existem outros métodos que buscam analisar as mesmas de um modo não convencional, como é o caso da altimetria (PIWOWAR, 2013).

A altimetria, segundo Priem (2011) e também Galligan e Dyas-Correia (2013), pode ser considerada uma bibliometria não tradicional, dado que foi proposta como uma alternativa ou complemento à indicadores tradicionais da bibliometria, e que portanto busca atuar em outras métricas alternativas em ambientes baseados na *Web*, o que gerou seu nome “métricas alternativas” ou *altmetrics*. O conjunto de métricas alternativas é um universo em constante expansão, dado que cada vez mais novas ferramentas ou repositórios são criados e conseqüentemente novas interações dos pesquisadores entre si e com seus meios de pesquisa ocorrem, fornecendo portanto novos meios de serem efetuadas análises (HUANG; WANG; WU, 2018). Piwowar (2013) reflete ainda sobre a questão de até quando uma métrica pode ser considerada alternativa dado que, pela definição sua definição, as métricas alternativas recebem esse nome pelo fato de serem compostas, como seu próprio nome retrata, de métricas que não são frequentemente exploradas, levantando a questão: seria então uma métrica alternativa frequentemente utilizada apenas uma métrica?

Algumas das principais métricas abordadas por estudos altimétricos, segundo Silva e Vanz (2019), são: quantidade de acessos a uma página, ferramenta ou publicação; quantidade de downloads de uma publicação; publicações favoritadas, salvas ou recomendadas e entre outras. Complementando os exemplos prévios, Piwowar (2013) elenca também potenciais vantagens da altimetria perante métricas convencionais:

- Melhores nuances do impacto de uma publicação: demonstrando quais produtos acadêmicos são mais lidos, discutidos, salvos, recomendados e citados.
- Dados mais oportunos: mostrando evidências de impacto em intervalos de dias ao em vez de anos.
- Mensuração do impacto de produtos acadêmicos como conjunto de dados, ferramentas, postagens em blogs, vídeos e entre outros.
- Mensuração do impacto em conjunto de pessoas diversos como por



exemplo acadêmicos, profissionais da área do estudo, educadores e também no público em geral.

Dado o exposto, fica claro a importância das métricas alternativas no contexto atual da *Web 4.0*, possibilitando mensurar elementos de extrema relevância, que muitas vezes são deixados de lado pelo pesquisador, e que em muitos casos poderiam ser utilizados lado a lado com as atuais métricas tradicionais.

## **2.2 Metadados e a interoperabilidade em ambientes digitais**

Uma das principais características de ambientes digitais atualmente, seja uma rede social, sistemas empresariais ou aplicativos de celular, é a necessidade da troca de recursos informacionais entre os mais diferentes mecanismos tecnológicos. Um recurso informacional trata-se de uma informação objetivada em um contexto digital e/ou analógico que contém um valor informacional que justifica sua concepção (SANTOS; SIMIONATO; ARAKAKI, 2014). Matias (2015) reflete sobre a questão relatando o fato de que um mesmo recurso informacional pode estar disponível em diferentes repositórios institucionais porém representado de diferentes maneiras. Essa questão pode ocasionar diversos problemas como a inviabilidade da interoperabilidade entre sistemas e o aumento demasiado da complexidade dos processos de busca, acesso e recuperação, fazendo com que seja necessária a definição de metadados e padrões de metadados.

Os metadados podem ser definidos de diferentes formas em diferentes contextos. Uma das primeiras definições foi feita na década de 1960 onde foi utilizada a expressão “dados sobre dados”, onde dados são utilizados para representar dados (POMERANTZ, 2015). Diversas outras definições surgiram como a de Gilliland (2016) que retrata os mesmos como a padronização de informações utilizadas com o objetivo de facilitar a organização, descrição, rastreabilidade e acessibilidade a recursos informacionais; a de Alves (2010) define os mesmos como “atributos que representam uma entidade (objeto do mundo real) em um sistema de informação”. Conforme expõe Angelozi e Martin (2011), não existe uma definição única para o conceito de metadados, o mesmo pode ter diferentes significados em diferentes áreas do conhecimento, sendo o cerne principal da questão a facilitação da interoperabilidade do recurso informacional em questão.

Os metadados podem ser representados segundo diversos formatos de organização. A existência de diversos formatos é devida aos diferentes contextos de uso dos metadados e também ao surgimento de formatos internos a determinadas comunidades de usuários, que se impõe como padrão para aquela comunidade, mas que se diferem de formatos surgidos em outras comunidades de usuários (DE BARROS CAMPOS, 2007).

De modo a exemplificar a representação de metadados em diferentes formatos note o *JavaScript Object Notation*<sup>7</sup> (JSON) e o *Extensible Markup Language*<sup>8</sup> (XML) presente na Figura 2. É possível notar que ambos representam a mesma informação, no caso em questão uma edição de um livro com título “O Senhor do Anéis”, porém representado de maneiras diferentes. Vale ressaltar que dentro do formato JSON e XML cada utilizador pode padronizar e adaptar sua própria versão, como por exemplo no lugar de “livro” e “autor” utilizar “book” e “author”, de acordo com suas necessidades de modo a facilitar a interoperabilidade entre os sistemas desejados. O formato JSON será aprofundado posteriormente.

**Figura 2** – Uma edição do livro “O Senhor do Anéis” representada nos formatos JSON e XML.

JSON	XML
<pre>{   "livro": {     "autor": "J.R.R. Tolkien",     "titulo": "O Senhor dos Anéis",     "dataLancamento": "14/04/2012"     "valor": 49.99   } }</pre>	<pre>&lt;livro&gt;   &lt;autor&gt;J.R.R. Tolkien&lt;/autor&gt;   &lt;titulo&gt;O Senhor do Anéis&lt;/titulo&gt;   &lt;data&gt;14/02/2012&lt;/data&gt;   &lt;valor&gt;49.99&lt;/valor&gt; &lt;/livro&gt;</pre>

Fonte: Elaborada pelo autor.

Dada a importância dos metadados e da interoperabilidade diversos formatos

<sup>7</sup> <<https://www.json.org/json-en.html>>

<sup>8</sup><<https://www.w3.org/TR/xml/>>

surgiram, não só para sistemas de informações comerciais, mas também para o estudo de publicações científicas disponíveis na rede mundial de computadores, a fim de que fosse possível facilitar a interoperabilidade entre repositórios de dados e programas que trabalham com esses tipos de dados. *Softwares* de referência como o Zotero<sup>9</sup>, BibDesk<sup>10</sup>, EndNote<sup>11</sup>, Citavi<sup>12</sup>, Papers<sup>13</sup> e RefWorks<sup>14</sup> são capazes de utilizar os mais diferentes tipos formatos de arquivos como o RIS, ISI-CE, BibTex<sup>15</sup>, *Dublin Core*<sup>16</sup> e entre muitos outros. A eficiência e eficácia de diferentes serviços de informação estão diretamente ligadas com a utilização adequada dos metadados e também com a quantidade de formatos suportados (SANTOS;SIMIONATO;ARAKAKI, 2014).

### 2.2.1 O formato RIS

Um dos principais formatos de metadados na temática de publicações científicas é o desenvolvido pela empresa *Research Information Systems*, que leva em seu nome a abreviatura do nome da empresa criadora, o formato RIS. O RIS é um formato padronizado de texto puro que pode representar diversas citações em um único arquivo por meio de sua estrutura de tagueamento, sendo capaz de permitir a diferentes programas que trabalham com citações a troca de informações. (CLARIVATE, 2021; MATIAS, 2015; THOMSON REUTERS, 2012).

O processo de tagueamento para uma publicação é feito por linha, onde cada linha é iniciada com uma tag padronizada iniciada com uma letra maiúscula, seguido de outra letra maiúscula ou um número, onde as duas letras ou a letra e número representam o elemento em questão, e seguida por dois espaços, um hífen e outro espaço, conforme pode ser visto no Quadro 2.

---

<sup>9</sup> <<https://www.zotero.org/>>

<sup>10</sup> <<https://bibdesk.sourceforge.io/>>

<sup>11</sup> <<https://endnote.com/>>

<sup>12</sup> <<https://www.citavi.com/>>

<sup>13</sup> <<https://www.papersapp.com/>>

<sup>14</sup> <<https://www.refworks.com/>>

<sup>15</sup> <<http://www.bibtex.org/>>

<sup>16</sup> <<https://dublincore.org/>>

**Quadro 2** – Ordem de tagueamento de uma linha no formato no formato RIS.

<b>Ordem</b>	<b>Caractere</b>
Primeiro caractere	Alfabético em letra maiúscula específico a cada tipo de campo
Segundo caractere	Alfabético em letra maiúscula ou numérico específico a cada tipo de campo
Terceiro e quarto caractere	Espaço no padrão ANSI 32
Quarto caractere	Traço no padrão ANSI 45
Quinto caractere	Espaço no padrão ANSI 42

Fonte: Traduzido de Clarivate (2021).

Após o tagueamento da linha é descrita a respectiva informação relacionada à tag, sendo um exemplo o nome do autor primário da publicação caso a tag seja “A1”, seguido de uma quebra de linha. Uma publicação sempre será iniciada pela tag “TY” que representa o tipo de publicação e sempre terminará com a tag “ER”, que representa o fim do registro (CLARIVATE, 2021). Clarivate (2021) ainda reitera que todas as outras tags não são obrigatórias, podendo aparecer ou não aparecer em uma publicação. A Figura 3 demonstra uma publicação científica disposta no formato RIS e o Quadro 3 mostra algumas das diversas tags disponíveis.

**Figura 3** – Exemplo de uma publicação científica no formato RIS.

```

TY - BOOK
AU - Behrens,J.
TI - History of the CDC
PY - 1999
CY - Chicago
PB - Parity Press
SP - 144
VL - 2nd
KW - Epidemiology
KW - U.S. Gov't
ER -

```

Fonte: Elaborada pelo autor.

Dada a importância do formato em questão diversas ferramentas e iniciativas têm oferecido suporte ao mesmo para que seja possível a troca de informação. Alguns dos principais exemplos são as bibliotecas digitais como a Scopus<sup>17</sup>, IEEE Xplore<sup>18</sup>, SpringerLink<sup>19</sup> e entre outras; ferramentas como o BibDesk, RefWorks, Zotero, Citavi, Papers, Mendeley<sup>20</sup>, EndNote e entre muitos outros. Além disso, algumas iniciativas têm adaptado o formato RIS para seus próprios usos como o caso do formato RIS adaptado de Matias (2015) e o formato ISI-CE da plataforma WoS. (MATIAS, 2015; PRINCETON UNIVERSITY LIBRARY, 2017).

**Quadro 3** – Algumas das principais tags disponíveis no formato RIS.

TY – Tipo de publicação	CY – Local de publicação	RN – Notas da pesquisa
A1 – Autores primários	DA – Data de publicação	SN – ISBN/ISSN
A2 – Autores secundários	DO – DOI	SP – Página inicial
A3 – Autores terciários	ID – ID de referência	T1 – Título primário
A4 – Autores subsidiários	DO – DOI	T2 – Título secundário
AB – Resumo	ET – Edição	T3 – Título terciário
AD – Endereço do autor	KW – Palavras chaves	TI – Título
AN – Número de acesso	LA – Linguagem	TT – Título traduzido
AU – Autor	LK – Endereço do <i>website</i>	UR – URL de acesso
AV – Local dos arquivos	NV – Número de volumes	VL – Número do volume
CN – Call Number	PB – Editora	ER – Fim do registro

Fonte: Adaptado e traduzido pelo autor de Thomson Reuters (2012).

<sup>17</sup> <<https://www.scopus.com/>>

<sup>18</sup> <<https://ieeexplore.ieee.org>>

<sup>19</sup> <<https://link.springer.com/>>

<sup>20</sup> <<https://www.mendeley.com/>>

### 2.2.2 O formato JSON

Conforme exposto previamente, o JSON é um dos inúmeros formatos existentes atualmente para troca de dados. O mesmo é amplamente utilizado por sistema de informação, dado que é um formato de padrão aberto (*open standard*) de formatação de arquivos e transferência de dados que se utiliza de uma linguagem de fácil leitura por seres humanos (ECMA, 2017; FRIESEN, 2019). É imperativo citar que apesar de o mesmo levar o nome de uma linguagem de programação em seu corpo (*JavaScript*) o mesmo pode e é amplamente utilizado por diversas outras (FRIESEN, 2019).

Quando se pondera sobre sua padronização, o formato faz uso da estrutura “chave”:“valor” para representar os dados a serem transmitidos entre diferentes sistemas, servidores e aplicações *web* (ECMA, 2017). Conforme pode ser visto na Figura 4, que busca representar um registro em uma lista telefônica, o valor por ser desde algo simples, como um nome, quanto também algo mais complexo, como por exemplo uma lista (*array*) de telefones, ou também outras informações, como seu endereço.

**Figura 4** – Um exemplo de registro telefônico no formato JSON.

```
{
  "nome": "Joaquim Ferreira",
  "endereço": {
    "rua": "Alameda sem saída",
    "numero": 1234,
    "cidade": "São Paulo",
    "cep": "15030-030",
  },
  "telefones": [
    {
      "numero": "9999-9999",
      "tipo": "celular"
    },
    {
      "numero": "3333-3333",
      "tipo": "residencial"
    }
  ]
}
```

Fonte: Elaborado pelo autor.

### 2.2.3 O formato ISI-CE

O ISI *common export format* (ISI-CE), conforme a própria tradução de seu nome diz, trata-se do formato de exportação padrão desenvolvido pelo Instituto para Informação Científica (ISI), sendo bastante semelhante ao formato RIS tradicional porém adaptado para um modo diferente de tagueamento (EDHLUND, 2007; FILE TYPE ADVISOR, 2021). Por padrão em banco de dados mantidos por Thomson Reuters, como é o caso da WoS, o formato padrão utilizado é o ISI-CE, comumente identificado pela extensão “.ciw” no fim de um arquivo, e não o formato RIS e sua extensão “.ris” (PRINCETON UNIVERSITY LIBRARY, 2017). Vale ressaltar que existem arquivos neste formato que não são terminados com a extensão “.ciw” mas sim em “.txt” onde ambos apresentam a mesma estrutura de tagueamento, não sendo portanto um problema para análises manuais dado que a estrutura se mantém a mesma. Um exemplo do mesmo é a exportação em texto plano na WoS onde é gerado um arquivo de publicações científicas com a extensão “.txt” mas contendo todas as regras do “.ciw”.

Conforme exposto previamente, o formato ISI-CE também se utiliza de uma estrutura de tagueamento para representação de publicações científicas que, apesar de semelhante a do formato RIS, tem suas particularidades. São três diferenças principais: (i) as “tags” utilizadas para representar a informação; (ii) a estrutura de tagueamento; (iii) a estrutura geral do arquivo.

Sobre as “tags” utilizadas, em comparação com as tags do RIS, existem algumas que são diferentes para representar a mesma informação, outras que são iguais e outras que existem em um mas não existem no outro (REFMAN, 2021; CLARIVATE, 2020). Um exemplo de tags diferentes que representam a mesma informação é o caso de *publisher* onde no formato RIS para se representar o mesmo é utilizado “PB” e no formato ISI-CE “PU”. A lista completa de tags utilizadas no formato ISI-CE pode ser vista no Quadro 4.

Já se tratando da estrutura de tagueamento, conforme pode ser visto no exemplo de publicação no formato ISI-CE presente na Figura 5, existem algumas diferenças, apesar de similar com o formato RIS. As diferenças são: (i) agora existe somente um espaço separando a tag e o seu valor; (ii) não existe mais o traço separador; (iii) em algumas tags que apresentarem mais de um valor faz com que o segundo seja disposto na linha seguinte, iniciando a mesma com três espaços seguido

do elemento, como o caso das tags “AU” e “AF” de autores (EDHLUND, 2007).

**Quadro 4** – As diversas tags disponíveis no formato ISI-CE.

FN – Nome do arquivo	RP – Endereço de reimpressão	PN – Número da parte
VR – Número da versão	EM – Endereços de e-mail	SU – Suplemento
PT – Tipo de publicação	RI – ResearcherID	MA – Resumo da reunião
AU – Nomes abreviados dos autores	OI – Identificador ORCID	BP – Página inicial
AF – Nomes completos dos autores	FU – Agência de financiamento da pesquisa	EP – Página final
BA – Nomes resumidos dos autores do livro	FX – Texto do financiamento da pesquisa	AR – Número do artigo
BF – Nomes completos dos autores do livro	CR – Referências citadas	DI – Digital Object Identifier (DOI)
CA – Grupo de autores	NR – Contagem de referências citadas	D2 – Book Digital Object Identifier (DOI)
GP – Grupos de autores do livro	TC – Contagem de citações na Web of Science Core Collection	EA – Data de acesso antecipado
BE – Editores	Z9 – Contagem de citações em outros índices da Web of Science	EY – Ano de acesso antecipado
TI – Título do documento	U1 – Quantidade de usos durante os últimos 180 dias	PG – Quantidade de páginas
SO – Nome da publicação	U2 – Quantidade de usos desde 2013	P2 – Quantidade de capítulos
SE – Título da série do livro	PU – Veículo de publicação	WC – Categorias da Web of Science
BS – Legenda da série do livro	PI – Cidade do veículo	SC – Áreas de pesquisa
LA – Linguagem	PA – Endereço do veículo	GA – Número de entrega do documento
DT – Tipo de documento	SN – International Standard Serial Number (ISSN)	PM – PubMed ID
CT – Título da conferência	EI – Electronic International Standard Serial Number (eISSN)	UT – Número de acesso
CY – Data da conferência	BN – International Standard Book Number (ISBN)	OA – Indicador de acesso aberto
CL – Local da conferência	J9 – Abreviação com 29 caracteres da fonte	HP – ESI Hot Papers
SP – Patrocinadores da conferência	J1 – Abreviação ISO da fonte	HC – ESI Highly Cited Paper
HO – Anfitrião da conferência	PD – Data de publicação	DA – Data que esse relatório foi gerado
DE – Palavras-chave do autor	PY – Ano de publicação	ER – Fim do registro
ID – <i>Keywords Plus</i> <sup>®</sup>	VL – Volume	EF – Fim do arquivo
AB – Resumo	IS – Edição	
C1 – Endereço dos autores	SI – Edição especial	

Fonte: Traduzido pelo autor de Clarivate (2020).



Já se tratando da estrutura de tagueamento, conforme pode ser visto no exemplo de publicação no formato ISI-CE presente na Figura 5, existem algumas diferenças, apesar de similar com o formato RIS. As diferenças são: (i) agora existe somente um espaço separando a tag e o seu valor; (ii) não existe mais o traço separador; (iii) em algumas tags que apresentarem mais de um valor faz com que o segundo seja disposto na linha seguinte, iniciando a mesma com três espaços seguido do elemento, como o caso das tags “AU” e “AF” de autores (EDHLUND, 2007).

Por fim, sobre a estrutura do arquivo em si, há diferenças no mesmo, conforme pode ser visto também na Figura 5. O arquivo sempre começa com uma linha contendo “FN Clarivate Analytics Web of Science”, seguido da sua versão “VR 1.0” e conseqüentemente são mostrados todos os registros, onde os mesmos se iniciam com a tag “PT” e finalizados com a tag “ER”; separados por uma linha em branco; e o arquivo finalizado com a tag “EF”.

**Figura 5** – Duas publicações científicas em um arquivo no formato ISI-CE.

```

FN Clarivate Analytics Web of Science
VR 1.0
PT J
AU Pulsiri, N
   Vatananan-Thesenvitz, R
AF Pulsiri, Nonthapat
   Vatananan-Thesenvitz, Ronald
TI Drones in Emergency Medical Services: A Systematic Literature Review with Bibliometric Analysis
SO INTERNATIONAL JOURNAL OF INNOVATION AND TECHNOLOGY MANAGEMENT
ER

PT J
AU Arbatov, A
AF Arbatov, Alexey
TI Saving Strategic Arms Control (vol 62, pg 79, 2020)
SO SURVIVAL
PY 2021
DI 10.1080/00396338.2021.1863617
EA DEC 2020
ER

EF

```

Fonte: Elaborado pelo autor.

## 2.2.4 O formato RIS modificado do syncLattes

O formato RIS modificado em que são representados os dados extraídos da PL por meio do *software* syncLattes desenvolvido por MATIAS (2015) é resultado da combinação de dois diferentes formatos de metadados: o RIS e o Dublin Core Qualificado. O mesmo se utiliza da estrutura do formato RIS, como por exemplo o tagging e a disposição dos metadados, em conjunto com as tags utilizadas pelo Dublin Core Qualificado (MATIAS, 2015).

As diferentes tags disponíveis no mesmo podem ser vistas por meio do Quadro 5. Na primeira coluna é exposto os campos fornecidos pelo Currículo Lattes em seu XML, quando obtidos por meio do extrator, e na segunda coluna é exposta a tag no formato Dublin Core Qualificado que representa a informação na coluna anterior. Por exemplo, as palavras-chave contidas no registro em questão serão mapeadas para o campo “dc.subject”.

**Quadro 5** – Mapeamento dos campos do XML do Lattes utilizados para a composição dos campos Dublin Core Qualificado no DSpace.

XML – Currículo Lattes	Dublin Core Qualificado
<TRABALHO-EM-EVENTOS> ou <ARTIGO-PUBLICADO>	dc.type
<AUTORES NOME-COMPLETO-DO-AUTOR=□/ >	dc.contributor.author
<DADOS-BASICOS-DO-TRABALHO ou -ARTIGO ANO-DO-TRABALHO ou -ARTIGO=□/ >	dc.date.issued
<DADOS-BASICOS-DO-TRABALHO ou -ARTIGO DOI=□/ >	dc.identifier.uri
<DADOS-BASICOS-DO-TRABALHO ou -ARTIGO HOME-PAGE-DO-TRABALHO=□/ >	dc.identifier.uri
<DETALHAMENTO-DO-ARTIGO ISSN=□/ >	dc.identifier.issn
<DETALHAMENTO-DO-TRABALHO ISBN=□/ >	dc.identifier.isbn
<DETALHAMENTO-DO-TRABALHO ou -ARTIGO VOLUME=□ FASCICULO=□ PAGINA-INICIAL=□ PAGINA-FINAL=□/ >	dc.identifier.citation
<DADOS-BASICOS-DO-TRABALHO ou -ARTIGO IDIOMA=□/ >	dc.language.iso
<DETALHAMENTO-DO-TRABALHO ou -ARTIGO TITULO-DOS-ANAIS-OU-PROCEEDINGS ou -DO-PERIODICO-OU-REVISTA=□/ >	dc.relation.ispartof
<PALAVRAS-CHAVE PALAVRA-CHAVE-n=□/ >	dc.subject
<AREA-DO-CONHECIMENTO NOME-x=□/ >	dc.subject.classification
<DADOS-BASICOS-DO-TRABALHO ou -ARTIGO TITULO-DO-TRABALHO ou -ARTIGO=□/ >	dc.title

Fonte: MATIAS (2015, p. 65).

Sobre a estrutura do arquivo, conforme pode ser visto na Figura 6, podem ser elencados os seguintes pontos:

- Não há uma tag que sinaliza o início ou fim de uma publicação, porém existe uma ordem para cada publicação que começa pelas tags “capes.qualis”, “dc.language.iso” e “dc.title”, na respectiva ordem, e termina com as tags “dc.type”, “dc.subject” e “dc.subject.classification”.
- As publicações são separadas entre si por uma linha vazia e não há uma tag em específico para delimitar o fim do arquivo;
- A tag e seu conteúdo e/ou valor são separadas por um espaço simples;
- Não há uma tag que sinaliza o início ou fim do arquivo;

**Figura 6** – Duas publicações científicas em um arquivo no formato RIS modificado disponibilizado pelo syncLattes.

```

dc.language.iso por
dc.title O Cinema Marginal e a política
dc.contributor.author AUTRAN, ARTHUR
dc.contributor.author.lattes 5055550097454248
dc.relation.ispartof Nava
dc.date.issued 2017
dc.identifier.citation Nava, v. 2, pp. 307-319, 2017
dc.identifier.issn 25257757
dc.identifier.issnl 25257757
dc.type article
dc.subject Política
      Brasil
      Cinema
dc.subject.classification Ciencias Sociais Aplicadas. Cinema

capes.qualis administracao_publica_e_de_empresas_ciencias_contabeis_e_turismo B5 2015
      comunicacao_e_informacao B5 2015
dc.language.iso por
dc.title A origem dos conhecimentos humanos no contexto dos estudos informacionais
dc.contributor.author GRACIOSO, LUCIANA DE SOUZA
dc.contributor.author.lattes 4898201916360294
dc.relation.ispartof LOGEION: FILOSOFIA DA INFORMAÇÃO
dc.date.issued 2019
dc.identifier.citation LOGEION: FILOSOFIA DA INFORMAÇÃO, v. 5, 2019, 2019
dc.identifier.issn 23587806
dc.identifier.issnl 23587806
dc.type article
dc.subject Epistemologia
      Linguagem
      pragmática
dc.subject.classification Ciencias Sociais Aplicadas

```

Fonte: Elaborado pelo autor.

### 2.2.5 O formato CSV

Outro formato de metadados bastante utilizado nos dias atuais, principalmente para importação e exportação de dados de um banco de dados e aplicações científicas é o: *comma-separated-values* (CSV). O CSV, conforme o próprio nome já propõe, trata-se de um formato em que os elementos em um arquivo são separados por vírgulas, sendo os mesmos representados de maneira tabular de modo que cada linha representa um elemento (SHAFRANOVICH, 2005). Apesar de o padrão ser a separação por vírgulas, o formato não é totalmente padronizado e variações existem como a utilização de “;” ou “!” para separação dos elementos e entre outras aplicações (RFC 4180).

Segundo a definição oficial (RFC 4180) as convenções mais utilizadas para um arquivo no formato CSV seguem a seguinte padronização:

- Cada registro é localizado em uma linha separada, terminada por uma quebra de linha. Exemplo: aaa, bbb, ccc (Quebra de linha)
- O último registro no arquivo pode ou não ter uma quebra de linha.
- Em alguns casos a primeira linha pode ser uma linha de cabeçalho que segue os padrões das demais linhas porém contendo o nome correspondente aos dados na coluna em questão.

No Quadro 6 é exposto a representação de um tabela que contém alguns automóveis em um arquivo em formato CSV seguindo a convenção demonstrada anteriormente.

**Quadro 6** – Representação de uma tabela em formato CSV.

Tabela			CSV
Ano	Modelo	Preço	Ano,Modelo,Preço
2000	Exemplo A1	50000.00	2000, Exemplo A1, 50000.00
2010		25000.00	2010,, 25000.00
2020	Exemplo A3	32000.00	2020, Exemplo A3, 32000.00

Fonte: Elaborado pelo autor.

### 2.3 Vocabulários controlados e tesouros

Além da necessidade de padrões de metadados, quando se discute a temática da análise de um conjunto de publicações científicas existentes e disponibilizadas em diferentes meios, há a necessidade de serem criados mecanismos de padronização do elemento a ser recuperado e/ou analisado (CUNHA; CAVALCANTI, 2008). Um exemplo disso é um mesmo artigo disponibilizado em diferentes bases de dados onde a organização acadêmica do autor é representada pelos metadados da plataforma ou ferramenta como “Universidade Federal de São Carlos” e em outra como “UFSCar”. Apesar de na cognição humana ambos representarem a mesma organização, uma máquina ou ferramenta que seja utilizada para interpretar tal campo não conseguirá realizar a mesma inferência sem que exista algum mecanismo que instrua a mesma sobre tal, considerando os metadados em questão organizações distintas. Tomando como exemplo um sistema de catálogo bibliográfico, dada a suposição que o mesmo tenha um mecanismo de buscar uma obra pela organização, caso a mesma não esteja preparada para a problemática em questão, quando o usuário buscar por “UFSCar”, não será possível retornar o outro exemplo, ou até mesmo seja possível, mas necessite de um maior processamento para identificar a questão.

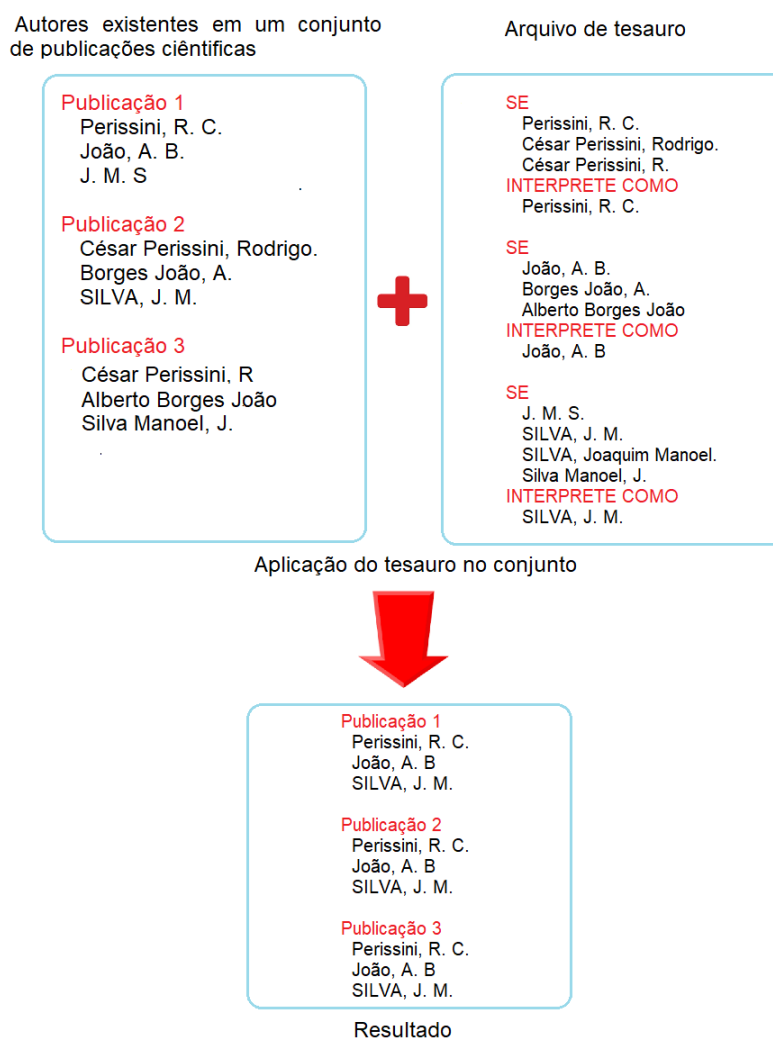
Os vocabulários controlados pode ser definidos como listas de palavras ou frases que são utilizadas para representar e/ou padronizar uma unidade de informação, resolvendo problemas de termos homógrafos, sinônimos e polissemes utilizando uma bijeção entre conceitos e termos autorizados (LEISE, 2008; MACULAN, 2011). Podem ser definidos também como um método de organização do conhecimento para buscas ou análises posteriores por meio da compatibilização da linguagem empregada por pesquisadores e sistemas de informação, sendo geralmente usados na indexação por assunto, tesouros, taxonomias e entre outros (CUNHA; CAVALCANTI, 2008). Em geral utilizam-se de um conjunto de termos pré-definidos e autorizados, diferente de vocabulários de linguagem natural que não apresentam tais restrições.

Os tesouros são um tipo de vocabulário controlado com o foco em eliminar ambiguidades de termos e facilitar o acesso à informação tanto para quem trabalha com a informação, como os sistema de informação, quanto para seu usuário, sendo portanto comumente utilizados no processamento da linguagem natural para atuar na problemática de ambiguidade no sentido de palavras e simplificação textual

(MOREIRA; ALVARENGA; OLIVEIRA, 2004; ANSI/NISO, 2013).

Voltando ao exemplo prévio da organização educacional de um autor, pode ser elaborado portanto um tesauro com todos os modos possíveis que pode ser representada a instituição organizacional de ensino superior “Universidade Federal de São Carlos” e fornecido a um sistema de informação o mesmo, especificando que todas essas diferentes variações da instituição devem ser interpretadas, por exemplo, como “Universidade Federal de São Carlos – UFSCar”. O mesmo poderia ser feito para autores, palavras-chaves e entre outros (ANSI/NISO, 2013). Na Figura 7 é exposto como poderia ocorrer a aplicação de um tesauro de autores em um sistema de informação onde três publicações distintas cujos autores têm seus nomes representados de diferentes modos em cada uma delas.

**Figura 7** – Aplicação de um arquivo de tesauro em um conjunto de três publicações que contemplam nomes de mesmos autores escritos de maneiras distintas



Fonte: Elaborada pelo autor.

Por fim, e não menos importante, outra aplicação dos tesouros é o seu uso em bases de dados como a *Library Information Science Abstracts* (LISA) e a *Library Information Science and Technology Abstracts* (LISTA). O tesouro nesse contexto é frequentemente utilizado com base em um panorama temático, tendo como objetivo facilitar a recuperação de informações relevantes em meio ao conjunto gigantesco de registros presentes nas bases de dados (GHENO, 2013). Em sua utilização o pesquisador utiliza-se de mecanismos disponibilizados pelo tesouro da base de dados para a recuperação do que lhe é relevante, onde tais mecanismos podem ser desde o fornecimento de termos similares, utilização de descritores e entre outros por parte do pesquisador (KUMAR, 2014).

## **2.4 Fontes de informação**

Conforme visto previamente, o estudo de publicações científicas é matéria de nítida relevância, porém, antes de se aprofundar nas ferramentas tecnológicas disponíveis para trabalhar com tal questão, é importante discutir previamente sobre os meios que fornecem tais publicações, as fontes de informação

As fontes de informação podem ser definidas como recursos, disponíveis em suportes físicos, digitais ou multimídia, que consigam atender às necessidades informacionais dos usuários, seja recriando ou disponibilizando conhecimento sobre um determinado assunto (ARAUJO; FACHIN, 2015). Conforme Blatmann (2015) e a dupla Caldeira e Campelo (2018), elas podem ser divididas em três diferentes tipos: (i) primárias: referem-se ao produto de informação produzido diretamente pelo autor como livros, dissertações e artigos; (ii) secundárias: trazem a informação organizada e revelam a participação de um segundo autor como é o caso de dicionários, periódicos de indexação, catálogos e bases de dados; (iii) terciárias: conforme as palavras de Blatmannn “biografias de bibliografias”, trazem uma síntese ou consolidação de informações, sendo exemplos guias e revisões de literatura, diretórios e entre outros.

Nesta seção o foco será nas fontes secundárias, mais especificamente na PL e na base de dados WoS, buscando entender o que as mesmas oferecem, como podem ser acessadas e também expondo algumas de suas particularidades.

### 2.4.1 Plataforma Lattes

A Plataforma Lattes<sup>21</sup>, também conhecida como Lattes ou Currículo Lattes, trata-se de um portal que agrupa os dados acadêmicos e científicos de pesquisadores de todo o Brasil e é mantida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico<sup>22</sup> (CNPq) (MATIAS, 2015). Conforme CNPq (2021a) a PL integra as bases de dados de currículos, grupos de pesquisa e instituições em um único sistema de informações, se tornando um padrão nacional no registro de informações tanto de discentes, docentes e instituições de ensino. Por meio dela e seus dados é possível fazer o mapeamento de toda a genealogia científica dos pesquisadores do país e, conseqüentemente, sendo possível também mapear toda sua produção (DAMACENO; ROSSI; MENA-CHALCO. 2017).

O acesso de suas informações é gratuito e pode ser efetuado de duas formas: o acesso tradicional por meio de sua interface web, onde é feita a consulta e extração de metadados de forma manual em formato RTF ou XML, e o acesso automatizado por meio de uma interface de programação (API), chamada de Lattes Extractor<sup>23</sup>, que fornece os dados em questão somente no formato XML (CNPQ, 2021b). Para efetuar o acesso manual via *web* não é necessária qualquer permissão especial, todavia para efetuar o acesso via Lattes Extractor é necessário efetuar um cadastro e solicitar seu acesso (CNPq, 2021b). Segundo Matias (2015) esse acesso às APIs é fornecido somente para instituições brasileiras de pesquisa e ensino superior, ressaltando ainda que o mesmo só é liberado para um endereço de IP por instituição.

Dada a relevância da PL e os dados disponibilizados diversas ferramentas surgiram para utilizar, processar e/ou mapear os dados disponibilizados por meio de sua API. Alguns exemplos são o syncLattes<sup>24</sup> de Matias (2015) que extrai os dados para popular um repositório institucional; o scripLattes<sup>25</sup> de Mena-Chalco e Cesar-Jr (2009) para extração e exportação dos dados; a Plataforma Acácia<sup>26</sup> de Damaceno, Rossi e Mena-Chalco (2017) para extração e documentação das relações formais de

---

<sup>21</sup> <<http://lattes.cnpq.br/>>

<sup>22</sup> <<https://www.gov.br/cnpq/pt-br>>

<sup>23</sup> <<https://www.gov.br/pt-br/servicos/obter-acesso-ao-extrator-da-plataforma-lattes>>

<sup>24</sup> <<https://github.com/nitmateriais/synclattes>>

<sup>25</sup> <<http://scriplattes.sourceforge.net/>>

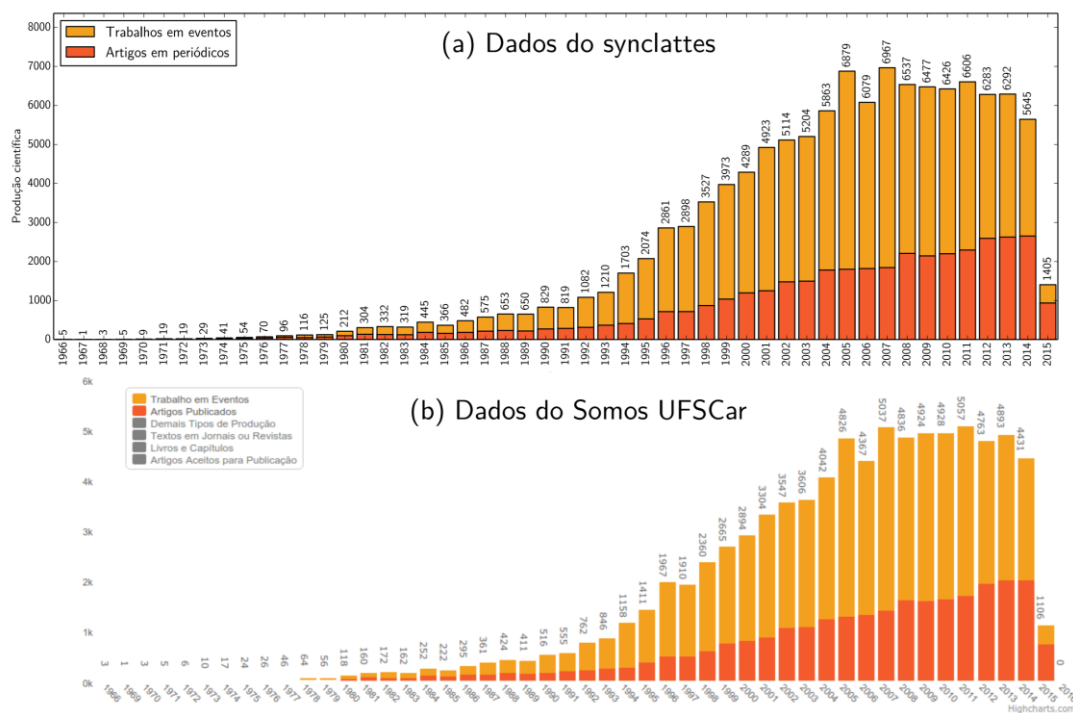
<sup>26</sup> <<http://plataforma-acacia.org/>>



orientações nos programas de pós graduação brasileiro; o lucyLattes<sup>27</sup> para manipulação e extração dos dados desenvolvido por Tieppo (2019); entre muitos outros.

Os dados extraídos da PL e processados por meio dos *softwares* previamente citados podem ser utilizados de modos mais diversos. A Figura 8 retrata um dos principais usos para dados processados, a criação de indicadores bibliométricos para análise das publicações científicas, sendo um dos gráficos gerado pela utilização dos dados processados do syncLattes e outra o gráfico gerado por meio dos dados processados do SOMOS UFSCar<sup>28</sup>. Um dos principais pontos a se observar é que por mais que ambas as ferramentas estejam trabalhando sobre o mesmo conjunto de dados, existem variações referentes às metodologias utilizadas no processamento dos dados. Dito isso, diversas outras análises e tipos de processamentos podem ser feitos sobre os dados da PL.

**Figura 8** – Comparação entre total de publicações por ano obtidas pelo projeto syncLattes em comparação com dados do Somos UFSCar



Fonte: MATIAS (2015).

<sup>27</sup> <<https://github.com/rafatiempo/lucylattes>>

<sup>28</sup> <<http://www.somos.ufscar.br/>>

## 2.4.2 Web of Science

A Web of Science<sup>29</sup>, também conhecida como ISI ou WoS, é um dos maiores bancos de dados multidisciplinar e independente de citação que conecta índices regionais, de especialidades, dados e patentes ao seu principal índice: *Web of Science Core Collection* (CLARIVATE, 2022; BIRKLE et al., 2019). De acordo com Clarivate Analytics (2021) a WoS permite que sejam pesquisados registros entre diversas disciplinas e intervalos temporais em uma plataforma contendo quase 1.9 bilhões de referências citadas em mais de 171 milhões de registros. Birkle et al. (2019) complementa ainda que a WoS “é a base de dados de publicações e citações mais antiga, usada e autoritativa”.

Uma das principais qualidades da mesma é que ela oferece acesso a diferentes índices de citação como por exemplo o *Web of Science Core Collection*, *BIOSIS Citation Index*, *Chinese Science Citation Database*, *Data Citation Index*, *Russian Science Citation Index*, *SciELO Citation Index* e *Arabic Citation Index* (CLARIVATE COVERAGE, 2021; BIRKLE et al., 2019). Também oferece acesso a índices regionais e por assunto e outros recursos como por exemplo: *Medline*, *Inspec*, *Korean Journal Database*, *Zoological Record*, *Derwent Innovation Index* e entre muitos outros (CLARIVATE COVERAGE, 2021).

O principal índice da base é o *Web of Science Core Collection*. O mesmo contém mais de 21.100 periódicos científicos revisados por pares e de alta qualidade publicados no mundo todo em mais de 250 áreas de pesquisa como ciências, ciências sociais, artes e humanidades (CLARIVATE, 2020). Segundo Liu (2019) e Clarivate (2020) o índice consiste de dez subconjunto de dados, sendo oito índices de citação e dois índices de química, conforme pode ser visto no Quadro 7 e no Quadro 8.

**Quadro 7** – Índices de química da Web of Science Core Collection

ÍNDICE DE QUÍMICA	ÁREA DE ATUAÇÃO	DATA DOS REGISTROS
Current Chemical Reactions (CCR-Expanded)	Novos métodos ou etapas para reações químicas	1985 até o presente
Index Chemicus (IC)	Estruturas e dados de suporte críticos para novos compostos orgânicos	1993 até o presente

Fonte: Elaborada pelo autor baseado em Clarivate Core Collection (2021).

<sup>29</sup> <<https://www.webofknowledge.com>>

Quando se volta para a forma de acesso à mesma, existem dois meios de acesso principais. O primeiro trata-se do acesso tradicional por meio do site/interface para a *Web* onde o pesquisador consegue acessar e buscar as informações desejadas de forma manual e interativa, podendo também efetuar alguns processamentos no conjunto de dados diretamente pela interface antes de fazer a exportação dos mesmos. O segundo meio de acesso é feito por meio de interfaces de programação (*Application Programming Interfaces - APIs*)<sup>30</sup> onde os dados são disponibilizados para que sistemas e programas consigam efetuar essa coleta dos dados desejados de maneira automatizada, sem a intervenção humana no processo de coleta, facilitando por exemplo a indexação de diversas publicações em um repositório institucional.

**Quadro 8** – Índices de citação da Web of Science Core Collection

ÍNDICE DE CITAÇÃO	ÁREA DE ATUAÇÃO	DATA DOS REGISTROS
Science Citation Index Expanded (SCIE)	Literatura de periódicos científicos multidisciplinares	1900 até o presente
Social Sciences Citation Index (SSCI)	Literatura de periódicos em ciências sociais	1900 até o presente
Arts & Humanities Citation Index (A&HCI)	Literatura de periódicos em artes e humanidades	1975 até o presente
Conference Proceedings Citation Index - Science (CPCI-S)	Literatura de conferências em todos os campos científicos e técnicos	1990 até o presente
Conference Proceedings Citation Index - Social Sciences & Humanities (CPCI-SSH)	Literatura de conferências em todos os campos das ciências sociais, artes e humanidades	1990 até o presente
Book Citation Index - Science (BKCI-S)	Livros acadêmicos que apresentam artigos referenciados ou revisões de literatura de ciência em geral	2005 até o presente
Social Sciences & Humanities (BKCI-SSH)	Livros acadêmicos que apresentam artigos referenciados ou revisões de literatura de ciências sociais e humanidades	2005 até o presente
Emerging Sources Citation Index (ESCI)	Literatura de periódicos internacionais e de escopo geral que oferecem cobertura regional ou especialidade	2005 até o o presente

Fonte: Elaborada pelo autor baseado em Clarivate Core Collection (2021).

Para que seja possível acessar suas diversas coleções e índices é necessário pagar uma assinatura pelo seu uso (JOB, 2018). A Capes atualmente financia um consórcio para que as universidades brasileiras pertencentes ao mesmo possam ter acesso ao conteúdo da WoS e de outras bases de dados (CAPES, 2021). Um ponto importante a ser ressaltado é que dependendo do tipo de contrato efetuado alguns

<sup>30</sup> <<https://developer.clarivate.com/apis>>

índices e seus sub índices podem ou não serem disponibilizados conforme pode ser visto na Figura 9 e na Figura 10. Conforme CAPES (2021) o consórcio oferece acesso a alguns dos índices da coleção principal da base de dados de acordo com o demonstrado na Figura 9, não oferecendo portanto acesso à totalidade desse índice. Ponto importante a ser ressaltado é que a permissão de acesso a um índice por meio da interface web não necessariamente implica na permissão de acesso à mesma API, sendo necessário entrar em contato com o suporte da base para contratação e obtenção da mesma .

**Figura 9** – Índices disponíveis na *Web of Science Core Collection* por meio do acesso Capes.

Principal Coleção do Web of Science: Índice de citações

- Science Citation Index Expanded (SCI-EXPANDED) --1945-presente
- Social Sciences Citation Index (SSCI) --1956-presente
- Arts & Humanities Citation Index (A&HCI) --1975-presente
- Conference Proceedings Citation Index - Science (CPCI-S) --1990-presente
- Conference Proceedings Citation Index - Social Science & Humanities (CPCI-SSH) -  
-1990-presente
- Emerging Sources Citation Index (ESCI) --2015-presente

Fonte: Elaborado pelo autor.

**Figura 10** – Índices disponíveis na *Web of Science Core Collection* para a Universidade de Manchester por meio do acesso completo.

Web of Science Core Collection: Citation Indexes

- Science Citation Index Expanded (SCI-EXPANDED) --1900-present
- Social Sciences Citation Index (SSCI) --1900-present
- Arts & Humanities Citation Index (A&HCI) --1975-present
- Conference Proceedings Citation Index- Science (CPCI-S) --1990-present
- Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH) --1990-present
- Book Citation Index- Science (BKCI-S) --2005-present
- Book Citation Index- Social Sciences & Humanities (BKCI-SSH) --2005-present
- Emerging Sources Citation Index (ESCI) --2015-present

Web of Science Core Collection: Chemical Indexes

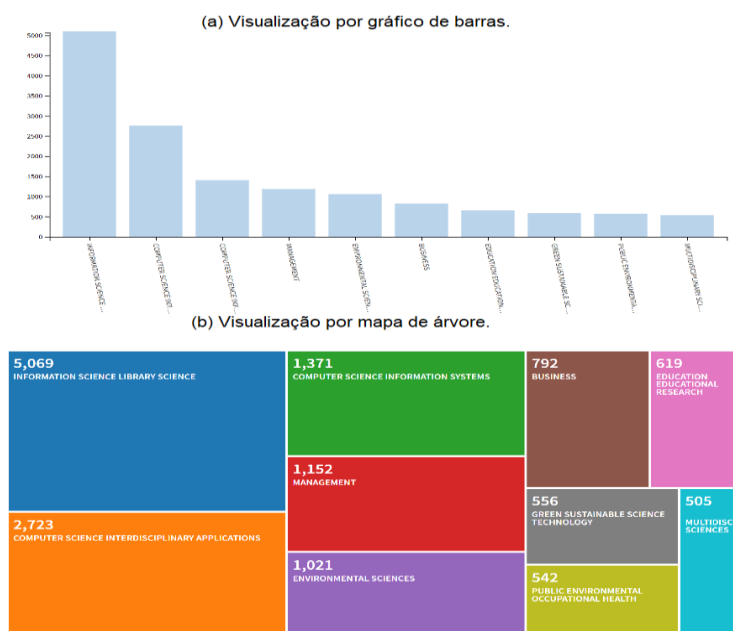
- Current Chemical Reactions (CCR-EXPANDED) --1993-present  
(Includes Institut National de la Propriete Industrielle structure data back to 1840)
- Index Chemicus (IC) --1993-present

Fonte: LIU, 2019.

Sob o prisma dos formatos disponibilizados para exportação das publicações também existem diferenças entre o mesmo quando efetuado pela interface *Web* e pela interface de programação. Pela interface *Web* é possível exportar para diversas ferramentas e formatos. Algumas das possibilidades de exportação da coleção principal da WoS: (i) EndNote Desktop; (ii) EndNote Online; (iii) Publons<sup>31</sup>; (iv) InCites<sup>32</sup>; (v) Outros formatos como HTML, texto plano, delimitado por tabulação e entre outros; (vi) RefWorks; (vii) BibTex; (viii) Impressão; (ix) Excel; (x) entre outros (CLARIVATE, 2022). Já para as APIs disponíveis o formato é único e padronizado, sendo utilizado o formato de metadados XML.

Por fim, é importante ressaltar que a WoS fornece mecanismos para se analisar os dados por meio da sua própria plataforma web. Apesar de simples, a mesma oferece dois tipos de visualização para os resultados de suas buscas, a visualização por gráficos de barras (a) e a visualização por mapa de árvore (b), conforme pode ser visto na Figura 11. Sobre o prisma de processamento, os únicos possíveis de serem feitos pela plataforma web é a filtragem de elementos por meio da exclusão de resultados indesejados ou efetuando uma busca dentro dos resultados da busca anterior.

**Figura 11** – Visualizando um conjunto de publicações na Web of Science



Fonte: Elaborada pelo autor.

<sup>31</sup> <<https://publons.com/>>

<sup>32</sup> <<https://incites.clarivate.com/>>

## 2.5 Ferramentas para auxílio em análises bibliométricas

Conforme fora apresentado no início desse capítulo, a bibliometria é área de notável importância para o avanço científico e, devido a isso, em conjunto com a gigantesca quantidade de dados disponíveis nos dias atuais nas bases de dados presentes na web, inúmeras ferramentas que auxiliam esse tipo de análise estão surgindo.

Nesta seção será apresentada algumas dessas ferramentas, sendo importante ressaltar que, pelo fato de existirem centenas de outras ferramentas, foram selecionadas as ferramentas que mais tinham relação com este trabalho e conhecidas no campo científico.

### 2.5.1 VOSviewer

O VOSviewer<sup>33</sup>, também chamado de VOS, pode ser considerado uma das principais ferramentas utilizadas em estudos bibliométricos onde o foco é o mapeamento da ciência. Ele é capaz de atuar em todas as três etapas de um *software* bibliométrico coletando dados diretamente de bases de dados, processando esses dados e fornecendo visualizações para análises por meio de redes (MORAL-MUÑOZ et al., 2020). Conforme definem os autores Van Eck e Waltman (2010, p. 524):

O VOSviewer é um programa que nós desenvolvemos para construção e visualização de mapas bibliométricos. O programa está disponível gratuitamente para a comunidade de pesquisa bibliométrica. O VOSviewer pode ser usado por exemplo para construir mapas de autores e jornais baseados em dados de citação ou construir mapas de palavras-chave baseados em dados de coocorrência. O programa oferece uma visualização que permite com que os mapas bibliométricos possam ser analisados em full detalhe. O VOSviewer consegue representar um mapa de diferentes maneiras, cada uma focando em um diferente aspecto.

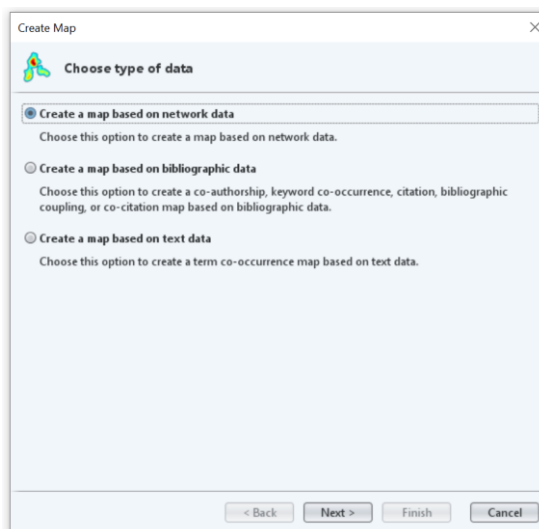
Explorando um pouco as características do mesmo, o primeiro ponto a ser analisado é sobre os dados utilizados em suas análises. Conforme Van Eck e Waltman (2018) expõem em seu manual técnico, e também pela Figura 12, a ferramenta

---

<sup>33</sup> <<https://www.vosviewer.com/>>

consegue trabalhar tanto com redes já processadas por outros *softwares*, sendo um exemplo o M2N (SARVO et al., 2020) que transforma matrizes em redes, ou com dados puros de publicações científicas a serem processadas.

**Figura 12** – Tipos de dados aceitos no *software* bibliométrico VOSviewer



Fonte: Elaborada pelo autor.

Consequente, para a obtenção de tais dados a serem analisados, é oferecida a possibilidade de obter os mesmos por meio das APIs disponibilizadas por algumas bases de dados (Figura 13), lembrando ser necessário ter uma conta com permissão na mesma para que isso seja possível, ou fornecendo arquivos obtidos das bases de dados WoS, Scopus, PubMed<sup>34</sup> e Dimensions<sup>35</sup> nos seus formatos padrões e também arquivos resultantes de softwares de referências como o EndNote entre outros (VAN ECK; WALTMAN, 2011; 2018). Para Moral-Muñoz *et al.* (2020, p. 1) o fato de o VOSviewer conseguir exportar e carregar informações de diferentes fontes é sua principal característica.

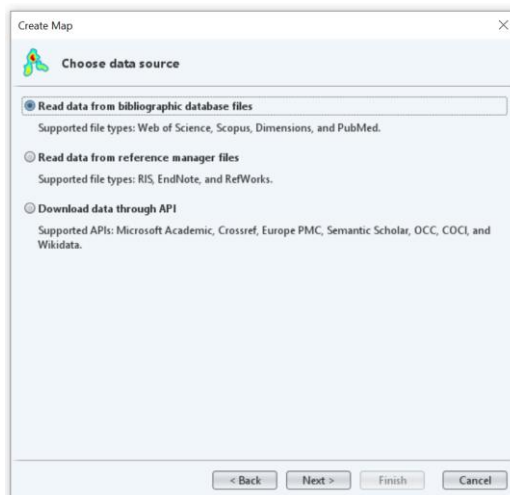
No quesito processamento dos dados bibliográficos, o principal ponto que há de ser relatado é que não há a possibilidade de efetuar o processamento por meio do mesmo e posteriormente utilizar as listas e matrizes processadas em outro software para diferentes análises, ficando as mesmas vinculadas ao VOSviewer e suas visualizações (VAN ECK; WALTMAN, 2011). As escolhas de processamento dos

<sup>34</sup> <<https://pubmed.ncbi.nlm.nih.gov/>>

<sup>35</sup> <<https://www.dimensions.ai/>>

dados se limitam à seleção e utilização de tesouros e a ocultação de um ou mais registros no processo de gerar a rede.

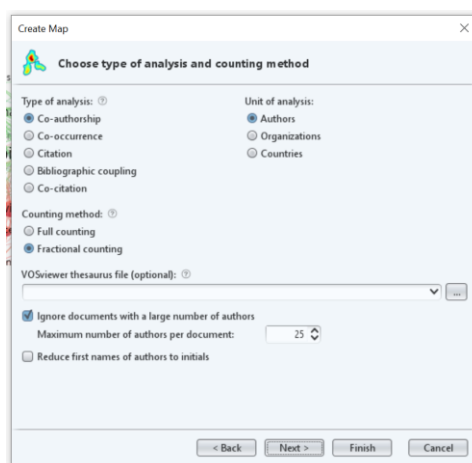
**Figura 13** – Escolhendo uma fonte de dados no *software* bibliométrico VOSviewer.



Fonte: Elaborada pelo autor.

Já sobre a geração das redes, o usuário tem duas opções: fornecer seus dados processados ou processar por meio do programa. Caso o usuário já tenha seus dados processados só é possível gerar uma rede com base em coocorrência, caso faça o processamento pelo *software* é oferecida uma gama muito maior de redes (Figura 14) como análises por coautoria, coocorrência, citação, acoplamento bibliométrico e cocitação, podendo ainda selecionar o tipo de contagem completa ou fracionária.

**Figura 14** – Diferentes tipos de análises oferecidas para dados processados no *software* VOSviewer.

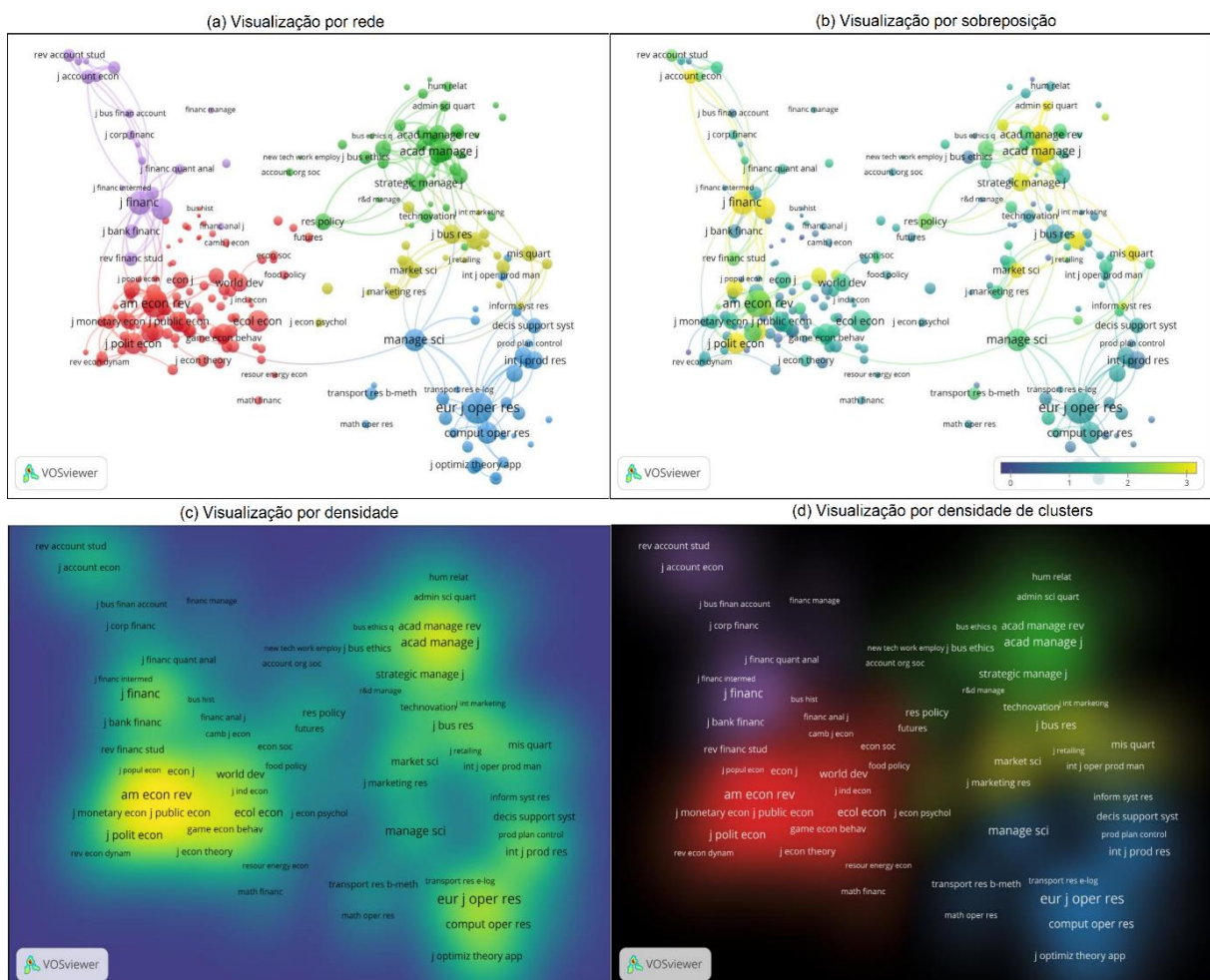


Fonte: Elaborada pelo autor.



Sob o prisma da visualização dos dados bibliométricos e suas redes, o *software*, de acordo com o exposto por Van-Eck e Waltman (2018) em seu manual, oferece três tipos diferentes de visualizações: visualização por redes, visualização por sobreposição e visualização por densidade. Vale ressaltar que a visualização por densidade apresenta duas variações, a variação tradicional e a visualização por densidade de clusters, conforme pode ser visto na Figura 15. O usuário ainda pode, de acordo com suas necessidade, alterar questões estéticas como as cores dos clusters, os tipos de linhas (retas ou curvas), o tamanho geral das ligações e elementos (círculos) e entre outras para facilitar a visualização (MORAL-MUÑOZ et al., 2020).

**Figura 15** – Os quatro diferentes tipos de visualizações disponíveis no programa VOSviewer.



Fonte: Adaptado de Van Eck e Waltman (2018).

Finalmente, podem ser exemplificados estudos que utilizaram a ferramenta para expor seus resultados. Alguns exemplos são o de Yu et al. (2020) que efetua uma análise bibliométrica em publicações sobre COVID-19 e utiliza a ferramenta para a construção de uma rede de palavras-chave relacionadas ao tema e o de Xie et. al. (2020) que utiliza a ferramenta para construção de diversos indicadores que retratam a temática de cirurgias da coluna antloxial, expondo a colaboração de países e universidades sobre a temática e também a relevância de palavras-chave ao longo dos anos.

### 2.5.2 syncLattes

Quando se busca medir o campo científico brasileiro, uma das principais bases de dados utilizadas é a PL. Dada a relevância da mesma, diversas pesquisas são feitas com base em seus dados e, conseqüentemente, a fim de facilitar tais estudos, existem ferramentas como o scripLattes e o syncLattes.

O syncLattes é uma ferramenta bibliométrica automatizada utilizada para extração de metadados provenientes da PL, sendo mais específico do Currículo Lattes, para povoamento de repositórios institucionais, composto por um conjunto de scripts em linguagem Python que atuam na extração, tratamento e sincronização dos metadados do Currículo Lattes com o DSpace<sup>36</sup> (MATIAS, 2015). É importante ressaltar o seguinte ponto apresentado por Soares (2019): “o syncLattes não tem por objetivo fornecer recursos visuais, mas sim extrair e normalizar os dados dos currículos”.

A ferramenta apresentava dois diferenciais na época em que foi elaborada, a capacidade de deduplicação e a utilização direta da API da PL. Sobre a deduplicação, era a capacidade da ferramenta de remover duplicatas de publicações durante o processamento dos currículos dos pesquisadores presentes no Currículo Lattes, garantindo portanto que exista somente um registro referente a uma publicação nos arquivos gerados, por mais que a mesma esteja presente no currículo de diversos pesquisadores (MATIAS, 2015). Já sobre a questão de utilização direta da API, essa foi outra questão de extrema importância, uma vez que as outras ferramentas da época (como o scripLattes que será explanado em seguida) extraíam diretamente da

---

<sup>36</sup> <<https://duraspace.org/dspace/>>



MACIEL et al., 2017) e também para avaliar a situação atual da produção científica brasileira (MACIEL, 2018).

### 2.5.3 scriptLattes

O scriptLattes<sup>37</sup> é uma ferramenta desenvolvida por Mena-Chalco e Cesar-Jr que, conforme os autores explanam (MENA-CHALCO; CESAR-JR, 2009), tem como objetivo a extração e processamento automático de dados de pesquisadores cadastrados na PL de produções bibliográficas, produções técnicas, produções artísticas, orientações, projetos, de pesquisa, prêmios e títulos, grafo de colaborações, mapa de geolocalização, coautoria e internacionalização.

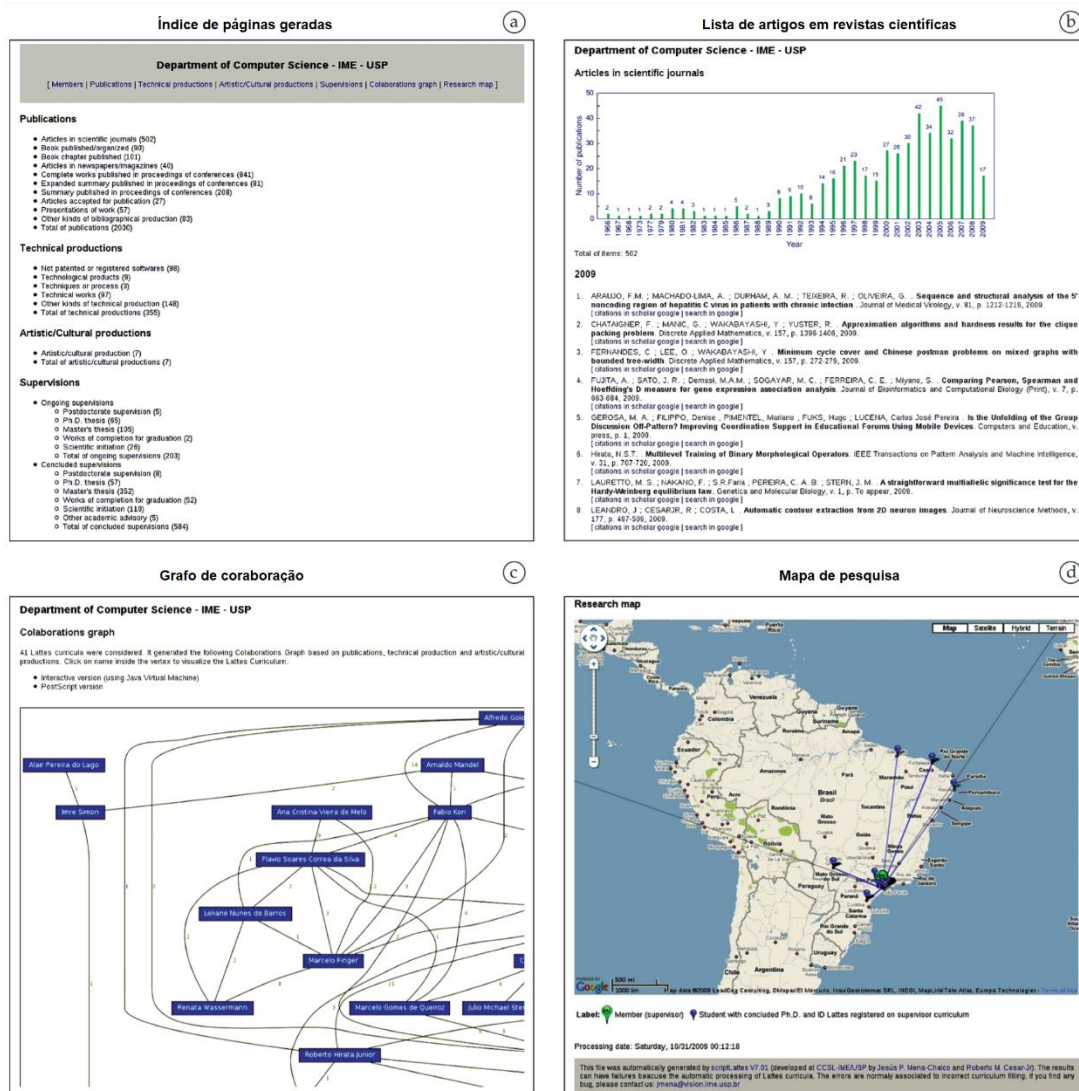
Sobre as saídas ou resultados do mesmo, diferentemente do syncLattes que tem como foco a população de um repositório institucional, conforme pôde ser visto, o scripLattes vai além de extrair os dados dos currículos dos pesquisadores, ele também fornece suas próprias análises com base nos dados extraídos por meio de redes, grafos e seus relatórios (Figura 17). É necessário enfatizar que as saídas da ferramenta scripLattes não se limitam somente aos relatórios, a mesma ainda fornece os registros baixados em diferentes formatos como o RIS, CSV e JSON (MENA-CHALCO; CESAR-JR, 2013).

Por fim, é válido ressaltar que, apesar de ser uma das primeiras ferramentas a serem desenvolvidas com o objetivo de abordar os dados presentes na PL, o mesmo ainda continua relevante até o presente momento, sendo utilizado de forma auxiliar, na extração e visualização de conteúdos da PL, ou objeto de análise reflexiva de diversas pesquisas (MOREIRA; MULLER, VILAIN FILHO, 2020; SOBRAL et al., 2020).

---

<sup>37</sup> <<http://scriptlattes.sourceforge.net/>>

Figura 17 – Exemplos de relatórios disponíveis ao utilizar a ferramenta scriptLattes



Fonte: Adaptado de Mena-Chalco e Cesar-Jr (2009).

## 2.5.4 VantagePoint

O VP é uma ferramenta bibliométrica para *desktop* com o foco em mineração textual, desenvolvida pela empresa *Search Technology, Inc*, capaz de analisar dados tanto de bancos de dados de patentes e quanto de literaturas (PORTER; CUNNINGHAM, 2004). Cobo et. al. (2011) ainda complementa que a ferramenta não só consegue trabalhar com dados de patentes e literaturas como também com qualquer tipo de dado estruturado. Isso se deve ao fato de, conforme a documentação técnica, a mesma oferecer um mecanismo dinâmico de importação onde o usuário, caso tenha o devido conhecimento de expressões regulares, é capaz de adaptar a importação de acordo com suas necessidades, apesar de ainda existir algumas

limitações (SEARCH TECHNOLOGY, INC. 2015).

Explorando as características do VP, o primeiro e principal ponto que necessita ser ressaltado é que, diferente das demais ferramentas expostas nesta pesquisa, o VP é uma ferramenta de uso comercial e, portanto, paga, sendo assim necessária uma assinatura para o acesso e utilização da mesma. Conforme Cobo *et al.* (2011) e a dupla Sangam e Mogali (2012), podem ser elencadas as algumas características relevantes dessa ferramenta:

- Apresenta 180 filtros de importação que permitem ao usuário importar dados de praticamente todos os bancos de dados de patentes e literaturas.
- Os campos mostrados pela ferramenta variam de acordo com o dado importado. Por exemplo, em dados de publicações científicas pode fornecer campos como autores, referências e entre outros enquanto para patentes pode fornecer o título da patente, ano, código e muito mais.
- A partir dos campos existentes nos registros informados é possível construir listas (Figura 18), como por listas de autores, que podem ser divididas e analisadas em diferentes grupos.

**Figura 18** – Uma lista de autores e uma matriz de coocorrência de autores gerados na ferramenta VP.

**Lista de Autores**

	# Records	# Instances	Autores (AU)
1	51	51	MERIGO, JM
2	30	30	COBO, MJ
3	28	28	BORNIMANN, L
4	25	25	HERRERA-VIEDMA, E
5	25	25	KUMAR, S
6	18	18	LEYDESORFF, L
7	17	17	YEUNG, AWK
8	16	16	XU, ZS
9	15	15	LI, J
10	15	15	SWELEH, WM

**Matriz de co-ocorrência de autores**

Autores (AU)	# Records	1	2	3	4	5	6	7	8	9	10
MERIGO, JM	51	51	4				6				
COBO, MJ	30	4	30				19				
BORNIMANN, L	28			28			12				
KUMAR, S	25				25						
HERRERA-VIEDMA, E	25	6	19		25						
LEYDESORFF, L	18			12		18					
YEUNG, AWK	17						17				
XU, ZS	16								16		
SWELEH, WM	15									15	
LI, J	15										15

Fonte: Elaborada pelo autor.

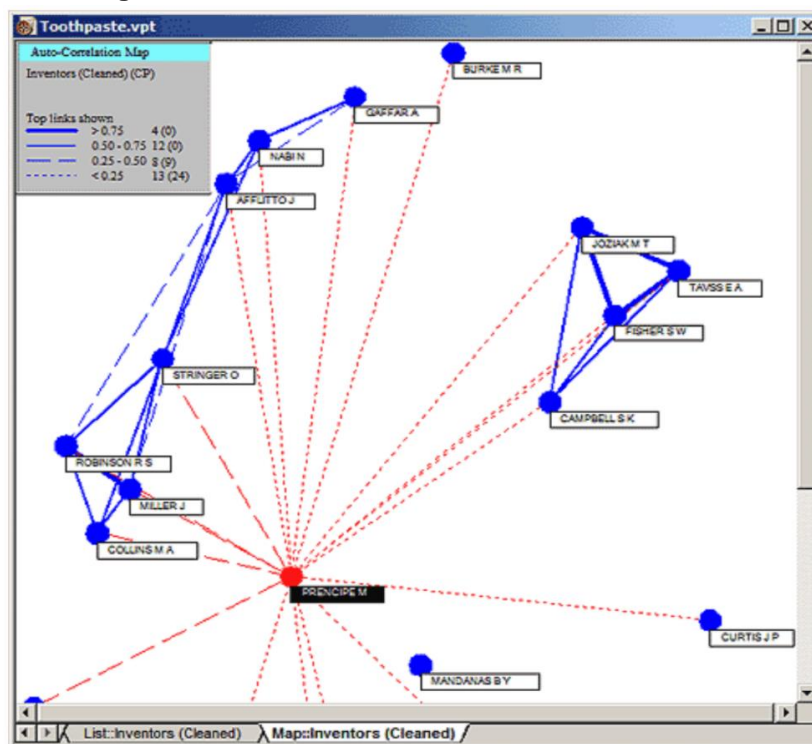
- O mecanismo de pré-processamento e manipulação dos dados importados é robusto, oferecendo mecanismos de limpar as listas geradas e removendo duplicatas por meio de técnicas *fuzzy* de programação e também pela aplicação de um arquivo de tesouro criado pelo usuário ou pela própria ferramenta de acordo com os comandos do mesmo.
- Além das listas, a ferramenta também é capaz de criar diversos tipos de



matrizes (Figura 18) como matrizes de coocorrência, autocorrelação, correlação, de fator e entre muitas outras.

- Com as matrizes criadas é possível criar também mapas visuais (redes) por meio da ferramenta (Figura 19).

**Figura 19** – Rede de autores criada na ferramenta VP.



Fonte: SANGAM; MOGALI, 2012.

Dada a relevância do VP, tanto para o meio científico quanto para o mercado privado, alternativas estão sendo desenvolvidas e apresentadas, como é o caso do Harpia desenvolvido por Sampaio (2021). Apesar de ainda se encontrar em estágios iniciais e não ter toda a robustez do VP, a ferramenta Harpia, que também é uma ferramenta executada diretamente no computador do pesquisador, busca ser uma alternativa ao mesmo solucionando a problemática do acesso, uma vez que é totalmente gratuita, tendo como foco a importação de registros da base de dados WoS e a geração de listas (SAMPAIO, 2021).

A relevância do VP é demonstrada também no gigantesco universo de publicações científicas que utilizam a ferramenta como mecanismo de mineração textual, extraíndo e analisando conteúdos de publicações. Exemplificando algumas de suas utilizações, pode-se citar alguns contextos como a análise de patentes (MA; PORTER, 2015), onde a ferramenta é utilizada para fazer mineração textual em

registros de patentes; a análise de publicações científicas, onde a mineração textual é aplicada em publicações contidas em diversas bases de dados (COBO et. al., 2011; ALENCAR; PORTER; ANTUNES, 2007)

### 2.5.5 Bibliometrix

O Bibliometrix é uma ferramenta bibliométrica instalável e *open source* desenvolvida na linguagem de programação R com o foco no mapeamento da ciência que fornece um conjunto de métodos para estudos quantitativos nas áreas da cientometria e bibliometria (ARIA; CUCCURULLO, 2017; 2021). O mesmo é muitas vezes referenciado como Biblioshiny, que é sua interface web para visualização de diferentes tipos de indicadores como gráficos de barras, redes de colaboração e muitas outras (PATIL, 2020). Moral-Muñoz *et al.* (2020) completa “o biblioshiny é o bibliometrix para não programadores”. Apesar de o mesmo ter uma interface web, o mesmo não é uma ferramenta que roda nesse ambiente, sendo necessária sua instalação e configuração na máquina do pesquisador.

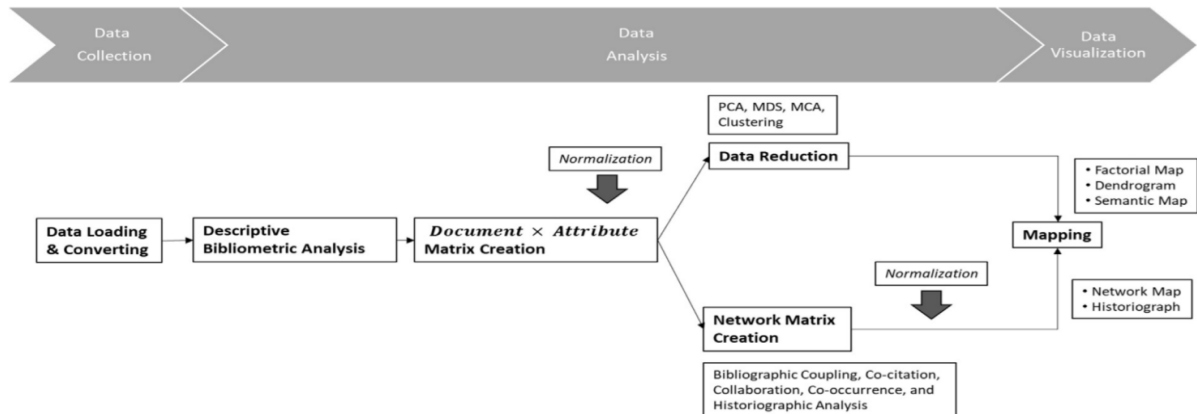
Conforme pode ser visto na Figura 20, o Bibliometrix atua em todas as três etapas de um *software* bibliométrico, primeiramente coletando e carregando os dados de publicações bibliográficas, posteriormente analisando e processando as mesmas e finalmente oferecendo formas de visualização das mesmas. Para Moral-Muñoz *et al.* (2020) o principal diferencial do bibliometrix / biblioshiny é sua capacidade de oferecer diferentes indicadores já que, por ser uma das ferramentas mais recentes, a mesma se preocupou em incorporar diversas das análises já oferecidas por outras ferramentas bibliométricas.

Aprofundando um pouco no processo de coleta e conversão de dados, o Bibliometrix, apesar de os desenvolvedores da ferramenta definirem em seu manual técnico que o mesmo efetua análises bibliométricas em conjuntos de dados da WoS e Scopus (ARIA; CUCCURULLO, 2021), Moral-Muñoz et. al. (2020) expõe que conjuntos de dados da Dimensions, The Lens, Cochrane e PubMed também podem também serem utilizados. Todos os dados a serem utilizados devem ser coletados diretamente do banco de dados selecionado pelo usuário e importados no Bibliometrix, com exceção da SCOPUS que pode ser feito tanto a coleta manual como a coleta por API. Com os dados fornecidos o *software* então faz a conversão em R



para que possam ser analisados e processados nas etapas posteriores (ARIA; CUCCURULO, 2021).

**Figura 20** – Bibliometrix e o fluxo de trabalho adequado para o mapeamento científico.



Fonte: ARIA; CUCCURULLO, 2017.

Mesmo sendo uma ferramenta recente, o Bibliometrix vem ganhando cada vez mais relevância no contexto da bibliometria, sendo utilizada por um número cada vez maior de pesquisadores tanto da área em questão quanto de outras (PERISSINI, FARIA; 2021). Apesar de atuar em todas as etapas de um software bibliométrico, grande parte das publicações que citam e utilizam o Bibliometrix, fazem o mesmo buscando utilizar algumas de suas visualizações disponíveis em sua interface *web*. Dois exemplos são a publicação de Campra, Esposito e Brescia (2020), que utiliza a ferramenta e sua visualização de ‘mapa-múndi’ para demonstrar visualmente os países com maior quantidade de publicações sobre COVID-19 e suas colaborações no tema, e a publicação de Rodríguez-Soler, Uribe-Toril e Valenciando (2020), que busca analisar tendências nas publicações sobre a temática do despovoamento rural por meio da utilização de outras visualizações disponíveis na ferramenta como redes de cocitação, evolução temática ao longo do tempo, gráficos de barras e entre outras.

## 2.5.6 Outras ferramentas

Conforme demonstrado, as ferramentas bibliométricas podem ser desenvolvidas para abordar diferentes questões como a extração, visualização, processamento ou até mesmo todos eles ao mesmo tempo. O Quadro 9 contém algumas das ferramentas bibliométricas já abordadas e também outras que são relevantes para os estudos métricos, detalhando algumas de suas características como o seu tipo de acesso, seus requisitos e sua crescente (ou não) utilização em estudos bibliométricos ao longo do tempo,

**Quadro 9** – Ferramentas bibliométricas, tipo de acesso, principais usos e publicações ao longo dos anos.

Ferramenta	Termos de busca	Tipo de acesso	Principais usos	N° de publicações de 1945 a 2020	N° de publicações de 1945 a 2020 em bibliometr*	Porcentagem de publicações bibliométricas no total	N° de publicações em 2011 e bibliometr*	N° de publicações em 2015 e bibliometr*	N° de publicações bibliometr* em 2020 e bibliometr*
VOSviewer	"vos viewer" OR vosviewer OR "vos-viewer"	Grátis	Geração e visualização de redes	917	756	82,44%	3	15	389
Pajek	pajek	Grátis	Geração e visualização de redes	184	40	21,74%	3	7	6
Bibliometrix	bibliometrix OR biblioshiny	Grátis	Geração e visualização de redes e indicadores diversos	76	76	100,00%	0	0	61
CiteSpace	citespace OR "CiteSpace"	Grátis	Geração e visualização de redes e indicadores diversos	721	388	53,81%	1	13	172
Gephi	gephi	Grátis	Geração e visualização de redes	275	47	17,09%	0	3	24
R-Studio	rstudio OR "r studio" OR "r-studio"	Grátis	Linguagem de programação focada em estatística e gráficos	470	14	2,98%	0	0	12
BibExcel	bibexcel OR "bib excel"	Grátis	Mineração textual, geração e visualização de redes	82	58	70,73%	0	4	18
Scimago	scimago	Grátis	Visualização de indicadores diversos	403	167	41,44%	5	14	21
HistCite	histcite	Pago	Mineração textual, geração e visualização de redes e indicadores diversos	108	74	68,52%	0	5	28
SciMAT	scimat	Grátis	Geração e visualização de redes	88	76	86,36%	0	0	43

Fonte: PERISSINI; FARIA, 2021.

Existem ainda inúmeras outras ferramentas, tanto existentes ou que estão em desenvolvimento, que poderiam ser objeto de reflexão desta pesquisa, porém, dado o escopo de um projeto de mestrado, torna-se inviável abordar todas elas no presente momento.

## **2.6 Democratização do acesso e conhecimento e seu contexto em ferramentas bibliométricas**

A palavra democratização e sua contextualização é tema de extrema relevância e, conforme sua definição linguística, apresenta duas interpretações: (i) pode ser interpretada como o ato de se tornar algo democrático segundo preceitos democráticos (contexto político) e (ii) pode ser interpretada como o ato de tornar algo acessível e popular (contexto da popularização). (DEMOCRATIZAÇÃO, 2022). No contexto dessa pesquisa a democratização é abordada sobre o seu contexto social e de caráter inclusivo da popularização, expandindo o conhecimento e o acesso a um número maior de pessoas.

A democratização do conhecimento científico é processo essencial para o constante avanço do estado da arte e para que os resultados das pesquisas possam ser usufruídos por aquele que necessita consumir o saber da mesma, uma vez que o conhecimento só é democratizado quando publicado (WITTER, 1997). No contexto dos estudos métricos, em pesquisas onde o foco é utilizar de metodologias bibliométricas para expor um problema ou solução (GLANZEL, 2003), é possível notar a importância da presença de ferramentas que auxiliem nesse tipo de estudo e suas análises; com elas estudos que poderiam ser caracterizados como inviáveis se tornam algo viável, possibilitando a análise de extensos conjuntos de publicações e seus metadados de modo automatizado como nos casos das pesquisas de Esposito e Brescia (2020), Sobra et al. (2020) e muitas outras. Perissini e Faria (2021) ainda expõem em seu estudo que pesquisas de cunho bibliométrico cada vez mais utilizam de ferramentas para auxiliar na sua elaboração com o passar dos anos, demonstrando a crescente relevância das mesmas para a área e viabilizando cada vez mais publicações, consequentemente auxiliando na democratização do conhecimento.

Apesar de as ferramentas terem notável relevância nesse contexto, existe um problema frequentemente enfrentado ao se tentar utilizar as mesmas: a

democratização de seu acesso. A democratização do acesso pode ser caracterizada como a “massificação” de algum objeto, onde o mesmo pode sofrer diversas transformações de modo a buscar com que o elemento de reflexão possa ser acessado e utilizado por um conjunto cada vez maior de pessoas (ALMEIDA et al., 2012). No contexto das ferramentas, é possível elencar barreiras no acesso das mesmas tanto do ponto de vista geral, sem colocar em foco as necessidades de nenhum grupo específico, quanto do ponto de vista específico, como o caso do campo científico brasileiro, o que dificulta sua popularização para um conjunto de pessoas.

Para o contexto geral, várias questões podem ser levantadas, como o fato de algumas das ferramentas (i) serem pagas ou terem algum tipo de licenciamento restritivo; (ii) necessitarem de instalação e/ou configuração; (iii) necessitarem de conhecimentos prévios de alguma área do conhecimento; (iv) não apresentarem documentação não técnica; (v) entre outros (COBO, et al. 2011; MORAL-MUÑOZ, 2020). Para o contexto específico do campo científico brasileiro, pode-se ponderar, a princípio, dois pontos sobre as ferramentas: (i) poucas dão suporte ao idioma “Português”; (ii) poucas oferecem mecanismos de se trabalhar com publicações e dados provenientes de bases de dados e plataformas brasileiras como a PL (PERISSINI; FARIA, 2022).

Algumas ferramentas até atendem alguns desses pontos como o VP que, conforme a documentação dos pesquisadores e desenvolvedores (SEARCH TECHNOLOGY, 2022), por meio de seus filtros dinâmicos é possível importar dados da PL e entre outras. Porém, a ferramenta em questão acaba pecando em outros pontos como até mesmo nos próprios filtros, onde é necessário que o pesquisador tenha um conhecimento específico para a correta configuração das expressões regulares e sua utilização; e também no fato de a mesma ser paga, de necessitar de uma infraestrutura para instalação e não oferecer suporte ao idioma “Português”. Essa dualidade pode ser vista em diversas outras ferramentas como o Bibliometrix que é grátis, possui extensa documentação, porém precisa ser instalado e apresenta certa complexidade em sua configuração (ARIA; CUCCURULLO, 2022). Por fim, vale ainda ser dito que existem ferramentas que estão buscando melhorar alguns desses pontos, como caso do VOSviewer que agora tem também uma versão “*web start*”, onde não é necessária sua instalação e configuração, o que facilita a sua utilização para o pesquisador sem conhecimento técnico de informática, mas não supre o problema de que o mesmo ainda necessitar de um computador/infraestrutura para executar a

mesma, diferente de uma ferramenta disponível diretamente em um servidor na *web*.

É inegável a importância e a relevância das ferramentas introduzidas na seção anterior, dado que elas contribuem de diferentes formas com a democratização de análises bibliométricas; porém, conforme exposto, nota-se também que existem pontos que poderiam ser melhorados para que as mesmas consigam atender uma gama ainda mais de pessoas, independente de sua condição financeira, intelectual e regional; e, conseqüentemente, proporcionar que ainda mais estudos bibliométricos possam ser conduzidos, contribuindo cada vez mais com a popularização do conhecimento e de análises bibliométricas.

### 3 MATERIAIS E MÉTODOS

Ao se desenvolver uma pesquisa científica uma série de procedimentos são realizados até que seja possível atingir o objetivo almejado. Porém, é essencial que, antes mesmo de se iniciar os mesmos, sejam definidos fatores intrínsecos à própria pesquisa como sua abordagem, tipologia e os materiais que possibilitaram sua execução.

#### 3.1 Abordagem e tipologia de pesquisa

A pesquisa em questão foi desenvolvida no núcleo de pesquisa NIT/Materiais da UFSCar, que engloba pesquisadores de dois departamentos de grande relevância na instituição, os departamentos de Engenharia de Materiais e Ciência da Informação. O mesmo e seus membros foram peças fundamentais no desenvolvimento da pesquisa, expondo sugestões, requisitos, soluções e conhecimento para que fosse possível a conclusão da mesma.

Com base nas teorias de Gil (2008) foram definidos a natureza, abordagem e procedimentos da pesquisa. Quanto sua natureza, trata-se de uma pesquisa aplicada, dado que busca desenvolver soluções práticas e aplicáveis à sociedade por meio da construção de uma ferramenta bibliométrica para a *web*, fazendo uso da abordagem empírica e do levantamento bibliográfico como procedimento para obtenção das teorias da temática e de ferramentas bibliométricas disponíveis, tendo como procedimentos a utilização da pesquisa-ação, onde membros do grupo de pesquisa NIT/Materiais foram responsáveis pela definição de requisitos necessários para que fosse possível construir uma ferramenta bibliométrica relevante e que contribuísse com democratização de análises bibliométricas.

De modo a analisar o cumprimento dos requisitos elencados a serem disponibilizados pela ferramenta foram realizadas várias validações, conforme pode ser visto no Quadro 10, onde são apresentados os objetivos das validações, sua metodologia, o conjunto amostral que foi utilizado e se houve a utilização de filtros quando comparado com o *software* VP. O motivo de parte das comparações terem sido realizadas com o mesmo se deve ao fato de o VP ser ferramenta já consolidada na área da bibliometria e também por oferecer mecanismos de se importar publicações da WoS em formato Excel e por conseguir trabalhar com publicações

obtidas do syncLattes por meio de seus filtros.

**Quadro 10 – Conjunto de validações a serem realizadas no BiblioAnalytics e as metodologias empregadas para geração de resultados**

Validação a ser realizada	Objetivo	Metodologia	Conjunto amostral da WoS	Conjunto amostral da PL	Filtro utilizado p/ WoS?	Filtro utilizado p/ PL?
Contagens após a extração dos metadados	Verificar a correta importação e extração dos metadados relevantes para a ferramenta, garantindo a consistência do conjunto de publicações importadas em análises posteriores	Comparar contagens e elementos distintos na geração das listas de "Publication Type" após a importação no BiblioAnalytics e VP	7104 publicações sobre "bibliometrics" obtidas por meio dos arquivos Excel de "Registro Completo" disponíveis na WoS	381 publicações do DCI da UFSCar obtidas da PL por meio da ferramenta syncLattes	Não foi utilizado	Não foi utilizado
Contagens após a realização de todos os processamentos	Demonstrar a importância da utilização de processamentos nas publicações e comparar as contagens com os processamentos feitos pelo BiblioAnalytics com as contagens oferecidas pelo VP e seu filtro	Gerar e comparar as contagens das listas de "Organizations" da WoS e de "Authors" da PL após todos os processamentos	7104 publicações sobre "bibliometrics" obtidas por meio dos arquivos Excel de "Registro Completo" disponíveis na WoS	381 publicações do DCI da UFSCar obtidas da PL por meio da ferramenta syncLattes	ISI – Web of Science v1.6	syncLattes v1.3
Contagens com a aplicação de um tesauro	Demonstrar a importância da utilização de tesouros e verificar se sua aplicação ocorre corretamente no BiblioAnalytics	Comparar a geração de uma lista de "Organizations" da WoS pelo BiblioAnalytics antes e após a aplicação do tesauro em uma instituição	7104 publicações sobre "bibliometrics" obtidas por meio dos arquivos Excel de "Registro Completo" disponíveis na WoS	Não foi utilizado	Não foi utilizado	Não foi utilizado
Contagens na geração de uma matriz	Verificar a correta contagem de coocorrências de elementos durante a geração de uma matriz	Gerar uma matriz de "Document Type" da WoS em ambas as ferramentas e comparar as contagens.	7104 publicações sobre "bibliometrics" obtidas por meio dos arquivos Excel de "Registro Completo" disponíveis na WoS	Não foi utilizado	ISI – Web of Science v1.6	Não foi utilizado
Ferramenta em um contexto geral	Garantir que possa ser possível acessar a ferramenta e utilizar suas funcionalidades tanto por um celular quanto computador	Disponibilizar a ferramenta em um ambiente online e acessar todas as suas páginas por meio do navegador Google Chrome	Não foi utilizado	Não foi utilizado	Não foi utilizado	Não foi utilizado
Ferramenta em contextos específicos	Validar a geração de listas e matrizes para um conjunto grande de dados	Gerar listas e matrizes no ambiente online verificando a viabilidade das mesmas para diferentes conjuntos de elementos.	26376 publicações contendo UFSCar como uma de suas organizações nos arquivos Excel de "Registro Completo" disponíveis na WoS	Não foi utilizado	Não foi utilizado	Não foi utilizado

Fonte: Elaborado pelo autor.

### 3.2 Requisitos e viabilização

Antes de adentrar no processo de construção de uma ferramenta bibliométrica e seu escopo tecnológico, é necessário primeiramente discutir sobre os requisitos da mesma e sua viabilização.

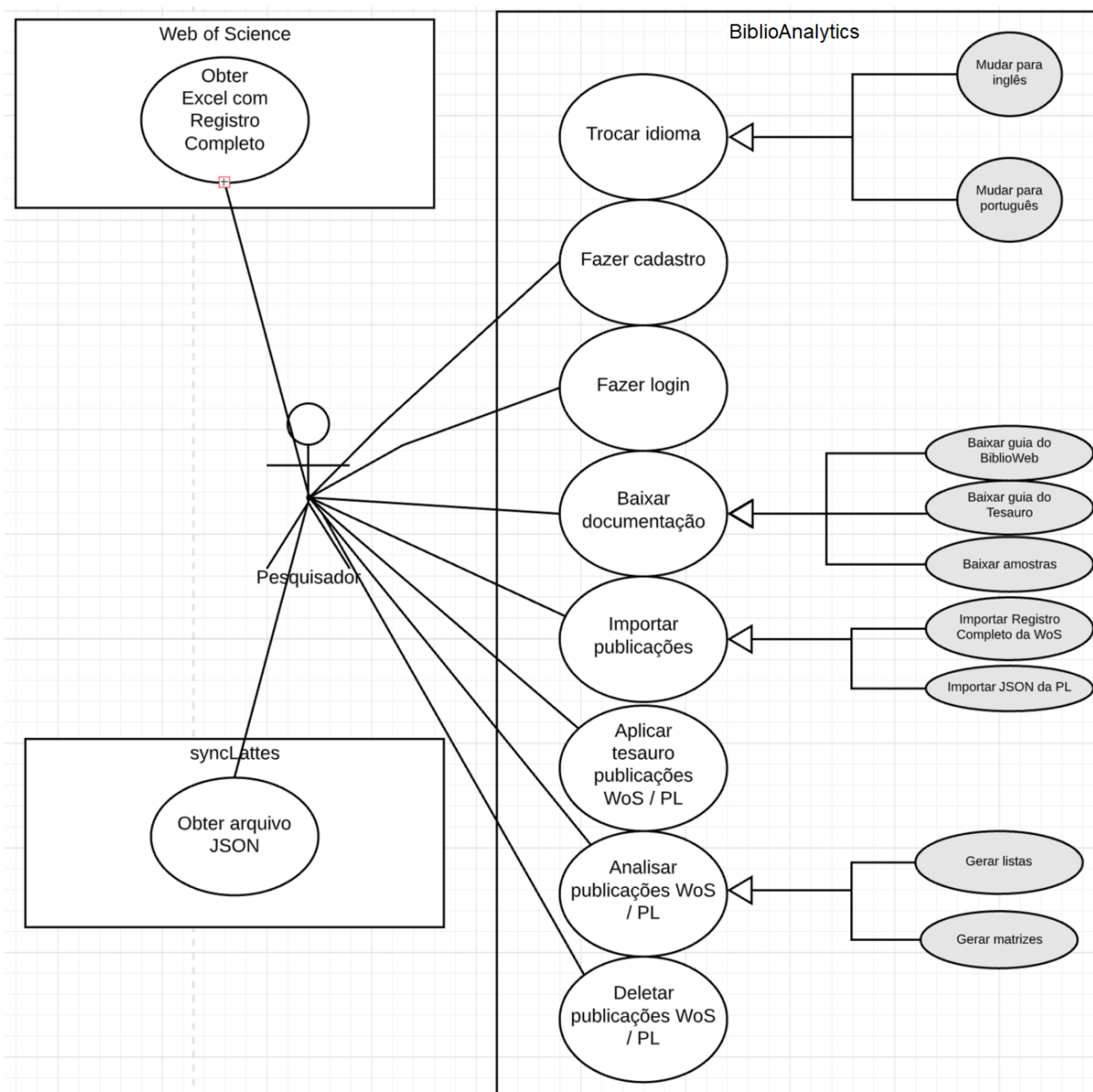
Iniciando pelos requisitos, os mesmos tiveram sua origem por meio da pesquisa-ação, onde os mesmos eram coletados por meio de discussões entre membros do grupo de pesquisa NIT/MATERIAIS da UFSCar. Por meio desse debate, com a finalidade de se ter uma ferramenta bibliométrica que pudesse contribuir com novos estudos bibliométricos e com a democratização de análises bibliométricas, pôde-se definir os seguintes requisitos para o *software* BiblioAnalytics: (i) deve ser capaz de trabalhar com publicações da WoS e PL; (ii) necessita de mecanismos de acesso, para que seja possível, em estudos posteriores serem, efetuadas análises altimétricas; (iii) estar disponível nos idiomas inglês e português, para que tanto pesquisadores que utilizem a base de dados WoS quanto pesquisador que utilizem a PL possam utilizar o BiblioAnalytics, sem o idioma ser uma restrição; (iv) fornecer a geração de listas e matrizes para que os usuários da ferramenta possam transformar suas publicações em indicadores; (v) possibilitar a aplicação de tesouros criados pelo autor, para que o mesmo possa refinar suas análises conforme julgar necessário; (vi) disponibilizar documentações para que o pesquisador e/ou usuário que não é da área possa compreender e utilizar a ferramenta; (vii) ser disponibilizado em um ambiente *web*, de modo que o utilizador não necessite ter ou configurar uma infraestrutura computacional para sua utilização. Os requisitos elencados e algumas de suas especificidades podem ser vistos no diagrama de caso de uso presente na Figura 21.

Além dos requisitos, é também necessário delimitar a infraestrutura em que o BiblioAnalytics será disponibilizado e viabilizado, dado que parte de seu objetivo é sua disponibilização e viabilização em um ambiente *web*. Para tal optou-se por utilizar plataformas que facilitam o processo técnico e complexo de disponibilização de um *software na web*, que tivessem uma versão de teste e/ou gratuita e que fossem amplamente utilizados no recorte temporal atual. Devido a isso, optou-se por utilizar a *Amazon Web Services (AWS)* para disponibilização da ferramenta e o *MongoDBAtlas* para a disponibilização do banco de dados, ambos em sua versão grátis. O motivo da escolha da utilização da versão grátis é para que, embora a expectativa seja que o BiblioAnalytics fique disponível em um local onde todos possam acessá-lo sem



restrição, pesquisadores contemplados com um maior conhecimento técnico que queiram alterar o mesmo e que não possuam condição financeira para utilizar a ferramenta em sua própria máquina, possam também disponibilizar sua versão do BiblioAnalytics nesses ambientes de modo gratuito.

**Figura 21** – Diagrama de caso de uso compreendendo os requisitos necessários para a ferramenta BiblioAnalytics.



Fonte: Elaborada pelo autor.

### 3.3 Escopo tecnológico para a construção de uma ferramenta bibliométrica

Uma ferramenta computacional pode apresentar diferentes características para que consiga atingir seu objetivo. Tais características podem levar em conta o contexto em que a ferramenta será utilizada, como o sistema operacional em que será executada, seus requisitos de hardware, o usuário que irá utilizar a ferramenta, a quantidade de dados e seus formatos que serão processados, se ela será uma aplicação exclusiva de *desktop* ou *mobile* e, finalmente, conforme já introduzido, a sua própria finalidade (SOMMERVILE, 2011).

As características se referem às tecnologias necessárias para a construção de uma ferramenta, como por exemplo sua arquitetura, seu funcionamento e comportamento, suas linguagens de programação e seus bancos de dados. Definir tais pontos antes da criação da mesma é primordial dado que, conforme exposto por Sommerville (2011), uma correta definição das mesmas é capaz de evitar diversos problemas posteriores.

#### 3.3.1 Aplicações para *web*

Quando se discute sobre aplicações para *web* (comumente chamadas de *web apps*), se refere a um sistema de informática projetado para ser utilizado por meio do navegador de um computador de mesa ou celular, sem que o usuário necessite ter conhecimentos de programação para sua utilização (SHKLAR; ROSEN, 2009). Nesse contexto, frequentemente existem a figura do cliente (*front end / site*), onde o usuário acessa e faz as requisições e operações que deseja, e a do *server (back end / servidor)*, responsável por receber as requisições do *front end* e efetuar e retornar os processamentos das requisições, e também acessar bancos possíveis bancos de dados (NATIONS, 2021). Nations (2021) e Scott (2016) enumeram alguns dos principais benefícios de se desenvolver um *web app*:

- **alcance:** em um *web app* convencional não há instalação ou configuração por parte do usuário, isso faz com que o mesmo não precise se preocupar com espaço no disco, em ter um dispositivo potente para acessar e pode acessá-la por qualquer dispositivo que tenha um navegador independente de sistema operacional ou marca.
- **facilidade de manutenção:** toda a disponibilização e manutenção da

ferramenta é feita somente nas máquinas que disponibilizam a aplicação, sendo, portanto, necessário somente a atualização ou modificação das mesmas para que todos os usuários possam ter acesso a uma atualização. Portanto, tanto quem fornece quanto quem consome a mesma têm custos reduzidos, não sendo necessário ter equipes para ir em cada um dos clientes efetuar uma nova instalação ou atualização da ferramenta.

- **simplificação do desenvolvimento:** aplicações convencionais e instaladas em um dispositivo necessita com que seu desenvolvimento seja adaptado para cada um dos sistemas e/ou dispositivos em que a ferramenta será fornecida, fazendo com que, conseqüentemente, sejam necessários diferentes profissionais com diferentes conhecimentos, habilidades e atitudes, o que encarece tanto o custo de desenvolvimento para quem desenvolve e fornece, quanto para quem consome, como por exemplo no caso de um ferramenta paga. Em um *web app*, ao ser desenvolvido adequadamente, é necessário focar somente em um conceito – que funcione em um navegador –, sendo necessárias ocasionalmente pequenas modificações para se adaptar a particularidades de um ou outro navegador.

Dados os pontos elencados, principalmente o alcance, é possível compreender a importância de uma aplicação para *web* no contexto da democratização. O fato de não ser necessária uma infraestrutura para a utilização da ferramenta faz com que uma gama muito maior de pessoas possam utilizar a mesma, diferente de ferramentas como o VP, VOSviewer e entre outras onde é necessário que a pessoa tenha um computador para sua utilização. Pelo fato de o pesquisador, se necessário, poder usar até mesmo seu aparelho celular para acessar a ferramenta, estudos que talvez anteriormente não eram possíveis devido à condição financeira do mesmo ou falta de máquinas por uma instituição de ensino, agora podem ser realizados.

### 3.3.2 Single Page Applications

Atualmente, quando o assunto são *web apps* e se deseja criar um novo, o principal conceito e metodologia utilizada para criar um *front end* é o de *Single Page Application* (SPA). O conceito de SPA, como o próprio nome já sugere, trata-se de um *web app* onde seu *front end* fornece somente uma página ao usuário que está acessando e a reescreve conforme ele interage, sem a necessidade de ficar realizando carregamentos constantes para cada interação (MIKOWSKI; POWELL, 2016). Para isso todo o código HTML, JavaScript e CSS pode tanto ser carregado de uma só vez ao acessar quanto paulatinamente, conforme a necessidade (FLANAGAN, 2020).

Pelo fato de as SPAs não realizarem carregamentos desnecessários durante a navegação, a tecnologia contribui com a democratização do acesso à uma ferramenta bibliométrica, diminuindo a quantidade de *internet* necessária para sua utilização. Aplicações *web* que não utilizam essa metodologia, a cada interação do usuário com a mesma utilizam uma banda muito maior de internet, uma vez que elementos já disponíveis são recarregados constantemente sem necessidade, sendo que em uma SPA isso não ocorre.

Ao se discutir sobre SPA, pode-se então elencar algumas vantagens de seu uso: não sobrecarregamento do servidor com requisições desnecessárias, transições entre conteúdos disponibilizados ocorrem mais suavemente, se assemelhando a aplicações nativas, e que o usuário não necessita de uma conexão muito potente para lidar com constantes carregamentos (MIKOWSKI; POWEL, 2016; FLANAGAN, 2020).

Para que esse tipo de *web app* seja desenvolvido, é importante a escolha de uma biblioteca JavaScript que atenda a esses princípios, facilitando seu desenvolvimento, como é o caso do Angular<sup>38</sup>, ReactJS<sup>39</sup>, VueJS<sup>40</sup> e entre outras. De modo a quantificar a importância dessas bibliotecas JavaScript, é possível citar alguns *sites* relevantes que utilizam essas bibliotecas: (i) Angular: IBM, Weather, PayPal, Forbes, The Guardian, Samsung, Deutsche Bank, Santander e muitos outros; (ii) ReactJS: Facebook, Instagram, New York Times, CNN, BBC, Netflix, Tesla e entre outros; (iii) VueJS: Apple, GitLab, Nintendo, BMW e ademais.

---

<sup>38</sup> <<https://angular.io/>>

<sup>39</sup> <<https://reactjs.org/>>

<sup>40</sup> <<https://vuejs.org/>>

### 3.3.3 Servidor de APIs

Partindo agora para a construção do *back end* de uma aplicação, local onde efetivamente serão processadas as operações de uma requisição e feito o acesso a bancos de dados. Em se tratando de um *web app*, é essencial a existência de *application programming interfaces* (APIs), que podem ser definidas tecnicamente como um conjunto de definições e protocolos responsáveis pela construção e integração de diferentes *softwares*, oferecendo serviços com diferentes finalidades a diferentes aplicações (SHKLAR; ROSEN, 2009; JIN;SAHNI; SHEVAT, 2018). Dentro de cada uma dessas APIs podem existir e serem executados inúmeros *scripts* para atender ao objetivo da mesma.

Uma analogia para facilitar o entendimento desse conceito é o de uma feira. Quando uma pessoa vai em uma feira ela pode ter várias finalidades como comprar feijão, comer um pastel, engraxar os sapatos e entre outras. Cada barraca de feira que oferece um desses serviços pode ser considerada uma API, onde cada uma delas oferece um serviço diferente de acordo com a necessidade de uma pessoa. Supondo ainda o caso de uma barraca de pastel, pode existir um funcionário responsável pelo preparo da comida e um para receber o pagamento, onde cada um desses funcionários seria um *script* diferente executado durante o acesso a essa API (ou no caso compra de um pastel na feira), iniciando com o primeiro funcionário recebendo seu pedido e informando ao outro para que faça seu pastel e que lhe entregue o mesmo posteriormente. Vale ressaltar que cada uma dessas barracas pode atender diferentes pessoas desde que as mesmas saibam se comunicar com elas, seja pela moeda de troca, linguagem correta ou pelo serviço oferecido. Diante desse contexto a feira pode ser considerada o servidor de APIs ou *API server*, que tem como objetivo fornecer diferentes serviços e produtos a quem tem conhecimento e permissão de utilização dos mesmos, lembrando que o “ser” que tem conhecimento pode ser tanto uma pessoa tentando acessar a mesma diretamente, indiretamente por meio de um site ou até mesmo outra aplicação.

Dada a explicação de um servidor de APIs, podem ser elencadas alguns dos principais benefícios de utilização de acordo com o exposto por Shklar e Rosen (2009); e também por Jin, Sahni e Shevat (2018):

- **automação:** uma API pode ser requisitada e ter seus *scripts* executados de

forma totalmente autônoma, sem a necessidade de intervenção de um desenvolvedor ou qualquer outra pessoa.

- **customização:** pelo fato de uma API ter um objetivo específico a ser atendido, elas podem ser construídas de acordo com diferentes necessidades, ponto essencial no mundo atual onde os usuários buscam por experiências individuais e de qualidade e não uma solução que tente atender a tudo e, conseqüentemente, não atende à excelência desejada.
- **integração:** dada a capacidade de automação de uma API, apesar de existir APIs que só podem ser utilizadas internamente dentro de um *web app*, a mesma pode ser desenvolvida de modo a atender diversos outros *web apps*. Além disso, uma API desenvolvida em uma linguagem de programação pode acessar outra, mesmo que desenvolvida em linguagem diferente. Um exemplo é a bandeira VISA que disponibiliza um conjunto de APIs para validação de transações bancárias que podem ser usadas por qualquer banco, onde a mesma e os bancos que a acessam tem suas APIs desenvolvidas em diferentes linguagens, conforme suas necessidades.

Atualmente existem diversas linguagens diferentes capazes de oferecerem mecanismos de construir um servidor de APIs. Alguns dos principais exemplos que podem ser citados são as seguintes linguagens, suas bibliotecas e empresas que utilizam uma delas: (i) JavaScript: NodeJS<sup>41</sup>, BackboneJS<sup>42</sup> e NextJS<sup>43</sup> (LinkedIn, Netflix, Uber, PayPal, eBay); (ii) Java: Spring<sup>44</sup>, Grails<sup>45</sup> e Hibernate<sup>46</sup> (LinkedIn, Citi Bank, Udemy, Via Varejo); (iii) Python: Flask<sup>47</sup>, Django<sup>48</sup> e Bottle<sup>49</sup> (Trivago, Samsung, Reddit, Patreon).

---

<sup>41</sup> <<https://nodejs.org/en/>>

<sup>42</sup> <<https://backbonejs.org/>>

<sup>43</sup> <<https://nextjs.org/>>

<sup>44</sup> <<https://spring.io/>>

<sup>45</sup> <<https://grails.org/>>

<sup>46</sup> <<https://hibernate.org/>>

<sup>47</sup> <<https://flask.palletsprojects.com/en/2.0.x/>>

<sup>48</sup> <<https://www.djangoproject.com/>>

<sup>49</sup> <<https://bottlepy.org/docs/dev/>>

### 3.3.4 Bancos de dados

Por fim, após serem expostos os conceitos de um *front end* e de um *back end*, é necessário abordar brevemente os bancos de dados. De acordo com Flejoles (2017) um banco de dados pode ser definido como um conjunto de dados armazenados organizadamente para sua posterior recuperação por um sistema de computação, onde os processos de recuperação e processamentos dentro do banco são realizados por um internamente por um Sistema de Gerenciamento de Banco de Dados (SGBD).

Existem inúmeros SGBDs disponíveis para as mais diferentes necessidades, porém os dois tipos mais comuns e utilizados são os relacionais e não relacionais. Os relacionais (MySQL<sup>50</sup>, PostgreSQL<sup>51</sup> e entre outros) têm como foco ter uma estrutura mais rígida e complexa de armazenamento com a presença de linhas e colunas, enquanto os não relacionais (MongoDB<sup>52</sup>, Redis<sup>53</sup> e ademais) apresentam estruturas de fácil alteração e escalabilidade, e trabalham com documentos (arquivos JSON) no lugar de linhas e colunas. (FLEJOLES; 2017).

### 3.3.5 Concatenação de tecnologias e a definição da arquitetura de um *web app*

Expostos os diferentes componentes necessários para a construção de um *web app* (*front end*, *back end* e banco de dados), necessita-se efetuar a seleção de tecnologias que serão utilizadas.

Começando pelo *front end*, dada a importância já exposta da utilização da linguagem JavaScript para construção de SPAs, foi selecionada a biblioteca Angular, desenvolvida pela Google. Alguns dos motivos que serviram de base para essa escolha são que o autor desta pesquisa já tem experiência prévia com a mesma, pelo fato dela ser atualmente uma das mais utilizadas e atualizadas, e também por ser amplamente recomendada para essa finalidade.

Para a construção do servidor de APIs optou-se pela utilização da biblioteca NodeJS, pelo fato de a mesma também ser uma biblioteca JavaScript, só que voltada para servidores. A escolha da mesma facilita com que outros pesquisadores que

---

<sup>50</sup> <<https://www.mysql.com/>>

<sup>51</sup> <<https://www.postgresql.org/>>

<sup>52</sup> <<https://www.mongodb.com/pt-br>>

<sup>53</sup> <<https://redis.io/>>

desejam contribuir com a ferramenta necessitem saber apenas uma linguagem de programação e não várias, e também é relevante por a mesma ser constantemente utilizada com aplicações que utilizam o Angular no front end.

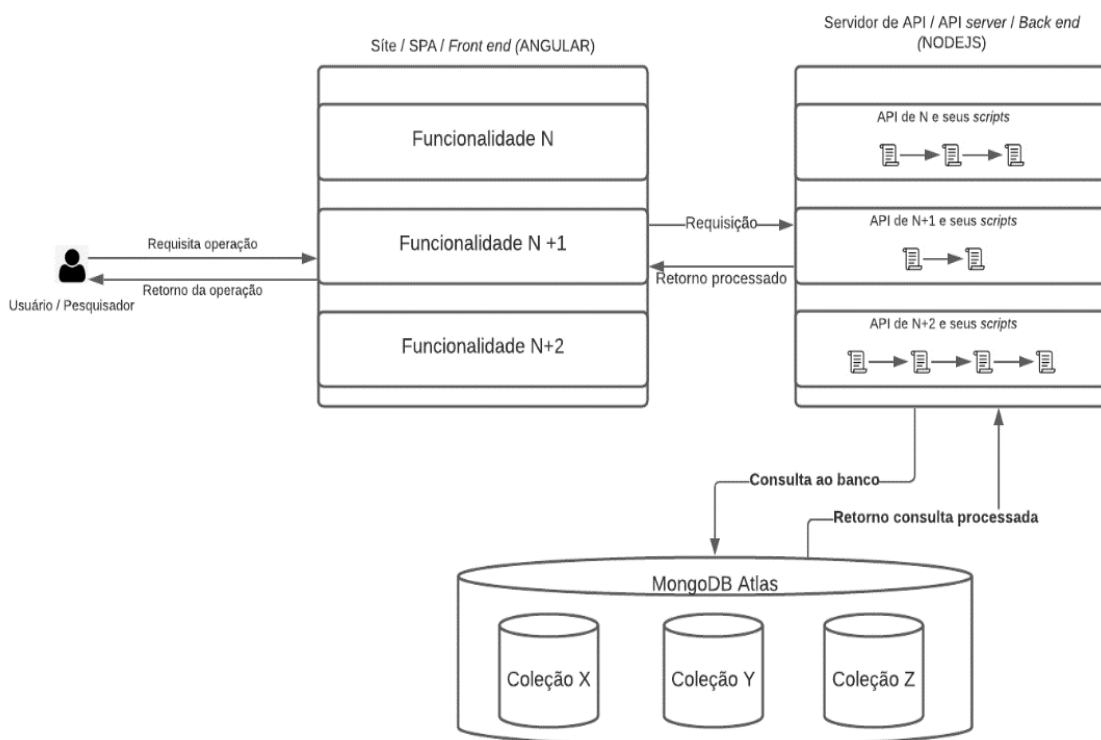
Por fim, para o banco de dados, se optou por um banco de dados não relacional chamado MongoDB, dado que a ferramenta desenvolvida neste estudo recebe um grande conjunto de dados em suas importações, a necessidade de constante alterações no mesmo conforme a ferramenta cresça e novos metadados passem a ser importados e utilizados e também pelo fato de o mesmo já trabalhar com arquivos no formato JSON diretamente. Sendo mais específico, optou pela versão “Atlas” do banco em questão, que é uma versão que fica disponibilizada no próprio servidor do Mongo, facilitando vários dos processos de manutenção de um banco de dados. Dentro desse banco de dados se criou três coleções: “*user*” para os usuários cadastrados, “*wos*” para as publicações da WoS e “*lattes*” para publicações da Plataforma Lattes.

A Figura 22 demonstra genericamente a concatenação de todas essas tecnologias para a construção do BiblioAnalytics. O *front end*, site onde o pesquisador visualiza e acessa as funções da ferramenta, fornece as requisições do mesmo para o servidor de API, que repassa a mesma para uma das APIs que conseqüentemente executam *scripts* para realização das operações necessárias. As APIs e *scripts* necessários para realizar as funções da ferramenta desenvolvida são explicados posteriormente.

Vale ressaltar alguns pontos sobre a estrutura elaborada: (i) o servidor de APIs consegue funcionar de modo totalmente independente do *front end*, portanto caso uma outra ferramenta queira acessar uma API em específico, basta saber as características da mesma e efetuar a requisição; (ii) caso se deseje no futuro refazer do zero o *front end* ou *back end* em outra linguagem e com outras características, qualquer um deles continuará funcionando normalmente, sendo somente necessário que o novo a ser criado se adapte às regras do antigo (i); (iii) o *back end* e o *front end* pode ser disponibilizados tanto da mesma máquina quanto de máquinas diferentes, apesar de o ideal ser que os mesmos estejam em máquinas diferentes para aumentar a disponibilidade da ferramenta.



**Figura 22** – Diagrama genérico representando a estrutura tecnológica do BiblioAnalytics



Fonte: Elaborada pelo autor.

### 3.4 Fontes de dados: seleção e utilização de fontes de informação online

Primeiramente, antes de aprofundar nas questões das fontes de informação, é importante iniciar este tópico enfatizando o motivo de terem sido escolhidas fontes *online* e não fontes físicas como catálogos de bibliotecas e entre outros.

O primeiro motivador é o fato de que tais fontes de informação, em geral, já trabalham e fornecem os registros que podem ser objetos de estudo de modo automático e formatado, facilitando a integração com ferramentas digitais e também a utilização da mesma pelo usuário (CAPES, 2021b; CLARIVATE, 2022). Isso pôde ser observado previamente quando foram discutidas as bases de dados, onde as mesmas fornecem metadados nos mais diferentes formatos como o RIS, JSON, XML e entre outros, se adaptando conforme a necessidade de seu usuário e dos sistemas de informação que poderão processar tais registros. Adentrando na questão dos sistemas, as bases de dados online também fornecem mecanismos de coleta automatizada (APIs) desses registros para integração com outros sistemas, sendo

alguns exemplos de APIs o Extrator Lattes oferecido pela PL e as diversas APIs disponibilizadas pela WoS.

O segundo ponto, que na verdade pode ser considerado uma dupla, é a enorme quantidade de registros constantemente disponibilizados por essas fontes de informação, como é o caso da WoS que indexa registros que datam desde 1900, e também seus poderosos e complexos motores de busca, que propiciam ao usuário uma maior precisão e rapidez na recuperação da informação que lhe é considerada relevante. Diferente de um ambiente físico, é possível que um usuário, seja ele um pesquisador ou até mesmo um sistema que está integrado com a mesma, tenha acesso a milhares de registros relevantes ao seu contexto em questão de segundos.

Dado os pontos elencados, foi então necessário efetuar a devida seleção de quais bases de dados iriam ser utilizadas inicialmente com o BiblioAnalytics. O ponto de partida foi a seleção de uma fonte de informação capaz de fornecer metadados de publicações científicas do campo brasileiro, no caso a PL, e outra que fornecesse metadados de publicações internacionais, como é o caso da WoS. Vale ressaltar que no futuro novas fontes de informações podem também ser incluídas conforme a necessidade científica.

No Quadro 11 são apresentados alguns dos fatores que motivaram a escolha das mesmas:

**Quadro 11** – As motivações que propiciaram a escolha da Plataforma Lattes e Web of Science como fontes de informação para o BiblioAnalytics.

<b>Plataforma Lattes</b>	<b>Web of Science</b>
<ul style="list-style-type: none"> <li>➤ Existência de ferramentas capazes de acessar e trabalhar com os registros da mesma (scriptLattes, syncLattes e etc)</li> <li>➤ Contém os currículos dos pesquisadores brasileiros, capazes de fornecer um mapa de toda a pesquisa científica brasileira.</li> <li>➤ Acesso gratuito.</li> <li>➤ Relevância nacional.</li> <li>➤ Frequentemente utilizada por pesquisas e estudos do NIT/Materiais da UFSCar.</li> </ul>	<ul style="list-style-type: none"> <li>➤ Existência de ferramentas capazes de acessar e trabalhar com os registros da mesma (Bibliometrix, VP, Vos Viewer e etc)</li> <li>➤ É uma das maiores bases de dados de citação, contendo registros desde 1900. <ul style="list-style-type: none"> <li>➤ Multidisciplinar.</li> <li>➤ Mult-iregional.</li> <li>➤ Independente.</li> </ul> </li> <li>➤ Relevância nacional e internacional.</li> <li>➤ Acessível a pesquisadores brasileiros por meio do consórcio da Capes.</li> <li>➤ Frequentemente utilizada por pesquisas e estudos do NIT/Materiais da UFSCar.</li> </ul>

Fonte: Elaborada pelo autor.

### 3.5 O processo de importação e tratamento de metadados

O processo de importação é crucial para o correto funcionamento da ferramenta uma vez que são por meio dos registros importados que serão obtidos os metadados responsáveis pela representação de cada publicação contida nos mesmos (FARIA, 2015). É importante reforçar que dependendo da base de dados escolhida, existem particularidades que necessitam ser trabalhadas separadamente, como é o caso dos diferentes formatos de metadados em que podem estar representados os registros em um arquivo, quais os metadados que serão extraídos e o processamentos dos mesmos para a correta representação e recuperação posterior (ARAUJO; FACHIN, 2015). Por fim, há que ser exposto também que esse é um processo em constante mudança e estudo em uma ferramenta, dado que por mais que o mesmo funcione para um conjunto de publicações, existe a possibilidade de as publicações terem seu formato alterado pela fonte de informação que fornece as mesmas e também de existirem casos não mapeados que podem ocasionar problemas durante a importação.

Dadas as diferenças presentes no processo de importação de publicações de ambas as fontes de informação, é necessário detalhar o que ocorre no servidor antes de as mesmas serem salvas no banco de dados.

#### 3.5.1 Web of Science

Começando pela WoS, conforme já exposto previamente, a mesma fornece os registros resultantes de uma pesquisa tanto por uma API quanto por meio de vários arquivos disponíveis em diferentes formatos (CLARIVATE, 2022). Devido às restrições impostas pela mesma ao acesso de suas APIs como por exemplo ser necessário uma assinatura, limite de acesso, limite de dados disponibilizados e entre outros, optou-se pela utilização dos arquivos obtidos diretamente pelo usuário da mesma, dado que tal acesso é possível graças ao consórcio da Capes.

Definido o modo com que serão obtidas as publicações a serem importadas é necessário selecionar qual o formato de metadados que será utilizado. A princípio havia sido escolhido os arquivos no formato ISI-CE (.ciw) para importação, dado que é o padrão utilizado pela base, porém após iniciar a extração dos campos relevantes

(que serão apresentados posteriormente) foi possível notar três problemas principais: (i) pelo fato de o arquivo estar em texto puro, é necessário percorrer todo o arquivo em busca dos campos desejados, o que faz com que tal operação possa ser demasiadamente demorada e custosa para a máquina; (ii) para cada campo selecionado para a extração durante a importação é necessário a criação de uma expressão regular para conseguir identificar qual linha do arquivo contém o mesmo, procedimento não trivial e que, dependendo do caso, pode ter alta complexidade; (iii) caso ocorra alguma mudança no formato ISI-CE, seja pela alteração de uma das *tags* ou da disposição das informações da respectiva tag, é necessário que a expressão regular seja adaptado para o caso em questão.

De modo a exemplificar o funcionamento de uma expressão regular, no Quadro 12 é exposta uma publicação no formato ISI-CE e um pseudocódigo que representa uma expressão regular para extração dos autores representados pela linha iniciada com “AU”. A expressão em questão capturara os seguintes elementos “Pulsiri, N\nVatananan-Thesenvitz, R” do registro. Os mesmos podem ser processados posteriormente de diversas formas posteriormente como por exemplo: padronizar tudo minúsculo, tudo maiúsculo e entre outros. Este é somente um exemplo, dos mais simples, de aplicação de uma expressão regular para extração de um campo desejado.

**Quadro 12** – Aplicação de uma expressão regular para extração dos autores (AU) de uma publicação no formato ISI-CE.

Publicação no formato ISI-CE	Pseudocódigo contendo expressão regular para encontrar os autores (AU)
FN Clarivate Analytics Web of Science VR 1.0 PT J <b>AU Pulsiri, N</b> <b>Vatananan-Thesenvitz, R</b> <b>AF</b> Pulsiri, Nonthapat Vatananan-Thesenvitz, Ronald TI Drones in Emergency Medical Services: A Systematic Literature Review with Bibliometric Analysis SO INTERNATIONAL JOURNAL OF INNOVATION AND TECHNOLOGY MANAGEMENT ER	Percorra a publicação procurando a seguinte expressão: → Obrigatoriamente começa com AU (Início da expressão) → Obrigatoriamente tem um espaço logo após o AU → Tem N ou mais caracteres, dígitos ou símbolos após o espaço. Capture os mesmos. → Termina com uma quebra de linha seguido por uma letra de A a Z em maiúsculo. (Fim da expressão)

Fonte: Elaborada pelo autor.

Dados os motivos apresentados, foi necessário buscar por um novo formato de arquivo para efetuar a importação e o escolhido foi o de planilha do Microsoft Excel, em específico o “Registro Completo” oferecido pela WoS. Pelo fato de se estar trabalhando com planilhas do Excel, os campos que anteriormente necessitavam da criação manual de expressões regulares para sua devida extração agora são disponibilizados por meio de diversas linhas e colunas. Com o auxílio da biblioteca *convert-excel-to-json*<sup>54</sup> o arquivo fornecido no formato Excel é convertido para o formato JSON, formato desejado para que seja salvo no banco de dados, sendo possível fazer quaisquer processamentos necessários antes de finalmente salvar no banco de dados. Vale ressaltar que embora a aplicação inicial de uma expressão regular para buscar o campo em específico não seja mais necessária, podem acontecer casos, como na extração de organizações, que ainda assim é necessária uma expressão regular interna para extração correta do nome da instituição dentro do campo C1 (no formato ISI-CE) ou *Address* (em uma planilha do Excel). Tais casos são aprofundados na temática dos processamentos específicos realizados pela ferramenta.

Os processamentos efetuados para os registros da WoS podem ser divididos em dois tipos: gerais e específicos. Os processamentos gerais são aqueles efetuados em mais de um campo extraído e os específicos, como o próprio nome já diz, tem como objetivo tratar especificidades de um campo específico.

Iniciando pelos processamentos gerais, conforme pode ser visto na Figura 23, o arquivo obtido da WoS contém diversos tipos de metadados para representar uma publicação, sendo que nem todos acabam sendo relevantes para uma análise bibliométrica. Portanto, o primeiro processamento geral a ser feito é a extração dos campos relevantes, processo que é realizado informando à biblioteca *convert-excel-to-json* quais as colunas relevantes e quais serão seus respectivos nomes no formato JSON a serem usados no BiblioAnalytics.

---

<sup>54</sup> <<https://www.npmjs.com/package/convert-excel-to-json>>

**Figura 23** – Um Registro Completo obtido da Web of Science no formato Microsoft Excel contendo várias publicações.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Publication Type	Authors	Abstract	Book Editors	Book Group	Author Full Names	Book Author I	Group Authors	Article Title	Source Title	Book Serie	Book Serie	Language	Document Cont
2	J	Li, GH; Fa	Ethnopharmacological relevance: The			Li, Guo-Hui; Fang,			Thesium chir	JOURNAL C			English	Review
3	J	Liu, SQ; Q	Ethnopharmacological relevance: Eigh			Liu, Siqi; Qiao, Sha			Intestinal bai	JOURNAL C			English	Article
4	J	Liu, Y; Ch	Ethnopharmacological relevance: Dan			Liu, Yan, Chang, M			Danggui Bux	JOURNAL C			English	Article
5	J	Mahnashi,	Ethnopharmacological relevance: Men			Mahnashi, Mater H			Phytochemi	JOURNAL C			English	Article
6	J	Mazumder	Ethnopharmacological relevance: Func			Mazumder, Kishor			Identification	JOURNAL C			English	Article
7	J	Sanap, A;	Ethnopharmacological relevance: The			Sanap, Avinash, Jo			Pre-conditio	JOURNAL C			English	Article
8	J	Sun, XT; Y	Ethnopharmacological relevance: Sopl			Sun, Xiangling, Yar			Evaluation of	JOURNAL C			English	Article
9	J	Tong, NN;	Ethnopharmacological relevance: Paet			Tong, Ning-Ning, Zi			A comprehen	JOURNAL C			English	Article
10	J	Van Puyve	Ethnopharmacological relevance: Tetr			Van Puyvelde, Luc;			Bioassay-gu	JOURNAL C			English	Article
11	J	Vieceli, P	Ethnopharmacological relevance: Phys			Vieceli, Paula Soch			Physalis ang	JOURNAL C			English	Article
12	J	Zhang, Z;	Ethnopharmacological relevance: Inon			Zhang, Zhe; Liang,			Effect of Ino	JOURNAL C			English	Article
13	J	Jin, RO; A	Agitation is a routine and increasingly i			Jin, Rowen O.; Ani			Exploring bis	ACADEMIC			English	Article; Ea
14	J	Royer, AS	Understanding how hospitalization affe			Royer, Ana Sanche			A systemati	EUROPEAN			English	Review; Ea
15	J	Matsumur				Matsumura, Norion			From geneti	INTERNATI			English	Editorial M
16	J	Pecere, T;	Aloe-emodin (1,8-dihydroxy-3-[hydrox			Pecere, Teresa; Pi			On the mech	INTERNATI			English	Article; Ea
17	J	Ingrey, SD	Aims Microbial endophytes produce st			Ingrey, S. D.; Pear			Australian bu	JOURNAL C			English	Article; Ea
18	J	Abdel-Rah	Aim: To assess the association of vac			Abdel-Rahman, On			Patterns anc	JOURNAL C			English	Article; Ea
19	J	Agar, OT;	Economical and fast analytical method			Agar, Osman Tunc			Developmen	JOURNAL C			English	Article; Ea
20	J	Ochmann,				Ochmann, Uta; Wik			Pregnant err	MEDIZINIS			German	Article; Ea
21	J	Brancato,	Background: Myocardial bridging is lar			Brancato, Federica			Complicator	PAEDIATRI			English	Article; Ea
22	J	Kim, M; Kr	Aim: There is a critical need for safe a			Kim, Mina; Knoepfl			Anticipated	REGENERAT			English	Article; Ea
23	J	Sun, F; Lu	Tissue engineering technology provide			Sun, Fei; Lu, Yi; W			Vascularizati	REGENERAT			English	Review; Ea
24	J	Flores, Al	Peripheral artery disease is an athero			Flores, Alyssa M.;			Leveraina I	CIRCULATI			English	Article

Fonte: Elaborada pelo autor.

A seleção dos campos contidos no Quadro 13 foi feita buscando atender às necessidades do NIT/Materiais e também fornece mecanismos de se efetuarem análises bibliométricas relevantes por meio da ferramenta. As colunas não informadas são descartadas pela ferramenta durante o procedimento para que a execução da mesma não seja prejudicada e que o banco não contenha informações desnecessárias. Conforme já exposto previamente, além de ser responsável pela extração dos campos relevantes de uma publicação contida no arquivo de “Registro Completo” do Excel, a biblioteca também converte a publicação para o formato JSON (Figura 24).

**Quadro 13** – Campos extraídos de um Registro Completo na Web of Science, suas colunas e suas contrapartes no BiblioAnalytics

Campo no Excel	Coluna no Excel	Campo no BiblioAnalytics
Publication Type	A	publicationType
Authors	B	authors
Title	I	title
Language	M	language
Document Type	N	documentType
Author Keywords	T	authorKeywords
Keywords Plus	U	keywordsPlus
Addresses	W	organizations
ORCIDs	AB	orcids
Funding Orgs	AC	fundingOrgs
Times Cited, WoS Core	AG	timesCited
Publication Year	AT	publicationYear
WoS Categories	BH	wosCategories

Fonte: Elaborado pelo autor

Os dois próximos processamentos gerais são realizados em conjunto, sendo eles a quebra de campos múltiplos e a remoção de duplicidades. A quebra de campos múltiplos é necessária em campos onde a célula do Excel pode conter um ou mais elementos. Um exemplo é o campo de “authors” presente na Figura 24 onde dois autores “Agar, OT” e “Demirezer, LO” são representados como uma coisa só “Agar, OT; Demirezer, LO”, sendo portanto necessário separá-los, processo que é feito pela “quebra” dos autores no caractere ponto e vírgula seguido por um espaço “; “. Já a remoção de duplicidades, é procedimento extremamente importante quando se lida com campos que podem apresentar múltiplos valores, uma vez que a depender do tipo de contagem utilizada (ALVARADO, 2002), diferentes resultados podem ser obtidos. Para exemplificar a importância desse processamento, ainda na Figura 24, note que no campo “organizations” a organização “Hacettepe Univ” aparece mais de uma vez, sendo necessário portanto remover a duplicidade em questão para a correta contagem.

**Figura 24** – Exemplo de JSON obtido após utilização da biblioteca convert-excel-to-json e feita a seleção dos campos necessários.

```
{
  "publicationType": "J",
  "authors": "Agar, OT; Demirezer, LO",
  "title": "Development of a new validated HPLC method for the chemical specification of Rosa damascena petals",
  "language": "English",
  "documentType": "Article; Early Access",
  "authorKeywords": "Fingerprinting; HPLC; method development; optimization; Rosa damascena; validation",
  "keywordsPlus": "OXIDATIVE STRESS; MEDICINAL-PLANTS; MILL.; EXTRACT; FLOWERS; FLAVONOIDS; FRACTIONS ; MIXTURES; EFFICACY",
  "orcid": "Demirezer, LO/0000-0000-0000-0000; Agar, OT/1111-1111-1111-1111",
  "fundingOrgs": "Hacettepe University Scientific Researches Coordination Unit Hacettepe University [THD-2017-15631]",
  "organizations": "[Agar, Osman Tuncay] Suleyman Demirel Univ, Fac Pharm, Dept Pharmacognosy, Isparta, Turkey; [Agar, Osman Tuncay] Hacettepe Univ, Fac Pharm, Dept Pharmaceut Bot, Ankara, Turkey; [Demirezer, L. Omur] Hacettepe Univ, Fac Pharm, Dept Pharmacognosy, Sihhiye Campus, TR -06100 Ankara, Turkey",
  "timesCited": "0",
  "publicationYear": "2021",
  "wosCategories": "Biochemical Research Methods; Chemistry, Analytical"
}
```

Fonte: Elaborada pelo autor.

Após a quebra dos campos múltiplos e a remoção de duplicidades, o último processamento geral a ser realizado consiste em um conjunto de procedimentos efetuados para normalizar o elemento da publicação. Os procedimentos aplicados no elemento são: (i) transformar em letras minúsculas; (ii) e (iii) remoção de diacríticos e Unicode (acentuações, cedilhas e outros sinais fonéticos textuais); (iv) remoção de pontos “.”. Esse conjunto de procedimentos é essencial em campos onde o autor de uma publicação tem liberdade para representar, por exemplo, uma palavra-chave, da maneira que quiser. Tomando como exemplo a palavra chave “Esperança”, a mesma pode ser representada de diferentes modos como por exemplo “esperança”, “esperanca” e “esperan\u00e7a”, sendo necessária então sua padronização como “esperanca” para que a contagem possa ser efetuada corretamente. Vale ressaltar que em certos casos tais procedimentos podem não ser necessários uma vez que algumas bases de dados já fazem suas próprias padronizações.

A Figura 25 contempla um exemplo de uma publicação em formato JSON após todos os processamentos gerais serem realizados. É possível notar a ausência de duplicidades, campos onde anteriormente seus elementos eram considerados uma coisa só agora encontram-se separados para a correta contagem e também o fato de que diversos campos foram normalizados para a contagem.

O Quadro 14 resume quais os processamentos gerais efetuados em cada campo da planilha do Excel obtida na WoS durante a importação no BiblioAnalytics. Nota-se que nem todos os processamentos são necessários para um mesmo campo, como a normalização que não é realizada para o campo “languages” dado que o mesmo já vem padronizado pela própria WoS, sendo os mesmos somente realizados nos campos necessários.

Finalizados os processamentos gerais é necessário tratar as especificidades que alguns dos campos contêm como é o caso dos campos “orcids”, “organizations” e “fundingOrgs”. As especificidades são tratadas por meio das expressões regulares já vistas anteriormente, para que conteúdos desnecessários sejam removidos dos campos em questão e que seja conseqüentemente aprimorada as contagens a serem realizadas. Os processamentos específicos realizados em cada um dos três campos são:

- (i) *orcids*: remoção do conteúdo anterior à “/”
- (ii) *organizations*: remoção dos autores contidos dentro dos caracteres “[“ e “]” e remoção do conteúdo presente após a primeira vírgula “,”



- (iii) *fundingOrgs*: remoção do número do processo de fomento contido entre os caracteres “[“ e ”]”

**Figura 25** – Exemplo de uma publicação em formato JSON após todos os processamentos gerais terem sido realizados.

```
{
  "publicationType": "J",
  "authors": [
    "agar, ot",
    "demirezer, lo"
  ],
  "title": "Development of a new validated HPLC method for the chemical specification of Rosa damascena petals",
  "language": "English",
  "documentType": [
    "article",
    "early access"
  ],
  "authorKeywords": [
    "fingerprinting",
    "hplc",
    "method development",
    "optimization",
    "rosa damascena",
    "validation"
  ],
  "keywordsPlus": [
    "oxidative stress",
    "medicinal plants",
    "mill",
    "extract",
    "flowers",
    "flavonoids",
    "fractions",
    "mixtures",
    "efficacy"
  ],
  "orcid": [
    "Demirezer, LO/0000-0000-0000-0000",
    "Agar, OT/1111-1111-1111-1111"
  ],
  "fundingOrgs": "hacettepe university scientific researches coordination unithacettepete university [thd-2017-15631]",
  "organizations": [
    "[agar, osman tuncay] suleyman demirel univ, fac pharm, dept pharmacognosy, isparta, turkey",
    "[demirezer, l omur] hacettepe univ, fac pharm, dept pharmacognosy, sihhiye Campus, tr-06100 ankara, turkey"
  ],
  "timesCited": "0",
  "publicationYear": "2021",
  "wosCategories": [
    "biochemical research methods",
    "chemistry, analytical"
  ]
}
```

Fonte: Elaborada pelo autor.

É necessário fazer um paralelo dos processamentos específicos realizados pelo BiblioAnalytics e a ferramenta VP. Esses processamentos e suas expressões regulares são feitos internamente pela ferramenta BiblioAnalytics, sem a necessidade de que o pesquisador tenha que criar seus próprios filtros e suas expressões regulares conforme é feito pelo VP (SEARCH TECHNOLOGY, 2015). Tal procedimento pode ser vantajoso para o usuário inexperiente dado a complexidade de se criar expressões regulares e ao mesmo tempo uma desvantagem caso o mesmo tenha o conhecimento técnico necessário, onde nesse caso o mesmo é obrigado a usar as expressões regulares já disponibilizadas pelo BiblioAnalytics.

**Quadro 14** – Processamentos gerais realizados pelo BiblioAnalytics nos campos extraídos de uma publicação da Web of Science

<b>Processamento gerais</b>	<b>Campo(s) no BiblioAnalytics</b>
Extração de campos relevantes e conversão de arquivo	<i>authors, publicationType, title, language, documentType, authorKeywords, keywordsPlus, organizations, orcid, fundingOrgs, timesCited, publicationYear, wosCategories;</i>
Quebra de campos múltiplos e remoção de duplicidades	<i>authors, documentType, authorKeywords, keywordsPlus, organizations, orcid, fundingOrgs, wosCategories</i>
Procedimentos de normalização	<i>authors, authorKeywords, keywordsPlus, wosCategories, documentType, organizations, fundingOrgs</i>

Fonte: Elaborado pelo autor.

Após todos esses processamentos a publicação está pronta para ser salva no banco de dados (Figura 26) e ser utilizadas em análises posteriores, sendo necessário apenas adicionar qual o usuário (*creator*) que importou a publicação e a data de importação (*dateCreated*), necessários para efetuar as contagens posteriormente. Vale ainda ressaltar que cada um dos processamentos gerais explanados é repetido para todas as publicações contidas em uma planilha importada.

Resumidamente, todas as etapas anteriormente explanadas são executadas no servidor por meio de uma API “*/wos/savePublications*” (que recebe a requisição) e de três scripts (que processam a requisição):

- **wosToJson**: recebe as publicações fornecidas pelo pesquisador e utiliza a biblioteca *convert-excel-to-json* para converter as publicações do formato de planilha do Excel para JSON. É nesse script que também é feito o processamento geral no qual são selecionados os metadados que serão salvos no banco de dados, processo que é feito informando internamente

as colunas a serem extraídas.

- **wosJsonFormat**: recebe as publicações convertidas da biblioteca em formato JSON e efetua os demais processamentos gerais (tratamento de campos múltiplos, duplicidade e normalização) e os específicos (aplicação de expressões regulares).
- **wosSavePublications**: salva todos os JSONs processados na coleção “wos” do banco de dados para posterior recuperação e análise durante a geração de listas.

**Figura 26** – JSON obtido após a realização de todos os tratamentos específicos e a adição do usuário que importou e a data de criação.

```
{
  "creator": "238",
  "dateCreated": "20220112",
  "publicationType": "J",
  "authors": [
    "agar, ot",
    "demirezer, lo"
  ],
  "title": "Development of a new validated HPLC method for the chemical specification of Rosa damascena petals",
  "language": "English",
  "documentType": [
    "article",
    "early access"
  ],
  "authorKeywords": [
    "fingerprinting",
    "hplc",
    "method development",
    "optimization",
    "rosa damascena",
    "validation"
  ],
  "keywordsPlus": [
    "oxidative stress",
    "medicinal plants",
    "mill",
    "extract",
    "flowers",
    "flavonoids",
    "fractions",
    "mixtures",
    "efficacy"
  ],
  "orcid": [
    "0000-0000-0000-0000",
    "1111-1111-1111-1111"
  ],
  "fundingOrgs": "hacettepe university scientific researches coordination unithacettepete university",
  "organizations": [
    "suleyman demirel univ",
    "hacettepe univ"
  ],
  "timesCited": "0",
  "publicationYear": "2021",
  "wosCategories": [
    "biochemical research methods",
    "chemistry, analytical"
  ]
}
```

Fonte: Elaborada pelo autor.

### 3.5.2 Plataforma Lattes

O processo de importação para publicações da PL é consideravelmente similar ao da WoS e também um pouco mais simples, dado que não é necessária a etapa de utilização de uma biblioteca para conversão de Excel para JSON e que não são necessários processamentos específicos. De modo introdutório, as principais diferenças são: (i) o arquivo de importação é o arquivo no formato JSON fornecido pela ferramenta syncLattes; (ii) os metadados relevantes selecionados e extraídos de cada publicação; (iii) não é necessária a quebra de campos múltiplos; (iv) não são necessários processamentos específicos.

Primeiramente, o foco foi a seleção do arquivo que seria utilizado na importação e o seu formato. Conforme já discutido previamente, a obtenção de informações da PL pode ser feita tanto de modo manual na página, processo mais trabalhoso para o usuário, ou por meio de sua API (Extrator Lattes), existindo algumas ferramentas que já realizam essa coleta automatizada na mesma como o caso do syncLattes e do scriptLattes (CNPQ, 2021b; MATIAS, 2015). Visando a praticidade do usuário, foi escolhido utilizar um arquivo extraído por meio de uma dessas ferramentas.

Para definir qual arquivo seria utilizado foi verificado os metadados contidos no arquivo RIS fornecido pelo exemplo contido no scriptLattes (versão 8.11) e também no arquivo RIS modificado do syncLattes (versão não encontrada). Conforme o Quadro 15 pode-se notar a presença de uma maior quantidade de metadados no arquivo RIS modificado do syncLattes como as áreas de pesquisa da Capes (*dc.subject.classification*), tópicos de pesquisa (*dc.subject*) e entre outros, que podem ser relevantes em estudos bibliométricos. Dado esse motivo e o fato de a ferramenta syncLattes ser desenvolvida por outros membros do NIT/Materiais da UFSCar, fazendo com que possíveis alterações na mesma para atender as necessidades do BiblioAnalytics ou até mesmo para realizar uma integração futura serem mais viáveis, foi selecionado o arquivo fornecido pelo syncLattes. Porém, assim como o caso do arquivo ISI-CE da WoS, conforme o desenvolvimento seguia utilizando o mesmo, notou-se que o processo seria custoso e complexo pela necessidade de se utilizar expressões regulares para extrair cada um dos campos relevantes, sendo necessário verificar a possibilidade de outro tipo de arquivo.

**Quadro 15** – Comparativo entre os arquivos no formato RIS expondo os metadados oferecidos pelos softwares scriptLattes e syncLattes.

Uma publicação no arquivo RIS do scriptLattes	Uma publicação no arquivo RIS modificado do syncLattes
<p>TY - CONF            AU - MATTOS, ANDRÉA BRITTO;            HERRMANN, RICARDO ; CARDONHA,            CARLOS ; GALLO, DIEGO ; AVEGLIANO,            PRISCILLA ; BORGER, SERGIO            T1 - Marker-assisted recognition of            dynamic content in public spaces            TI - the 11th Web for All Conference,            2014            VL -            SP - 1            EP -            PY - 2014            L2 -            http://dx.doi.org/10.1145/2596695.259            6722            ER -</p>	<p>capex.qualis educacao C 2012            letras_linguistica C 2012            ensino B3 2015            letras_linguistica B1 2015            letras_linguistica B1 2015            dc.language.iso por            dc.title Os percursos éticos do cantar no discurso de denúncia            social em três canções de protesto            dc.contributor.author BASTOS, N.            MIOTELLO, VALDEMIR            dc.contributor.author.lattes nao identificado            9131819326282708            dc.relation.ispartof Diálogo das Letras            dc.date.issued 2017            dc.identifier.citation Diálogo das Letras, v. 6, pp. 284-302, 2017            dc.identifier.issn 23161795            dc.identifier.issnl 23161795            dc.identifier.uri            http://periodicos.uern.br/index.php/dialogodasletras/article/view            /2690            dc.type article            dc.subject Ética            Canção de protesto            dc.subject.classification Linguística Letras E Artes.            Filosofia da Linguagem</p>

Fonte: Elaborada pelo autor.

Buscando outras alternativas, foi observado que o syncLattes também fornece as publicações em outro formato: o formato JSON. Pelo fato de as mesmas já estarem nesse formato, não seria necessário utilizar uma biblioteca como a *convert-excel-to-json* para adaptar as publicações para o mesmo, facilitando ainda mais o processo de importação. Com isso fica sendo necessário somente extrair os campos relevantes da publicação em JSON (Quadro 16) e efetuar os processamentos necessários para esses metadados.

**Quadro 16** – Campos relevantes extraídos do JSON do syncLattes e sua nova nomenclatura no BiblioAnalytics.

<b>Campo no JSON do syncLattes</b>	<b>Campo no BiblioAnalytics</b>
dc.data.issued	publicationYear
dc.language.iso	language
dc.contributor.author	authors
dc.contributor.author.authority	authorsLattes
dc.type	publicationType
dc.subject	researchTopics
dc.subject.classification	capesResearchAreas

Fonte: Elaborado pelo autor.

Na temática dos processamentos efetuados para as publicações da PL, os mesmos são bem similares aos efetuados para publicações da WoS. No caso do primeiro processamento geral, como não há mais a necessidade de se converter a publicação para um outro formato, uma vez que a mesma já se encontra em formato JSON, é necessário apenas extrair os metadados priorizados. O segundo processamento geral efetuado para as publicações da WoS (quebra dos elementos múltiplos) não é necessário, uma vez que os elementos já se encontram separados adequadamente no JSON do syncLattes. O terceiro e quarto processamento geral (remoção de duplicidades e normalização) ocorrem de modo idêntico ao efetuado na WoS, sendo o quarto processamento extremamente importante para publicações da PL uma vez que diversas palavras da língua portuguesa apresentam acentuação e diátricos.

No Quadro 17 é apresentado um resumo de todos os processamentos gerais aplicados durante a importação de publicações no formato JSON da ferramenta syncLattes no BiblioAnalytics. A escolha dos campos relevantes a serem extraídos seguiu o mesmo critério definido para os campos da WoS e os demais processamentos gerais foram aplicados em campos que se beneficiam dos mesmos, sendo exemplos a remoção de duplicidades em campos múltiplos e a normalização de um elemento em campos que poderiam conter variações.

Nas Figuras 27 e 28 são mostradas duas publicações, uma sem os processamentos e uma com os processamentos, onde podem ser visualizadas mudanças muito nítidas entre uma e outra, sendo possível citar duas delas: (i) a síntese dos campos julgados relevantes, que exclui os que não são relevantes, e a (ii)

normalização efetuada na palavra “Ética” e entre outras, onde na primeira figura a mesma é representada como “\u00c9tica” e após os processamentos se transforma em “etica”.

**Quadro 17** – Processamentos gerais efetuados em publicações do formato JSON do syncLattes durante uma importação do BiblioAnalytics e os campos em que as mesmas são aplicadas.

<b>Processamento geral</b>	<b>Campo(s) no BiblioAnalytics</b>
Extração de campos relevantes	<i>publicationYear, language, authors, authorsLattes, publicationType, researchTopics e capesResearchAreas</i>
Remoção de duplicidades	<i>authors, authorsLattes, researchTopics e capesResearchAreas</i>
Procedimentos de normalização	<i>language, authors, publicationType, researchTopics e capesResearchAreas</i>

Fonte: Elaborado pelo autor.

Finalmente, assim como no processo para publicações da WoS, após todos os processamentos são adicionados o usuário que está fazendo a importação e a data de importação (Figura 28), ficando as publicações em questão disponíveis para a realização de análises bibliométricas futuras no BiblioAnalytics.

Diferentemente do processo de importação para publicações da WoS, no caso da PL, foi necessário apenas uma API “*/lattes/savePublications*” e dois scripts para efetuar todas as etapas da importação. Isso se deve ao fato de a mesma já trabalhar com publicações em formato JSON que, conforme já exposto, facilita a importação das mesmas. Os scripts desenvolvidos foram:

- **lattesJsonFormat**: recebe as publicações em formato JSON e efetua todos os processamentos gerais.
- **lattesSavePublications**: salva todos os JSONs processados na coleção “lattes” do banco de dados para posterior recuperação e utilização em análises bibliométricas.

Figura 27 – Publicação contida em um arquivo fornecido pelo *software* syncLattes no formato JSON

```
{
  "dc.identifier.issn":[
    {
      "value":"23161795"
    }
  ],
  "capes.qualis":[
    {
      "value":"educacao C 2012"
    },
    {
      "value":"letras_linguistica C 2012"
    },
    {
      "value":"ensino B3 2015"
    },
    {
      "value":"letras_linguistica B1 2015"
    },
    {
      "value":"letras_linguistica B1 2015"
    }
  ],
  "dc.subject.classification":[
    {
      "lang":"por",
      "_origem":"area-conhecimento",
      "value":"Linguística Letras E Artes. Filosofia da Linguagem"
    },
    {
      "lang":"por",
      "_origem":"area-conhecimento",
      "value":"Linguística Letras E Artes. Ling\u00fc\u00edstica. Estudos do discurso"
    }
  ],
  "dc.identifier.issn1":[
    {
      "value":"23161795"
    }
  ],
  "dc.identifier.citation":[
    {
      "value":"Di\u00e1logo das Letras, v. 6, pp. 284-302, 2017"
    }
  ],
  "_lattes.natureza":[
    {
      "value":"COMPLETO"
    }
  ],
  "dc.identifier.uri":[
    {
      "value":"http://periodicos.uern.br/index.php/dialogodasletras/article/view/2690"
    }
  ],
  "dc.type":[
    {
      "value":"article"
    }
  ],
  "dc.date.issued":[
    {
      "value":"2017"
    }
  ],
  "dc.relation.ispartof":[
    {
      "value":"Di\u00e1logo das Letras"
    }
  ],
  "_lattes.relevancia":[
    {
      "value":true
    }
  ],
  "dc.language.iso":[
    {
      "value":"por"
    }
  ],
  "dc.subject":[
    {
      "value":"\u00c9tica"
    },
    {
      "value":"Can\u00e7\u00e3o de protesto"
    }
  ],
  "dc.contributor.author":[
    {
      "_nomecompleto":"Nathan Bastos de Souza",
      "value":"BASTOS, N."
    },
    {
      "confidence":500,
      "authority":"9131819326282708",
      "value":"MIOTELLO, VALDEMIR",
      "_nomecompleto":"Valdemir Miotello"
    }
  ],
  "dc.title":[
    {
      "lang":"por",
      "value":"Os percursos \u00e9ticos do cantar no discurso de den\u00fancia social em tr\u00eas can\u00e7\u00f5es de protesto"
    }
  ]
}
```

Fonte: Elaborada pelo autor



**Figura 28** – Publicação no formato JSON após todos os processamentos feitos pelo BiblioAnalytics

```
{
  "creator": "60ad894000f98e31e067c71f",
  "timeCreated": "20210810",
  "publicationYear": "2017",
  "language": "por",
  "authors": [ "bastos, n", "miotello, valdemir" ],
  "authorsLattes": [ "9131819326282708" ],
  "publicationType": "article",
  "researchTopics": [ "etica", "cancao de protesto" ],
  "capesResearchAreas": [
    "linguistica letras e artes filosofia da linguagem",
    "linguistica letras e artes linguistica estudos do discurso"
  ]
}
```

Fonte: Elaborada pelo autor.

### 3.6 Analisando metadados: construção de listas e matrizes

Assim como a importação, o processo de construção de listas e matrizes é uma das principais funções da ferramenta, dado que as mesmas podem ser fonte de informações para construção de indicadores de impacto, redes de colaboração e entre outros (FARIA, 2015). A construção de listas e matrizes têm como finalidade transformar os metadados das publicações previamente importadas, no caso desta pesquisa das fontes de informação WoS ou PL, em elementos capazes de serem interpretados e utilizados para geração de conhecimento (GUEDES, 2012). Essa facilitação da compreensão dos metadados presentes em uma publicação científica é extremamente relevante para a democratização do conhecimento, dado que diversos estudos podem ser feitos por meio de listas e matrizes obtidas por meio da análise de publicações e seus metadados, gerando publicações científicas capazes de expandir o conhecimento de temas muitas vezes complexos.

De modo a exemplificar e diferenciar uma lista de uma matriz, considere um universo de publicações sobre uma temática a ser analisada. A lista consiste na contagem de elementos/campos previamente selecionados desse conjunto, buscando adquirir a quantidade total dos objetos de reflexão nesse conjunto, como por exemplo a quantidade total de publicações sobre bibliometria (tema) ao longo dos anos (elemento/campo). Já as matrizes também trabalham com a contagem, sendo a sua principal diferença que a mesma busca contar as coocorrências de elementos distintos ou iguais, sendo exemplos uma matriz de “Autor x Ano”, onde são contabilizadas quantas publicações cada autor efetuou em cada ano, e também a de uma matriz

quadrada “Autor x Autor”, onde são contabilizadas quantas publicações cada autor faz em conjunto com outros autores. Listas e matrizes são elementos essenciais em análises bibliométricas automatizadas e, devido a isso, muitas ferramentas dão suporte a esse tipo de análise direta ou indiretamente (COBO, *et. al.* 2011; MORAL-MUÑOZ, 2020).

### 3.6.1 Construção de listas

Focando em específico no processo de geração de uma lista, é necessário detalhar as listas trabalhadas, qual o processo de contagem utilizado e como as listas são retornadas para o pesquisador.

Primeiramente, sobre as listas trabalhadas, as mesmas são geradas e referentes aos metadados selecionados e extraídos anteriormente durante a importação, portanto um pesquisador pode, por exemplo, requisitar uma lista de publicações por ano, por autores e entre outras. É importante expor que as listas também são referentes a uma fonte de informação e a um usuário, fazendo com que a geração das mesmas só utilize publicações da fonte definida pelo mesmo e que também foram importadas por tal. Isso garante com que as contagens sejam feitas corretamente e que não sejam contabilizadas publicações da PL em uma lista da WoS ou que sejam contabilizadas publicações importadas por outros pesquisadores.

Sobre a contagem, foi utilizada a contagem completa exposta por Alvarado (2002), para todos os tipos de listas e metadados, sendo assim não havendo distinções entre os elementos a serem contados; e a mesma é feita diretamente no banco de dados da ferramenta. O motivo de escolha desse tipo de contagem se deve ao fato de outras ferramentas já consolidadas como o VP fornecerem a mesma por padrão, conforme exposto em sua documentação técnica (SEARCH TECHNOLOGY, 2015), e também por ser a contagem frequentemente utilizada em várias pesquisas bibliométricas (FARIA, 2001; JOB, 2008; PERISSINI; FARIA, 2021). A justificativa para a realização da contagem diretamente no banco de dados e não pelo servidor do BiblioAnalytics é para evitar que um grande número de dados seja transportado entre o banco e o servidor desnecessariamente em cada análise.

Por fim, após efetuadas as contagens, é necessário retornar as mesmas para a pessoa que as requisitou. De modo a facilitar a compreensão das listagens, optou-

se por retornar uma planilha do Microsoft Excel contendo a lista requerida, por ser um formato de arquivo bastante conhecido por pesquisadores e pessoas em geral; e também por ser de fácil manipulação. A construção desse arquivo é feita por meio da utilização da biblioteca *json2xls*<sup>55</sup> que recebe o JSON com as contagens efetuadas no banco de dados e o converte em um arquivo de extensão “.xlsx”. Seria possível retornar as mesmas diretamente em formato JSON, sem a necessidade da biblioteca, o que seria mais interessante para utilização por outros sistemas e ferramentas, porém, a princípio, de modo a expandir o acesso aos estudos métricos buscou-se facilitar a utilização do usuário.

De modo a exemplificar o processo exposto de requisição de uma lista, a seguir é detalhado o mesmo em um caso hipotético:

1. Pesquisador informa no site (*front end*) qual a lista que ele deseja e qual a fonte de informação das publicações. Por exemplo: uma lista de publicações por ano da base de dados WoS.
2. Site informa à API da WoS no servidor o tipo de lista necessária e qual o pesquisador que está requisitando essa lista.
3. Servidor se comunica com o banco de dados pedindo uma lista de publicações por ano da base de dados WoS para o pesquisador em questão.
4. Banco de dados busca na coleção de publicações da WoS as publicações do pesquisador e efetua, a partir dessas publicações, a contagem por ano.
5. Banco de dados retorna ao servidor um arquivo no formato JSON contendo a lista com a contagem de publicações por ano.
6. Servidor então, por meio da utilização da biblioteca *json2xls*, converte o arquivo em formato JSON para uma planilha do Excel e retorna a mesma para o *front end*.
7. Finalmente, o site então inicia o download da planilha do Excel contendo a lista requisitada.

Para que fosse possível atingir essas etapas, foi necessária a construção de duas APIs no servidor “*/wos/getFrequency*” e “*/lattes/getFrequency*” para receber a requisição do pesquisador contendo o tipo de lista que deseja ser gerada, a fonte de informação e o pesquisador que está requisitando a lista. Para realização da consulta no banco, efetuar a contagem e posterior geração do arquivo contendo a lista de acordo com a requisição recebida nessas duas APIs, foi necessária a criação de dois scripts: ***wosGetFrequency*** e ***lattesGetFrequency*** que efetuam, ao serem executados, os passos 3, 4, 5 e 6. Ambos os scripts e APIs são similares, se

---

<sup>55</sup> <<https://www.npmjs.com/package/json2xls>>

diferenciando somente sobre a coleção a ser consultada para contagem e criação das listas, no caso de um para publicações da WoS e do outro para a PL.

### 3.6.2 Construção de matrizes

Antes de iniciar a construção das matrizes foi necessário primeiramente definir o tipo de matriz a ser disponibilizada e trabalhada pelo BiblioAnalytics. Uma das matrizes que frequentemente se utiliza em estudos bibliométricos e também constantemente disponibilizada por ferramentas bibliométricas, como o VOSviewer e Gephi (BASTIAN; HEYMANN; JACOMY, 2009; VAN ECK; WALTMAN, 2010), é a matriz de coocorrência, sendo mais específico as matrizes quadradas, que também pode receber outros nomes dependendo de seu contexto de utilização, como por exemplo matriz de coautoria (no contexto da análise de autores), cocitação (no contexto da análise de citações), colaboração (no contexto da análise de colaboração entre instituições) e entre outras (FARIA, 2015). Essa matriz retrata a coocorrência de elementos de um mesmo contexto (como por exemplos Autores x Autores), ou no contexto da ferramenta campos que apresentam múltiplos elementos, em um conjunto de publicações a serem analisada, sendo frequentemente utilizada em indicadores de ligação ou relacional (CALLON et al,1993; OKUBO, 1997).

Dada a relevância da matriz em questão na área da bibliometria e sua fácil compreensão, optou-se por disponibilizar a mesma no BiblioAnalytics, evitando conceitos e teorias mais complexas tanto para o usuário que irá utilizar a ferramenta quanto na questão do desenvolvimento da mesma. Vale ainda abrir um parêntesis, destacando que a opção de disponibilização de um arquivo GDF no lugar de matrizes também foi considerada, mas após ser investigado o mesmo se mostrou demasiadamente complexo para o escopo da pesquisa em questão, sendo válido citar a inexistência de bibliotecas JavaScript que auxiliem na geração do mesmo.

Assim como na geração de listas, durante a elaboração de uma matriz é necessário também a realização de contagens para a verificação da quantidade de coocorrências entre os elementos. Para tal, de modo análogo à contagem das listas e também aos argumentos expostos durante a explicação das mesmas, optou-se pela utilização da contagem completa exposta por Alvarado (2002), onde no caso da matriz os vértices da diagonal principal da mesma contêm a quantidade total de publicações que têm o elemento que está sendo correlacionado, e nos demais a quantidade de

coocorrência entre elementos (Figura 29). Vale ressaltar que a soma contida na diagonal principal não necessita ser necessariamente igual à soma de suas coocorrências; exemplificando, o “Author 3” por exemplo apresenta sete publicações totais no conjunto a ser analisado, mas somente duas delas com outros autores (no caso o “Author 3”), o que significa que suas outras cinco publicações não houveram coautoria.

**Figura 29** – Exemplo de uma matriz de coautoria

	Author 1	Author 2	Author 3	Author 4	Author 5
Author 1	5		2	3	
Author 2		2			
Author 3	2		7		1
Author 4	3			3	2
Author 5			1	2	4

Fonte: Elaborada pelo autor.

Diferente da geração das listas, além das contagens, são necessárias outras etapas durante a geração de uma matrix no BiblioAnalytics, como a organização e ordenação dos elementos da matriz. Para facilitar o processo de criação de uma matriz no BiblioWeb, suponhamos um exemplo hipotético no Quadro 18, onde se tem um conjunto de seis publicações e suas ‘wosCategories’ (i); primeiro obtem-se a lista com a contagem das ‘wosCategories’ (ii); depois obtem-se a lista com a contagem das coocorrências (iii); ademais junta-se a lista com os vértices de coocorrência, onde a lista é a diagonal principal da matriz e os demais vértices (iii) gerando o JSON (iv); por fim o JSON (iv) é passado para a biblioteca ‘json2xls’ que gera a matriz (v). Vale ressaltar que devido à característica da biblioteca em questão, é necessário em (iv) que o primeiro elemento do JSON contenha todas as ‘wosCategories’ para a geração do cabeçalho da matriz.

Conforme pode ser visto, nota-se que a geração de matrizes é um processo de maior complexidade que a geração de listas, o que significa que quanto mais elementos (no exemplo, ‘wosCategories’) são objetos de análise, maior a complexidade computacional e a quantidade de operações necessária para sua geração e organização, diferentemente das listas que acaba sendo linear. De modo a exemplificar esse processo do ponto de vista do pesquisador/usuário que utiliza a

ferramenta, o mesmo segue as mesmas etapas de 1 a 7 explanadas na seção 3.5.1 para a geração de uma lista, com a diferença que agora se trata-se de uma matriz, e portanto é necessário que diversas outras operações sejam feitas no servidor, apesar de serem invisíveis para quem usa a ferramenta.

**Quadro 18** – Exemplos de JSONs durante a geração de uma matriz na ferramenta BiblioAnalytics e sua matriz resultante.

(i) 6 publicações hipotéticas e suas wosCategories	(ii) Lista de wosCategories	(iii) JSON com vértices da coocorrência de wosCategories	(iv) JSON da matrix no formato necessário para funcionar com a biblioteca json2xls																									
<pre>[ {   'wosCategories': [     'computer science',     'information science'   ] }, {   'wosCategories': [     'information science',     'computer science'   ] }, {   'wosCategories': [     'mathematics',     'information science'   ] }, {   'wosCategories': [     'information science',     'mathematics',     'geology'   ] }, {   'wosCategories': [     'geology',     'information science'   ] }, {   'wosCategories': [     'mathematics',     'information science'   ] } ]</pre>	<pre>[ {   _id: 'information science',   sum: 6 }, {   _id: 'mathematics',   sum: 3 }, {   _id: 'geology',   sum: 2 }, {   _id: 'computer science',   sum: 2 } ]</pre>	<pre>[ {   _id: {     a: 'mathematics',     b: 'information science'   },   n: 3 }, {   _id: {     a: 'information science',     b: 'geology'   },   n: 2 }, {   _id: {     a: 'information science',     b: 'computer science'   },   n: 2 }, {   _id: {     a: 'mathematics',     b: 'geology'   },   n: 1 } ]</pre>	<pre>{   '': 'computer science',   'computer science': 2,   'geology': '',   'information science': 2,   'mathematics': '' }, {   '': 'geology',   'information science': 2,   'mathematics': 1,   'geology': 2 }, {   '': 'information science',   'mathematics': 3,   'geology': 2,   'computer science': 2,   'information science': 6 }, {   '': 'mathematics',   'information science': 3,   'geology': 1,   'mathematics': 3 }</pre>																									
(v) <b>Matriz gerada dentro do arquivo Excel</b>																												
<table border="1" style="margin-left: auto; margin-right: auto;"> <thead> <tr> <th></th> <th>computer science</th> <th>geology</th> <th>information science</th> <th>mathematics</th> </tr> </thead> <tbody> <tr> <th>computer science</th> <td>2</td> <td></td> <td>2</td> <td></td> </tr> <tr> <th>geology</th> <td></td> <td>2</td> <td>2</td> <td>1</td> </tr> <tr> <th>information science</th> <td>2</td> <td>2</td> <td>6</td> <td>3</td> </tr> <tr> <th>mathematics</th> <td></td> <td>1</td> <td>3</td> <td>3</td> </tr> </tbody> </table>					computer science	geology	information science	mathematics	computer science	2		2		geology		2	2	1	information science	2	2	6	3	mathematics		1	3	3
	computer science	geology	information science	mathematics																								
computer science	2		2																									
geology		2	2	1																								
information science	2	2	6	3																								
mathematics		1	3	3																								

Fonte: Elaborado pelo autor.

Por fim, no ponto de vista dos scripts e APIs necessários para sua disponibilização, as matrizes utilizam dos mesmos utilizados pelas listas, duas APIs “/wos/getFrequency” e “/lattes/getFrequency” para receber a requisição do pesquisador, sendo que neste caso é informado que o tipo de análise é “matriz”, qual o campo a ser retratado, a base de dados e o usuário que requisitou a análise. Ainda no servidor são repassados esses dados para os scripts **wosGetFrequency** e **lattesGetFrequency** que efetuam as contagens e geram a matriz solicitada.

### 3.7 Refinamento das análises por meio da utilização de tesouros

Quando são feitas análises, sejam elas a construção de listas ou matrizes, sem qualquer preparação dos dados das publicações, podem existir casos onde um mesmo campo e seu metadado representam a mesma informação, como por exemplo um nome de autor, representado de maneiras diferentes em cada publicação. Conforme já foi amplamente explorado, os vocabulários controlados, no caso deste trabalho os tesouros, podem ser de grande ajuda em aprimorar as contagens fornecidas pelo BiblioAnalytics.

Para a disponibilização de tal funcionalidade na ferramenta foi necessário primeiramente analisar como outras ferramentas que dispunham de tal mecanismo ofereciam o mesmo a seus usuários. Com isso procurou-se por ferramentas amplamente utilizadas em estudos bibliométricos nacionais e internacionais que forneciam tal operação, como o caso dos *softwares* VOSviewer e VP. Por meio da análise da análise da documentação técnica dos autores (VAN-ECK; WALTMAN, 2018; SEARCH TECHNOLOGY, 2015), notou-se que os mesmos permitem que o usuário forneça seu próprio arquivo de tesouro durante a utilização (Quadro 19), um arquivo de texto formatado de maneiras diferentes, sendo possível elencar algumas diferenças entre os mesmos:

- **VantagePoint**: tesouro é criado com um arquivo de texto onde a linha iniciada com “\*\*\*” indica o elemento que será aplicado e as demais linhas sub que serão substituídas se iniciam com “100 1” seguido do elemento a ser substituído. Vale ressaltar que o elemento em questão pode ser uma expressão regular e que a separação dos elementos é feita por um espaço simples (SEARCH TECHNOLOGY, 2015).

- **VOSviewer:** tesauro é criado com um arquivo de texto onde a primeira linha define duas colunas chamadas de “label e “replace by”, indicando quais elementos serão procurados e posteriormente trocados. As linhas subsequentes contêm os elementos que serão substituídos na primeira coluna e os elementos que irão substituir na segunda coluna, sendo que os elementos em questão são separados por uma tecla “TAB” (VAN-ECK; WALTMAN, 2018).

**Quadro 19** – Exemplo de tesauro em arquivo de texto para as ferramentas VP e VOSviewer

Tesauro no VP	Tesauro no VOSviewer
**BRAZIL	label replace by
100 1 brasil	brasil BRAZIL
100 1 brazil	brazil BRAZIL
100 1 Brasil	Brasil BRAZIL
100 1 ^Brasil\$	United States of America USA
**USA	Estados Unidos da America USA
100 1 United States of America	Estados unidos america USA
100 1 Estados Unidos da America	
100 1 estados unidos américa	

Fonte: Elaborado pelo autor.

Conforme pode ser notado, ambas as ferramentas cumprem seus objetivos, sendo possível apontar que a aplicação de um tesauro pelo VP, apesar de ser mais robusta por permitir a utilização de expressões regulares, acaba sendo mais complexa tanto na questão de viabilização da funcionalidade e sua construção, quanto na visão do usuário da ferramenta que pode não entender corretamente sua finalidade e modo de utilização. Dada a simplicidade do formato de arquivo para um tesauro do VOSviewer, o tesauro desenvolvido no BiblioAnalytics buscou seguir sua estrutura, sendo possível sua utilização tanto no BiblioAnalytics quanto no VOSviewer.

Para que o mesmo pudesse ser feito no BiblioAnalytics foi necessária a construção de duas APIs no servidor “/wos/applyThesaurus” e “/lattes/applyThesaurus” para receber a requisição do pesquisador contendo: (i) o seu arquivo de tesauro; (ii) a fonte de informação em que o mesmo será aplicado; (iii) pesquisador que está requisitando a aplicação do tesauro; (iv) o campo em que o tesauro será aplicado. O servidor necessita de quatro *scripts* (dois para cada fonte de informação) para receber essa requisição e aplicar o tesauro, sendo eles:



- **processThesaurusFile**: responsável por ler o arquivo de tesauro fornecido e organizar os valores que serão substituídos e suas substituições. Essa organização depende da correta formatação do arquivo, por isso é extremamente importante que ao se criar o tesauro, o usuário separe os valores com a tecla “TAB” e o arquivo seja um arquivo de texto.
- **applyThesaurus**: responsável por receber os valores já organizados e efetuar a aplicação do tesauro das publicações importadas pelo usuário no banco de dados. Como o tesauro depende do usuário para sua correta criação, caso o mesmo confunda os campos que deseja substituir, o mesmo pode ser prejudicial para suas análises.

### 3.8 Possibilitando análises altimétricas por meio de mecanismos de acesso

Dada a relevância crescente dos estudos métricos, neste caso referindo-se especificamente aos estudos métricos alternativos, é importante que a ferramenta explorada neste estudo ofereça mecanismos para que esse tipo de métrica possa ser avaliada. De modo a oferecer uma fundação para que os mesmos possam ser explorados no futuro, é necessário haver um mecanismo de identificar os usuários/pesquisadores que utilizam a ferramenta: um sistema de autenticação (*login*).

O sistema de autenticação aplicado na ferramenta necessita que o pesquisador efetue um cadastro antes de utilizar efetivamente a ferramenta. Sempre pensando na simplicidade, para efetuar o cadastro o mesmo só necessita informar o seu e-mail e uma senha de acesso, com isso o mesmo está apto para utilizar as funcionalidades da ferramenta. Pelo fato de o pesquisador estar cadastrado, durante o processo de importação, conforme exposto previamente, é registrado qual pesquisador importou cada publicação. Por meio do e-mail, caso o pesquisador utilize seu e-mail institucional, e sabendo quais suas publicações importadas, abre-se um leque gigantesco de análises que poderiam ser feitas tanto por instituição quanto por pesquisador, exemplificando algumas: (i) publicações mais importadas; (ii) áreas de pesquisa mais pesquisadas; (iii) tipo de análise mais efetuada; (iv) horários de maior acesso; (v) diversas outras.

O domínio de estudos das métricas alternativas pode ser expandido ainda mais

no futuro caso sejam fornecidas outras informações durante o cadastro do mesmo, como por exemplo sua localização (estado, cidade e etc), titulação do pesquisador, área de atuação e entre outras. Por meio destas informações extras seria possível até fazer um mapeamento regional e/ou por titulação dos temas apresentados anteriormente e entre outros.

Para que fosse possível atender os pontos relacionados à autenticação em questão foi necessária a criação de duas APIs no servidor (“/login” e “/signup”), responsáveis por receber as requisições de login e de cadastro no servidor, e três scripts:

- **checkAuth:** intercepta todas as requisições feitas por um usuário por meio da interface (*front end*) ou por uma máquina diretamente em qualquer API da ferramenta e valida se a requisição em questão está sendo realizada por um usuário autenticado na mesma por meio da verificação de sua chave de acesso no banco de dados. Caso não esteja, não prossegue o fluxo e informa ao pesquisador ou máquina que se autentique. Esse script é essencial para que pessoas maliciosas não consigam acessar as publicações de outros usuários.
- **createUser:** recebe os dados informados no formulário de cadastro (e-mail e senha) disponíveis no *front end* e efetua o devido cadastro do mesmo, caso não exista, na coleção de *users* do banco de dados.
- **userLogin:** recebe os dados informados, no *front end*, pelo usuário que deseja se autenticar e valida os mesmos no banco de dados e, caso estejam válidos, efetiva o acesso do mesmo retornando sua chave de acesso que fica válida por uma hora, sendo necessário novo login após o término do prazo em questão.

### 3.9 Mecanismos extras para auxiliar na democratização de análises bibliométricas

Apesar de a ferramenta desse estudo ter sido desenvolvida de modo a ser disponibilizada na *web*, sem a necessidade que o pesquisador necessite ter condição financeira e conhecimentos de computação para sua instalação, outros elementos podem ser de extrema importância para a democratização do conhecimento, acesso e, conseqüentemente, da bibliometria. Nas ferramentas previamente citadas e exploradas neste estudo, nota-se constantemente a falta de instrumentos que atendam a possíveis necessidades de pesquisadores brasileiros na utilização da mesma como a inexistência do idioma português e de mecanismos que possibilitem a análise bibliométrica de publicações do campo brasileiro, pontos que muitas vezes inviabilizam a utilização das mesmas por esse nicho de pesquisadores. Em um contexto mais amplo, sem considerar o caso dos pesquisadores brasileiros, é possível citar também questões que poderiam ter sido atendidas para contribuir com a democratização de análises bibliométricas, como por exemplo a disponibilização da ferramenta gratuitamente, a inexistência de documentação não técnica e técnica (código fonte) e a necessidade de instalação de algumas ferramentas. Nesta seção serão explorados alguns desses pontos ainda não abordados, como a documentação e a disponibilização de múltiplos idiomas durante a utilização do BiblioAnalytics.

De modo a atender algumas das questões elencadas, iniciando com a temática da documentação, foi criado dentro da própria ferramenta BiblioAnalytics uma seção onde são disponibilizados guias não técnicos (um sobre como utilizar a ferramenta e outro de como utilizar o tesouro) para que o usuário leigo possa entender como funciona a ferramenta e como utilizá-la, em conjunto com amostras de publicações da WoS e PL e também um exemplo de tesouro. Além de detalhar um passo a passo de utilização da ferramenta, são disponibilizadas explicações de alguns conceitos técnicos da área, como os tesouros. Sobre a documentação técnica da ferramenta, a mesma será disponibilizada no GitHub do NIT/MATERIAIS<sup>56</sup>, para aqueles que tenham conhecimento técnico consigam contribuir e gerar novos estudos a partir da versão da ferramenta disponibilizada nesta pesquisa.

Para a questão do idioma, foi criada uma configuração na ferramenta onde o

---

<sup>56</sup> <https://github.com/nitmateriais>

pesquisador pode alternar entre os idiomas “Inglês” e “Português”. O motivo da escolha do idioma “Inglês” se pauta no fato de que o mesmo é o idioma padrão e amplamente utilizado no contexto de publicações científicas e também, conforme pôde ser visto sobre as ferramentas bibliométricas, em ferramentas que trabalham com a temática dos estudos métricos. A escolha do segundo idioma ser o “Português” se deve ao fato de que, como o BiblioAnalytics trabalha com publicações da PL, que é em sua maior parte acessada por pesquisadores fluentes na língua em questão e por isso é interessante a disponibilização da língua em questão para que não se limite a utilização da ferramenta somente por pesquisador que falam a língua inglesa. Vale reforçar que o “Português” que é referido neste estudo se trata do “Português” do Brasil, dado que é a variação utilizada por uma maior quantidade de pessoas.

Por meio da disponibilização de documentações e de uma segunda linguagem (além de outras questões já exploradas neste estudo como a importância de uma ferramenta *web*, sua gratuidade e entre outras), espera-se que o BiblioAnalytics possa contribuir com a acessibilidade a uma temática tão importante como a bibliometria e suas análises para um número ainda mais de pessoas, sejam elas pesquisadores da área ou não.

## 4 RESULTADOS

De modo a atender os objetos previamente definidos, foi desenvolvido BiblioAnalytics uma ferramenta bibliométrica onde o usuário e/ou pesquisador por meio de uma SPA em Angular, um servidor de APIs em NodeJS e um banco de dados MongoDB, consegue: se registrar, efetuar o *login*, importar publicações das bases de dados WoS e PL, fazer análises de tais publicações por meio da construção de listas e matrizes, aplicar um tesouro, deletar tais publicações e ter acesso a documentos e amostras que facilitem sua utilização da ferramenta.

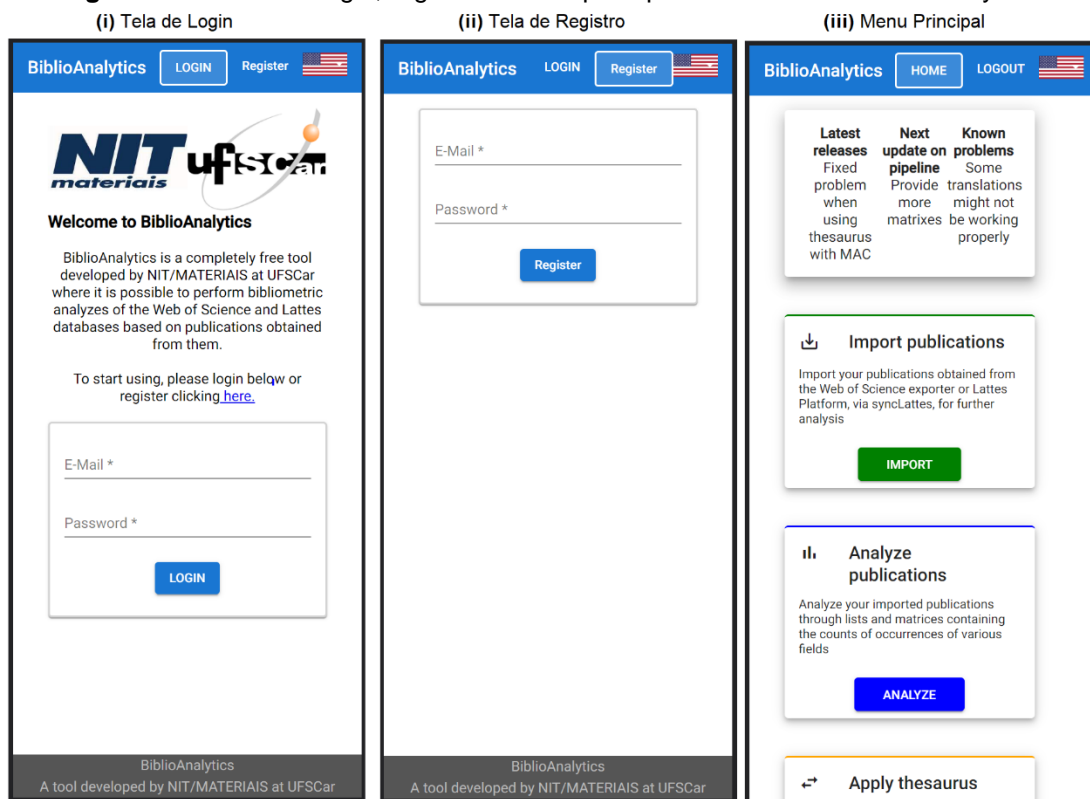
### 4.1 A ferramenta BiblioAnalytics e seu funcionamento

Pelas Figuras 30, 31 e 32 pode ser visto com detalhes as telas onde o usuário pode efetuar as operações disponíveis na ferramenta, neste caso do ponto de vista de utilização de um aplicativo móvel. Conforme exposto anteriormente, pelo fato de se utilizar tecnologias como Angular, as mesmas podem também ser acessadas e visualizadas de modo satisfatório por meio de dispositivos móveis e por meio de um computador tradicional, oferecendo as mesmas funcionalidades independente do acesso efetuado, seja em um computador de mesa quanto em um aparelho celular. Resumidamente, cada uma das telas em questão, independente do modo de acesso, com base na numeração apresenta nas figuras oferecem ao usuário da ferramenta:

- (i) **tela de login**: tela inicial da ferramenta onde o usuário é redirecionado em seu primeiro acesso, caso não tenha “entrado” ainda na ferramenta, dando um breve resumo da ferramenta e requisitando o e-mail e senha cadastrado; a mesma também oferece meios de o usuário ir até a tela de registro caso a pessoa em questão ainda não esteja cadastrada;
- (ii) **tela de cadastro**: local onde é possível com que seja realizado um novo cadastro na ferramenta, informando apenas o e-mail e a senha desejadas para acessar a mesma.
- (iii) **tela principal para navegação**: expõe o menu da ferramenta contendo as diversas opções disponíveis na mesma como importar, analisar e deletar publicações; aplicar tesouros e verificar os guias e amostras da ferramenta; vale reforçar que a tela em questão é a tela principal do sistema e,

portanto, caso o usuário já tenha feito a operação de login, é onde o mesmo será redirecionado.

**Figura 30** – Telas de login, registro e menu principal da ferramenta BiblioAnalytics.



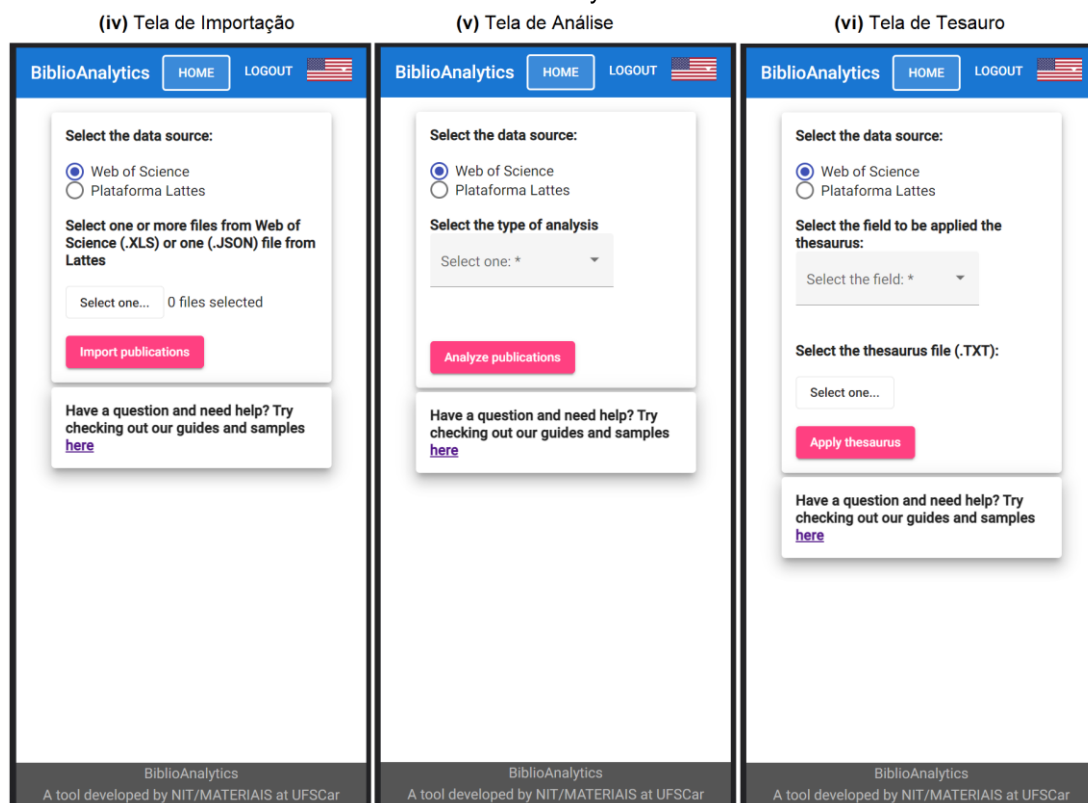
Fonte: Elaborada pelo autor

- (iv) **tela de importação:** utilizada para importação das publicações das duas fontes, sendo necessário selecionar a fonte em questão, informar as publicações que serão importadas (".xls" para WoS e .json para Lattes); vale ressaltar que há uma limitação por importação de dez arquivos para a WoS e de um arquivo para a PL, totalizando 50mb por importação, para que não seja sobrecarregada a ferramenta e a conexão do pesquisador, podendo o processo em questão ser repetido para quantas publicações forem necessárias
- (v) **tela de análise de publicações:** oferece ao usuário opções de análises de suas publicações importadas, sendo necessário informar a fonte de informação em que será feita a análise (WoS ou PL), o tipo de análise (listas ou matrizes) e o metadado que será analisado entre as opções disponíveis; tem como resultado um arquivo em formato de planilha do Microsoft Excel

contendo a lista ou matriz gerada;

- (vi) **tela de aplicação de tesouro**: provê mecanismos de o pesquisador fornecer seu próprio arquivo de tesouro para refinar suas análises do jeito que lhe achar relevante; basta que o mesmo siga a formatação descrita nos guias do BiblioAnalytics.

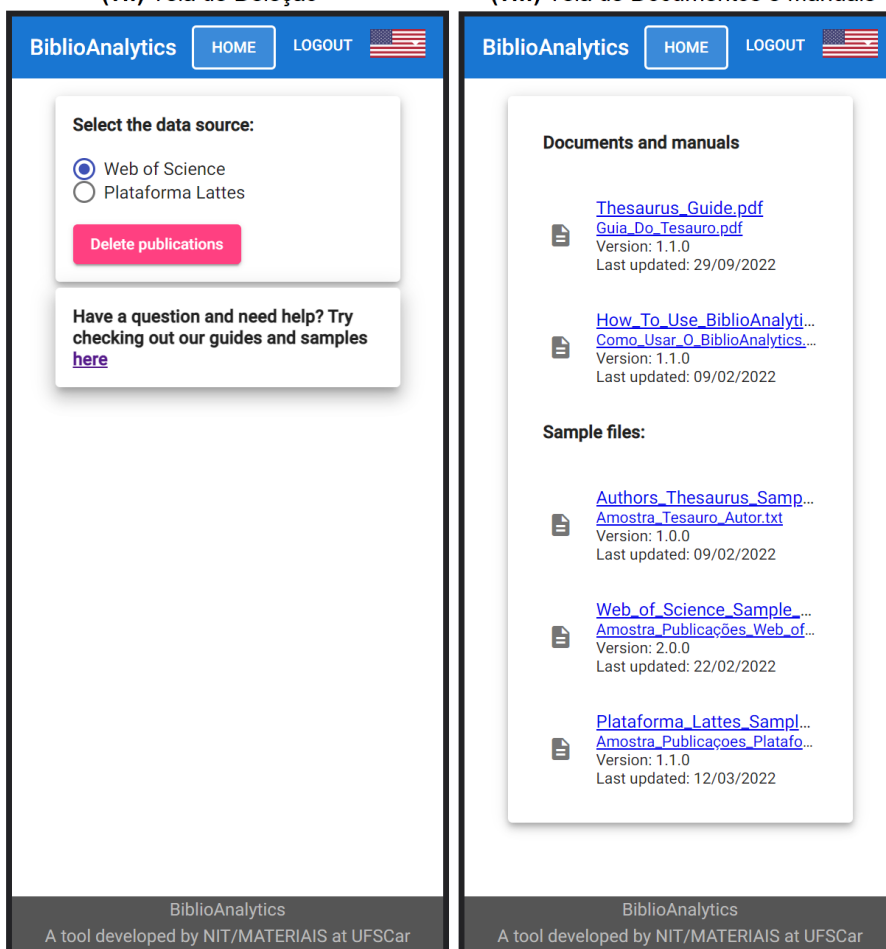
**Figura 31** – Telas de importação, análise e aplicação de tesouro da ferramenta BiblioAnalytics



Fonte: Elaborada pelo autor.

- (vii) **tela de deleção**: fornece o mecanismo de o pesquisador excluir suas publicações importadas para que novas análises possam ser efetuadas, sendo necessário apenas informar a fonte de informação em que deseja efetuar a operação em questão.
- (viii) **tela de documentos e manuais**: fornece ao usuário da ferramenta guias e manuais explicando como funciona a ferramenta e suas funções e como utilizar a mesma adequadamente; nela é possível também ter acesso a amostras de publicações de ambas as fontes de informação e um exemplo de tesouro

**Figura 32 – Telas de deleção e de documentos e manuais**  
 (vii) Tela de Deleção (viii) Tela de Documentos e Manuais



Fonte: Elaborada pelo autor.

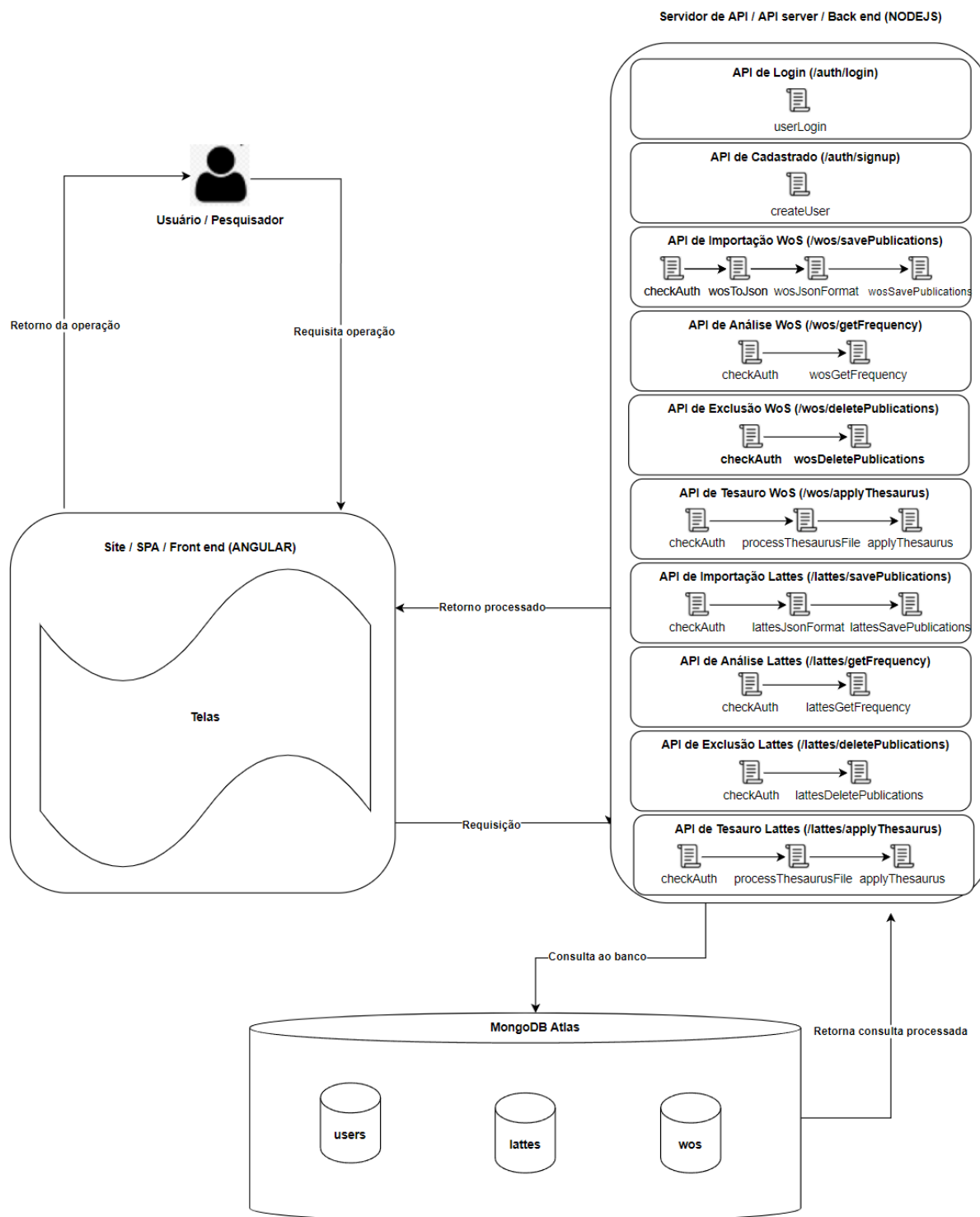
Todas as telas em questão, quando requisitado pelo pesquisador uma operação e, independente do dispositivo em que se está acessando a página, acessa o servidor de APIs (Figura 33) para a execução da mesma por meio dos *scripts* e APIs desenvolvidos durante esta pesquisa. O mesmo durante a utilização da ferramenta pode ainda selecionar, clicando na bandeira presente na parte de cima da ferramenta, o idioma que deseja utilizar.

Por fim, a ferramenta e suas páginas estão acessíveis no ambiente *web* e a mesma já encontra disponível<sup>57</sup> para utilização de toda a comunidade científica, sendo necessário ressaltar que a mesma trata-se de um projeto em constante evolução e, portanto, problemas poderão ser encontrados e trabalhados com o passar do tempo.

<sup>57</sup> <http://biblioanalytics.s3-website-us-east-1.amazonaws.com>



**Figura 33 – Visão técnica integral do BiblioAnalytics**



Fonte: Elaborada pelo autor

## **4.2 Análise das contagens oferecidas pelo BiblioAnalytics por meio de listas e matrizes em diferentes cenários**

Para validar as contagens efetuadas pelo BiblioAnalytics, foram selecionados alguns dos metadados extraídos das publicações para serem contados e comparados durante a geração de listas e matrizes, conforme cada caso a ser considerado. Sobre as publicações utilizadas nesta seção, para a WoS, foi utilizado um conjunto de 7104 publicações sobre “bibliometrics” extraídas da WoS por meio do Registro Completo da mesma em formato Excel, enquanto para a PL é utilizado um conjunto de 381 publicações do DCI da UFSCar obtidas pelo syncLattes, quando necessário.

### **4.2.1 Contagens após a extração de metadados**

A primeira validação a ser realizada foi a contagem dos elementos contidos no Excel da WoS e do JSON da PL. Esse processo buscou verificar, antes de processamentos mais complexos serem aplicados pela ferramenta e a geração das listas e matrizes, se a mesma está trabalhando com o conjunto de metadados correto e que nenhum desses se perdeu durante sua etapa inicial de extração.

Para isso foram selecionados dois campos, um da WoS e outro da PL, onde foi verificada a quantidade de elementos presentes em cada um dos arquivos e se sua contagem está correta. A ferramenta VP possui mecanismo de importação de arquivos em formato Excel sem a aplicação de filtros e, devido a isso, foi utilizada para efetuar essa validação preliminar dos arquivos da WoS. Já para a PL, se desconhece de uma ferramenta bibliométrica que aceite a importação de publicações em formato JSON fornecidas pelo syncLattes e, devido a isso, foi necessária a conferência manual.

Conforme pode ser visto pelo Quadro 20, o campo escolhido tanto para a WoS quanto para a PL foi o de “Publication Type”. Sua escolha se deve ao fato de o mesmo conter poucas variações em ambas as fontes de informação, facilitando a conferência manual nessa primeira verificação.

Analisando a contagem, pode então ser constatado pelo quadro que a extração efetuada pelo BiblioAnalytics ocorre corretamente, dado que tanto a quantidade de elementos diferentes (tipos de publicações, no caso da WoS: J, C, S e B) quanto suas contagens apresentaram valores iguais tanto entre ferramentas quanto com a contagem manual.

**Quadro 20** – Comparação da contagem do campo “Publication Type” para a WoS e PL.

<b>Comparação WoS</b>			
<b>BiblioAnalytics</b>		<b>VantagePoint</b>	
J	6427	J	6427
C	671	C	671
S	4	S	4
B	2	B	2
<b>Comparação PL</b>			
<b>BiblioAnalytics</b>		<b>Manual</b>	
article	227	article	227
conferenceobject	136	conferenceobject	136
book	14	book	14
bookedt	4	bookedt	4

Fonte: Elaborado pelo autor.

#### 4.2.2 Contagens após todos os processamentos: a geração de uma lista

A próxima contagem a ser explorada é a realizada após todos os processamentos efetuados pelo BiblioAnalytics, como as listas oferecidas pela ferramenta. Pelo fato de os campos escolhidos na validação passada serem pouco afetados pelos processamentos, uma vez que não há a necessidade de os mesmos serem normalizados, terem sua acentuação e conteúdos desnecessário removidos e entre outros, optou-se para, neste caso, utilizar os campos de “Organizations” para a WoS e o de “Authors” para a PL para tentar demonstrar os impactos que os processamentos podem ter.

Conforme pode ser observado no Quadro 21, nota-se nas “Organizations” da WoS a presença de inúmeros elementos desnecessários para a identificação da instituição de ensino, como por exemplo os nomes dos autores e endereços, além de também ser possível observar que alguns desses elementos contêm mais de uma organização, como o caso do ítem contendo a contagem (6), sendo necessária a sua devida separação. Para os “Authors” da PL, pode-se observar a inexistência de um padrão das letras, onde alguns autores são representados em sua totalidade por letras maiúsculas enquanto outros não, podendo também citar que alguns deles apresentam acentuação, como o caso do autor “MARTINEZ, Vinício Carrilho”.

**Quadro 21** – Os dez elementos com maior coocorrência para os campos “Organizations” e “Authors” no conjunto de publicações da WoS e da PL após a extração

<b>“Organizations” da WoS</b>	
[Prathap, Gangan] APJ Abdul Kalam Technol Univ, Thiruvananthapuram 695016, Kerala, India	9
Drexel Univ, Coll Informat Sci & Technol, Philadelphia, PA 19104 USA	7
CSIR Natl Inst Sci Commun & Informat Resources, New Delhi 110012, India	7
[Bornmann, Lutz] Max Planck Gesell, Adm Headquarters, Div Sci & Innovat Studies, Hofgartenstr 8, D-80539 Munich, Germany; [Haunschild, Robin] Max Planck Inst Solid State Res, Heisenbergstr 1, D-70569 Stuttgart, Germany	6
[Prathap, Gangan] Vidya Acad Sci & Technol, Trichur 680501, India; [Prathap, Gangan] APJ Abdul Kalam Technol Univ, Thiruvananthapuram 695016, Kerala, India	5
[Bornmann, Lutz] Max Planck Gesell, Adm Headquarters, Div Sci & Innovat Studies, Hofgartenstr 8, D-80539 Munich, Germany	4
Off Naval Res, Arlington, VA 22217 USA	4
Adm Headquarters Max Planck Soc, Div Sci & Innovat Studies, D-80539 Munich, Germany	4
Hebrew Univ Jerusalem, Sch Lib Arch & Informat Studies, IL-91904 Jerusalem, Israel	3
German Federal Ministry for Education and Research (BMBF) Federal Ministry of Education & Research (BMBF) [01PQ08004A]	3
<b>“Authors” da PL</b>	
MARTINEZ, Vinício Carrilho	80
Hayashi, Maria Cristina Piumbato Innocentini	32
PEDRO, Wilson José Alves	31
GRACIOSO, L. S.	23
GRACIOSO, LUCIANA DE SOUZA	22
AMARAL, R. M.	18
DONADONE, J. C.	18
ZANIN, M.	17
FURNIVAL, A. C. M.	15
HOFFMANN, W. A. M.	14

Fonte: Elaborado pelo autor.

Para a validação e comparação das contagens obtidas pelo BiblioAnalytics após todos os processamentos, foram obtidas as contagens fornecidas pela ferramenta VP para os mesmos conjuntos de publicações, com a utilização de filtros utilizados por membros do NIT/MATERIAIS da UFSCar, no caso o “ISI – Web of Science v1.6” para a WoS e o “syncLattes v1.3” para a PL. Vale ressaltar que pelo fato de o VP, na versão utilizada durante esta pesquisa, não apresentar filtros para a importação de planilhas do Excel da WoS e também para arquivos JSON do syncLattes, foram utilizados os arquivos ISI-CE e RIS modificados contendo o mesmo conjunto de publicações.

Ao se analisar o Quadro 22 é possível notar primeiramente, principalmente para o campo de “Organizations”, a importância dos processamentos para uma contagem mais precisa e adequada. Quando comparadas as contagens feitas por ambas as ferramentas, nota-se resultados iguais entre ambas no campo “Organizations” da

WoS, mostrando como o BiblioAnalytics é capaz de obter resultados próximo ao VP, ferramenta já consolidada a anos na área. Já para o campo “Authors” da PL, é possível ver que, apesar de as contagens em grande parte serem iguais, alguns elementos apresentaram contagens diferentes, como é o caso dos autores “martinez, inicio carrilho” e “pedro, wilson jose alves”. Isso mostra como os processamentos efetuados pelo BiblioAnalytics podem gerar contagens mais precisas se comparadas com as contagens do filtro em questão utilizado no VP, sendo que essa discrepância se deve, provavelmente, ao fato de ambos os autores apresentarem acentuação em seus nomes, onde os mesmos são removidos no BiblioAnalytics e não são removidos no filtro da outra ferramenta.

**Quadro 22** – Comparativo da contagem dos campos “Organizations” da WoS e “Authors” da PL entre as ferramentas BiblioAnalytics e o VantagePoint após todos os processamentos na geração de uma lista

<b>Campo “Organizations” da WoS</b>			
<b>BiblioAnalytics</b>		<b>VantagePoint</b>	
univ granada	125	univ granada	125
univ roma tor vergata	112	univ roma tor vergata	112
univ valencia	109	univ valencia	109
univ sao paulo	87	univ sao paulo	87
max planck gesell	81	max planck gesell	81
wuhan univ	81	wuhan univ	81
chinese acad sci	81	chinese acad sci	81
univ chile	75	univ chile	75
univ politecn valencia	68	univ politecn valencia	68
natl res council italy	67	natl res council italy	67
<b>Campo “Authors” da PL</b>			
<b>BiblioAnalytics</b>		<b>VantagePoint</b>	
martinez, inicio carrilho	82	MARTINEZ, Vinício Carrilho	80
hayashi, maria cristina piumbato		Hayashi, Maria Cristina Piumbato	
innocentini	34	Innocentini	34
pedro, wilson jose alves	32	PEDRO, Wilson José Alves	31
gracioso, l s	23	GRACIOSO, L. S.	23
gracioso, luciana de souza	22	GRACIOSO, LUCIANA DE SOUZA	22
donadone, j c	18	AMARAL, R. M.	18
amaral, r m	18	DONADONE, J. C.	18
zanin, m	17	ZANIN, M.	17
furnival, a c m	15	FURNIVAL, A. C. M.	15
hoffmann, w a m	14	HOFFMANN, W. A. M.	14

Fonte: Elaborado pelo autor.

Finalmente, é importante ainda ressaltar que todos os processamentos efetuados para esses dois campos exemplificados (e também para os demais trabalhados pelo BiblioAnalytics) são feitos internamente na ferramenta, sem a necessidade de o pesquisador ter conhecimento algum de expressões regulares e/ou criação de filtros, o que facilita com que o mesmo consiga utilizar a ferramenta e obtenha análises precisas sem ter conhecimento técnico algum, contribuindo com a democratização da ferramenta.

#### 4.2.3 Contagens com a aplicação de um tesauro

De modo a verificar a utilização de tesouros no BiblioAnalytics, foi criado um tesauro para a organização *Chinese Academy of Sciences* e aplicado o mesmo no conjunto de publicações da WoS, onde todos os elementos que representam a instituição em questão sejam substituídos por “*chinese academy of science*” (Figura 34).

**Figura 34** – Tesauro criado para padronização da instituição *Chinese Academy of Sciences* no BiblioAnalytics

label	replace by
chinese acad sci	chinese academy of science
univ chinese acad sci	chinese academy of science
univ chinese acad sci cas	chinese academy of science
univ chinese acad	chinese academy of science

Fonte: Elaborada pelo autor

Para criação desse tesauro e obtenção dos termos a serem substituídos foi necessário primeiramente a geração de uma lista de “Organizations” no BiblioAnalytics onde foi possível observar as seguintes representações para a instituição em questão e suas contagens: *chinese acad sci* (81), *univ chinese acad sci* (32), *univ chinese acad sci cas* (1) e *univ chinese acad* (1). Ao se aplicar o tesauro nesse conjunto de publicações é esperado que “*chinese academy of science*” passe a ter 115 ocorrências (a soma), o que ocorreu conforme pode ser visto no Quadro 23.

**Quadro 23** – Comparativo de uma lista de “Organizations” no BiblioAnalytics antes e após a aplicação de um tesauro para a *Chinese Academy of Sciences*

Antes do tesauro		Após o tesauro	
univ granada	125	univ granada	125
univ roma tor vergata	112	<b>chinese academy of science</b>	<b>115</b>
univ valencia	109	univ roma tor vergata	112
univ sao paulo	87	univ valencia	109
max planck gesell	81	univ sao paulo	87
wuhan univ	81	wuhan univ	81
<b>chinese acad sci</b>	<b>81</b>	max planck gesell	81
univ chile	75	univ chile	75
univ politecn valencia	68	univ politecn valencia	68
natl res council italy	67	natl res council italy	67

Fonte: Elaborado pelo autor.

Concluindo, do ponto de vista técnico, é possível observar que o tesauro se comporta adequadamente, realizando a troca dos elementos a serem substituídos pelo desejado. Do ponto de vista científico, nota-se também a importância de se ter um mecanismo como os tesouros em uma ferramenta, dado que a instituição em questão, antes da aplicação do tesauro, era somente a sétima colocada dentre as dez mais relevantes e, após a aplicação do tesauro, a mesma foi para a segunda colocação, o que demonstra que a mesma era muito mais relevante do que aparentava ser. Essa mudança de relevância da instituição poderia também ocorrer com outras, caso o tesauro tivesse sido aplicado nas mesmas.

#### 4.2.4 Contagens na geração de uma matriz

A verificação das contagens de uma matriz se deu com a comparação de uma matriz quadrada de “Document Type” gerada tanto pelo BiblioAnalytics quanto pelo VP para o conjunto de publicações da WoS. Nela foram verificados os valores em sua diagonal principal, que representam a quantidade total de ocorrência no conjunto de publicações que se está analisando (o mesmo valor que é mostrado na geração de uma lista), e também a contagem de suas cocorrências (contagens que não estão na diagonal principal), que representam a quantidade de publicações em colaboração com outras instituições.

Conforme pode ser visto na Figura 35, ambas as contagens nas matrizes geradas pelas duas ferramentas são idênticas, podendo assim concluir que o

BiblioAnalytics efetua as mesmas corretamente. Ponto que vale ser ressaltado, é que o BiblioAnalytics já efetua um pré-processamento na construção matriz, removendo elementos sem coocorrência (que acabam sendo desnecessários nesse tipo de análise), conforme pode ser visto o tipo de documento “Bibliography” que está presente na matriz do VP e não na do BiblioAnalytics.

**Figura 35 – Comparativo das matrizes de “Document Type” ou “Tipo de Documento” geradas por meio das ferramentas BiblioAnalytics e VP**

(i) Gerada pelo VantagePoint

Reset	Tipo de documento (DT)		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
		# Records	5220	839	825	205	182	76	57	13	7	6	5	5	3	2	1	1	1	
		Show Values >= 1																		
		Cooccurrence # of Records																		
		# Records	Article	Review	Proceedings Paper	Editorial Material	Early Access	Letter	Book Review	Meeting Abstract	Data Paper	Note	Correction	Biographical-Item	Reprint	News Item	Correction, Addition	Bibliography	Retracted Publication	
1	5220	Article	5220		154		142				7									1
2	839	Review		839			39													
3	825	Proceedings Paper	154		825															
4	205	Editorial Material				205														
5	182	Early Access	142	39			182					1								
6	76	Letter						76												
7	57	Book Review							57											
8	13	Meeting Abstract								13										
9	7	Data Paper	7								7									
10	6	Note										6								
11	5	Correction					1						5							
12	5	Biographical-Item												5						
13	3	Reprint													3					
14	2	News Item														2				
15	1	Correction, Addition															1			
16	1	Bibliography																1		
17	1	Retracted Publication	1																	1

(ii) Gerada pelo BiblioWeb

	A	B	C	D	E	F	G	H
1		Article	Correction	Data Paper	Early Access	Proceedings Paper	Retracted Publication	Review
2	Article	5220		7	142	154		1
3	Correction		5		1			
4	Data Paper	7		7				
5	Early Access	142	1		182			39
6	Proceedings Paper	154				825		
7	Retracted Publication	1						1
8	Review				39			839
9								

Fonte: Elaborada pelo autor.



### 4.3 Validação da ferramenta

Apesar de ter sido demonstrada a importância e precisão das contagens oferecidas pelo BiblioAnalytics durante a geração de uma lista, de uma matriz ou na aplicação de um tesouro, é essencial validar que ferramenta desenvolvida neste estudo será capaz de ser utilizada em um cenário real. De pouco adianta que a ferramenta faça suas contagens com maestria se as mesmas não puderem ser realizadas por meio de sua efetiva utilização por uma pessoa.

Dado o exposto, e pelo fato de o BiblioAnalytics se tratar de uma ferramenta disponibilizada na *web* e poder ser acessada tanto por computadores de mesa quanto dispositivos móveis, é necessário validar que a mesma funcione adequadamente em ambos os casos, em um contexto geral. Além do contexto geral, durante a utilização da ferramenta existem casos onde uma grande quantidade de dados necessitam ser processados, sendo também necessário sua validação neste contexto específico.

#### 4.3.1 Contexto geral

De modo a validar a ferramenta no contexto geral, foi acessada a mesma e suas páginas tanto por um *desktop* quanto por um aparelho celular por meio do navegador Google Chrome, dado que o mesmo é o navegador mais utilizado<sup>58</sup>, e testado se seu funcionamento ocorria adequadamente. Empiricamente foi constatado por meio da navegação pela ferramenta, que todas suas páginas e funcionalidades eram apresentadas e funcionavam adequadamente, em sua plenitude, não apresentando erros ou fatores que pudessem gerar uma indisponibilidade da mesma

#### 4.3.2 Contagens

O contexto específico que pode demandar mais processamento por parte da ferramenta e, conseqüentemente, gerar uma indisponibilidade da mesma é a geração de listas e matrizes e suas contagens. É importante ressaltar que, apesar de a importação ser um processo que também demanda tempo considerável, pelo fato de

---

<sup>58</sup> <<https://gs.statcounter.com/browser-market-share>>

a mesma ser um procedimento onde a quantidade de dados pode ser limitada sem prejuízos às análises (o que já acontece na ferramenta), optou-se por analisar somente as contagens, dado que elas não podem ser limitadas sem prejuízos ao pesquisador. Exemplificando, durante a importação, caso necessário, é possível limitar a quantidade de arquivos enviados e processados durante as importações conforme necessário, tendo como “problema” somente o fato de que o pesquisador que utiliza a ferramenta poderá ter mais trabalho tendo que efetuar o processo em questão mais vezes para todo seu conjunto de publicações. Já para as contagens que ocorrem durante a geração de listas e matrizes, a mesma não pode ocorrer uma limitação da quantidade de elementos a serem analisados sem que análise seja prejudicada; seria incoerente realizar a geração de uma lista de autores somente para os 1000 primeiros autores de um conjunto que pode mais de dezenas de milhares de autores.

O principal fator gerador de muitos ou poucos processamentos em uma análise efetuada na ferramenta é a quantidade de elementos distintos, como por exemplo uma palavra-chave, presentes nas publicações e não a quantidade de publicações. Isso se deve ao fato de que para cada novo elemento, por exemplo um autor, em um conjunto de publicações, uma nova contagem deverá ser feita na construção de uma lista e diversas novas contagens necessitam ser feitas para obtenção das coocorrências na geração de uma matriz.

Para validação do contexto em questão foi portanto coletado e importado no BiblioAnalytics (disponibilizado nos servidores gratuitos já explanados) um conjunto de publicações que pudessem representar um uso comum por um pesquisador, nesse caso 26376 publicações que tinham como uma de suas instituições a UFSCar, e verificado como a ferramenta se comportava conforme a quantidade de elementos diferentes aumentava na geração de listas e matrizes. Em requisições HTTP (por exemplo a realizada quando o pesquisador faz uma requisição na ferramenta gerando uma lista ou matriz) tem se como média um tempo máximo que pode variar de 60 e 120 segundos<sup>59</sup> a depender da aplicação, apesar de não ser incomum existirem requisições que possam demorar bem além do tempo recomendado. Constatado tal

---

<sup>59</sup> <<https://aws.amazon.com/pt/blogs/aws/elb-idle-timeout-control/>>  
<<https://docs.microsoft.com/en-us/dotnet/api/system.net.http.httpclient.timeout?view=net-6.0>>  
<<https://www.ibm.com/docs/en/sva/7.0.0?topic=cpc-timeout-settings-http-https-communication>>

fato, seria interessante a ferramenta BiblioAnalytics conseguisse respeitar a questão e portanto buscou-se analisar como a mesma se comporta nessa questão.

Conforme exposto no Quadro 24, iniciando pelas listas, nota-se que o processo de geração de uma lista é realizado rapidamente, notando que mesmo para quase 50000 elementos diferentes (neste caso 49085 autores contidos nas 26376 publicações), não passou de dez segundos. Com isso pode-se concluir que, em condições normais, dificilmente o BiblioAnalytics teria problemas de efetuar a geração de listas e respeitar a média de 60 a 120 segundos.

**Quadro 24** – Comparativo de tempo entre a geração de uma lista e de uma matriz

<b>Quantidade de elementos distintos</b>	<b>Tempo para geração da lista</b>	<b>Tempo para geração da matriz</b>
17	1 segundo	1 segundo 430 milissegundos
233	1 segundo	3 segundos 920 milissegundos
5305	1 segundo 300 milissegundos	100 segundos 860 milissegundos
7127	1 segundo 400 milissegundos	474 segundos 170 milissegundos*
15570	2 segundos 200 milissegundos	Não foi possível validar
49085	7 segundos 800 milissegundos	Não foi possível validar

Fonte: Elaborado pelo autor

Já para a geração de matrizes, é necessário efetuar considerações e elencar alguns problemas encontrados. Primeiramente, conforme pode ser observado no Quadro 24, quando cresce o número de diferentes elementos na criação de uma matriz, seu tempo de geração da matriz também sobe consideravelmente, sendo que com 7127 já nota-se que o tempo necessário para a criação da mesma excede, e muito, o recomendado. Durante esse processo foi possível também presenciar problemas relevantes durante a geração de matrizes do BiblioAnalytics que devem ser relatados: (i) a construção das matrizes ocorre inadequadamente quando os elementos que compõem a matriz apresentam outros caracteres além de letras em sua composição, por exemplo uma palavra-chave “Oforce”, quebrando a formatação

da matriz; (ii) foi descoberta a existência de uma limitação do número de colunas por parte das planilhas Excel que, conforme a documentação oficial da Microsoft<sup>60</sup>, não aceita mais que 16.384 colunas, o que é um problema caso o conjunto de publicações a ser analisada e seu metadado, em um caso hipotético, apresente um número maior que esse de elementos na geração de uma matriz; (iii) ainda no quadro pode ser visto que não foi possível verificar a geração de matrizes para grandes conjuntos de elementos distintos, onde até mesmo no caso de 7127 elementos houveram problemas, isso se deve ao fato que, durante a verificação dos mesmos, notou-se que a ferramenta apresentava problemas no ambiente da AWS, onde foi acusado um erro de falta de memória.

Ainda sobre os problemas elencados durante a validação das matrizes, é necessário efetuar algumas ponderações em relação aos mesmos. Sobre o ponto (i) nota-se que não se pode garantir que a construção das matrizes sejam realizadas corretamente caso os elementos a serem utilizados na construção das matrizes apresentem algum caractere que não seja letra em seu início. Para o ponto (ii) nota-se que existe uma limitação intransponível na geração das matrizes geradas pelo BiblioAnalytics, dado que o arquivo utilizado para geração das mesmas apresenta um limite de colunas. Finalmente, sobre o ponto (iii), nota-se que o BiblioAnalytics, caso for ser disponibilizado em ambientes gratuitos como os utilizados nesta pesquisa, sofrerá de uma limitação na geração de matrizes por causa da memória da máquina utilizada, dado que a mesma só disponibiliza 1gb de memória.

Por fim, ao se falar sobre as matrizes, dados os problemas e limitações elencados, fica clara a necessidade de se realizar novos estudos sobre o tema e repensar as abordagens e metodologias empregadas na construção das mesmas nesta ferramenta, para tentar ao menos amenizar alguns dos problemas em questão. As mesmas, no presente momento, são viáveis somente para campos que apresentam poucas variações e, devido às limitações e problemas elencados, optou-se por disponibilizar somente a geração de matrizes de “Document Type” e “WoS Categories” para a WoS e de “Capes Research Areas” para a PL (Quadro 25), já que as mesmas apresentam um conjunto de elementos pequenos (as vezes finitos) e sem a presença de caracteres que não sejam letras.

---

<sup>60</sup> <https://support.microsoft.com/en-us/office/excel-specifications-and-limits-1672b34d-7043-467e-8e27-269d656771c3>

**Quadro 25** – Listas e matrizes disponíveis para geração durante a utilização do BiblioAnalytics.

	<b>Web of Science</b>	<b>Plataforma Lattes</b>
<b>Lista</b>	<ul style="list-style-type: none"> <li>• Authors</li> <li>• Author Keywords</li> <li>• Document Type</li> <li>• Funding Orgs</li> <li>• Keywords Plus</li> <li>• Language</li> <li>• ORCIDs</li> <li>• Organizations</li> <li>• Publication Type</li> <li>• Publication Year</li> <li>• WoS Categories</li> </ul>	<ul style="list-style-type: none"> <li>• Authors</li> <li>• Authors Lattes</li> <li>• Capes Reserach Areas</li> <li>• Language</li> <li>• Publication Type</li> <li>• Publication Year</li> <li>• Research Topics</li> </ul>
<b>Matriz</b>	<ul style="list-style-type: none"> <li>• Document Type</li> <li>• WoS Categories</li> </ul>	<ul style="list-style-type: none"> <li>• Capes Research Areas</li> </ul>

Fonte: Elaborado pelo autor.

É válido reforçar que novos campos podem ser disponibilizados em estudos posteriores, possibilitando a geração de outros tipos de listas e matrizes conforme as limitações elencadas puderem ser sanadas e novos estudos forem conduzidos abordando a questão.

## 5 CONSIDERAÇÕES FINAIS

Conforme apresentado, nota-se a relevância da bibliometria e da existência de ferramentas que auxiliem nesse tipo de estudo. Para tal existem inúmeras ferramentas tanto em construção quanto já prontas que buscam auxiliar nos estudos métricos. Todavia, foi possível observar a falta de preocupação com a democratização das mesmas, existindo questões não exploradas que mereciam ser abordadas para que fosse possível contribuir com a popularização de análises bibliométricas a um conjunto ainda maior de pessoas, como para os pesquisadores brasileiros. Questões como necessidade de infraestrutura, condição financeira e intelectual, conhecimento da língua inglesa e entre outras podem ser amenizadas com a disponibilização e utilização da ferramenta BiblioAnalytics criada neste estudo.

Portanto, de modo a atender os objetivos da pesquisa, foi feito: (i) uma revisão da literatura a cerca de temas relevantes para a construção de uma ferramenta; (ii) foi proposta uma ferramenta por meio da coleta de requisitos e criação de modelos e casos de uso especificando necessidades a serem abordadas para contribuir com a popularização de análises bibliométricas; (iii) foi implementada a ferramenta, de modo a atender à temática e requisitos abordados no estudo; (iv) foi validado o correto funcionamento da mesma e verificadas suas características e particularidades.

Conforme fora observado, a ferramenta apresenta limitações como o processo de criação de matrizes, do modo em que o tema foi abordado neste estudo, o que pode ser fator decisivo pela não utilização do BiblioAnalytics para alguns tipos de estudos, pelo menos até o presente momento, limitando a quantidade de pessoas que possam usar a ferramenta. É importante discorrer que a ferramenta elaborada neste estudo não tem como finalidade ser um *software* comercial e, devido a isso, tanto as limitações elencadas quanto problemas não previstos podem acontecer, e inclusive são esperados durante a utilização da mesma, dado que se trata de uma prova de conceito.

Dadas as limitações apresentadas e a gama de elementos disponíveis em uma ferramenta bibliométrica, diversos estudos futuros podem ser conduzidos para analisar os métodos e características da ferramenta (possivelmente solucionando algumas das limitações expostas) e, principalmente, propor novas possibilidades de expansão do BiblioAnalytics para outras questões como sua utilização para efetuar

estudos altimétricos, a disponibilização de diversos tipos de visualizações com base nas listas e matrizes geradas pela mesma, ampliação das fontes de informação e formatos de metadados suportados e entre muitos outros. Além de estudos reflexivos sobre a expansão e melhoria da ferramenta, estudos futuros podem também ser realizados utilizando a ferramenta para auxiliar em análises bibliométricas e também para analisar o comportamento da mesma perante outras.

Concluindo, a ferramenta em seu estado atual é viável, e dados os fatores elencados sobre sua contribuição com a democratização de análises bibliométricas, principalmente pelo fato de poder ser acessada por um dispositivo de celular, pode ser utilizada até mesmo em salas de aulas por discentes que estejam aprendendo sobre os estudos métricos ou até mesmo por pesquisadores durante suas pesquisas, contribuindo cada vez mais com a democratização de análises bibliométricas e com a área da Ciência da Informação e demais áreas que possam utilizá-la por meio da viabilização da elaboração e publicação de estudos que, sem a mesma, poderiam não ser possíveis.

## REFERÊNCIAS

ALVARADO, R. U. A bibliometria no Brasil. **Ciência da informação**, v. 13, n. 2, 1984.

ALVARADO, R. U. A Lei de Lotka na bibliometria brasileira. **Ciência da informação**, v. 31, p. 14-20, 2002.

ALVARADO, R. U. A produtividade dos autores sobre a Lei de Lotka. **Ciência da Informação**, v. 37, n. 2, p. 87-102, 2008.

ALVARADO, R.U. A bibliometria no Brasil. **Ciência da Informação**, Brasília, v. 13, n. 2, p. 91-105, 1984. Disponível em: <<http://revista.ibict.br/ciinf/index.php/ciinf/article/view/1444>>. Acesso em 15 fev. 2021.

ALENCAR, M. S. M.; PORTER, A. L.; ANTUNES, A. M. S. Nanopatenting patterns in relation to product life cycle. **Technological Forecasting and Social Change**, v. 74, n. 9, p. 1661-1680, 2007.

ALMEIDA, L. et al. Democratization of access and success in higher education: a reflection from the realities of Portugal and Brazil. **Avaliação: Revista da Avaliação da Educação Superior (Campinas)**, v. 17, p. 899-920, 2012.

ALVES, R. C. V. **Metadados como elementos do processo de catalogação**. 2010. 132 f. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2010. Disponível em: <[http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciaDaInformacao/Dissertacoes/alves\\_rcv\\_do\\_mar.pdf](http://www.marilia.unesp.br/Home/Pos-Graduacao/CienciaDaInformacao/Dissertacoes/alves_rcv_do_mar.pdf)>. Acesso em: 25 fev. 2021

ANGELOZZI, S. M.; MARTÍN, S. G. **Metadatos para la descripción de recursos electrónicos en línea: análisis y comparación**. Buenos Aires: Alfagrama, 2010.

ANSI/NISO **Z39.19-2003**: guidelines for the construction, format, and management of monolingual thesauri. Bethesda: NISO, 2003.

ARAÚJO, R. F.; ALVARENGA, L. A bibliometria na pesquisa científica da pós-graduação brasileira de 1987 a 2007. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 16, n. 31, p. 51-70, 2011.

ARAUJO, N. C.; FACHIN, J. **Evolução das fontes de informação**. 2015.

ARIA, M.; CUCCURULLO, C.; bibliometrix: An R-tool for comprehensive science mapping analysis. **Journal of informetrics**, v. 11, n. 4, p. 959-975, 2017.

ARIA, M.; CUCCURULLO, C.; **Package 'bibliometrix'**. 2021.



AULER, D.; DELIZOICOV, D. Alfabetização científico-tecnológica para quê?. **Ensaio Pesquisa em Educação em Ciências (Belo Horizonte)**, v. 3, n. 2, p. 122-134, 2001.

BASSOLI, M. **Avaliação do Currículo Lattes como fonte de informação para construção de indicadores**: o caso da UFSCar. 2017. 84 p. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) – Universidade Federal de São Carlos, São Carlos, 2017.

BASTIAN, M; HEYMANN, S; JACOMY, M. Gephi: an open source software for exploring and manipulating networks. In: **Third international AAI conference on weblogs and social media**. 2009.

BLATTMANN, U. Fontes de Informação: Primárias, Secundárias e Terciárias. 2015. Disponível em: <<http://bib-ci.wikidot.com/fontes-primarias>> . Acesso em: 23 jul. 2015

BORNMANN, L.; DANIEL, H. D. What do we know about the h index?. **Journal of the American Society for Information Science and technology**, v. 58, n. 9, p. 1381-1385, 2007.

BOURDIEU, P. O campo científico. In: ORTIZ, R. (Org). Pierre Bourdieu: sociologia. São Paulo: Ática, p. 122-155, 1983.

CALLON, et. al.. The use of patent titles for identifying the topics of invention and forecasting trends. **Scientometrics**, v. 26, n. 2, p. 231-242, 1993.

CAMPELLO, B.; CALDEIRA, P. T. **Introdução às fontes de informação**. Autêntica, 2018.

CAMPRA, M.; PAOLO, E.; BRESCIA, VC. State of the art of COVID-19 and business, management, and accounting sector. A bibliometrix analysis. 2021.

CAPES. **Acesso livre**. 2021. Disponível em: < [https://www-periodicos-capes-gov-br.ez1.periodicos.capes.gov.br/index.php?option=com\\_pcollection&mn=70&smn=79&cid=81&](https://www-periodicos-capes-gov-br.ez1.periodicos.capes.gov.br/index.php?option=com_pcollection&mn=70&smn=79&cid=81&)>. Acesso em: 12 mar. 2021.

CORRÊA, T. S. et al. O fim do scriptLattes? Uma análise de suas funcionalidades, alternativas para o presente e perspectivas para o futuro. **Revista do EDICC-ISSN 2317-3815**, v. 3, n. 3, 2017.

CLARIVATE. **Web of Science**: Confident research begins here. [S. l.], 2022. Disponível em: <https://clarivate.com/webofsciencegroup/solutions/web-of-science/>. Acesso em: 14 mar. 2022.

CLARIVATE. **Web of Science Core Collection Field Tags**. [S. l.], 16 jan. 2020. Disponível em:

<[https://images.webofknowledge.com/images/help/WOS/hs\\_wos\\_fieldtags.html](https://images.webofknowledge.com/images/help/WOS/hs_wos_fieldtags.html)>. Acesso em: 19 maio 2021.

CLARIVATE. **EndNote: Directions for implementing EndNote Direct Export**. Clarivate. 2021. Disponível em: <[https://support.clarivate.com/Endnote/s/article/EndNote-Directions-for-implementing-EndNote-Direct-Export?language=en\\_US](https://support.clarivate.com/Endnote/s/article/EndNote-Directions-for-implementing-EndNote-Direct-Export?language=en_US)>. Acesso em: 11 abr. 2021.

CNPQ. **História do surgimento da plataforma Lattes**. [S. l.], 2021a. Disponível em: <<http://lattes.cnpq.br/>>. Acesso em: 16 fev. 2021.

CNPQ. **Extração de Dados**. [S. l.], 2021b. Disponível em: <<http://memoria.cnpq.br/web/portal-lattes/extracoes-de-dados>>. Acesso em: 25 fev. 2021.

COBO, M. J. *et al.* Science mapping software tools: Review, analysis, and cooperative study among tools. **Journal of the American Society for information Science and Technology**, v. 62, n. 7, p. 1382-1402, 2011.

CUNHA, M. B.; CAVALCANTI, C. R. O. **Dicionário de biblioteconomia e arquivologia**. Briquet de Lemos, 2008.

DAMACENO, R. J. P.; ROSSI, L.; MENA-CHALCO, J. P. (2017). Identificação do grafo de genealogia acadêmica de pesquisadores: Uma abordagem baseada na Plataforma Lattes. In **Proceedings of the 32nd Brazilian Symposium on Databases** (pp. 76-87).

DE BARROS CAMPOS, L. F. Metadados digitais: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 12, n. 23, p. 16-46, 2007.

DEMOCRATRIZAÇÃO. In: DICIO, Dicionário Online de Português. Porto: 7Graus, 2022. Disponível em: <<https://www.dicio.com.br/democratizacao/>>. Acesso em: 22/03/2022.

DORTA-GONZÁLEZ, P.; DORTA-GONZÁLEZ, M. I. Indicador bibliométrico basado en el índice h. **Revista española de Documentación Científica**. Madrid, v. 33, n. 2, p. 225-245, 3 maio 2010. Disponível em: <<http://redc.revistas.csic.es/index.php/redc/article/view/553/627>>. Acesso em 21 de Out. 2019.

ECMA. Standard ECMA-404 The JSON Data Interchange Syntax. 2017.

EDHLUND, B. M. **Manuscript Writing Using EndNote and Word: A User's Guide that Makes Your Scientific Writing Easier**. 2. ed. [S.l.]: lulu.com, 2007. p. 191.

ENGE, E. **Mobile vs. Desktop Usage in 2020**. [S. l.], 23 mar. 2021. Disponível em: <<https://www.perficient.com/insights/research-hub/mobile-vs-desktop-usage>>. Acesso em: 20 jul. 2021.

FARIA, L. I. L. **Bibliometria**. [S. l.], 2015. (Apostila de Apoio).

FARIA, L. I. L. **Prospecção tecnológica em materiais: aumento da eficiência do tratamento bibliométrico**. Aplicação na análise de tratamentos de superfície resistentes ao desgaste. 2001.

FILE TYPE ADVISOR. **What is the .CIW file type?**. [S. l.], 2021. Disponível em: <<https://www.filetypeadvisor.com/extension/ciw>>. Acesso em: 11 maio 2021.

FLANAGAN, D. **JavaScript: The Definitive Guide**. 7. ed. [S. l.]: O'Reilly Media, Inc., 2020. 708 p. ISBN 9781491952023.

FLEJOLE, R. P. **Database Theory and Application**. 1. ed. [S. l.]: Arcler Press, 2017. v. 352. ISBN 978-1773612218.

FRIESEN, J. Introducing json. In: **Java XML and JSON**. Apress, Berkeley, CA, 2019. p. 187-203.

GALLIGAN, F.; DYAS-CORREIA, S. Altmetrics: Rethinking the way we measure. **Serials review**, v. 39, n. 1, p. 56-61, 2013.

GHENO, T. C. et al. Estrutura de tesauro em Ciência da Informação: análise dos tesouros das bases de dados LISA e LISTA. 2013.

GILLILAND, A. J. **Setting the Stage**. In: BACA, Murtha (Org.). **Introd. Metadata**. 3. ed. Los Angeles: Getty Research Institute, 2016. Disponível em: <<http://www.getty.edu/publications/intrometadata/>>. Acesso em: 28 abr. 2021.

GLÄNZEL, W. **Bibliometrics as a research field: a course on theory and application of bibliometric indicators**. Bélgica: [s.n.], 2003. Disponível em <[https://onderwijsaanbod.kuleuven.be/syllabi/e/D0N24AE.htm#activetab=doelstelling\\_en\\_idm4718704](https://onderwijsaanbod.kuleuven.be/syllabi/e/D0N24AE.htm#activetab=doelstelling_en_idm4718704)>.

GUEDES, V. L. S. A bibliometria e a gestão da informação e do conhecimento científico e tecnológico: uma revisão da literatura. **Ponto de Acesso**, v. 6, n. 2, p. 74-109, 2012.

GUEDES, V. L. S.; BORSCHIVER, S. Bibliometria: uma ferramenta estatística para a gestão da informação e do conhecimento, em sistemas de informação, de comunicação e de avaliação científica e tecnológica. In: CIFORM - ENCONTRO NACIONAL DE CIÊNCIA DA INFORMAÇÃO, 6., 2005, Salvador. **Anais...** Salvador:

ICI/UFBA, 2005. Disponível em: <[http://www.cinform.ufba.br/vi\\_anais/docs/VaniaLSGuedes.pdf](http://www.cinform.ufba.br/vi_anais/docs/VaniaLSGuedes.pdf)>. Acesso em: 11 fev. 2021.

GRÁCIO, Maria Cláudia Cabrini. Colaboração científica: indicadores relacionais de coautoria. **Brazilian Journal of Information Science: research trends**, v. 12, n. 2, 2018.

HUANG, W.; WANG, P.; WU, Q. A correlation comparison between Altmetric Attention Scores and citations for six PLOS journals. **PloS one**, v. 13, n. 4, p. e0194962, 2018.

INEFUKU, H. W. Globalization, open access, and the democratization of knowledge. **Educause Review**, v. 52, n. 4, p. 62, 2017

JIN, B.; SAHNI, S.; SHEVAT, A. **Designing Web APIs: Building APIs That Developers Love**. 1. ed. [S. l.]: O'Reilly Media, 2018. 237 p.

JOB, I. **Bibliometria aplicada aos estudos do campo da Educação Física: confiabilidade, qualidade e relevância das publicações**. 2018. Disponível em: <<https://periodicos.ufsc.br/index.php/motrivivencia/article/view/2175-8042.2018v30n54p18>>. Acesso em: 25 abr. 2021.

KROEFF, M. S. et al. Análise de citações dos artigos publicados em periódicos da área da Ciência da Informação que versam sobre gestão da informação. **Revista Brasileira de Biblioteconomia e Documentação**, v. 11, n. 1, p. 41-65, 2015.

KUMAR, K. A scientometric study of digital literacy in online library information science and technology abstracts (LISTA). **Library Philosophy and Practice**, v. 1, p. 1-13, 2014.

LAKATOS, E. M.; MARCONI, M. A. Fundamentos da metodologia científica. In: **Fundamentos da metodologia científica**. 2010. p. 320-320.

LEISE, F. Controlled vocabularies: an introduction. **Indexer**, v. 26, n. 3, 2008.

LIMA, R. A; VELHO, L. M. L. S; FARIA, L. I. L. Bibliometria e "avaliação" da atividade científica: um estudo sobre o índice h. **Perspectivas em Ciência da Informação**, v. 17, p. 03-17, 2012.

MA, J.; PORTER, A. L. Analyzing patent topical information to identify technology pathways and potential opportunities. **Scientometrics**, v. 102, n. 1, p. 811-827, 2015.

MACIAS-CHAPULA, C. A. O papel da informetria e da cienciometria e sua perspectiva nacional e internacional. **Ciência da Informação**, v. 27, n. 2, p. 134-140, 1998.

MACIEL, R. S. et al. A Plataforma Lattes como recurso estratégico para a gestão dos programas de Pós-Graduação. In: **XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (XVIII ENANCIB)**. 2017.

MACIEL, R. S. et al. Efeito Qualis e a produção científica dos Programas de Pós-Graduação da Universidade Federal de São Carlos. **Em Questão**, v. 24, p. 88-110, 2018.

MACULAN, B.C. M. S. Taxonomia facetada e navegacional - um mecanismo de recuperação. Universidade federal de Minas Gerais, 2011. Disponível em: <[http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/ECID-8LAN5K/maculan\\_taxonomia\\_facetada\\_navegacional.pdf?sequence=1](http://www.bibliotecadigital.ufmg.br/dspace/bitstream/handle/1843/ECID-8LAN5K/maculan_taxonomia_facetada_navegacional.pdf?sequence=1)>. Acesso em: 7 jul. 2021.

MATIAS, M. S. O. **Base referencial para povoamento de repositórios institucionais: coleta automatizada de metadados da Plataforma Lattes**. 2015. 94f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de São Carlos, São Carlos, 2015.

MENA-CHALCO, J. P.; CESAR-JR, R.M. scriptLattes: An open-source knowledge extraction system from the Lattes platform. **Journal of the Brazilian Computer Society**, vol. 15, n. 4, p. 31-39, 2009.

MENA-CHALCO, J. P.; CESAR-JR, R.M. Prospecção de dados acadêmicos de currículos Lattes através de scriptLattes. **Bibliometria e Cientometria: reflexões teóricas e interfaces**. São Carlos: Pedro & João, p. 109-128, 2013.

MIKOWSKI, M. C.; POWELL, J. C. **Single Page Web Applications**: JavaScript end-to-end. 1. ed. [S. l.]: Manning Publications, 2013. 432 p.

MILANEZ, D. H. **Elaboração de indicadores de ciência e tecnologia para o monitoramento de avanços tecnológicos em nanocelulose**. 2016. 187 f. Tese (Doutorado em Ciência e Engenharia de Materiais) – Universidade Federal de São Carlos, 2016. Disponível em: <<https://repositorio.ufscar.br/bitstream/handle/ufscar/737/6613.pdf>>. Acesso em: 25 fev. 2021.

MORAL-MUÑOZ, J. A. *et al.* Software tools for conducting bibliometric analysis in science: An up-to-date review. 2020.

MOREIRA, A.; ALVARENGA, L.; OLIVEIRA, A. P. O nível do conhecimento e os instrumentos de representação: tesouros e ontologias. **DataGramZero-Revista de Ciência da Informação**, v. 5, n. 6, p. 1-25, 2004.

MOREIRA, P. S.; GUIMARÃES, A. J. R.; TSUNODA, D. F. Qual ferramenta bibliométrica escolher?. **P2P E INOVAÇÃO**, v. 6, p. 140-158, 2020.

MOREIRA, J. R.; MUELLER, S. P. M.; VILAN FILHO, J. L. Produção científica dos membros dos grupos de pesquisa das áreas de Informação no Brasil. **Informação & Informação**, v. 25, n. 1, p. 1-20, 2020.

NATIONS, D. **What Is a Web Application?:** Improve your understanding of web-based application programs. [S. l.], 24 jun. 2021. Disponível em: <<https://www.lifewire.com/what-is-a-web-application-3486637>>. Acesso em: 25 ago. 2021.

OKUBO, Y. **Bibliometric indicators and analysis of research** systems: methods and examples. Paris: OECD, 1997.

PATIL, S. B. Global Library & Information Science Research seen through Prism of Biblioshiny. **Stud. Indian Place Names**, v. 40, p. 158-170, 2020.

PERISSINI, R. C.; FARIA, L. I. L.; A presença de ferramentas analíticas em estudos bibliométricos. In: **XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO (XXI ENANCIB)**. 2021. Disponível em: <<http://hdl.handle.net/20.500.11959/brapci/193219>>. Acesso em: 06 abr. 2022.

PERSOON, O.; DANELL, R.; SCHNEIDER, J. W. How to use Bibexcel for various types of bibliometric analysis. **Celebrating scholarly communication studies: A Festschrift for Olle Persson at his 60th Birthday**, v. 5, p. 9-24, 2009.

PIWOWAR, H. Introduction altmetrics: What, why and where?. **Bulletin of the American Society for Information Science and Technology**, v. 39, n. 4, p. 8-9, 2013.

POMERANTZ, J. **Metadata**. Cambridge, Massachusetts; London, England: The MIT Press, 2015.

PORTER, A. L.; CUNNINGHAM, S. W. **Tech mining: exploiting new technologies for competitive advantage**. John Wiley & Sons, 2004.

PRIEM, J. et al. Altmetrics: A manifesto. 2011.

PRINCETON UNIVERSITY LIBRARY. **Using EndNote at Princeton: Thomson Reuters, incl. Web of Science (ISI)**. 2017. Disponível em: <<https://libguides.princeton.edu/c.php?g=84090&p=544629>>. Acesso em: 31 mar. 2021.

PRITCHARD, Alan. Statistical bibliography or bibliometrics. **Journal of documentation**, v. 25, n. 4, p. 348-349, 1969.

RAO, I. K. **Métodos quantitativos em biblioteconomia e ciência da informação**. Brasília: ABDF, 1986.

RAVELLI, A. P. X. et al. A produção do conhecimento em enfermagem e envelhecimento: estudo bibliométrico. **Texto & Contexto-Enfermagem**, v. 18, n. 3, p. 506-512, 2009.

RODRÍGUEZ-SOLER, R.; URIBE-TORIL, J.; VALENCIANO, J. P. Worldwide trends in the scientific production on rural depopulation, a bibliometric analysis using bibliometrix R-tool. **Land Use Policy**, v. 97, p. 104787, 2020.

SANTOS, B. S. Da sociologia da ciência à política científica. *Revista Crítica de Ciências Sociais*, Coimbra, p. 11-56, jun. 1978.

SANTOS, P. L. V. A. C.; SIMIONATO, A. C.; ARAKAKI, F. A. Definição de metadados para recursos informacionais: apresentação da metodologia beam. **Informação & Informação**, v. 19, n. 1, p. 146-163, 2014. Disponível em: <<http://hdl.handle.net/11449/114736>>. Acesso em: 28 abr. 2021.

SARVO, D. O. et al. M2N – Matrix to Network: software de conversão de matrizes para Gephi e VOSviewer. In: ENCONTRO BRASILEIRO DE BIBLIOMETRIA E CIENTOMETRIA - EBBC, 7., 2020, Salvador. **Anais...** Salvador, UFBA, 2020. Disponível em: <<https://repositorio.ufba.br/ri/handle/ri/32385>>. Acesso em: 27/07/2021.

SCOTT, A. D. **Building Web Apps for Everyone**. 1. ed. [S. l.]: O'Reilly, 2016. 62 p.

SEARCH TECHNOLOGY, INC. **VantagePoint User's Guide**. Northcross: [s. n.], 2015. 200 p. Disponível em: [https://www.thevantagepoint.com/\\_Analyst\\_Guide\\_Online\\_/Users\\_Guide/VP%20Users%20Guide.pdf](https://www.thevantagepoint.com/_Analyst_Guide_Online_/Users_Guide/VP%20Users%20Guide.pdf). Acesso em: 10 mar. 2022.

SHAFRANOVICH, Y. Common format and MIME type for comma-separated values (CSV) files. 2005.

SHKLAR, L.; ROSEN, R. **Web Application Architecture:: Principles, Protocols and Practices**. 2. ed. [S. l.]: Wiley, 2009. 440 p.

SILVA, R. C; VANZ, S. A. S. Impacto de altmetrics sobre a visibilidade de artigos em acesso aberto da enfermagem brasileira: um estudo de caso. **Transinformação**, v. 31, 2019.

SOARES, V. R. M. **Interface interativa para visualização de indicadores de produção científica a partir de dados bibliométricos processados da Plataforma Lattes**. 2019. 97f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de São Carlos, São Carlos, 2019.

SOBRAL, N. V. et al. Redes de colaboração científica na produção de conhecimento

em doenças tropicais negligenciadas no Brasil: estudo a partir da plataforma LATTES do CNPq. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 25, p. 1-22, 2020.

SOMMERVILLE, I. S.: Software Engineering. **Monthly Notices of....** <https://doi.org/10.1111/j>, p. 1365-2362.2005, 2011.

THOMSON REUTERS. **RIS File Format**. ResearcherId. 2012. Disponível em: <[http://www.researcherid.com/resources/html/help\\_upload.htm](http://www.researcherid.com/resources/html/help_upload.htm)>. Acesso em: 22 abr. 2021.

TIEPPO, R. 2019. **lucyLattes script para a extração e compilação de dados do currículo Lattes**. 10.5281/zenodo.2591748.

VAAN RAN, A. F. J. Scientometrics: State-of-the-art. **Scientometrics**, v. 38, n. 1, p. 205-218, 1997.

VAN ECK, N. J.; WALTMAN, L. Software survey: VOSviewer, a computer program for bibliometric mapping. **Scientometrics**, v. 84, n. 2, p. 523-538, 2010.

VAN ECK, N. J.; WALTMAN, L. Text mining and visualization using VOSviewer. **arXiv preprint arXiv:1109.2058**, 2011.

VAN ECK, N.J.; WALTMAN, L. VOSviewer manual. **Leiden: Univeristeit Leiden**, v. 1, n. 1.6.8, p. 1-51, 2018.

XIE, L. et al. Bibliometric and visualized analysis of scientific publications on atlantoaxial spine surgery based on Web of Science and VOSviewer. **World neurosurgery**, v. 137, p. 435-442. e4, 2020.

XU, C.; OUYANG, F.; CHU, H. The academic library meets Web 2.0: applications and implications. **The journal of academic librarianship**, v. 35, n. 4, p. 324-331, 2009.

YU, Y. et al. A bibliometric analysis using VOSviewer of publications on COVID-19. **Annals of translational medicine**, v. 8, n. 13, 2020.

ZHU, D. et al. A process for mining science & technology documents databases illustred for the case of knowledge discovery and data mining. **Ciência da Informação**, Brasília, DF, v. 28, n.1, jan. 1999.

WITTER, G. P. **Produção científica**. Átomo, 1997.