

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

**INTERFACE INTERATIVA PARA VISUALIZAÇÃO DE INDICADORES
DE PRODUÇÃO CIENTÍFICA A PARTIR DE DADOS
BIBLIOMÉTRICOS PROCESSADOS DA PLATAFORMA LATTES**

VINÍCIUS RAFAEL MICALI SOARES

SÃO CARLOS - SP

2019

VINÍCIUS RAFAEL MICALI SOARES

**INTERFACE INTERATIVA PARA VISUALIZAÇÃO DE INDICADORES
DE PRODUÇÃO CIENTÍFICA A PARTIR DE DADOS
BIBLIOMÉTRICOS PROCESSADOS DA PLATAFORMA LATTES**

Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Informação, do Centro de Educação e Ciências Humanas, da Universidade Federal de São Carlos, como parte dos requisitos para a obtenção do título de Mestre em Ciência da Informação.

Orientador: Prof. Dr. Leandro Innocentini Lopes de Faria

SÃO CARLOS - SP

2019



UNIVERSIDADE FEDERAL DE SÃO CARLOS

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

Folha de Aprovação

Assinaturas dos membros da comissão examinadora que avaliou e aprovou a Defesa de Dissertação de Mestrado do candidato Vinicius Rafael Micali Soares, realizada em 20/08/2019:

Prof. Dr. Leandro Innocentini Lopes de Faria
UFSCar

Prof. Dr. Jesús Pascual Mena Chalco
UFABC

Prof. Dr. Douglas Henrique Milanez
UFSCar

Certifico que a defesa realizou-se com a participação à distância do(s) membro(s) Jesús Pascual Mena Chalco e, depois das arguições e deliberações realizadas, o(s) participante(s) à distância está(ao) de acordo com o conteúdo do parecer da banca examinadora redigido neste relatório de defesa.

Prof. Dr. Leandro Innocentini Lopes de Faria

PUBLICAÇÃO RELACIONADA À DISSERTAÇÃO

Trabalho completo apresentado e publicado no 16º Congresso Nacional de Pesquisadores, promovido pelo Centro Universitário Central Paulista (UNICEP):

- SOARES, V. R. M; FARIA, L. I. L.; SIMIONATO, A. C. Importância dos metadados da Plataforma Lattes para a avaliação da ciência. In: CONGRESSO NACIONAL DE PESQUISADORES (CONAPE-UNICEP), 16., São Carlos, nov. 2018, p. 169-176. **Anais...** Disponível em: <<https://www.unicep.edu.br/arquivos/anais-conape-2018-2.pdf>>. Acesso em: 29 mar. 2019.

Dedico este trabalho à minha avó, Maria Amélia Passarelli Micali (*in memoriam*)

AGRADECIMENTOS

Agradeço ao prof. Dr. Leandro Innocentini Lopes de Faria pela dedicada orientação e pelas oportunidades que compartilhou durante o período do mestrado.

Aos professores e colegas do PPGCI pelos debates proporcionados. Ao Dr. Douglas Henrique Milanez e ao professor da Universidade Federal do ABC (UFABC), Dr. Jesús Pascual Mena-Chalco, pelas valiosas contribuições no exame de qualificação e na defesa. Aos professores Dr. Roniberto Morato do Amaral e Dr. Paulo Matias pelas discussões durante a realização do trabalho.

Ao colega Luís Gustavo Maschietto pela parceria firmada.

Aos amigos Mesailde Souza de Oliveira Matias, Tadeu Borges de Abreu Sampaio, José Eduardo dos Reis, Bruno Cesar Previero e Elizeu Rogério Furtado, que compartilharam comigo os seus conhecimentos.

Aos colegas do NIT/Materiais, da SIn/UFSCar e do UNICEP, pelos momentos de troca de ideias, trabalho e também descontração.

Aos gestores da SIn/UFSCar por me liberarem para participar das atividades acadêmicas.

À minha esposa Silmara Corrêa da Silva Micali e aos meus familiares e amigos, por me incentivarem e estarem sempre do meu lado.

Por fim, agradeço à Deus pela minha vida e por ter colocado todas essas pessoas no meu caminho.

Muito obrigado a todos!

*"A certeza da imortalidade ganhou minha mente meu irmão,
eis porque já não temo o futuro e vou sair por aí,
bem tranquilo, sem pressa de chegar ao meu destino [...]"*

Danilo Micalli

RESUMO

Os metadados da Plataforma Lattes vêm sendo utilizados recorrentemente em pesquisas científicas e relatórios técnicos de agências de fomento, como também em instituições de ensino superior e por empresas da iniciativa privada, visando a elaboração de indicadores de Ciência e Tecnologia. Por sua vez, a ferramenta automatizada synclattes tem se mostrado relevante para a extração de dados dos currículos Lattes de pesquisadores. Nesse contexto, o presente trabalho teve por objetivo o desenvolvimento de uma interface interativa para a visualização de indicadores de produção científica a partir do uso do synclattes. Trata-se de uma pesquisa de natureza aplicada e exploratória, tendo como unidade-caso a Universidade Federal de São Carlos – UFSCar. Com base nos conceitos da área de visualização de informação, foi desenvolvido um sistema para a representação visual de dados bibliométricos, contendo gráficos diversos como barras, barras empilhadas, nuvem de etiquetas, grafo e matriz de colaboração. Sendo assim, por meio de uma expressiva fonte de dados como a Plataforma Lattes, foi possível realizar um estudo a respeito da produção científica da UFSCar, considerando os 50 pesquisadores que mais produzem artigos em eventos e em periódicos. Ademais, este trabalho mostrou a viabilidade da implementação de uma interface interativa para a visualização de indicadores bibliométricos de Institutos de Ciência e Tecnologia brasileiros, a partir de dados extraídos pelo synclattes.

Palavras-chave: Interface interativa. Visualização de informação. Indicadores bibliométricos. Plataforma Lattes. Synclattes.

ABSTRACT

Lattes Platform metadata has been used frequently in scientific researches and technical reports from funding agencies, as also in higher education institutions and by private companies, aiming elaboration of Science and Technology indicators. In turn, the synclattes automated tool has been relevant for extracting data from the researchers' Lattes curriculum. In this context, the goal of this project was the development of an interactive interface for the visualization of scientific production indicators by use of the synclattes. It is a research of applied and exploratory nature, having as case unit the Federal University of São Carlos – UFSCar. Based on the concepts of information visualization area, a system was developed for the visual representation of data bibliometric, containing several graphs such as bars, stacked bars, tag cloud, graph and matrix of collaboration. Therefore, through a relevant data source such as the Lattes Platform, it was possible to elaborate a study about UFSCar's scientific production, considering the 50 researchers who more produce articles in events and in journals. Furthermore, this scientific work showed the viability of implementing an interactive interface for the visualization of Brazilian's Institutes of Science and Technology bibliometric indicators, from data extracted by the synclattes.

Keywords: Interactive interface. Information visualization. Bibliometric indicators. Lattes Platform. Synclattes.

LISTA DE FIGURAS

Figura 1 - Produção científica da UFSCar entre os anos de 1994 a 2014.....	30
Figura 2 - Interface de pesquisa do Somos UFSCar.....	31
Figura 3 - Página inicial do Portal Brasileiro de Dados Abertos.....	32
Figura 4 - Representação gráfica de Playfair de 1805.....	33
Figura 5 - Tonalidade e intensidade.....	36
Figura 6 - Visualização de densidade.....	37
Figura 7 - Comprimento, largura e tamanho.....	38
Figura 8 - Orientação.....	38
Figura 9 - Formato.....	38
Figura 10 - Posição 2-D.....	39
Figura 11 - Gráfico de dispersão.....	39
Figura 12 - Exemplo de gráfico com escala.....	40
Figura 13 - Diferença de proporções de altura e largura.....	41
Figura 14 - Três elementos essenciais de bons gráficos de informação.....	42
Figura 15 - Comparação entre dois gráficos de colunas.....	43
Figura 16 - GEOCAPES: distribuição dos docentes das IES em 2017.....	46
Figura 17 - Painel Lattes: total de currículos cadastrados até novembro de 2016.....	47
Figura 18 - VisPublica: aprovação no Ensino Fundamental na Zona Urbana de São Paulo em 2011.....	48
Figura 19 - DataViva: cursos oferecidos pela UFSCar em 2017.....	49
Figura 20 - Mapa do TCU: Fragilidade de Controles X Poder Econômico (UFSCar).....	50
Figura 21 - Ligação entre as pesquisas.....	53
Figura 22 - Layout da interface interativa.....	55
Figura 23 - Código-fonte do programa de montagem de gráficos (Frontend).....	56
Figura 24 - Código-fonte do programa de controle da aplicação (Backend).....	57
Figura 25 - Fluxo de execução dos programas (Frontend e Backend).....	58
Figura 26 - Arquivo JSON de número de publicações por pesquisador.....	60
Figura 27 - Gráfico de barras: quantidade de publicações por pesquisadores.....	61
Figura 28 - Gráfico de barras empilhadas: artigos em periódicos e artigos em eventos.....	62
Figura 29 - Nuvem de etiquetas: pesquisadores com mais publicações.....	63
Figura 30 - Arquivo JSON de coautoria entre pesquisadores (“nós”).....	64
Figura 31 - Arquivo JSON de coautoria entre pesquisadores (“arestas”).....	65
Figura 32 - Grafo de colaboração entre pesquisadores.....	67
Figura 33 - Sub-rede de colaboração entre pesquisadores.....	68
Figura 34 - Código-fonte do programa de exibição do grafo de colaboração.....	69
Figura 35 - Matriz de colaboração entre pesquisadores - ordenada por nome.....	70
Figura 36 - Matriz de colaboração entre pesquisadores - ordenada por departamento.....	71
Figura 37 - Matriz de colaboração entre pesquisadores - ordenada por frequência de publicações.....	72
Figura 38 - Matriz de colaboração entre pesquisadores - ordenada por frequência de coautoria.....	73
Figura 39 - Código-fonte do programa de montagem da matriz de colaboração.....	75

LISTA DE QUADROS

Quadro 1 - Caracterização dos tipos de dados.	34
Quadro 2 - Regras da visão.	36
Quadro 3 - Categorização de Dashboard.	45
Quadro 4 - Customizações e aprimoramentos nas visualizações.	59

LISTA DE TABELAS

Tabela 1 - Presença dos indicadores bibliométricos de produção em RIs.....	26
Tabela 2 - Presença dos indicadores bibliométricos de citação em RIs.....	26

LISTA DE ABREVIATURAS E SIGLAS

AJAX	<i>Asynchronous JavaScript and XML</i>
API	<i>Application Programming Interface</i>
CNPq	Conselho Nacional de Desenvolvimento Científico e Tecnológico
CSS	<i>Cascading Style Sheets</i>
CSV	<i>Comma Separated Values</i>
D3	<i>Data Driven Documents</i>
DOI	<i>Digital Object Identifier</i>
HTML	<i>Hyper Text Markup Language</i>
ICT	Institutos de Ciência e Tecnologia
IES	Instituição de Ensino Superior
JSON	<i>JavaScript Object Notation</i>
PHP	<i>PHP Hypertext Preprocessor</i>
PL	Plataforma Lattes
PPG	Programa de Pós-Graduação
RI	Repositório Institucional
SVG	<i>Scalable Vector Graphics</i>
XML	<i>eXtensible Markup Language</i>

ANEXOS

ANEXO 1 - JSON DE NÚMERO DE PUBLICAÇÕES POR PESQUISADOR.	85
ANEXO 2 - JSON DE COAUTORIA ENTRE PESQUISADORES.	90

SUMÁRIO

1. INTRODUÇÃO	16
1.1 Justificativa	17
1.2 Objetivos gerais e específicos.....	19
1.3 Estrutura da dissertação	20
2. COMUNICAÇÃO CIENTÍFICA E VISUALIZAÇÃO DE INFORMAÇÃO	21
2.1 Avaliação da ciência	21
2.2 Metadados em ambientes digitais	28
2.3 Conceitos fundamentais de visualização de informação	33
2.4 Interfaces para visualização de indicadores	43
3. MÉTODO E DESENVOLVIMENTO	51
3.1 Abordagem e tipologia da pesquisa	51
3.2 Processo de desenvolvimento da pesquisa.....	52
3.3 Tecnologias utilizadas na interface interativa	54
4. RESULTADOS E DISCUSSÕES	60
5. CONSIDERAÇÕES FINAIS	76
REFERÊNCIAS	78

1. INTRODUÇÃO

A partir da explosão informacional do século XX surgiu a preocupação de se estabelecer uma ciência que fosse responsável pela organização do conhecimento e disseminação da informação (SILVA; FREIRE, 2012). Freire (2006) argumenta que por meio da atividade científica, o relacionamento entre as forças produtivas e quem produz conhecimento científico tornou-se cada vez mais objetivo e concreto, desse modo, o fenômeno da informação ganhou relevância nas trocas econômicas e culturais da sociedade capitalista. Ainda segundo o autor, a Ciência da Informação (CI) tem por um de seus objetivos contribuir para a informação se tornar um elemento de inclusão social, visando auxiliar pessoas e nações.

De acordo com Dorta-González e Dorta-González (2010), a pesquisa científica possui influência no desenvolvimento econômico e social, uma vez que a ciência pode ser utilizada pelo homem para conhecer o mundo e encontrar respostas para diversas situações e fenômenos. Entretanto, os recursos estão cada vez mais disputados e o aumento da demanda por financiamentos torna imprescindível a avaliação da atividade científica.

O processo de construção do conhecimento e da ciência é dependente da divulgação e publicação de resultados de pesquisas. Uma das formas encontradas para se realizar a avaliação da ciência é pela análise de indicadores, e é por meio da Bibliometria que os estudos quantitativos relacionados às publicações científicas são realizados. Segundo Bush (1945), o pesquisador é influenciado pelas conclusões e descobertas de outros pesquisadores. Porém, por haver milhares de trabalhos disponíveis, falta tempo para selecionar e aproveitar o resultado dessas pesquisas. Okubo (1997) argumenta que os indicadores de produção científica se tornaram elementos indispensáveis no debate a respeito das relações entre os avanços da ciência e da tecnologia, como também da economia e do progresso social.

Por meio de técnicas de visualização de informação é possível exibir dados bibliométricos graficamente. Com base no trabalho de Freitas et al. (2001), a visualização de informação é a área que estuda as principais formas de representações gráficas para apresentação de informações, podendo combinar aspectos de computação gráfica, interfaces homem-máquina e mineração de dados, auxiliando no entendimento da informação, além de ajudar na percepção do usuário a fim de deduzir novos conhecimentos por meio do que está sendo apresentado.

Reis (2016) aponta que a área de desenvolvimento de software tem contribuído para a visualização de informação ao implementar componentes como plug-ins, extensões e API's (pacote composto por linhas de código), com o objetivo de acrescentar funcionalidades aos sistemas informatizados que trabalham com grandes volumes de dados (VIEIRA; CORREA, 2011).

Por sua vez, a Plataforma Lattes destaca-se como base de dados da produção científica brasileira e possibilita a construção de indicadores, já que reúne currículos da academia (estudantes, professores e pesquisadores), entre outros. Para tanto, torna-se necessária a realização de estudos sobre como mostrar a informação por meio de indicadores quantitativos e como representá-la num formato visual, visando facilitar a sua compreensão.

Nesse contexto, apresenta-se a seguinte questão de pesquisa: Como otimizar a visualização dos indicadores bibliométricos de produção científica de uma ICT?

Mediante pesquisa no Núcleo de Informação Tecnológica em Materiais (NIT-Materiais) da Universidade Federal de São Carlos (UFSCar), foi feita a extração de dados da produção científica da UFSCar na Plataforma Lattes por meio da ferramenta automatizada synclattes, que exporta arquivos de texto em diferentes formatos, inclusive JSON. Em seguida, Luís Gustavo Maschietto, discente do Programa de Pós-Graduação em Ciência, Tecnologia e Sociedade (PPGCTS) da UFSCar, usou a extração do synclattes para popular uma base de dados não relacional (NoSQL) com o uso da ferramenta Elasticsearch. Bases não relacionais têm por característica a facilidade de programação e a busca de texto livre. Após, desenvolveu um procedimento para tratamento bibliométrico de dados com o uso da linguagem Python, o que possibilitou a exportação de arquivos no formato JSON por meio de consultas via programação. Esses arquivos foram utilizados como fonte de dados para a interface interativa que foi desenvolvida neste projeto, ao qual disponibilizou gráficos que podem ser usados para a análise de indicadores de produção científica de ICTs.

1.1 Justificativa

Diversas iniciativas têm procurado explorar os metadados da Plataforma Lattes para a produção de indicadores de Ciência e Tecnologia, concentrando-se na superação de diferentes desafios. Como produto dessas pesquisas, algumas

ferramentas têm sido desenvolvidas, por ex., o scriptLattes (MENA-CHALCO; CESAR-JR, 2013), o Somos (CTIT UFMG, 2019) e o synclattes (MATIAS, 2015).

O grupo NIT-Materiais há alguns anos tem realizado pesquisas sobre a PL, como a criação de uma base referencial para o povoamento de repositórios institucionais (SARVO, 2018), o uso dos dados dos CVs Lattes como recurso estratégico para a gestão dos PPGs (MACIEL, 2018), e até mesmo a caracterização e discussão da produção científica de uma IES privada (VIDOTTI, 2016).

Uma das ferramentas bastante utilizada para a extração de dados dos currículos Lattes é o scriptLattes, um script robô *open source* com Licença Pública Geral escrito em linguagem de programação Python que extrai informações dos currículos formatados em HTML que estão disponíveis publicamente no site da Plataforma Lattes (MENA-CHALCO; CESAR-JR, 2013). Já o Somos é um sistema Web que apresenta indicadores em forma de gráficos, o que possibilita uma melhor análise e compreensão, disponibilizando esses recursos para toda comunidade acadêmica, bem como indivíduos externos. No entanto, apresenta informações pertinentes apenas a uma determinada instituição, como por exemplo, por meio do sistema Somos UFSCar. Por outro lado, a ferramenta synclattes não tem por objetivo fornecer recursos visuais, mas sim extrair e normalizar os dados dos currículos.

Para contagem de metadados e geração de visualizações podem ser usados programas bibliométricos, aplicações estatísticas ou até planilhas eletrônicas. No entanto, um desafio é gerar visualizações dinâmicas na Web utilizando dados já padronizados a partir dos metadados extraídos da Plataforma Lattes, o que compreende o objetivo deste trabalho.

Apesar de existirem sistemas capazes de disponibilizar recursos interativos na Web, é possível constatar falhas na extração dos dados via Plataforma Lattes. Com o uso do DOI desde setembro de 2008 pelo CNPq, esse problema foi minimizado em relação às publicações, principalmente as de periódicos, uma vez que basta ao pesquisador indicar o DOI do artigo a ser cadastrado no currículo que os metadados são obtidos diretamente do sistema DOI (mantido pela International DOI Foundation - IDF), o que aumenta a qualidade dos metadados para a geração de indicadores. Contudo, para publicações sem DOI os dados são informados manualmente pelos pesquisadores, o que pode acarretar diversos erros, como nome de pesquisador informado de várias maneiras, publicações com títulos diferentes em

currículos distintos, ausência de dados, etc. Além disso, há o desafio de desduplicar publicações extraídas de forma automatizada, uma vez que um mesmo artigo publicado em coautoria entre pesquisadores está nos currículos de cada autor.

A ferramenta synclattes foi desenvolvida para resolver os problemas mencionados, em que segundo Matias (2015), fatores como a falta de uma sistemática para a coleta, tratamento e importação dos metadados dificultam o aproveitamento do potencial da Plataforma Lattes. Superados esses desafios, no entanto, há ainda a necessidade de se ter uma interface de visualização de indicadores de produção científica por meio dos dados extraídos pelo synclattes, corroborando com o objetivo deste trabalho. Sendo assim, a interface interativa apresentada neste projeto integra-se à saída gerada pelo synclattes, ou seja, não é necessário o uso de outros softwares para importação de arquivos texto (por ex., Microsoft Excel). Ainda por parte do usuário, torna-se dispensável o conhecimento prévio sobre o uso de softwares de bibliometria, qualquer pessoa com nível básico em utilização de computadores consegue manipular a interface Web para realizar a análise de indicadores.

1.2 Objetivos gerais e específicos

O objetivo geral desta pesquisa é desenvolver uma interface interativa para visualização de indicadores de produção científica das ICTs brasileiras a partir de dados da Plataforma Lattes, tendo como unidade-caso a UFSCar.

Esse objetivo geral pode ser desdobrado nos seguintes objetivos específicos:

- Analisar a conjuntura de pesquisas sobre visualização de indicadores de produção científica a partir de dados da Plataforma Lattes;
- Explorar e selecionar ferramentas de visualização de informação para sistemas Web, ajustando parâmetros e aprimorando funcionalidades, quando necessário;
- Construir uma interface de acordo com as melhores práticas de visualização de informação.

1.3 Estrutura da dissertação

Esta dissertação está organizada em cinco seções. A primeira compreende a introdução. A segunda seção abrange o referencial teórico, contendo uma breve revisão da literatura acerca dos paradigmas da comunicação científica e dos estudos métricos, como também abordando temas referentes à avaliação da ciência, aos metadados em ambientes digitais e à visualização de informação. A terceira seção apresenta o método e o desenvolvimento da pesquisa. Já na quarta seção há a discussão dos resultados. Por fim, a quinta seção compreende as principais contribuições da pesquisa para a área de bibliometria, uma vez que foi desenvolvida uma interface interativa que permite a visualização de indicadores de produção científica a partir de dados da Plataforma Lattes extraídos pela ferramenta synclattes.

2. COMUNICAÇÃO CIENTÍFICA E VISUALIZAÇÃO DE INFORMAÇÃO

A comunicação científica ocorre pela disponibilização e circulação da informação produzida principalmente em artigos publicados em periódicos e em eventos científicos. Por sua vez, a visualização de informação facilita o entendimento de dados bibliométricos, por exemplo, da produção científica de universidades e de institutos de ciência e tecnologia.

2.1 Avaliação da ciência

De acordo com Silva e Freire (2012), mesmo sendo considerada uma área científica recente, é difícil precisar o surgimento da Ciência da Informação. O ser humano no decorrer da história vem tentando encontrar formas de classificar, registrar, organizar e difundir a informação em seus mais variados campos, porém, havia a necessidade de criação de uma área específica para tratar de problemas relativos à informação enquanto um fenômeno social.

Para Borko (1968), a Ciência da Informação é uma disciplina que investiga as propriedades e o comportamento informacional, além de lidar com as forças que governam os fluxos de informação e com os significados do processamento da informação. Ainda de acordo com o autor, a CI se preocupa com o corpo de conhecimentos sobre a origem, coleção, organização, armazenamento, recuperação, interpretação, transmissão, transformação e utilização da informação.

"Um propósito da ciência da informação é o de conhecer e fazer acontecer o tênue fenômeno de percepção da informação pela consciência, percepção que conduz ao conhecimento do objeto percebido. A essência do fenômeno da informação é a sua intencionalidade, ou seja, uma mensagem de informação deve ser direcionada, arbitrária e contingente para atingir o seu destino e produzir, para esse fim, tensão de competências distintas existentes em dois mundos: o do gerador e o do receptor da informação para onde o conhecimento se destina, e esta é a qualidade e a característica contida no fluxo de informação, que por isto é raro e surpreendente. [...] estruturas simbolicamente significantes, codificadas de forma socialmente decodificável e registradas (para garantir permanência no tempo e portabilidade no espaço) e que apresentam a competência de gerar conhecimento para o indivíduo e para o seu meio. Estas estruturas significantes são estocadas em função de

um uso futuro, causando a institucionalização da informação” (SMIT; BARRETO, 2002, p. 22).

A informação como “coisa” é importante no estudo de sistemas de informação, já que estes tratam a informação nesse sentido, ou seja, bibliotecas com livros, museus com objetos, e bases de dados informatizadas com dados na forma física de bits e bytes (BUCKLAND, 1991). Segundo Carvalho e Crippa (2013), dizer que o objeto da Ciência da Informação é a informação causa mais dúvidas e problemas epistemológicos do que soluções, pois se trata de algo demasiado abrangente. Sendo assim, um dos caminhos encontrados foi a afirmação do caráter interdisciplinar da área. Ainda de acordo com as autoras, a apropriação de termos, conceitos e métodos de outras áreas para a CI não é sinônimo de que realmente acontece uma relação dialógica, ou seja, as disciplinas envolvidas devem possuir a própria identidade a fim de, por diferentes pontos de vista, chegarem à compreensão de um objeto comum, fazendo com que a interdisciplinaridade se torne mais concreta.

Para Borko (1968), a Ciência da Informação é uma ciência interdisciplinar derivada de campos relacionados, tais como a Matemática, Lógica, Linguística, Psicologia, Ciência da Computação, Engenharia de Produção, Artes Gráficas, Comunicação, Biblioteconomia, Administração, e outros campos científicos semelhantes, além disso, a CI possui componentes de ciência pura, uma vez que investiga seu objeto sem considerar sua aplicação; e um componente de ciência aplicada, pois desenvolve produtos e serviços, possibilitando que os pesquisadores da área enfatizem um ou outro aspecto, dependendo do interesse ou prática. Segundo Arakaki, Simionato e Santos (2017), as tecnologias utilizadas no campo da informação assumiram um papel relevante nos processos sociais e econômicos, gerando oportunidades aos profissionais que possuem competências relacionadas ao tratamento e à representação da informação.

De acordo com Freire (2006), a Ciência da Informação tem como um de seus objetivos contribuir para a informação se tornar um elemento de inclusão social, trazendo desenvolvimento para as pessoas e nações. Sendo assim, quando cientistas e profissionais da informação organizam textos ou documentos, deveriam acreditar que as informações serão úteis para seus usuários potenciais e que delas resultarão em benefícios para a sociedade. Para Saracevic (1996), a Ciência da

Informação influenciou a forma como a informação é manipulada na sociedade, inclusive pela tecnologia, além disso, a CI possibilita uma melhor compreensão de diversos processos, estruturas e problemas associados ao conhecimento, à informação e ao comportamento humano.

Freire (2006) argumenta que a criação da tecnologia de impressão foi bastante útil para o desenvolvimento das forças produtivas, pois facilitou a circulação da informação científica, e a partir disso surgiram os primeiros periódicos científicos e o processo de comunicação científica se formalizou. Porém segundo Bush (1945), o crescimento em quantidade de publicações científicas trouxe uma dificuldade a respeito do uso eficaz do conhecimento gerado, pois o processo ao qual o ser humano faz uso desses conhecimentos é mais importante que a consulta para a extração de dados.

Segundo Engwall, Blockmans e Weaire (2014), as instituições acadêmicas possuem duas obrigações básicas, a de disseminar conhecimento por meio do ensino e a de gerar conhecimento por meio da pesquisa, sendo que em ambas as atividades há a significativa tarefa de se medir a qualidade. Sobre a geração de conhecimento, Santos e Mugnaini (2019) apontam que o periódico é um veículo de publicação utilizado pela comunidade científica que geralmente é focado em uma área específica, no qual reflete as características dessa área. Os periódicos científicos

“têm como função, além de certificação da ciência, estabelecer e implementar critérios de qualidade para a realização e divulgação de pesquisas, ajudar a consolidar as áreas de pesquisa, constituir-se como depósito das informações de interesse internacional, nacional ou regional, treinar revisores e autores em análise e crítica, melhorando a qualidade da ciência” (ERDMANN et al., 2009, p. 3).

De acordo com Bourdieu (1983), o pesquisador depende de sua reputação perante os colegas cientistas para obter fundos para pesquisas e atrair bons estudantes, como também para conseguir convites para eventos, ser chamado para consultorias e receber distinções (por ex., o Prêmio Nobel). Além disso, as contribuições acadêmicas individuais servem de alicerce para outras medições de impacto acadêmico (ROEMER; BORCHARDT, 2015), como por exemplo, para a geração de indicadores de produtividade de uma instituição específica. Para Vanz e

Silva Filho (2019, p. 19), "a Ciência é feita por um grande grupo de pessoas que partilham entre si o trabalho investigativo, mas também fiscalizam permanentemente as contribuições de cada um dos componentes do grupo".

A ciência é um dos recursos que o homem faz uso para conhecer o mundo e encontrar as respostas para inúmeras situações e fenômenos. Contudo, o aumento da demanda por financiamentos levou à necessidade de se avaliar a atividade científica e seus resultados (DORTA-GONZÁLEZ; DORTA-GONZÁLEZ, 2010). Para Alves (2015), uma das formas encontradas para se realizar a medição da produção acadêmica é por meio de indicadores, que são classificados como quantitativos ou qualitativos. Os indicadores de modo geral podem ser compreendidos como dados estatísticos utilizados para medir algo intangível e ilustrar aspectos de uma realidade multifacetada (GREGOLIN et al., 2005). Segundo Hicks et al. (2015), os indicadores quantitativos podem ser utilizados em conjunto com a avaliação por pares, evitando assim que tendências enviesadas por parte dos avaliadores prejudiquem a análise.

Spinak (1998) aponta que a avaliação da ciência é um componente da política pública de um país, uma vez que objetiva mensurar o desempenho científico e averiguar se o esforço em pesquisas e consequentemente em publicações resultou em uma contribuição real para o progresso da sociedade, tendo como referência as metas de política científica e tecnológica definidas para o país ou região. Em relação à produção científica das universidades no âmbito nacional e internacional, a análise dos seus indicadores estimula a realização de pesquisas e auxilia nas tomadas de decisões, permitindo a criação de novas estratégias para, entre outras possibilidades, alavancar estudos e melhorar o posicionamento (ranking) das instituições.

Segundo Bassoli (2017), a validação das pesquisas, dos pesquisadores e das instituições dentro do campo científico ocorre regularmente e é possível por meio de mecanismos de avaliação da qualidade, do volume, da relevância e do alcance do principal capital científico, que são as publicações. Targino (2000) argumenta que "a comunicação científica é indispensável à ciência, pois permite somar os esforços individuais dos membros das comunidades científicas".

Lima (2007, p. 29) argumenta que "a publicação científica tem sido, historicamente, a fonte de dados mais utilizada para gerar indicadores que permitam analisar os resultados e a qualidade da produção científica e, ainda, estimar o impacto científico". O avanço dos recursos computacionais possibilitou o aumento da

disponibilidade de bases de dados eletrônicas de periódicos e facilitou a extração, o armazenamento e o tratamento de dados, além disso, fez com que os artigos científicos se tornassem o tipo de fonte mais usado a nível mundial na construção de indicadores (FARIA et al., 2011).

Os meios digitais alteraram os processos utilizados na criação de novos conhecimentos, sendo que

“a linearidade e a sequencialidade, inerentes ao modelo tradicional, foram substituídas por um fluxo de comunicação ágil, rápido, dinâmico e, por vezes, interativo, desenvolvido no espaço virtual criado pela Internet. O novo fluxo permite a convergência entre autores, revisores e editores (leitores e pesquisadores) e estimula o compartilhamento de ideias e experiências. A comunicação se dá por meio de mensagem e arquivos digitais transferidos automaticamente de uma etapa a outra, que podem estar visíveis e acessíveis a vários desses autores simultaneamente, independentemente de distâncias físicas” (CASTRO, 2006, p. 60).

Alves (2015) aponta que os estudos quantitativos relacionados às publicações científicas são realizados por meio da Bibliometria, em que análises são feitas levando em consideração diversos parâmetros (coautoria, citações, etc.), constituindo uma forma de medir indiretamente a ciência, sua estrutura e seus resultados, provendo informações sobre a orientação científica e dinamismo de um país, como também ajudando no processo de tomada de decisões.

Segundo Okubo (1997), o termo Bibliometria foi usado pela primeira vez em 1969 pelo pesquisador Alan Pritchard em um *paper*, no qual procurava naquela época um nome para uma área de estudo que já existia por décadas. Ainda de acordo com o autor, a Bibliometria se fundamenta no cômputo e na análise estatística da produção científica em diversos formatos, como artigos e patentes, e também do levantamento de citações da comunicação científica. Van Raan (2014) argumenta que para saber se a análise bibliométrica é aplicável a um campo específico é necessário levantar onde os artigos da área costumam ser publicados, por exemplo, se os periódicos internacionais são os principais meios de comunicação em um campo, então na maioria dos casos a análise bibliométrica é aplicável.

De acordo com Davyt e Velho (2000), os principais indicadores são derivados das publicações de artigos científicos e são divididos em três categorias: os

indicadores de produção, que medem a quantidade de trabalhos publicados; os indicadores de citação, que mensuram o alcance das publicações por meio das citações em outros trabalhos (impacto); e os indicadores de ligação, que avaliam as relações entre os coautores dos trabalhos, seus grupos de pesquisa e instituições. Segundo Narin, Olivastro e Stevens (1994), os indicadores de ligação medem as coocorrências de autores, afiliação, palavras e referências.

Para Okubo (1997), os indicadores de produção científica são produzidos por meio da contagem do número de publicações por tipo de documento (livros, artigos, publicações científicas, relatórios, etc.), por instituição, área de conhecimento, país, autor, dentre outros. A Tabela 1 mostra a presença dos indicadores bibliométricos de produção em repositórios institucionais.

Tabela 1 - Presença dos indicadores bibliométricos de produção em RIs.

INDICADORES BIBLIOMÉTRICOS DE PRODUÇÃO												
Autor	Área do conhecimento	Ano da Publicação	País	Idioma	Comunidade / coleção	Instituição, Unidade ou Departamento	Tipo de documento (artigo, livro, tese, vídeo, etc)	Agência de fomento	Título	Assunto / palavra-chave	Resumo	Número de Referências Bibliográficas
91,4%	16,0%	87,7%	8,6%	32,1%	74,1%	27,2%	44,4%	17,3%	87,7%	93,8%	34,6%	4,9%

Fonte: Reis (2016).

Como pode ser visto na Tabela 1, o indicador bibliométrico de produção “Autor” é o que mais aparece em repositórios institucionais. Por meio da Tabela 2 é possível visualizar a presença dos indicadores bibliométricos de citação em repositórios institucionais.

Tabela 2 - Presença dos indicadores bibliométricos de citação em RIs.

INDICADORES BIBLIOMÉTRICOS DE CITAÇÃO							
Fator de Impacto	Índice de imediatez	Vida Média	Citações concedidas	Citações recebidas	Auto citação	Índice h	Mediana h
3,7%	2,5%	2,5%	4,9%	4,9%	4,9%	2,5%	2,5%

Fonte: Adaptada de Reis (2016).

De acordo com Reis (2016), a carência de indicadores bibliométricos de citação em repositórios institucionais pode ter relação com a complexidade dos algoritmos computacionais que são necessários para a produção desse tipo de indicador.

Matias (2015) argumenta que além dos indicadores disponíveis em bases de dados estabelecidas e reconhecidas pela comunidade científica, como por exemplo, a Scopus e a Web of Science, existe também a possibilidade de se obter indicadores de outras bases. Um exemplo de base de dados da produção científica brasileira que se destaca é a Plataforma Lattes, um sistema de informação na Web sob responsabilidade do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), que possibilita a integração das bases de dados de Currículos, Grupos de Pesquisa e Instituições, reunindo currículos de estudantes e pesquisadores do país (AUTRAN et al., 2015).

A Base de Currículos Lattes (BCL) é utilizada pelas principais universidades, institutos e fundações de amparo à pesquisa dos estados como instrumento para a avaliação de professores e alunos, estabelecendo-se como um grande repositório da produção científica de pesquisadores de todo o país, sendo caracterizada pela livre inserção de dados. Além disso, tornou-se estratégica não só para as atividades de planejamento e gestão, mas também para a formulação das políticas do Ministério da Ciência, Tecnologia, Inovações e Comunicações, e de outros órgãos governamentais da área de Ciência, Tecnologia e Inovação (CNPq, 2017).

O CNPq (2017) aponta que a BCL possui milhões de currículos que organizam referências a documentos (alguns públicos e outros privados) do arquivo pessoal ou institucional dos pesquisadores, formando um histórico das atividades científicas, acadêmicas e profissionais de pesquisadores cadastrados, como artigos em revistas internacionais, com fator de impacto e indexadas em bases de dados; artigos em revistas locais, sem DOI, de pequena circulação; livros, capítulos de livros, trabalhos e resumos em anais de congressos; demais publicações existentes. Lane (2010) em seu artigo publicado na revista Nature aponta o Currículo Lattes como uma das bases de dados de pesquisa mais completas e qualificadas que existem.

A disponibilização pública dos dados da Plataforma Lattes na Internet proporciona transparência e confiabilidade às atividades de fomento do CNPq, bem como para as agências que utilizam a plataforma. Além disso, fortalecem o intercâmbio entre pesquisadores e instituições e é fonte inesgotável para estudos e pesquisas. Na medida em que suas informações são recorrentes e cumulativas, têm também o papel de preservar a memória da atividade de pesquisa no país (CNPq, 2017).

O crescimento da produção científica e o desenvolvimento de tecnologias para a divulgação dos resultados da ciência contribuíram para o aumento do volume de dados e informações disponibilizadas à sociedade. Por sua vez, a publicação eletrônica de revistas científicas na Internet possibilita

"[...] que os artigos estejam disponíveis imediatamente após aprovação pelos editores. Essa modalidade de publicação contribui para aumentar a visibilidade dos resultados de pesquisa e diminuir o tempo entre a aprovação dos trabalhos e sua publicação em formato impresso. O artigo científico passa a ser uma unidade informacional independente, embora reunido posteriormente em fascículos, enquanto permanecerem vigentes os princípios tradicionais. Informações complementares e versões em outros idiomas podem ser incluídas nos sites das revistas, constituindo valor agregado à forma impressa. A publicação em papel e a organização em fascículos no novo fluxo de comunicação científica passam a ser subprodutos do formato eletrônico" (CASTRO, 2006, p. 60).

Diante da sobrecarga de informação em meio digital, torna-se necessária a realização de estudos sobre como mostrá-la por meio de indicadores quantitativos e representá-la num formato visual de modo a facilitar a sua compreensão. Apesar de existirem diversos estudos envolvendo a elaboração de indicadores bibliométricos a partir da análise da produção científica indexada em bases de dados, há a necessidade de investigar soluções tecnológicas que viabilizem a visualização desses indicadores, objetivando ampliar a compreensão a respeito dos resultados das atividades de pesquisa (REIS, 2016).

2.2 Metadados em ambientes digitais

A disseminação da informação em ambientes digitais requer a adesão a padrões para sua representação. Para tanto, existem os metadados, que para Hillmann (2005) são compostos por atributos ou elementos que descrevem um recurso a ser recuperado por meio de uma busca.

Baptista e Machado (2001) argumentam que apesar da alta tecnologia envolvida nos sistemas de indexação e de recuperação da informação, a relevância e a precisão nas respostas que eles retornam aos usuários ainda não alcançaram níveis satisfatórios. Então, tornou-se necessário melhorar a eficácia e a eficiência

dos serviços de informação e assim foram criados os chamados metadados semânticos.

Para Angelozzi e Martín (2010), diferentes esquemas de metadados irão coexistir e cada modelo de metadado será usado para um determinado fim, sendo importante pensar na questão da interoperabilidade pois permite o intercâmbio entre esquemas. Segundo Baptista (2007), a discussão sobre metadados inclui: os aspectos conceituais e tecnológicos; a diversidade de objetos que são transformados em recursos informacionais; os diferentes atores participantes do processo, tais como produtores de informação (autores), desenvolvedores de software, profissionais da informação (principalmente bibliotecários), agências normativas e os usuários da informação.

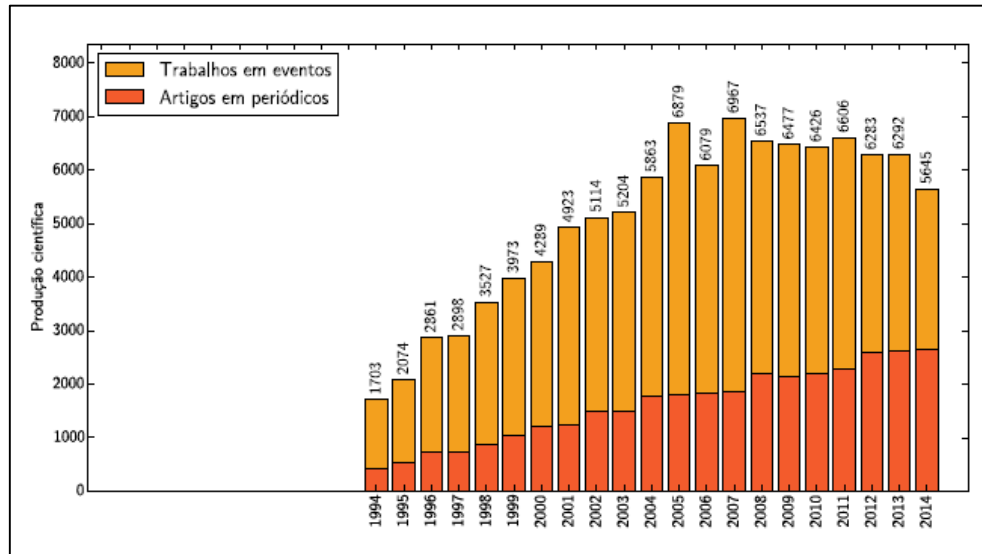
Uma notação que vem sendo largamente utilizada no desenvolvimento de sistemas para Web chama-se JSON (*JavaScript Object Notation*)¹, por ser um formato leve de troca de dados e de fácil compreensão.

O CNPq permite por meio de um convênio que as instituições de ensino, pesquisa e inovação do país obtenham cópias dos currículos Lattes de seus alunos e professores, no qual cada ICT pode utilizar os dados de acordo com suas necessidades. Esse convênio torna possível a implementação de ferramentas como o synclattes, um conjunto de scripts implementados em linguagem Python por Matias (2015), que entre outras coisas, realiza a extração de metadados dos currículos Lattes por meio da utilização do *Web Service* oficial da PL.

A Figura 1 apresenta a produção científica da UFSCar extraída pelo synclattes, levando em consideração os trabalhos em eventos e os artigos em periódicos.

¹ *JavaScript Object Notation* - <<https://www.json.org>>

Figura 1 - Produção científica da UFSCar entre os anos de 1994 a 2014.



Fonte: Adaptada de Matias, 2015.

É possível observar na Figura 1 que entre os anos de 1994 a 2005 houve um aumento gradativo na produção científica, em 2006 teve uma queda de aproximadamente 12% comparado a 2005, nos anos seguintes ocorreram oscilações até chegar no ano de 2014, que teve uma baixa considerável se comparar com o ano de 2013 (11%).

De acordo com Bassoli (2017), a coleta automatizada de grandes quantidades de metadados sobre as publicações científicas foi relevante para a ampliação de estudos bibliométricos. Como exemplo, a Universidade Federal de Minas Gerais (UFMG) desenvolveu o sistema Somos, uma interface que tem, entre outros objetivos, facilitar o mapeamento de competências por meio da identificação dos pesquisadores, de suas especialidades e publicações. O Somos foi implantado na UFSCar², sendo que “todas as informações apresentadas foram retiradas da Plataforma Lattes com autorização do CNPq” (CTIT UFMG, 2019). A Figura 2 apresenta a interface de pesquisa do sistema Somos UFSCar.

² Somos UFSCar - <<http://somos.ufscar.br>>

Figura 2 - Interface de pesquisa do Somos UFSCar.

The screenshot shows the Somos UFSCar search interface. At the top, there is a search bar with the letter 'a' entered. Below the search bar, the results are organized into four columns:

- ESPECIALIDADES:** Lists 14 results from 5,775 possible. Examples include 'A FIGURA FEMININA NO CINEMA BRASILEIRO / (CINEMA)', 'ABORDAGEM SOFT DA PESQUISA OPERACIONAL / (PESQUISA OPERACIONAL)', and 'ACID CATALYSIS / (QUÍMICA INORGÂNICA)'.
- PALAVRAS CHAVE:** Lists 20 results from 80,299 possible. Examples include 'A ÁGUA COMO BEM PÚBLICO DE VALOR ECONÔMICO', 'A ÁGUA COMO DIREITO HUMANO', and 'A FABRICA DE SOFTWARE'.
- PROFESSORES:** Lists 20 results from 1,299 possible. Examples include 'ADAIL RICARDO LEISTER GONCALVES', 'ADALBERTO PICININ', and 'ADRIANA MATTAR MAAMARI'.
- UNIDADES ACADÊMICAS:** Lists 9 results from 52 possible. Examples include 'CENTRO DE CIÊNCIAS AGRÁRIAS - CCA', 'CENTRO DE CIÊNCIAS BIOLÓGICAS E DA SAÚDE - CCBS', and 'DEPARTAMENTO DE CIÊNCIAS AMBIENTAIS - DCA'.

Fonte: Somos UFSCar, busca realizada pelo autor.

Como pode-se observar na Figura 2, ao digitar a letra “a” na barra de busca do sistema Somos UFSCar foram exibidos resultados sobre especialidades, palavras-chave, professores e unidades acadêmicas (separadas por centros e por departamentos).

A Universidade de São Paulo (USP) possui uma interface de geração de indicadores científicos chamada weR_USP³, ao qual trabalha com diversas fontes de informação, desde bases de sistemas internos da própria instituição até dados da Plataforma Lattes.

Outra solução que merece destaque é o e-Lattes⁴, que é um projeto que se apresenta como a evolução do scriptLattes, tendo por objetivo disponibilizar de maneira aberta e fácil os indicadores sobre a produção acadêmica brasileira, por meio de dados oriundos da Plataforma Lattes. O e-Lattes foi desenvolvido por pesquisadores da Fiocruz Brasília, da Universidade de Brasília e da Universidade Federal do ABC, sendo que os pesquisadores Ricardo Sampaio e Jesús Mena-Chalco lideram a equipe de desenvolvimento do projeto.

Empresas da iniciativa privada também desenvolvem soluções por meio da utilização dos dados da Plataforma Lattes, que é o caso da empresa Indeorum,

³ weR_USP - <<https://uspdigital.usp.br/datausp/publico/indicador/indicadores.jsp>>

⁴ e-Lattes - <<http://elattes.com.br>>

desenvolvedora do sistema Cientum⁵, que tem o objetivo de realizar a gestão de pesquisadores e disponibilizar análises gráficas de maneira qualitativa, quantitativa e temporal.

Por fim, é importante destacar o relevante papel do Portal Brasileiro de Dados Abertos⁶ para os estudos métricos, que é uma ferramenta que o governo federal disponibiliza para que qualquer cidadão encontre e utilize facilmente dados e informações da administração pública (PORTAL BRASILEIRO DE DADOS ABERTOS, 2019). Na Figura 3 é mostrada a página inicial do portal.

Figura 3 - Página inicial do Portal Brasileiro de Dados Abertos.

The screenshot shows the homepage of the Brazilian Open Data Portal (dados.gov.br). The header is green and contains the logo, the text 'dados.gov.br', and 'PORTAL BRASILEIRO DE DADOS ABERTOS'. There are navigation links for 'Dados', 'Organizações', 'Aplicativos', 'Inventários', 'Concursos', 'INDA', 'Perguntas frequentes', 'Contato', and 'Sobre o portal'. A search bar at the top right says 'Pesquisar conjuntos de dados...'. The main content area is titled 'Conjuntos de dados' and features a sidebar with a list of organizations and their respective dataset counts. The main area shows search results for 'Pareceres Jurídicos (Ementas) do Serviço de Limpeza Urbana' and 'Indicadores Nacionais de Ciência, Tecnologia e Inovação'. The search results indicate that 6,767 datasets were found, ordered by relevance.

Fonte: Portal Brasileiro de Dados Abertos, busca realizada pelo autor.

Como pode ser visto na Figura 3, há 6.767 conjuntos de dados na plataforma. Esses dados são categorizados por: Organizações; Grupos (Política, Saúde, Educação, etc.); Etiquetas (por ex., Indicador); Formatos (CSV, PDF, HTML, JSON, entre outros); e Licenças (inclusive Creative Commons).

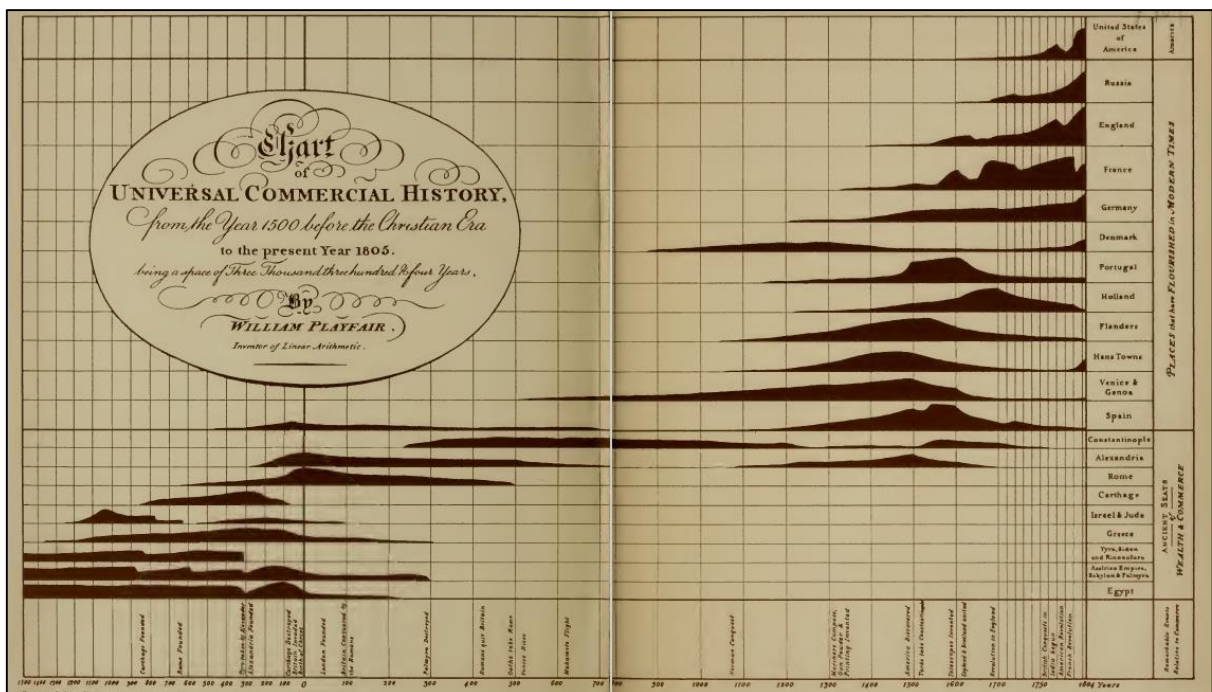
⁵ Cientum - <<https://cientum.indeorum.com>>

⁶ Portal Brasileiro de Dados Abertos - <<http://dados.gov.br>>

2.3 Conceitos fundamentais de visualização de informação

A preocupação em mostrar a informação de forma visual é objeto de estudo de séculos atrás, o que foi constatado por Brinton (1939), que relatou o fato de William Playfair ter construído os seus primeiros gráficos a partir do ano de 1786, sendo que um deles pode ser visto na Figura 4. Esse gráfico foi revolucionário pois representou uma grande quantidade de informação em um formato visual que até então não tinha sido utilizado.

Figura 4 - Representação gráfica de Playfair de 1805.



Fonte: Brinton 1939.

Segundo Alves (2015), a área de visualização de informação pode contribuir para a compreensão dos resultados das iniciativas científicas e tecnológicas em que os pesquisadores brasileiros estão envolvidos, uma vez que, com a sobrecarga de informações, o estudo de como representá-las de uma forma visual que facilite sua compreensão tem sido uma das principais preocupações dos estudos da área de visualização de informação.

Freitas et al. (2001) apontam que a visualização de informação pode ser compreendida como o estudo das principais formas de representações gráficas para apresentação de informações, utilizando aspectos de computação gráfica, interfaces homem-computador e mineração de dados. Ainda segundo os autores, é necessário

prover formas de manipulação dos conjuntos de dados, tanto geométrica (com rotações e zoom) como analítica (por meio de redução ou ampliação, de acordo com algum critério escolhido pelo usuário), de modo a permitir alterações na representação visual para que novos aspectos do conjunto possam ser observados.

De acordo com Card, Mackinlay e Shneiderman (1999), a visualização de informação apresenta as seguintes características:

- Suporte a computador: a visualização é exibida por meio da tela de um computador;
- Interatividade: a visualização pode ser manipulada de maneira simples e direta, possibilitando a filtragem dos dados;
- Representação visual: a informação é representada por meio de atributos como localização, comprimento, forma, cor e tamanho;
- Abstração: informações como dados quantitativos, processos ou relações não têm um aspecto físico natural, portanto torna-se necessário transformá-los em formas e cores, de modo a ficarem mais perceptíveis;
- Ampliação da cognição: visualizações gráficas possibilitam que o cérebro humano consiga compreender a informação de maneira mais fácil, além de auxiliar a memória e conseqüentemente a retenção da informação.

Yamaguchi (2010) argumenta que o tipo de dado que será manipulado deve ser levado em consideração para a escolha de uma técnica de visualização de informação. A autora criou um sumário da caracterização dos dados tendo por base os trabalhos de Keim (2002), Freitas et al. (2001) e Shneiderman (1996), que pode ser visto no Quadro 1.

Quadro 1 - Caracterização dos tipos de dados.

Critério	Classe	Exemplos
Classe de informação	Texto/Web	Documentos/Textos na Web
	Hierárquico e grafos	Chamadas telefônicas e páginas da Internet
	Algoritmos e softwares	Dados de depuração de código-fonte, como logs

Natureza do domínio	Qualitativos nominais	Gênero e estado civil
	Qualitativos ordinais	Nível de conhecimento (básico, intermediário e avançado)
	Quantitativos discretos	Total de vendas de um produto
	Quantitativos contínuos	Pressão arterial
Dimensão	1-D	Dados com característica de tempo, como dados temporais
	2-D	Dados geográficos, mapas e superfícies de terrenos
	3-D	Dados do cotidiano, como dados médicos
	n-D	Tabelas de base de dados

Fonte: Adaptado de Yamaguchi (2010).

De acordo com Nascimento e Ferreira (2005), por meio da visualização gráfica é possível sintetizar uma grande quantidade de informações. Por sua vez, a visão é o sentido humano que tem a maior capacidade de assimilação de informações por unidade de tempo e, além disso, o sistema cognitivo do cérebro humano trabalha com mais facilidade quando se depara com artefatos visuais, como por exemplo os diagramas. Silva (2006, p. 7) argumenta que é por meio da visão que “o ser humano pode reparar em detalhes e detectar padrões em dados representados pelos artefatos, trazendo à tona informações que seriam difíceis de se analisar sem o auxílio de representações de dados”.

“O interesse sobre a maneira como artefatos visuais e interativos podem ser desenvolvidos visando auxiliar esse processo de cognição externa facilitado por computadores deu origem à área de pesquisa denominada Visualização de Informação (*Information Visualization*), que estuda o uso de representações visuais e interativas de dados abstratos e não baseados em aspectos físicos, com o propósito de ampliar a cognição” (CARD; MACKINLAY; SHNEIDERMAN, 1999).

Few (2009) argumenta que para que os dados tenham mais sentido, de modo que o processo de tomada de decisão seja racional e sustentável, se faz necessário

conhecer e aplicar as regras da visão, que estão distribuídas em quatro categorias, as quais podem ser vistas no Quadro 2.

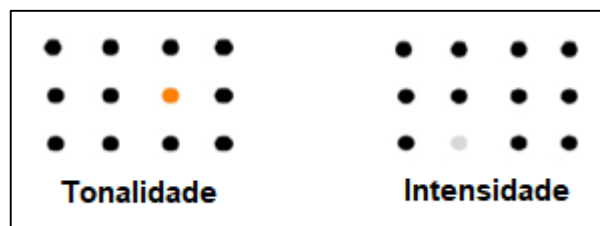
Quadro 2 - Regras da visão.

Categoria	Atributo	Quantitativo
Cor	Tonalidade	Não
	Intensidade	Sim, mas limitado
Forma	Comprimento	Sim
	Largura	Sim, mas limitado
	Tamanho	Sim, mas limitado
	Orientação	Não
	Formato	Não
Posição	2-D	Sim
Movimento	Direção	Sim, mas limitado (“velocidade”)

Fonte: Adaptado de Few, 2006.

A cor está relacionada a tonalidade (colorido) e a intensidade (fraco/forte). Juntamente com a forma, a cor é uma regra da visão de fácil percepção. Na Figura 5 podem ser vistos exemplos de tonalidade e intensidade.

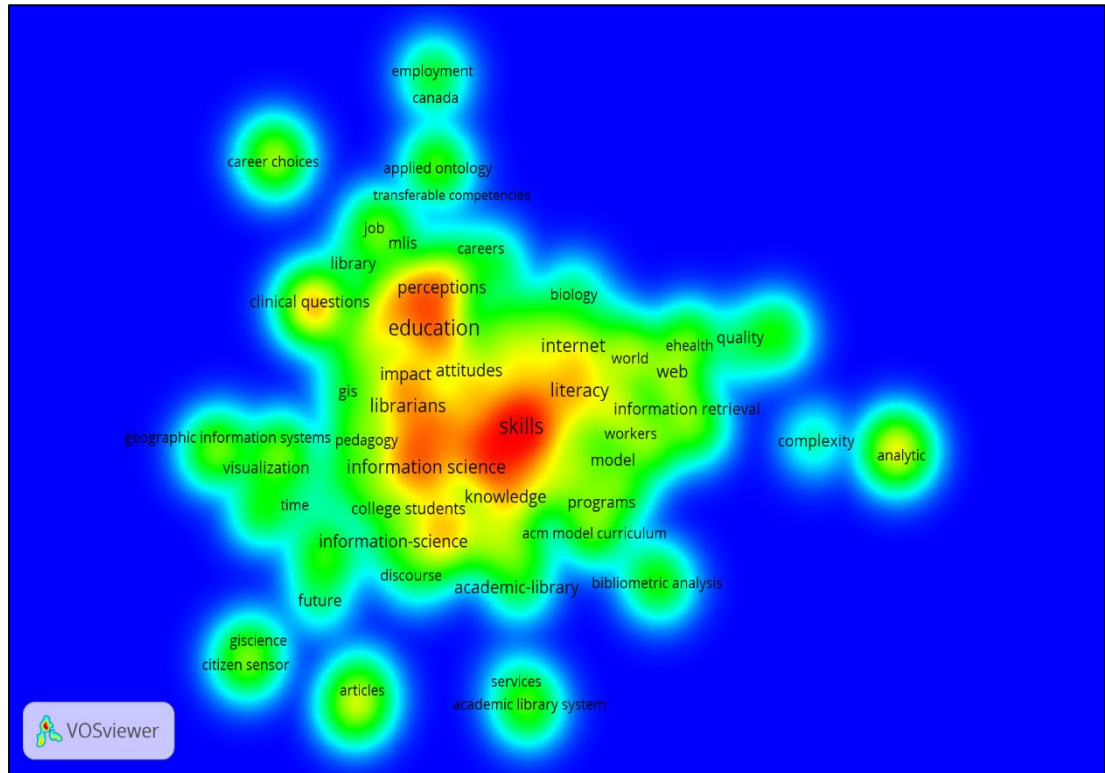
Figura 5 - Tonalidade e intensidade.



Fonte: Adaptada de Few, 2009.

A Figura 6 mostra a importância de tonalidade e intensidade, a partir de uma visualização de densidade.

Figura 6 - Visualização de densidade.



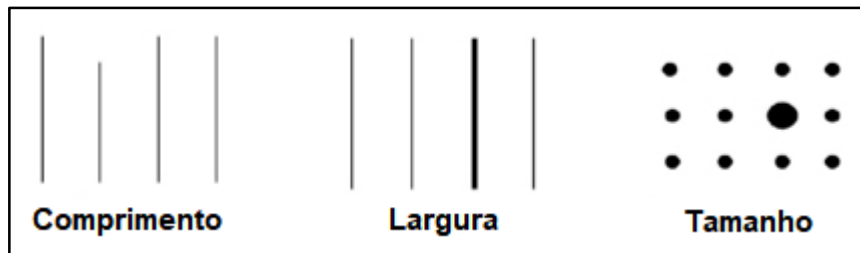
Fonte: Elaborada pelo autor.

A visualização de densidade da Figura 6 foi gerada a partir de uma busca por artigos na base de dados Web of Science, considerando a relação entre as palavras "Ciência da Informação", "Competência" e "Habilidades". A busca retornou um arquivo no formato "txt", ao qual foi importado no software VOSviewer⁷, que por sua vez considerou a coocorrência de palavras-chave (do autor e *plus*) presentes nos artigos. No caso dessa visualização, "cores quentes" significam forte ligação entre palavras-chave.

É importante lembrar que alguns cuidados devem ser tomados em relação à cor: algumas pessoas não conseguem enxergar determinadas cores; há cores que carregam um significado implícito, como o vermelho, que geralmente está relacionado com "perigo".

A forma é outra regra da visão descrita por Few (2006), ao qual está vinculada a atributos como comprimento, largura e tamanho (Figura 7), orientação (Figura 8) e formato (Figura 9).

⁷ O VOSviewer é um software para construção e visualização de redes bibliométricas. Possui funcionalidade de mineração de texto que pode ser usada para construir e visualizar redes de coocorrência de termos extraídos de literatura científica. <<https://www.vosviewer.com>>

Figura 7 - Comprimento, largura e tamanho.

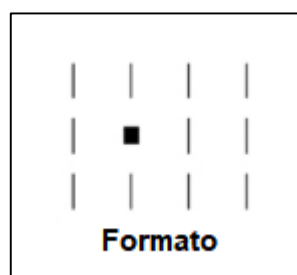
Fonte: Adaptada de Few, 2009.

O comprimento é utilizado, por exemplo, no gráfico de barras empilhadas. A largura pode ser encontrada nos grafos, no qual a espessura das arestas varia conforme a relação entre os nós. O tamanho também é visto em grafos, em que os círculos (nós) possuem diâmetros que variam de acordo com um valor quantitativo.

Figura 8 - Orientação.

Fonte: Adaptada de Few, 2009.

A respeito do que foi mostrado na Figura 8, é possível citar os gráficos de barras e de colunas, uma vez que a diferença entre eles é a orientação, sendo horizontal no gráfico de barras e vertical no gráfico de colunas.

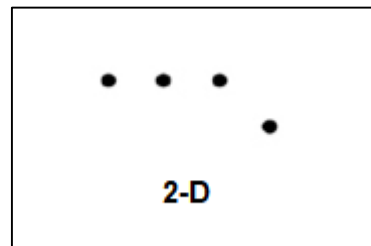
Figura 9 - Formato.

Fonte: Adaptada de Few, 2009.

Sobre o atributo formato apresentado na Figura 9, o grafo é uma visualização que utiliza elementos de diferentes formatos, como círculos (nós) e linhas (arestas).

A posição é uma regra da visão que é percebida quantitativamente com alto grau de precisão por disponibilizar a informação em eixos (vertical e horizontal), como também possibilita a identificação de padrões, tendências e exceções. Na Figura 10 é mostrado um exemplo de posição 2-D.

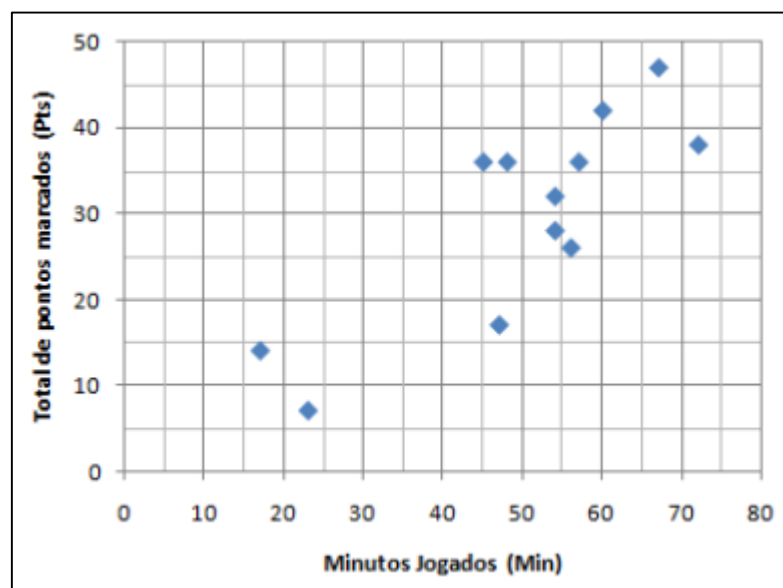
Figura 10 - Posição 2-D.



Fonte: Adaptada de Few, 2009.

Já a regra de visão movimento está relacionada com o atributo direção, sendo que ao observar a direção dos dados em uma imagem torna-se possível perceber as mudanças que os dados sofrem. O gráfico de dispersão (Figura 11) é um exemplo de visualização que utiliza essa regra da visão.

Figura 11 - Gráfico de dispersão.



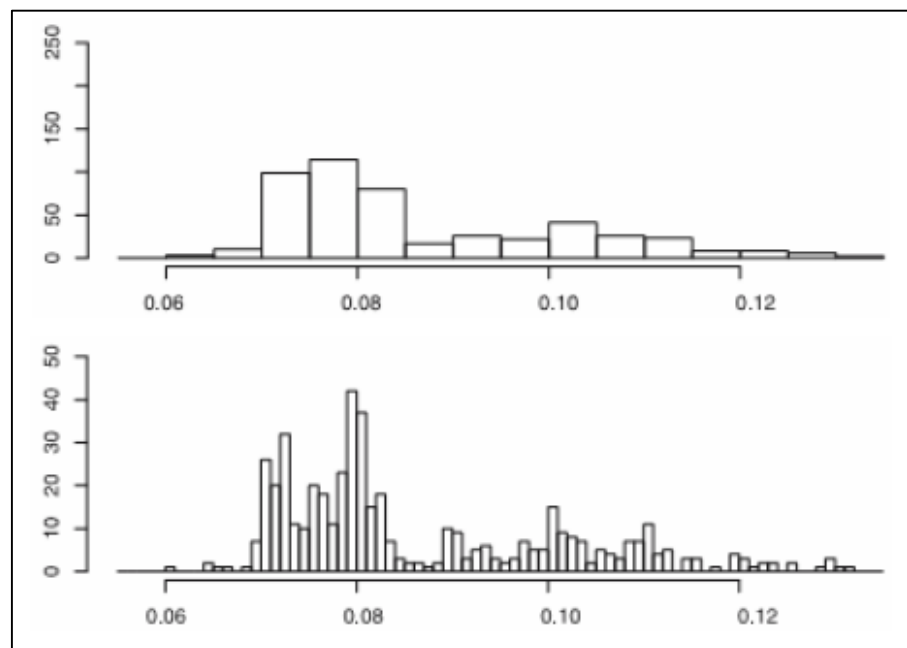
Fonte: Adaptada de Mugabe, Fernandes e Correia (2012).

Segundo Yamaguchi (2010), com a utilização do gráfico de dispersão é possível identificar a correlação entre as variáveis, podendo ser positiva (eixo y cresce quando eixo x cresce), negativa (eixo y decresce quando eixo x cresce), como também pode não haver uma correlação, que é o caso apresentado no gráfico da Figura 11, em que são mostrados dados de jogadores de um time de Basquete (total de pontos marcados X minutos jogados).

Unwin (2008) apresenta uma discussão a respeito dos fatores que influenciam na leitura e construção de um gráfico, como: (a) escala; (b) ordenação e organização dos valores que são desenhados no gráfico; (c) legenda; (d) posição do gráfico em relação ao texto; (e) tamanho, estrutura e proporção do gráfico.

a) Escala: é importante para a organização de dados categóricos e sua definição é mais difícil quando os dados são contínuos. Embora seja comum, não é obrigatório ter o valor “zero” como o ponto base. É considerada boa prática a utilização de valores arredondados nos eixos e também números que são múltiplos de uma constante. É importante definir rótulos nos eixos, sendo que muitos nomes podem causar confusão e poucos nomes podem dificultar o entendimento dos valores. Na Figura 12 são mostrados dois gráficos para o mesmo conjunto de dados, porém com escalas diferentes.

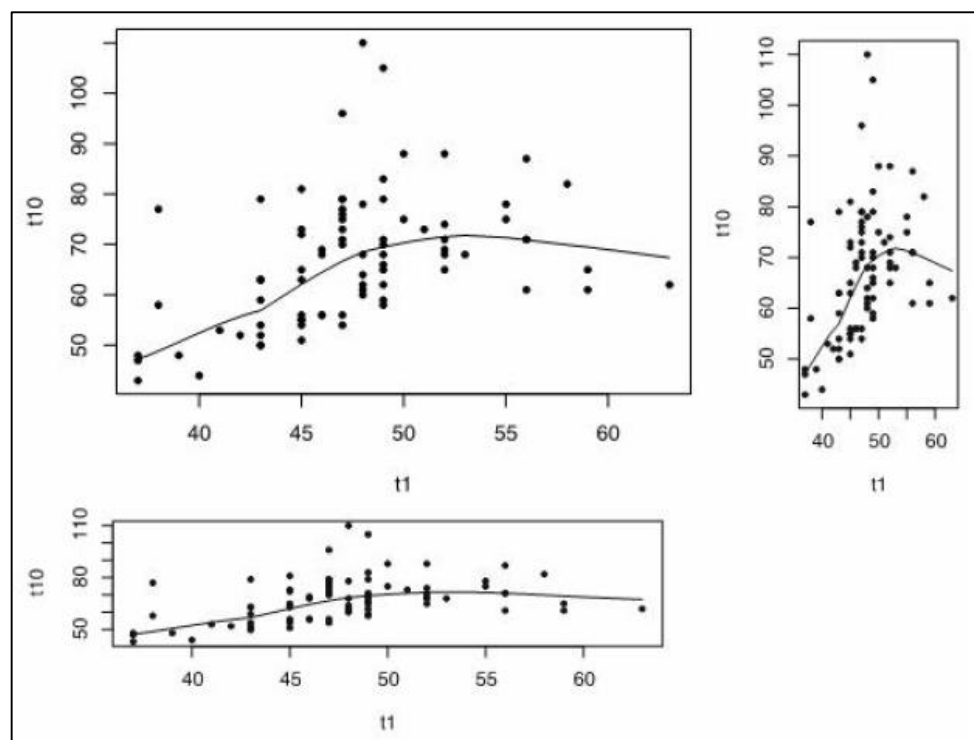
Figura 12 - Exemplo de gráfico com escala.



Fonte: Unwin, 2008.

- b) Ordenação e organização: a ordem e a posição em que os atributos dos dados são dispostos no gráfico interferem no resultado final da visualização. Quando o atributo não possui uma ordem natural (por ex., alfabética), a recomendação é que a ordenação seja de acordo com algum tipo de categoria.
- c) Legenda: deve acompanhar e explicar o gráfico de modo completo, conciso e claro, objetivando a sintetização da informação.
- d) Posição: como o gráfico ilustra uma ideia exposta por um texto, se possível, deve-se posicioná-lo na mesma página, visando facilitar a leitura.
- e) Tamanho, estrutura e proporção: o gráfico deve possuir medidas de modo que as informações sejam exibidas nitidamente. Em relação à estrutura, molduras em torno do gráfico podem ser utilizadas para delimitar a sua área. A proporção entre a altura e a largura da imagem é importante para o entendimento do gráfico, como pode ser visto na Figura 13, em que os mesmos dados são utilizados em três gráficos com altura e largura em proporções diferentes.

Figura 13 - Diferença de proporções de altura e largura.



Fonte: Unwin, 2008.

Shneiderman (1996) classifica as técnicas de visualização de acordo com os tipos de dados e com as tarefas que são realizadas pelo usuário, sendo:

- Visão geral sobre todos os dados;
- Foco (zoom) nos itens de interesse;
- Filtragem para descarte de itens que não interessam;
- Detalhe sob demanda (*details-on-demand*) de modo a selecionar um item ou um grupo visando a obtenção de mais detalhes;
- Relacionamento entre itens para encontrar ligações entre eles;
- Registro do histórico das ações de forma a tornar possíveis as opções de desfazer, refazer e refinar progressivamente;
- Extração de subconjuntos de dados.

Wong (2010) apresenta na Figura 14 os três elementos essenciais de bons gráficos de informação.

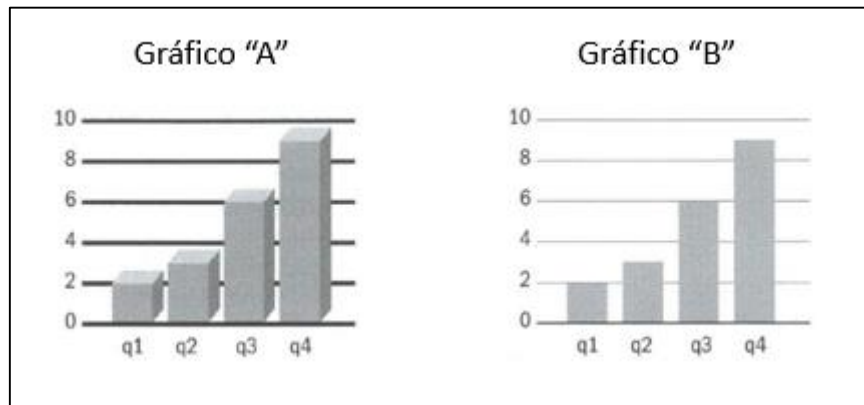
Figura 14 - Três elementos essenciais de bons gráficos de informação.



Fonte: Adaptada de Wong, 2010.

Segundo Wong (2010), o conteúdo rico ajuda a trazer significado à visualização. A execução sofisticada faz com que o gráfico seja associado a elementos do cotidiano, facilitando o entendimento dos dados disponibilizados. Por meio da visualização convidativa é possível destacar para o leitor a essência da informação, sendo esse elemento exemplificado na Figura 15.

Figura 15 - Comparação entre dois gráficos de colunas.



Fonte: Adaptada de Wong, 2010.

Como pode ser visto na Figura 15, o Gráfico "A" contém linhas de grade pesadas e é plotado em 3-D, o que obscurece os dados e desvia a atenção do usuário. Já o Gráfico "B" é limpo e nítido, ou seja, o foco está nos dados e não nos elementos visuais.

2.4 Interfaces para visualização de indicadores

Para Vieira e Correa (2011), uma interface baseada na visualização permite a compreensão sobre a organização da informação disponibilizada por um sistema e proporciona uma visão geral dos assuntos tratados, possibilitando a dedução de novos conhecimentos. Ainda segundo os autores, a utilização de recursos gráficos na apresentação de informações faz com que a mensagem transmitida seja compreendida de forma mais natural, ou seja, com menos esforço cognitivo.

Segundo Wong (2010), a melhor prática para a construção de gráficos envolve quatro etapas:

1) Pesquisa

- i. Pesquisar por fontes autorizadas e atualizadas;
- ii. Usar fontes de dados independentes, ou seja, que não são vinculadas a organizações de direito privado;
- iii. Obter autorização para a utilização dos dados (quando necessário).

2) Edição

- i. Identificar a mensagem principal que se quer transmitir;

- ii. Escolher a melhor série de dados para ilustrar o objetivo, por exemplo, fatia de mercado X receita total;
- iii. Filtrar e simplificar os dados visando entregar para o público-alvo a essência dos dados;
- iv. Fazer ajustes numéricos nos dados brutos para facilitar o entendimento, por exemplo, valores absolutos X porcentagem.

3) Plotagem

- i. Escolher o tipo de gráfico correto para apresentar os dados;
- ii. Configurar o gráfico de forma apropriada;
- iii. Rotular o gráfico por meio da utilização de título, legenda, etc.;
- iv. Trabalhar com cores e tipografia para destacar a mensagem principal a ser transmitida.

4) Revisão

- i. Comparar os dados plotados com a fonte de dados;
- ii. Avaliar se o gráfico tem significado e é compreensível;
- iii. Olhar o gráfico com a perspectiva do leitor;
- iv. Validar os dados por meio de outras fontes e consultar especialistas em situações como o aparecimento de *outlier*.

Pela técnica de visualização de informação dados abstratos são convertidos em gráficos ou imagens de modo a diminuir o esforço cognitivo necessário para a compreensão de um determinado assunto. Few (2006) argumenta que recursos como o Dashboard (“painel”), uma ferramenta gráfica que possibilita a exibição visual de informações, auxiliam a visão e o cérebro humano no processo de extração de informações relevantes. A representação de informações por meio de gráficos complexos foi possível devido a evolução dos equipamentos de processamento de imagens.

Para Few (2006), são características do Dashboard:

- O painel deve ser capaz de exibir informações provenientes de diversas fontes;

- As visualizações do painel devem ser projetadas de forma a caber em uma única tela, possibilitando ao usuário rápida absorção do conteúdo exibido;
- Se não houver a necessidade de monitoramento da informação em tempo real (tráfego aéreo, etc.), a base de dados pode ser atualizada com uma frequência pré-determinada, por exemplo, uma vez por mês;
- O painel tem que possuir mecanismos de exibição pequenos, concisos, claros e intuitivos.

O Quadro 3 apresenta as características da ferramenta Dashboard, com suas variáveis e possíveis valores.

Quadro 3 - Categorização de Dashboard.

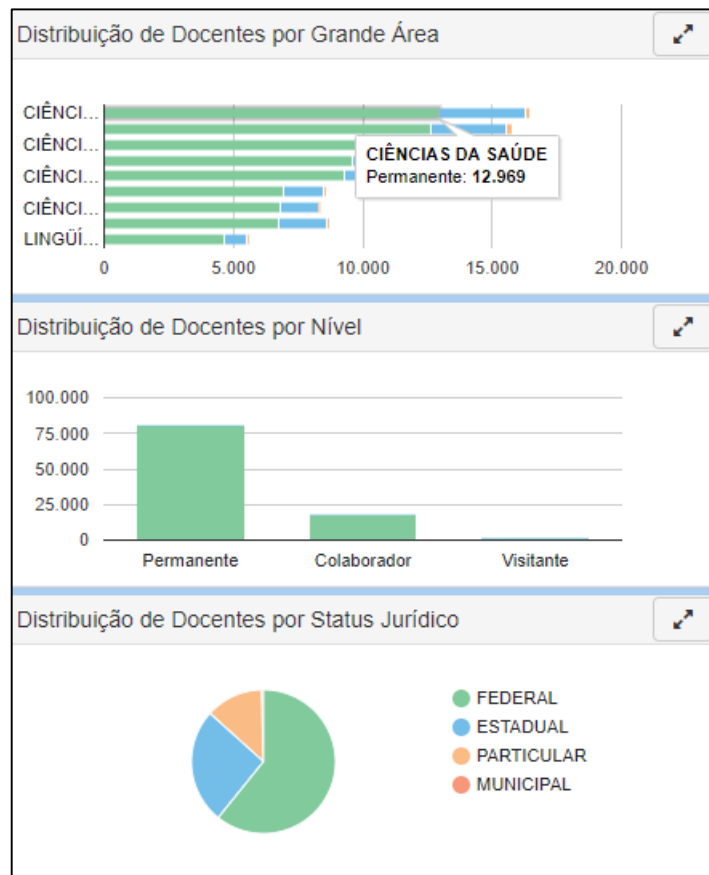
Variável	Valor
Papel	Estratégico Analítico Operacional
Tipo de dados	Quantitativo Não quantitativo
Domínio dos dados	Vendas Finanças Marketing Manufatura Recursos humanos
Tipo de medidas	Balanced Scorecard (BSC) Seis Sigma Sem mensuração de desempenho
Alcance dos dados	Toda a empresa Departamental Individual
Frequência de atualização	Mensal Semanal Diária De hora em hora Tempo real ou próximo de tempo real
Interatividade	Interface estática Interface interativa (com filtros, etc.)
Mecanismos de exibição	Principalmente gráficos Principalmente textuais Gráficos e textuais
Funcionalidades de portal	Canal para dados adicionais Não há funcionalidades de portal

Fonte: Adaptado de Few, 2006.

A visualização de informação ainda é pouco explorada para auxiliar na análise e compreensão de indicadores relacionados ao desempenho de atividades científicas e tecnológicas, mas existem algumas iniciativas que merecem ser destacadas:

- a. GEOCAPES⁸: é um sistema de informações georreferenciadas que oferece visualizações interativas com dados sobre os seguintes indicadores: concessão de bolsas de pós-graduação da Capes no Brasil; distribuição de bolsistas da Capes no exterior; distribuição de discentes de pós-graduação no Brasil; distribuição de programas de pós-graduação no Brasil; distribuição de docentes das IES (por grande área, por nível/tipo de vínculo e por status jurídico da instituição); acessos ao Portal de Periódicos da Capes; e investimentos da Capes em bolsas e fomento (GEOCAPES, 2019). A Figura 16 apresenta a distribuição dos docentes das IES em 2017.

Figura 16 - GEOCAPES: distribuição dos docentes das IES em 2017.



Fonte: GEOCAPES, busca realizada pelo autor.

⁸ GEOCAPES - <<https://geocapes.capes.gov.br>>

- b. Painel Lattes⁹: é um projeto do CNPq que oferece ao usuário informações qualificadas e atualizadas sobre a atuação de pesquisadores em Ciência, Tecnologia e Inovação, provenientes dos currículos Lattes. As informações são extraídas de currículos atualizados nos últimos 48 meses e estão dispostas em forma de painéis com dados qualificados por geografia, sexo, faixa etária, etc. (PAINEL LATTES, 2019). Por meio da Figura 17 pode-se observar o total de currículos cadastrados na Plataforma Lattes até novembro de 2016.

Figura 17 - Painel Lattes: total de currículos cadastrados até novembro de 2016.



Fonte: Painel Lattes, busca realizada pelo autor.

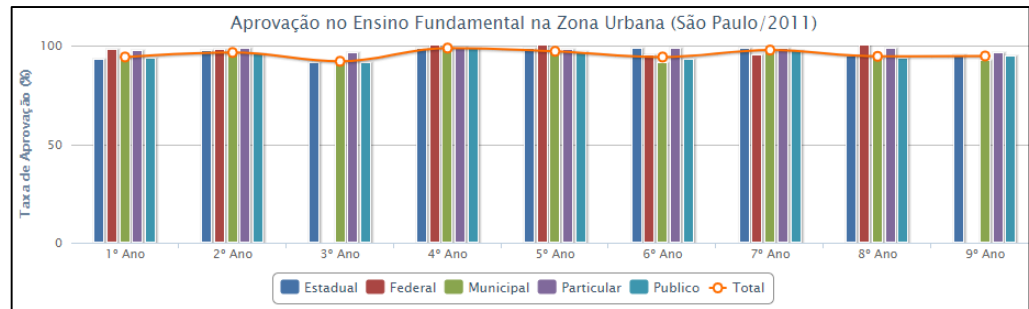
- c. VisPublica¹⁰: é um projeto desenvolvido em parceria entre o COPPE/UFRJ e a UNIFEI (Universidade Federal de Itajubá) por meio de acordo de cooperação técnica com o Ministério do Planejamento, Desenvolvimento e Gestão, que tem o objetivo de investigar de que maneira as técnicas de visualização de informação podem ser aplicadas para facilitar a transparência de dados públicos e a tomada de decisão, além de utilizar de forma eficiente essas técnicas para a massificação dos dados públicos (VISPUBLICA, 2019). A Figura 18

⁹ Painel Lattes - <<http://estatico.cnpq.br/painelLattes>>

¹⁰ VisPublica - <<http://vispublica.gov.br>>

mostra a aprovação no Ensino Fundamental na Zona Urbana de São Paulo em 2011.

Figura 18 - VisPublica: aprovação no Ensino Fundamental na Zona Urbana de São Paulo em 2011.

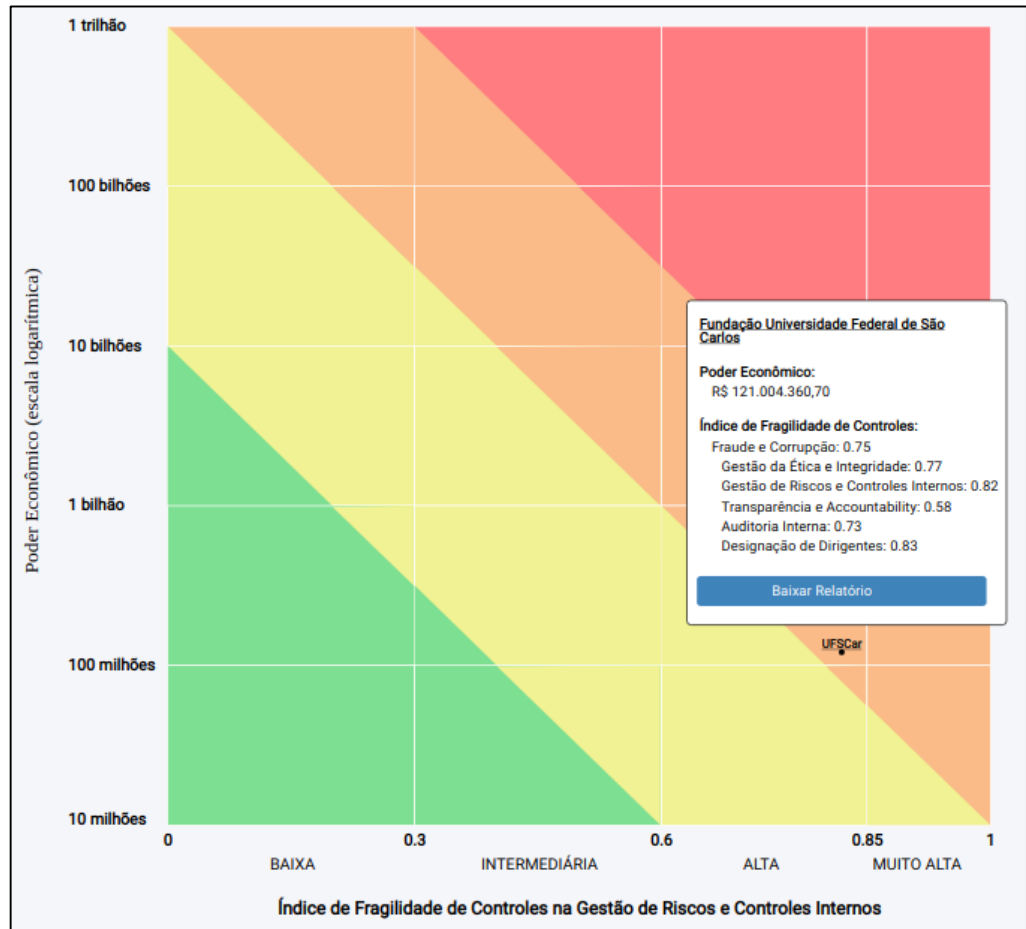


Fonte: VisPublica, busca realizada pelo autor.

- d. DataViva¹¹: é uma iniciativa do Governo de Minas Gerais e da Agência de Promoção de Investimento e Comércio Exterior de Minas Gerais (INDI), com apoio da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG). Constitui-se uma plataforma aberta de pesquisa construída com software livre que possibilita ao usuário o acesso a mais de 1 bilhão de visualizações com dados socioeconômicos dos mais de 5 mil municípios brasileiros, incluindo dados oficiais sobre exportações, atividade econômica, educação e ocupações de todo o Brasil (visualizadas por localidade), que tem por objetivo contribuir para a implementação de políticas públicas, possibilitar investimentos públicos e privados, e auxiliar na realização de estudos acadêmicos (DATAVIVA, 2019). Na Figura 19 podem ser vistos os cursos oferecidos pela UFSCar em 2017.

¹¹ DataViva - <<http://dataviva.info>>

Figura 20 - Mapa do TCU: Fragilidade de Controles X Poder Econômico (UFSCar).



Fonte: Mapa do TCU, busca realizada pelo autor.

3. MÉTODO E DESENVOLVIMENTO

3.1 Abordagem e tipologia da pesquisa

A pesquisa foi realizada no NIT-Materiais¹³ da Universidade Federal de São Carlos¹⁴, uma unidade do departamento de Engenharia de Materiais vinculada ao departamento de Ciência da Informação, além de ser reconhecido como um programa de extensão da UFSCar. Foi utilizado como procedimento o levantamento bibliográfico, com uma abordagem quali-quantitativa, sendo a UFSCar a unidade de caso delimitada para esta pesquisa.

O trabalho iniciou por meio da análise de pesquisas sobre a geração de indicadores de produção científica a partir de dados da Plataforma Lattes. A seguir, foi feita uma extração de dados da PL com o uso da ferramenta synclattes. Após, foram realizadas pesquisas na Web visando encontrar e estudar ferramentas de visualização de informação e também algumas aplicações que trabalham com indicadores, como a plataforma Somos UFSCar e os portais VisPublica e DataViva.

Foi desenvolvida uma interface interativa para a visualização da produção científica das ICTs brasileiras, de modo que os dados extraídos dos currículos Lattes dos pesquisadores sejam representados em forma de gráficos diversos. Nesse sentido, este projeto é uma pesquisa de natureza aplicada e exploratória (GIL, 2008), sendo que sua aplicação ocorre na unidade-caso UFSCar com o apoio do NIT-Materiais. A UFSCar foi fundada em 1968 e atualmente é formada pelos campi São Carlos, Araras, Sorocaba e Lagoa do Sino (em Buri), tratando-se de uma instituição pública de grande relevância no cenário nacional, única instituição federal de ensino superior localizada no interior do Estado de São Paulo, destacando-se pelo alto nível de qualificação de seu corpo docente, que em sua maioria desenvolve atividades de ensino, pesquisa e extensão em regime de dedicação exclusiva. A finalidade principal da UFSCar é a formação acadêmica, produção e disseminação de conhecimento e divulgação científica, tecnológica, cultural e artística.

¹³ O Núcleo de Informação Tecnológica em Materiais (NIT-Materiais) atua na pesquisa de prospecção tecnológica e inteligência competitiva, suas metodologias, ferramentas e aplicações para suporte ao desenvolvimento sustentável de empresas, arranjos empresariais e instituições públicas. <<http://www.nit.ufscar.br>>

¹⁴ Universidade Federal de São Carlos (UFSCar) - <<http://www.ufscar.br>>

3.2 Processo de desenvolvimento da pesquisa

Para alcançar os objetivos geral e específicos desta pesquisa seu desenvolvimento envolveu três macroatividades: 1] Revisão bibliográfica, visando a construção do embasamento teórico necessário para o desenvolvimento do projeto; 2] Pesquisa por componentes e recursos gráficos usados em sistemas Web; e 3] Criação de uma interface interativa para visualização de indicadores de produção científica. Como resultado, alguns indicadores bibliométricos da unidade-caso são apresentados a título de compreensão do potencial da ferramenta desenvolvida.

Este projeto tem ligação com o que foi desenvolvido por Matias (2015) e com o que foi feito por Maschietto (2019). A partir de arquivos no formato JSON que foram gerados por meio da ferramenta synclattes, foi implementado por Maschietto um procedimento para tratamento bibliométrico de dados. Para isso ele povoou uma base de dados não relacional e criou métodos parametrizados de busca que por sua vez geram outros arquivos JSON (Anexo 1 e Anexo 2).

As ligações entre os projetos de Matias (2015), Maschietto (2019) e o presente trabalho são descritas a seguir:

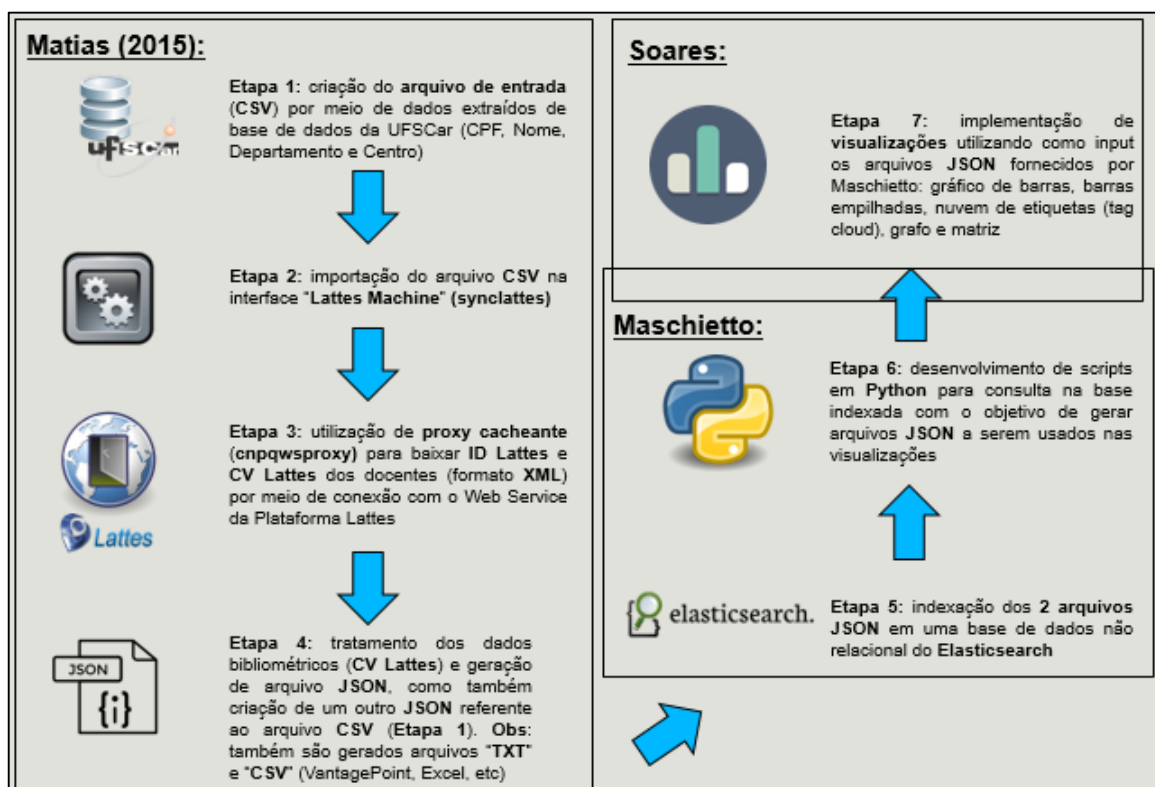
1. Criação de um arquivo de entrada (extensão CSV) por meio de dados extraídos de uma base de dados da UFSCar (CPF, Nome, Departamento e Centro). Esse processo é executado por Mesailde Souza de Oliveira Matias, servidora técnico-administrativa da Secretaria Geral de Informática (SIn) da UFSCar. Após, é realizada a importação do arquivo CSV na ferramenta synclattes por meio de sua interface Web chamada Lattes Machine;
2. O synclattes realiza o *download* dos ID Lattes e CVs dos docentes (formato XML) por meio de um proxy cacheante (cnpqwsproxy) que faz a conexão com o *Web Service* da Plataforma Lattes, em seguida, executa um tratamento dos dados bibliométricos e gera um arquivo no formato JSON. Também é criado um outro arquivo JSON com os dados do arquivo de entrada (CSV). Em relação à produção científica, o synclattes na versão atualmente disponível se limita a extrair dados referentes a artigos em eventos e em periódicos;
3. Maschietto realiza a indexação dos dois arquivos JSON em uma base de dados não relacional (Elasticsearch) e para realizar consultas e

extrair indicadores da base indexada ele executa os scripts que desenvolveu com a linguagem de programação Python. Esses scripts geram outros arquivos JSON (Anexo 1 e Anexo 2) que são enviados de maneira manual para serem utilizados como *input* na interface interativa, ou seja, os campos presentes nesses arquivos JSON foram previamente discutidos entre os discentes envolvidos no desenvolvimento do protótipo. Para facilitar o envio desses arquivos, optou-se por trabalhar com os 50 pesquisadores da UFSCar que mais publicam, ou seja, eles ficaram leves (tamanho físico em kilobytes). Futuramente esses arquivos JSON (Anexo 1 e Anexo 2) estarão disponíveis por intermédio de API's, dessa forma não haverá limites em relação a obtenção de pesquisadores;

4. Por meio do presente projeto foi desenvolvida uma interface interativa que disponibiliza visualizações para análise de indicadores bibliométricos, as quais utilizam como dados de entrada os arquivos JSON (Anexo 1 e Anexo 2) fornecidos por Maschietto.

O referido processo é descrito em sete etapas na Figura 21.

Figura 21 - Ligação entre as pesquisas.



Fonte: Elaborada pelo autor.

3.3 Tecnologias utilizadas na interface interativa

Para a implementação da interface interativa optou-se pela utilização de tecnologias gratuitas e bibliotecas gráficas de código *open source*, ou seja, que podem ser adaptadas para diferentes fins.

A respeito da parte técnica pertinente a criação da interface interativa, foram executados os seguintes procedimentos:

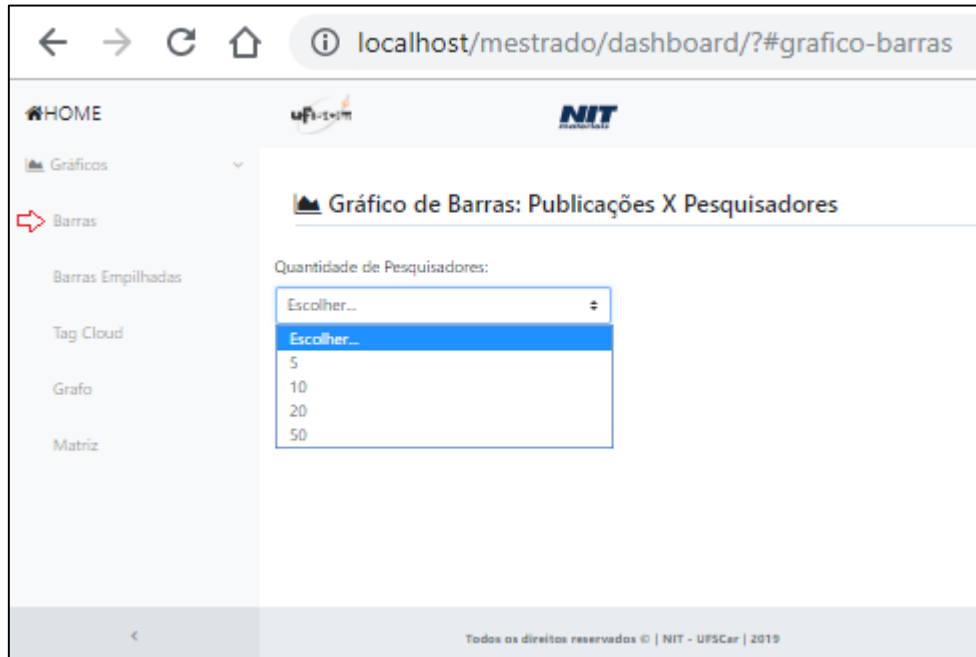
1. *Download* e customização de um *template* (“Dashboard”) construído com o *framework* Bootstrap¹⁵, utilizado para desenvolvimento de componentes de interface para aplicações Web por meio do uso das linguagens HTML, CSS e JavaScript, tendo por característica principal possibilitar uma melhor experiência do usuário em um site amigável e responsivo (que se adapta ao tamanho da tela que está sendo exibido);
2. Construção de uma arquitetura de software baseada no conceito SPA (*Single Page Application*), utilizando as linguagens de programação JavaScript (Frontend) e PHP (Backend), visando facilitar a integração de novos componentes gráficos;
3. Estudo de bibliotecas gráficas (coleções de subprogramas) e implementação de visualizações utilizando como *input* os arquivos JSON (Anexo 1 e Anexo 2) fornecidos por Maschietto (gráfico de barras, barras empilhadas, nuvem de etiquetas/tag cloud, grafo e matriz). Para a implementação do grafo e da matriz (ambos relacionados a coautoria), foi utilizada a D3 (*Data Driven Documents*), uma biblioteca JavaScript de código livre criada para manipulação de documentos com paradigma de programação centrado em dados. Essa tecnologia suporta vários conjuntos de dados e componentes dinâmicos de interação e animação, para isso ela utiliza HTML, CSS e SVG. Já para os gráficos de barras e barras empilhadas foi usada a biblioteca Google Charts, que além de gratuita, é simples de usar. A nuvem de etiquetas foi implementada com jQCloud, que por sua vez utiliza jQuery, a mais popular das bibliotecas JavaScript, que foi

¹⁵ O Bootstrap é um kit de ferramentas de código aberto para desenvolvimento com HTML, CSS e JS. <<https://getbootstrap.com>>

desenvolvida para simplificar os scripts interpretados nos navegadores Web.

A Figura 22 apresenta o layout da interface interativa.

Figura 22 - Layout da interface interativa.



Fonte: Elaborada pelo autor.

Como pode-se observar na Figura 22, com o *framework* Bootstrap foi criado um menu com o rótulo “Gráficos”, com 5 links: “Barras”, “Barras Empilhadas”, “Tag Cloud”, “Grafo” e “Matriz”. Cada um dos links redireciona para uma página com a respectiva visualização gráfica.

Na Figura 23 é apresentada uma parte do código-fonte de um programa escrito em linguagem JavaScript que possibilita a montagem do gráfico de barras, do gráfico de barras empilhadas e da nuvem de etiquetas. Esse programa pertence à camada de visão da interface interativa, também conhecida como “Frontend”.

Figura 23 - Código-fonte do programa de montagem de gráficos (Frontend).

```

1 $.get("controller/monta_json.php?tipoGrafico="+tipoGrafico+"&qtdeItens="+qtdeItens, function(dados) {
2     if (dados) {
3         switch (tipoGrafico) {
4             case 'barras':
5                 arrayValores = [];
6                 arrayValores.push(['PESQUISADOR', 'Qtde de artigos']);
7                 dados['valores'].forEach(function(objValores) {
8                     arrayValores.push([objValores['nome'], objValores['artigos_eventos'] + objValores['artigos_periodicos']]);
9                 });
10                google.charts.load('current', {'packages':['bar'], 'language': 'pt'});
11                google.charts.setOnLoadCallback(drawChartBarras);
12                break;
13            case 'barras-empilhadas':
14                arrayValores = [];
15                arrayValores.push(['PESQUISADOR', 'Qtde de artigos em periódicos', 'Qtde de artigos em eventos']);
16                dados['valores'].forEach(function(objValores) {
17                    arrayValores.push([objValores['nome'], objValores['artigos_periodicos'], objValores['artigos_eventos']]);
18                });
19                google.charts.load('current', {'packages':['corechart'], 'language': 'pt'});
20                google.charts.setOnLoadCallback(drawChartBarrasEmpilhadas);
21                break;
22            case 'tagcloud':
23                var words = [];
24                dados['valores'].forEach(function(objValores) {
25                    words.push({text: objValores['nome'], weight: objValores['artigos_eventos'] + objValores['artigos_periodicos'], autoResize: true,
26                        handlers: {
27                            click: function() { window.open("http://lattes.cnpq.br/"+objValores['idlattes']); },
28                            mouseover: function() { $(this).css('cursor','pointer'); }
29                        }
30                    });
31                });
32                $('#cloud').empty();
33                $('#cloud').jqCloud(words);
34            }
35        }
36    });

```

Fonte: Elaborada pelo autor.

Na primeira linha do código-fonte a quantidade de pesquisadores que foi selecionada pelo usuário na interface é passada por parâmetro via método “Get”/AJAX (JQUERY, 2018) para um programa escrito na linguagem de programação PHP, que é dedicado à camada de controle da aplicação (Backend). A montagem dos gráficos é realizada por meio da transformação de um JSON retornado pelo programa PHP (linha 2) em um arranjo estruturado de dados (linhas 8, 17 e 25) que armazena uma coleção de pesquisadores, sendo por sua vez enviado para as bibliotecas gráficas (Google Charts nas linhas 11 e 20; jqCloud/jQuery na linha 33), ao qual irão disponibilizar as visualizações dos indicadores. Na Figura 24 pode ser visto o código construído em PHP que é referente ao Backend da interface interativa.

Figura 24 - Código-fonte do programa de controle da aplicação (Backend).

```

1 <?php
2 $qtdeItens = ( isset($_GET['qtdeItens']) ? $_GET['qtdeItens'] : null );
3
4 if (empty($qtdeItens) || !is_numeric($qtdeItens) || $qtdeItens < 2) {
5     ob_end_clean();
6     return;
7 }
8
9 $arrRetorno = array();
10 $valores = array();
11 switch($_GET['tipoGrafico']) {
12     default:
13         foreach (json_decode(file_get_contents("publicacoes_article_conference.json"), true) as $key => $value) {
14             $valores[$key] = array('nome' => $value['nome'], 'artigos_eventos' => $value['artigos_eventos'],
15                 'artigos_periodicos' => $value['artigos_periodicos'], 'idlattes' => $value['idlattes']);
16             if ($key == $qtdeItens-1) {
17                 break;
18             }
19         }
20     }
21 $arrRetorno['valores'] = $valores;
22
23 ob_end_clean();
24 header('Content-type: application/json');
25 echo json_encode($arrRetorno);
26 return;

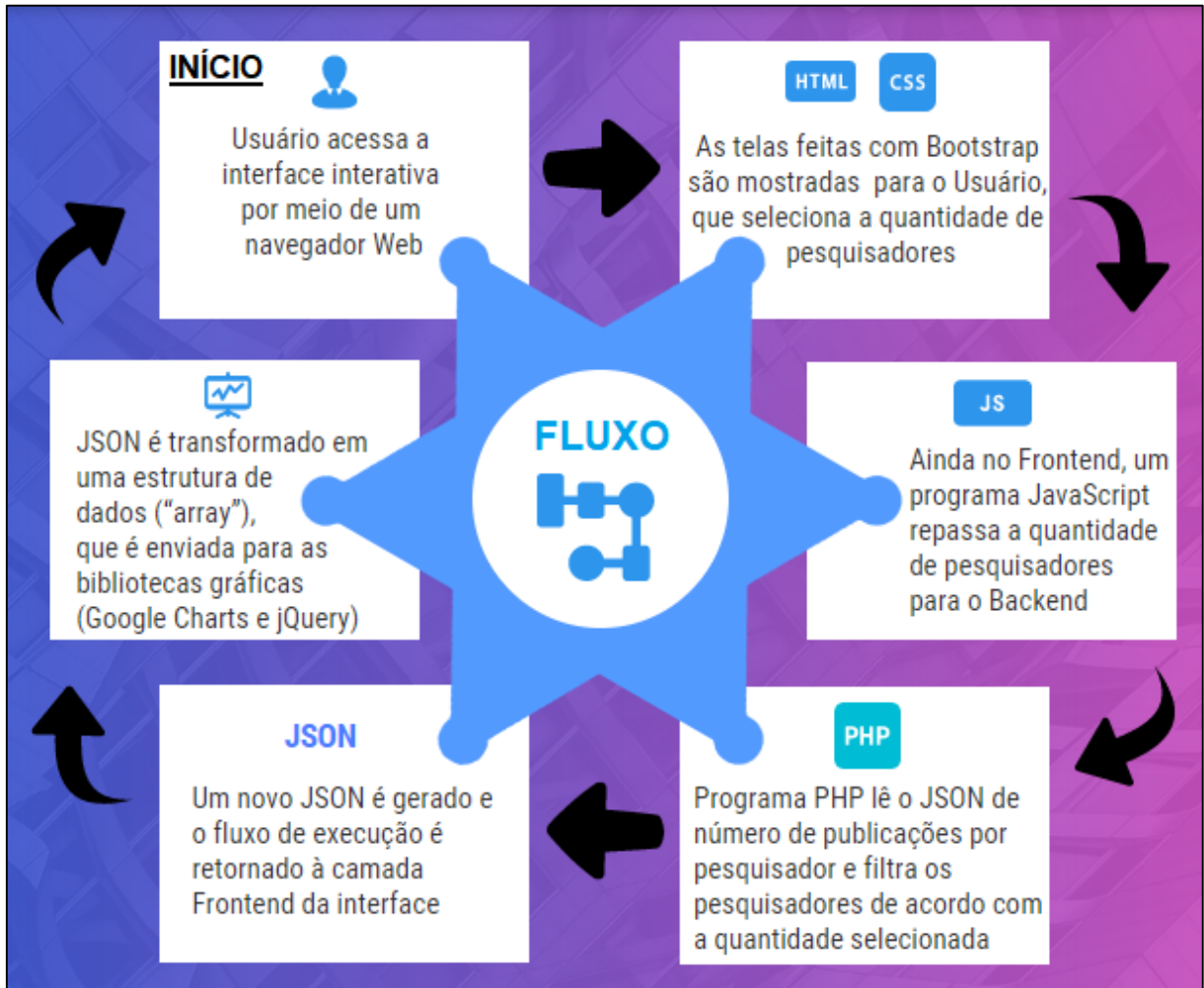
```

Fonte: Elaborada pelo autor.

O programa mostrado na Figura 24 é responsável por ler (linha 13) o arquivo JSON de número de publicações por pesquisador (Anexo 1) que foi criado e enviado por Maschietto, como também filtrar os pesquisadores de acordo com o valor que foi selecionado pelo usuário da aplicação (linha 16). Ele faz o recorte de dados do arquivo JSON original e gera um novo JSON em tempo real pelo software.

O fluxo de execução dos programas (Frontend e Backend) pode ser visto na Figura 25.

Figura 25 - Fluxo de execução dos programas (Frontend e Backend).



Fonte: Elaborada pelo autor.

O Quadro 4 apresenta as customizações e os aprimoramentos que foram realizados nos componentes gráficos utilizados para a implementação das visualizações (gráfico de barras, barras empilhadas, nuvem de etiquetas, grafo e matriz de colaboração), que foram disponibilizadas na interface interativa.

Quadro 4 - Customizações e aprimoramentos nas visualizações.

Visualização	Customização	Aprimoramento	Documentação
Gráfico de barras	Orientação: de coluna (vertical) para barra (horizontal)	-	https://developers.google.com/chart/interactive/docs/gallery/barchart
Barras empilhadas	Posicionamento da legenda; Tamanho da fonte do nome dos pesquisadores	-	https://developers.google.com/chart/interactive/docs/gallery/barchart
Nuvem de etiquetas	-	Link para o CV Lattes ao clicar no nome do pesquisador	https://mistic100.github.io/jQCloud/
Grafo	Espaçamento entre nós; Tamanho dos nós; Largura das arestas; Texto ao posicionar o mouse sobre os nós	Mostrar o número de coautoria ao posicionar o mouse sobre as arestas	https://github.com/evelinag/StarWars-social-network
Matriz	Alteração da categoria de cores dos departamentos; Retirada do temporizador das ordenações	Criação da ordenação por departamento; Atribuição da cor cinza na diagonal principal	https://bost.ocks.org/mike/miserables

Fonte: Elaborado pelo autor.

A escolha por cada uma das bibliotecas gráficas foi baseada de acordo com a complexidade das visualizações a serem implementadas, ou seja, para a construção de componentes gráficos mais simples foram utilizadas ferramentas mais populares (com vasta documentação). Isso vai ao encontro do que vem sendo divulgado no portal VisPublica, que apresenta ferramentas e técnicas de visualização de informação, além de mostrar a aplicação no contexto governamental, com exemplos de uso de diversas bibliotecas. De acordo com Autran et al. (2015), as plataformas de governo eletrônico (eGOV) possuem diversos objetivos, como auxiliar no processo de tomada de decisão, permitir que a gestão pública se torne mais transparente por meio da Lei de Acesso à Informação, elaborar planos e políticas governamentais e contribuir na construção de indicadores de ciência, tecnologia e inovação.

Diferentemente dos portais governamentais, o protótipo considera apenas os 50 pesquisadores da UFSCar que mais publicam artigos em eventos e em periódicos, o que representa aproximadamente 4% do total de professores ativos da universidade. A partir do momento que forem implementadas API's para recuperar todos os pesquisadores da instituição, a filtragem de pessoas a serem mostradas nas visualizações passará a ser realizada exclusivamente pela interface interativa.

4. RESULTADOS E DISCUSSÕES

Após realizar uma extração de dados da produção científica da UFSCar na Plataforma Lattes por meio da ferramenta automatizada synclattes, com base nos conceitos da área de visualização de informação foi desenvolvida uma interface para a representação visual dos indicadores de produção científica de ICTs.

A Figura 26 apresenta uma parte do arquivo JSON (Anexo 1) que é utilizado no desenvolvimento do gráfico de barras, barras empilhadas e da nuvem de etiquetas. Nela, encontram-se os principais pesquisadores da UFSCar em termos de total de publicação de artigos em eventos e em periódicos.

Figura 26 - Arquivo JSON de número de publicações por pesquisador.

```

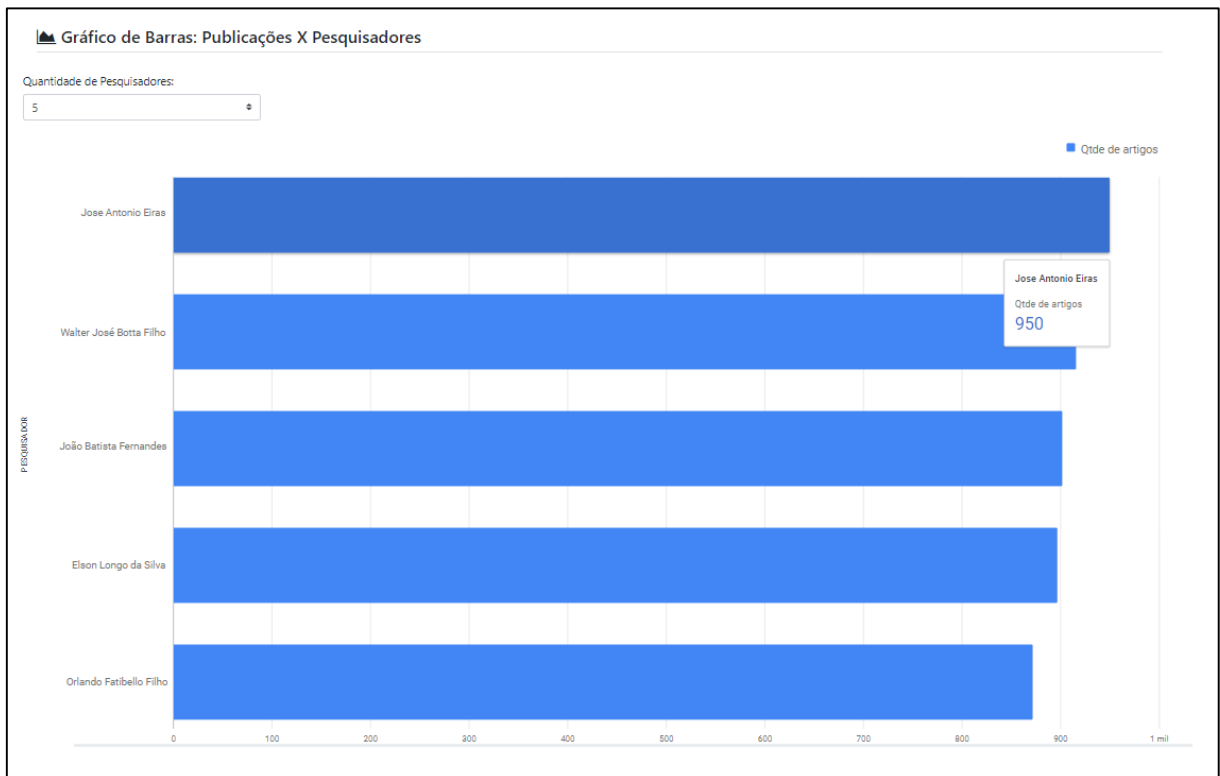
▼ 0:
  nome: "Jose Antonio Eiras"
  artigos_eventos: 581
  artigos_periodicos: 369
  idlattes: "6653223814715498"
▼ 1:
  nome: "Walter José Botta Filho"
  artigos_eventos: 545
  artigos_periodicos: 371
  idlattes: "8956458007749112"
▼ 2:
  nome: "João Batista Fernandes"
  artigos_eventos: 604
  artigos_periodicos: 298
  idlattes: "2691336077459602"
▼ 3:
  nome: "Elson Longo da Silva"
  artigos_eventos: 214
  artigos_periodicos: 683
  idlattes: "9848311210578810"
▼ 4:
  nome: "Orlando Fatibello Filho"
  artigos_eventos: 529
  artigos_periodicos: 343
  idlattes: "9859737944357808"

```

Fonte: Elaborada pelo autor.

Como pode ser visto na Figura 26, há no arquivo JSON os seguintes campos: “nome”, usado para representar o nome do pesquisador; “artigos_eventos”, que armazena o total de publicações em eventos; “artigos_periodicos”, que guarda o total de publicações em periódicos; e “idlattes”, utilizado para acessar o currículo Lattes do pesquisador (link para a PL). Na Figura 27 é apresentado o gráfico de barras criado a partir desse JSON.

Figura 27 - Gráfico de barras: quantidade de publicações por pesquisadores.



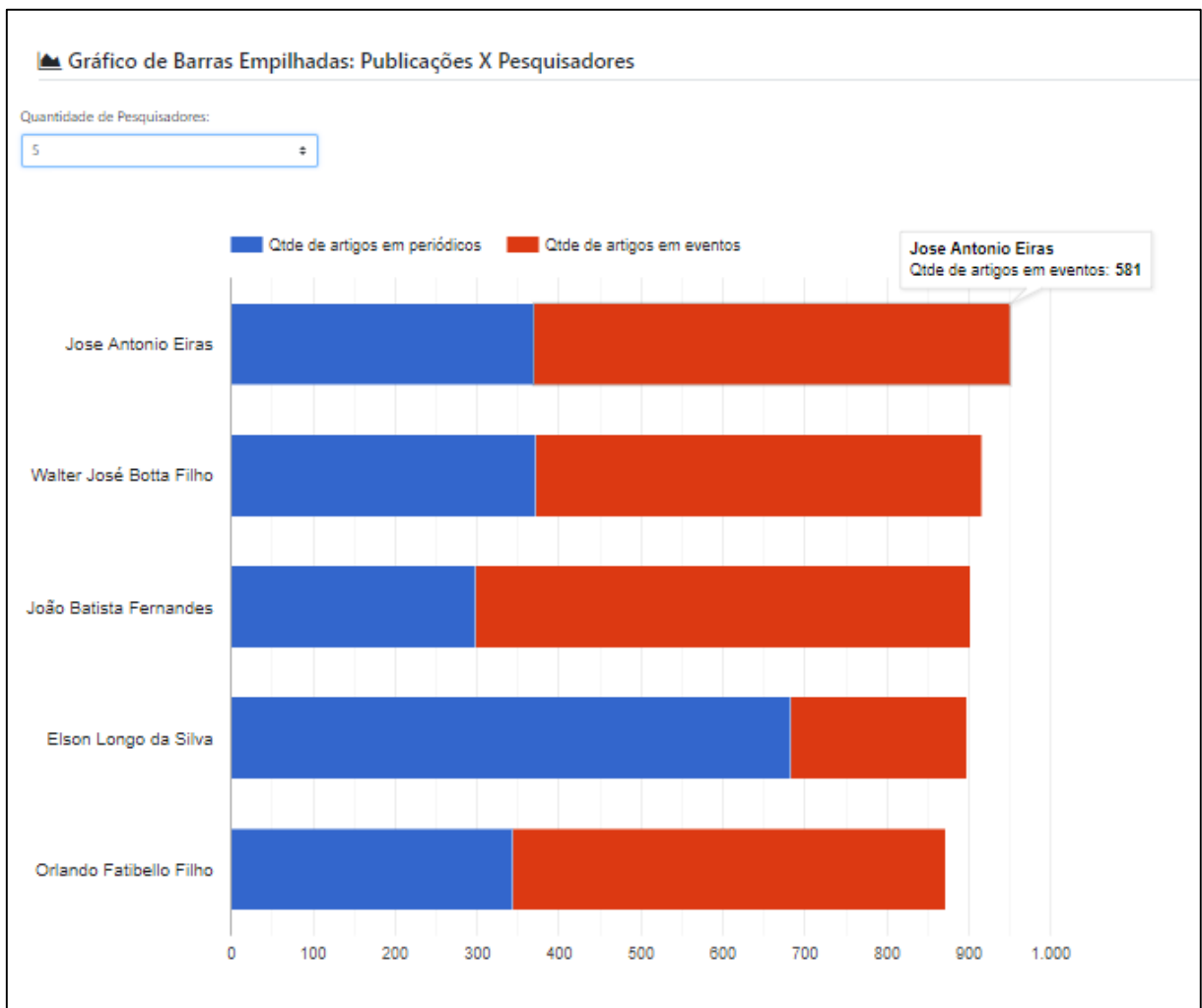
Fonte: Elaborada pelo autor.

O gráfico de barras é uma visualização que representa uma série de dados por meio do uso de retângulos, apresentados horizontalmente (em barras), em que os retângulos possuem a mesma altura e seu comprimento é proporcional ao dado representado. As categorias são dispostas ao longo do eixo vertical (y) e os valores, ao longo do eixo horizontal (x). No gráfico da Figura 27, os pesquisadores se encontram no eixo y e o número de publicações está no eixo x.

A biblioteca gráfica utilizada para a construção do gráfico de barras segue as orientações de Wong (2010) sobre simplificação visual (visualização convidativa), ou seja, procura diminuir a distração e concentrar o foco do usuário no essencial, que são os dados.

Na Figura 28 é mostrado o gráfico de barras empilhadas, contendo duas variáveis, pesquisador e tipo de publicação. Para artigos em eventos foi utilizada a cor vermelha e para artigos em periódicos a cor azul. Essa visualização também é gerada a partir do arquivo JSON de número de publicações por pesquisador (Anexo 1).

Figura 28 - Gráfico de barras empilhadas: artigos em periódicos e artigos em eventos.



Fonte: Elaborada pelo autor.

Como pode ser visto no gráfico de barras empilhadas, dos cinco pesquisadores que mais produzem, quatro deles possuem mais publicações de artigos em eventos do que publicações de artigos em periódicos. Autran et al. (2015) argumenta que o artigo de periódico é o principal instrumento de comunicação científica, porém Meadows (1999) apud Autran et al. (2015) aponta que as

comunicações apresentadas em eventos são mais comuns na área das engenharias, em que os trabalhos apresentados são avaliados pelos pares e dessa forma têm importância semelhante aos artigos de periódicos.

Em relação às boas práticas de visualização de informação defendidas por Wong (2010), as legendas estão dispostas na mesma ordem em que aparecem as informações nas barras, ou seja, primeiro os artigos em periódicos e depois os artigos em eventos. Ainda sobre as legendas, elas estão posicionadas na parte de cima do gráfico, que é outro requisito importante em uma visualização gráfica.

A Figura 29 apresenta a nuvem de etiquetas, mais uma visualização desenvolvida a partir do JSON de número de publicações por pesquisador (Anexo 1), onde para cada pesquisador foi realizada a soma entre artigos publicados em eventos e artigos publicados em periódicos.

Figura 29 - Nuvem de etiquetas: pesquisadores com mais publicações.



Fonte: Elaborada pelo autor.

Na nuvem de etiquetas o tamanho da fonte é variável. No caso, pesquisadores com mais publicações aparecem com o tamanho da fonte maior. Vieira e Correa (2011) apontam que as etiquetas facilitam a navegação e a seleção de documentos relacionados, ou seja, ao clicar em uma palavra outro site pode ser acessado, assim sendo, optou-se por abrir o currículo Lattes do pesquisador, o que corrobora com a argumentação de Wong (2010), ao qual considera a associação do conteúdo do gráfico com componentes do cotidiano como um dos elementos essenciais das boas visualizações gráficas (execução sofisticada).

É possível notar que o indicador referente ao total de publicações por autor é melhor visualizado por meio do gráfico de barras, uma vez que com esse tipo de componente gráfico é possível ver a quantidade exata de publicações de cada

pesquisador, como também realizar a comparação entre cada uma das barras (número de publicações do autor) de forma rápida.

A Figuras 30 e 31 apresentam partes do arquivo JSON de coautoria entre pesquisadores (Anexo 2) que foi utilizado para a criação do grafo e da matriz de colaboração.

Figura 30 - Arquivo JSON de coautoria entre pesquisadores (“nós”).

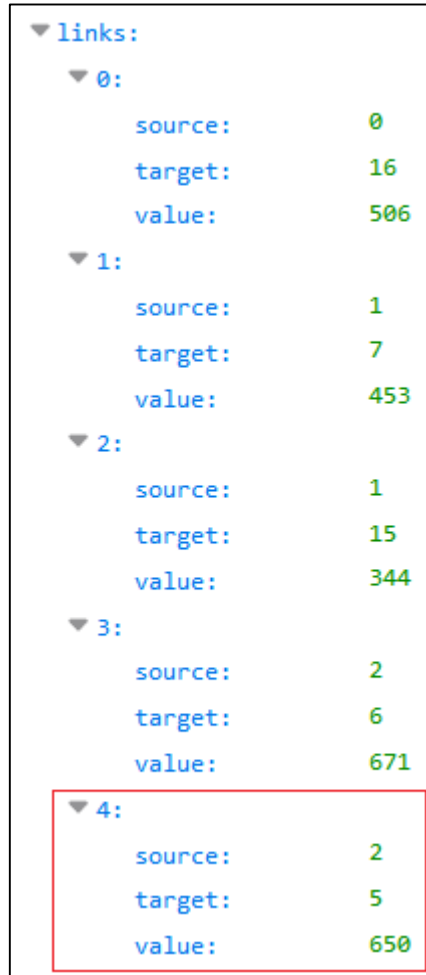
```
▼ nodes:
  ▼ 0:
    nome: "Jose Antonio Eiras"
    qtde_artigos: 950
    idlattes: "6653223814715498"
    depto: "Departamento de Física"
  ▼ 1:
    nome: "Walter José Botta Filho"
    qtde_artigos: 916
    idlattes: "8956458007749112"
    depto: "Departamento de Engenharia de Materiais"
  ▼ 2:
    nome: "João Batista Fernandes"
    qtde_artigos: 902
    idlattes: "2691336077459602"
    depto: "Departamento de Química"
  ▼ 3:
    nome: "Elson Longo da Silva"
    qtde_artigos: 897
    idlattes: "9848311210578810"
    depto: "Departamento de Química"
  ▼ 4:
    nome: "Orlando Fatibello Filho"
    qtde_artigos: 872
    idlattes: "9859737944357808"
    depto: "Departamento de Química"
  ▼ 5:
    nome: "Paulo Cezar Vieira"
    qtde_artigos: 785
    idlattes: "8923844563320377"
    depto: "Departamento de Química"
```

Fonte: Elaborada pelo autor.

Pode-se observar na Figura 30 que há no arquivo JSON o campo “qtde_artigos”, que armazena a quantidade de artigos publicados por pesquisador

(somatório de artigos em eventos e artigos em periódicos). A continuação desse arquivo JSON é apresentada na Figura 31.

Figura 31 - Arquivo JSON de coautoria entre pesquisadores (“arestas”).



Fonte: Elaborada pelo autor.

Por meio da Figura 31 é possível observar a relação entre os pesquisadores, por exemplo, no índice “4” há uma ligação representada por “source = 2”, e “target = 5”, ou seja, entre os autores “João Batista Fernandes” e “Paulo Cezar Vieira” há 650 artigos em coautoria.

Van Eck et al. (2010) argumentam que nos campos da bibliometria e da cienciometria a ideia da construção de mapas científicos (ou redes bibliométricas) baseados em dados bibliográficos intrigou pesquisadores por várias décadas. Diversos tipos de mapas foram estudados, os quais mostram relações entre, por exemplo, autores, documentos, periódicos, ou palavras-chave, e eles geralmente têm sido construídos com base em citação, cocitação ou acoplamento bibliográfico,

como também com base em dados de coocorrências de elementos bibliométricos (palavras-chave, nome, etc.) em documentos.

Para Franco, Sarvo e Faria (2018), nos dias de hoje o arranjo organizacional está estruturado em redes, um conjunto de pontos ou nós contendo atributos próprios que se ligam a outros pontos por meio de linhas ou links, sendo que quanto mais espessas forem as linhas, maior é a qualidade da relação entre os nós. Fontes e Stelzig (2004) argumentam que rede é um conceito utilizado para a compreensão dos processos estruturadores da sociedade de forma micro e macro, uma vez que possibilita a identificação tanto de ações individuais, como também permite o estudo de comportamentos em um contexto estrutural amplo.

Boutin et al. (1996) apontam que a colaboração científica é evidenciada por meio da análise de coautoria existente entre pesquisadores, instituições e países. De acordo com Alves, Yanasse e Soma (2011), sem que se tenha auxílio de desenvolvedores de software experientes, não é uma tarefa trivial descobrir os relacionamentos que existem entre os pesquisadores. Porém, a interface interativa desenvolvida neste trabalho permite a visualização de indicadores sem o apoio de profissionais de Tecnologia da Informação, como também não é exigido conhecimento prévio sobre o uso de softwares bibliométricos e estatísticos.

Na Figura 32 é mostrado o grafo de colaboração entre os 50 pesquisadores da UFSCar que mais publicam. Para gerá-lo, foi utilizado o arquivo JSON de coautoria entre pesquisadores (Anexo 2).

Figura 32 - Grafo de colaboração entre pesquisadores.



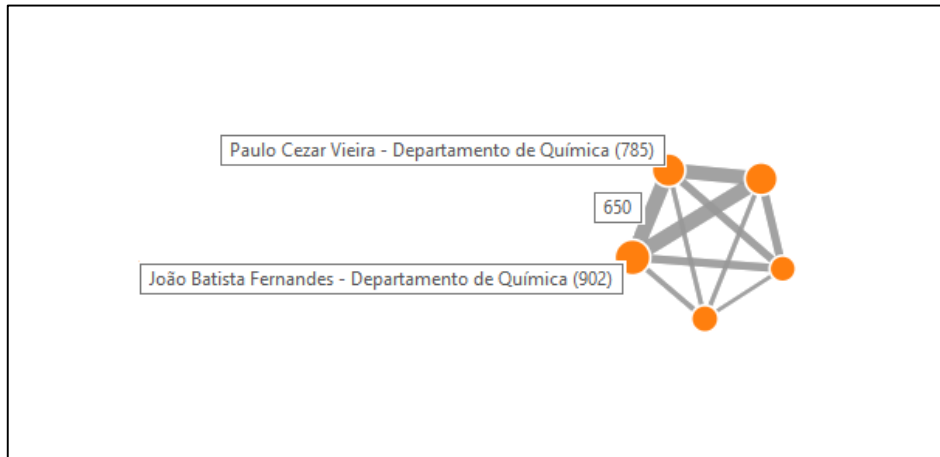
Fonte: Elaborada pelo autor.

Um grafo é uma estrutura abstrata de dados que possibilita a representação de diferentes elementos, como vértices e arestas, e seus respectivos relacionamentos (ROSSI; FREIRE; MENA-CHALCO, 2017). No presente trabalho, a palavra “nó” é utilizada como sinônimo de “vértice”. No grafo de colaboração mostrado na Figura 32, os nós representam pesquisadores e as arestas indicam a relação entre os pares de nós. O tamanho do nó varia de acordo com a quantidade de publicações do pesquisador e a espessura da aresta indica a força da coautoria entre os pesquisadores. Por sua vez, a cor do nó representa o departamento do pesquisador (por ex., Departamento de Química é representado pela cor laranja).

Na Figura 33 optou-se por destacar uma sub-rede do grafo que foi apresentado anteriormente, de modo a explicar a interação proporcionada pela biblioteca gráfica D3. Pelo navegador Web, ao posicionar o mouse em cima de um nó são mostrados o nome e o departamento do pesquisador, bem como a quantidade de publicações do pesquisador (entre parênteses). Já ao colocar o

mouse sobre uma aresta, é apresentada a quantidade de publicações em coautoria entre dois pesquisadores. Essas funcionalidades interativas são baseadas no conceito de conteúdo rico (WONG, 2010).

Figura 33 - Sub-rede de colaboração entre pesquisadores.



Fonte: Elaborada pelo autor.

A Figura 34 mostra uma parte do código-fonte do programa responsável pela exibição do grafo de colaboração.

Figura 34 - Código-fonte do programa de exibição do grafo de colaboração.

```

1  d3.json("controller/coautoria_pesquisadores.json", function(error, graph) {
2    if (error) throw error;
3    force
4      .nodes(graph.nodes)
5      .links(graph.links)
6      .start();
7
8    var link = svg.selectAll(".link")
9      .data(graph.links)
10     .enter().append("line")
11     .attr("class", "link")
12     .style("stroke-width", function(d) { return Math.sqrt(d.value)*1.1; });
13    link.append("title")
14     .text(function(d) { return d.value; });
15
16    var node = svg.selectAll(".node")
17     .data(graph.nodes)
18     .enter().append("circle")
19     .attr("class", "node")
20     .style("fill", function (d) { return color(d.depto); })
21     .attr("r", function (d) { return 1.1*Math.sqrt(d.qtde_artigos) + 1.5; })
22     .call(force.drag);
23    node.append("title")
24     .text(function(d) { return d.nome + ' - ' + d.depto + ' (' + d.qtde_artigos + ')'; });
25
26    force.on("tick", function() {
27      link.attr("x1", function(d) { return d.source.x; })
28        .attr("y1", function(d) { return d.source.y; })
29        .attr("x2", function(d) { return d.target.x; })
30        .attr("y2", function (d) { return d.target.y; })
31        .attr("value", function (d) { return d.value; });
32      node.attr("cx", function(d) { return d.x; })
33        .attr("cy", function(d) { return d.y; });
34    });
35  });

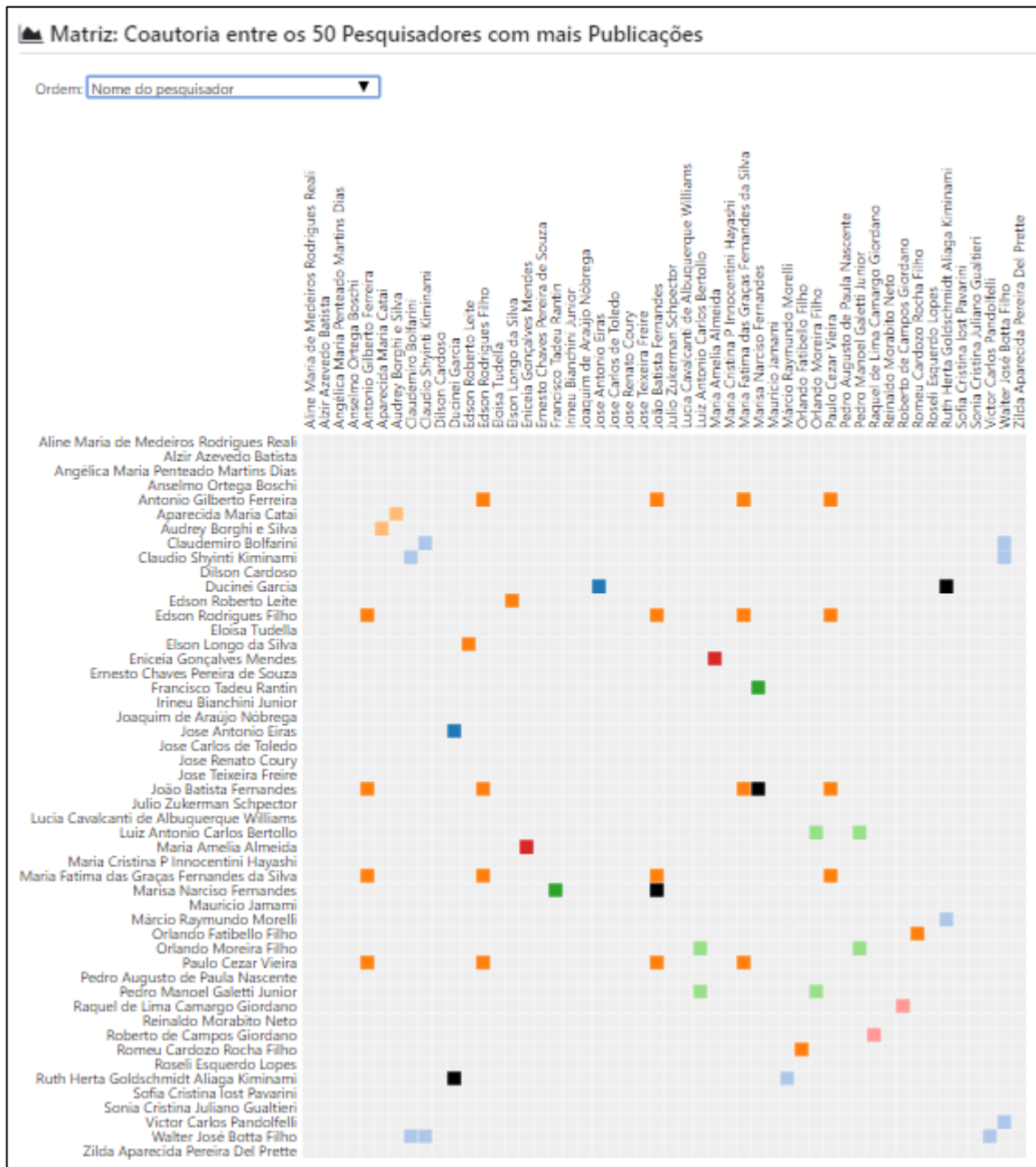
```

Fonte: Elaborada pelo autor.

Pode ser visto no código-fonte do grafo de colaboração que o programa inicia com a leitura do arquivo JSON de coautoria entre pesquisadores (Anexo 2). Em seguida são criados links/arestas (linhas 8 a 12) e nós/pesquisadores (linhas 16 a 22). O trecho de código das linhas 13 a 14 é responsável por mostrar o número de artigos em coautoria ao colocar o mouse sobre uma aresta. Já nas linhas 23 e 24 é definido o que será mostrado na tela ao colocar o mouse em cima de um nó.

Herman, Melançon e Marshall (2000) argumentam que é mais fácil analisar um grafo quando o volume de dados é pequeno. Neste trabalho, optou-se por disponibilizar mais de uma visualização para representar a colaboração entre os 50 pesquisadores da UFSCar que mais publicam. No caso, além do grafo, foi implementada uma matriz que também utiliza o arquivo JSON de coautoria entre pesquisadores (Anexo 2), ao qual é apresentada nas Figuras 35, 36, 37 e 38.

Figura 35 - Matriz de colaboração entre pesquisadores - ordenada por nome.

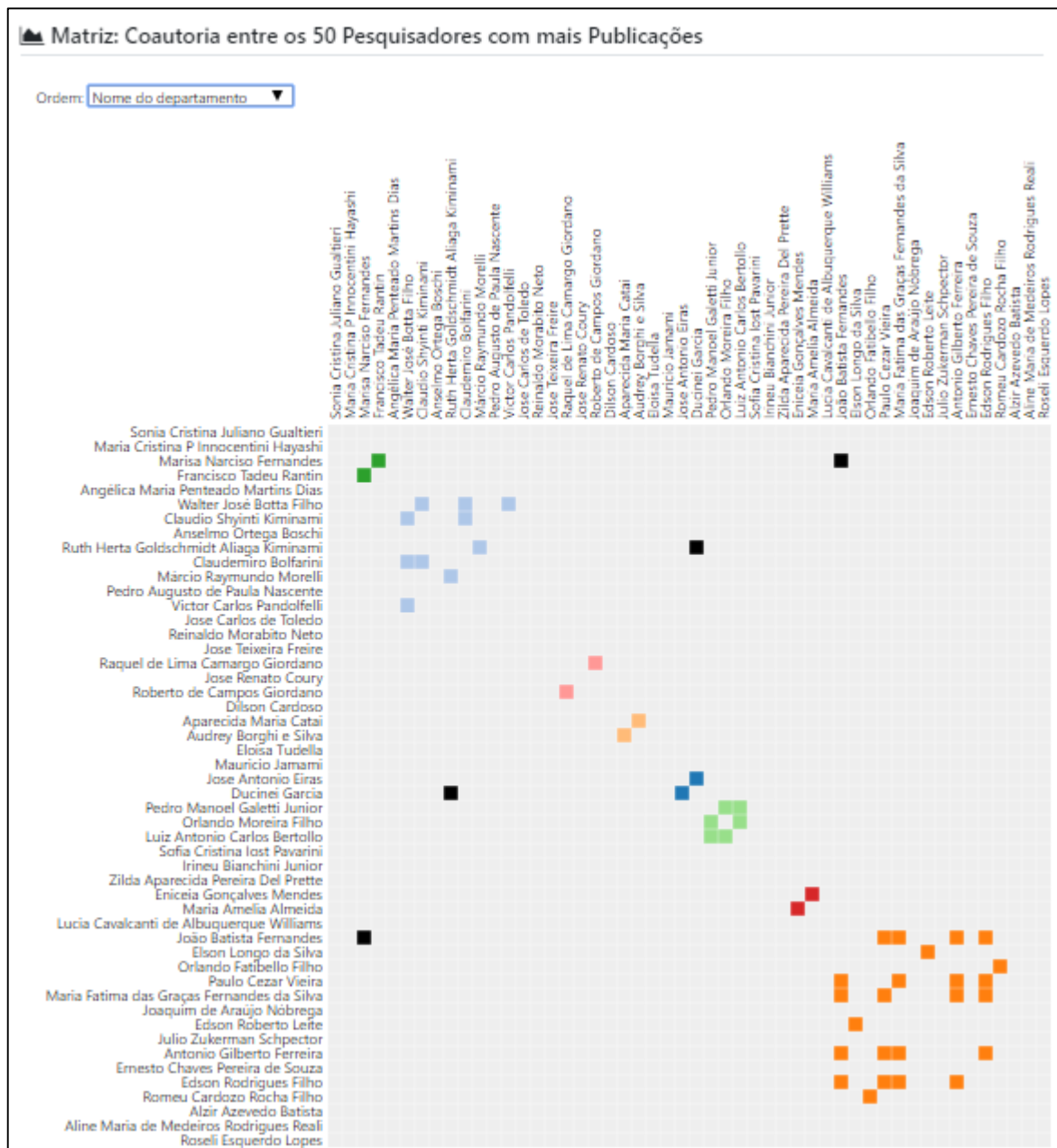


Fonte: Elaborada pelo autor.

Os pesquisadores são mostrados nas linhas e colunas e as interseções entre eles nas células da matriz. Assim como no grafo, na matriz de colaboração as cores representam os departamentos. Quando pesquisadores de departamentos diferentes possuem relação (publicação em coautoria), a célula de intersecção entre eles é preenchida pela cor preta (por padrão). É importante ressaltar que as cores utilizadas no grafo não têm relação com as cores que são usadas na matriz, ou seja, um determinado departamento pode ter uma cor no grafo e outra cor na matriz.

Caso haja a necessidade de trabalhar com mais do que 50 pesquisadores, torna-se necessário customizar o componente utilizado para a geração da matriz, como por ex., aumentando a largura e a altura da matriz.

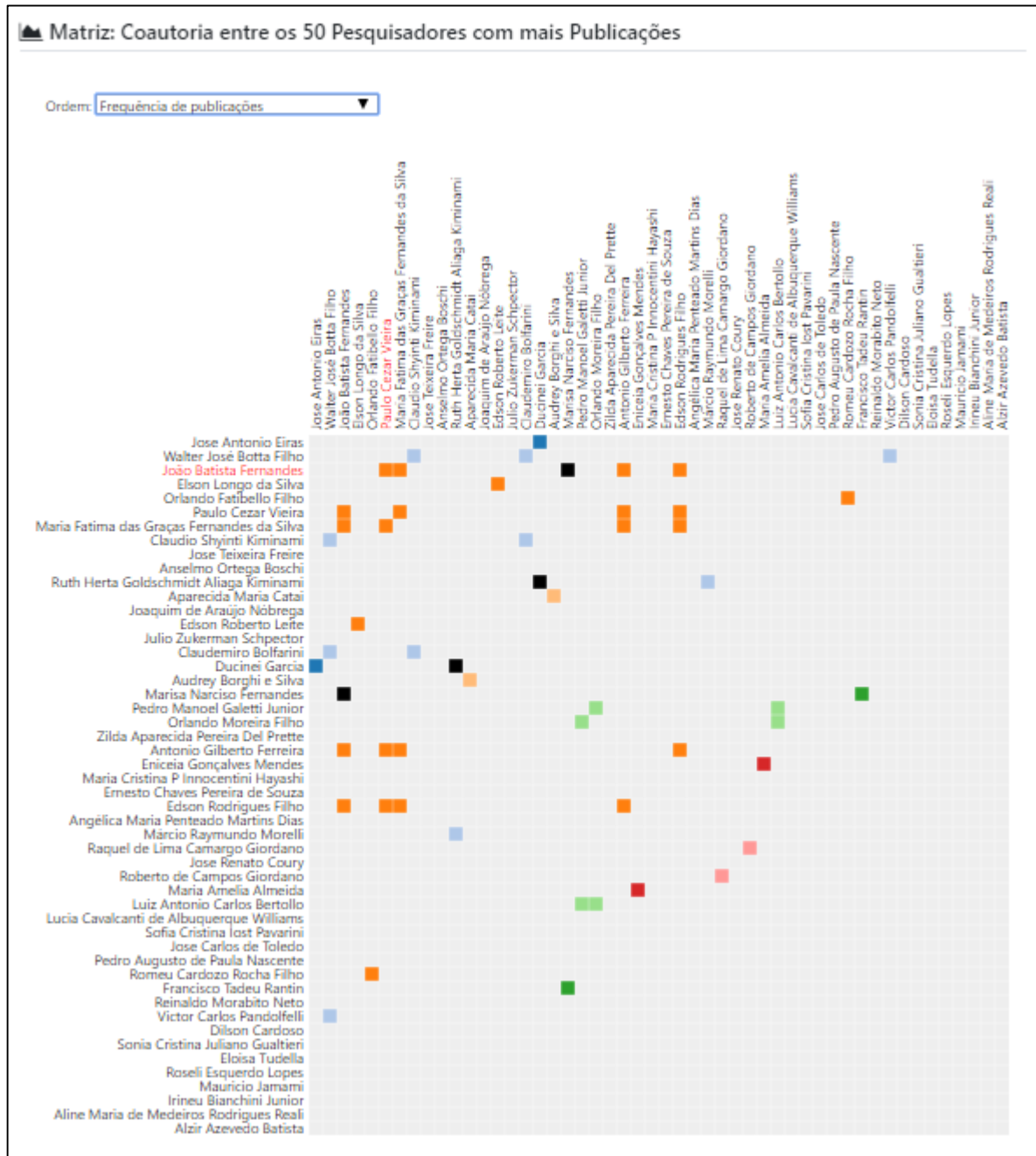
Figura 36 - Matriz de colaboração entre pesquisadores - ordenada por departamento.



Fonte: Elaborada pelo autor.

A visualização mostrada na Figura 36 destaca as colaborações interdepartamentais. Contudo, com matriz é possível representar outros níveis de afiliação, por exemplo, por instituição, país, grupo de pesquisa, etc.

Figura 37 - Matriz de colaboração entre pesquisadores - ordenada por frequência de publicações.

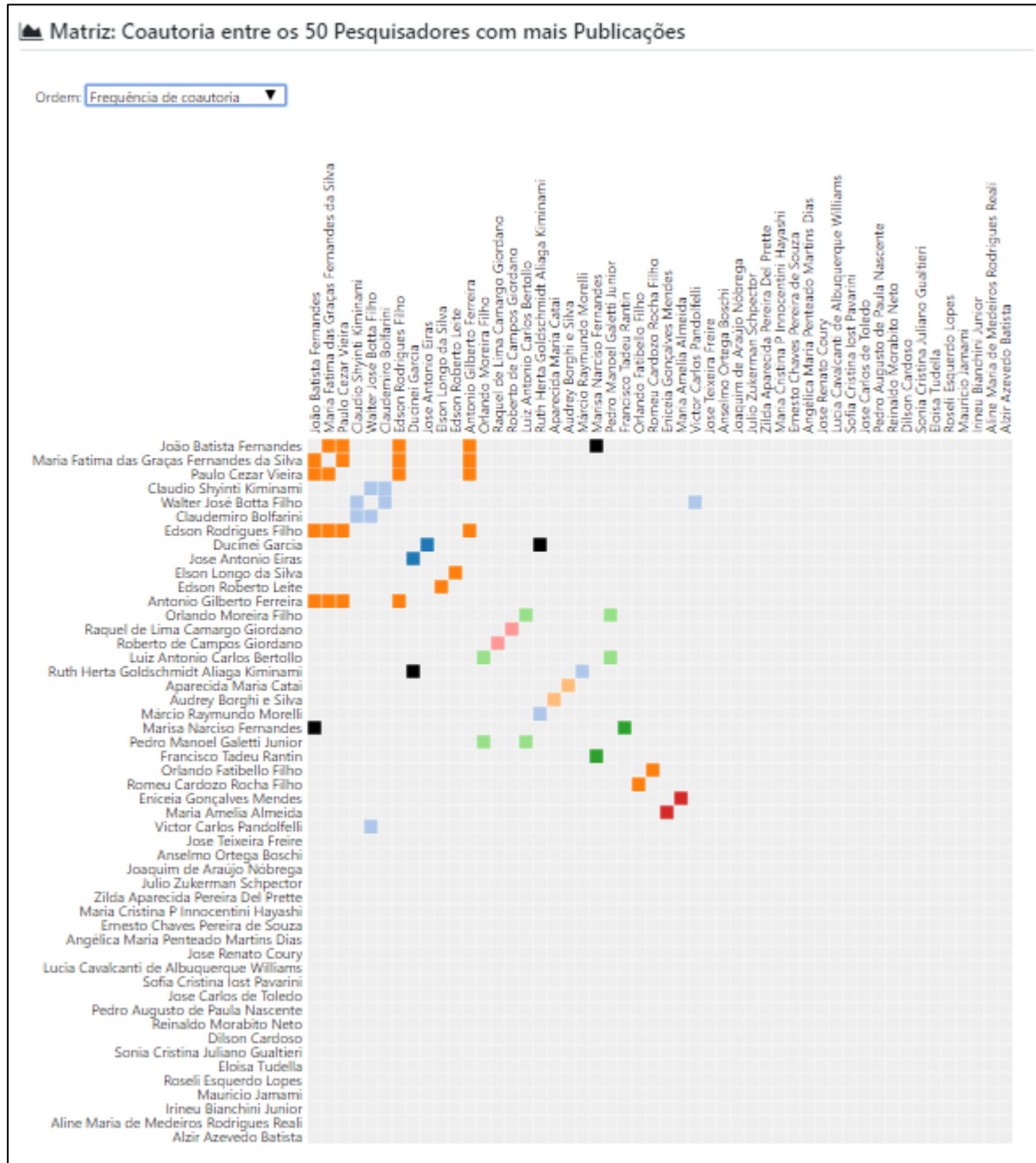


Fonte: Elaborada pelo autor.

Como pode ser visto na Figura 37, ao colocar o mouse em cima de uma célula de intersecção entre os pesquisadores, os nomes dos autores são destacados (linha e coluna ficam na cor vermelha), esse recurso é similar ao que foi apontado por Wong (2010) para legendas, quando o autor argumenta que uma legenda longe de um dado exige um trabalho extra dos leitores para a interpretação da informação

disponível no gráfico, ou seja, esse cuidado foi tomado pela biblioteca gráfica que foi utilizada no desenvolvimento da presente visualização.

Figura 38 - Matriz de colaboração entre pesquisadores - ordenada por frequência de coautoria.



Fonte: Elaborada pelo autor.

Unwin (2008) ressalta que a ordenação e a organização da informação interferem na leitura de um gráfico. Para o presente trabalho, optou-se por disponibilizar quatro formas de ordenação da matriz:

- 1) Por ordem alfabética do nome do pesquisador (Figura 35), que serve para encontrar um determinado autor mais facilmente;
- 2) Por ordem alfabética do nome do departamento (Figura 36), sendo que a ordem dos pesquisadores de um mesmo departamento ocorre de acordo com aparecimento do autor no arquivo JSON (no caso, está ordenado pelo número de publicações, quem publica mais aparece primeiro). Essa ordenação é interessante pois agrupa pesquisadores que são de um mesmo departamento;
- 3) Por frequência de publicações (Figura 37), em que a ordenação é feita considerando a quantidade de publicações dos pesquisadores, o que torna fácil localizar os autores que mais publicam;
- 4) Por frequência de coautoria (Figura 38), ou seja, para cada um dos 50 pesquisadores que mais publicam na instituição, é realizada a soma das coautorias que ele teve com os outros 49 pesquisadores que mais publicam. Essa ordenação ajuda a encontrar os autores que mais colaboram.

Na Figura 39 são apresentados trechos do código-fonte do programa referente à montagem da matriz de colaboração.

Figura 39 - Código-fonte do programa de montagem da matriz de colaboração.

```

1  d3.json("controller/coautoria_pesquisadores.json", function(coautoria) {
2  var matrix = [],
3  nodes = coautoria.nodes,
4  n = nodes.length;
5
6  nodes.forEach(function(node, i) {
7  node.index = i;
8  node.frequency_coauthorship = 0;
9  matrix[i] = d3.range(n).map(function(j) { return {x: j, y: i, z: 0}; });
10 });
11
12 coautoria.links.forEach(function(link) {
13 matrix[link.source][link.target].z += link.value;
14 matrix[link.target][link.source].z += link.value;
15 matrix[link.source][link.source].z += link.value;
16 matrix[link.target][link.target].z += link.value;
17 nodes[link.source].frequency_coauthorship += link.value;
18 nodes[link.target].frequency_coauthorship += link.value;
19 });
20
21 var orders = {
22 name: d3.range(n).sort(function(a, b) { return d3.ascending(nodes[a].nome.toLowerCase(), nodes[b].nome.toLowerCase()); }),
23 departament: d3.range(n).sort(function(a, b) { return d3.ascending(nodes[a].depto.toLowerCase(), nodes[b].depto.toLowerCase()); }),
24 frequency_publications: d3.range(n).sort(function(a, b) { return nodes[b].qtde_artigos - nodes[a].qtde_artigos; }),
25 frequency_coauthorship: d3.range(n).sort(function(a, b) { return nodes[b].frequency_coauthorship - nodes[a].frequency_coauthorship; });
26 };
27
28 function row(row) {
29 var cell = d3.select(this).selectAll("cell")
30 .data(row.filter(function(d) { return d.z; })))
31 .enter().append("rect")
32 .attr("class", "cell")
33 .attr("x", function(d) { return x(d.x); })
34 .attr("width", x.rangeBand())
35 .attr("height", x.rangeBand())
36 .style("fill-opacity", function(d) { return z(d.z); })
37 .style("fill", function(d) {
38 var color;
39 if (nodes[d.x].depto == nodes[d.y].depto) {
40 if (d.x == d.y) { //diagonal principal
41 color = "#EEE"; //cinza claro
42 } else {
43 color = c(nodes[d.x].depto);
44 }
45 } else {
46 color = null;
47 }
48 return color;
49 })
50 .on("mouseover", mouseover)
51 .on("mouseout", mouseout);
52 }

```

Fonte: Elaborada pelo autor.

Como pode-se ver no código da Figura 39, o programa é dividido em quatro blocos principais, no qual o primeiro (linhas 6 a 10) é referente à montagem dos pesquisadores, o segundo (linhas 12 a 19) é responsável pela criação dos links entre os pesquisadores (coautorias), o terceiro (linhas 21 a 26) provê os quatro métodos de ordenação da matriz de colaboração, e o quarto (linhas 28 a 52) é relacionado ao visual do gráfico, sendo que nas linhas 40 e 41 está presente o código encarregado de atribuir a cor cinza nas células da diagonal principal da matriz (um autor X ele mesmo).

Por fim, os programas desenvolvidos neste trabalho podem ser encontrados na página da organização NIT-Materiais no GitHub¹⁶.

¹⁶ GitHub do NIT-Materiais/UFSCar - <<https://github.com/nitmateriais/dashboard>>

5. CONSIDERAÇÕES FINAIS

Em um mundo em que cada vez mais as pessoas necessitam ter acesso à informação por meio de diferentes fontes, a Ciência da Informação torna-se essencial no processo de descoberta de novas formas de pensar e agir. Por sua vez, a interdisciplinaridade da CI possibilita a atuação de profissionais das mais diversas áreas do conhecimento, tanto teóricos quanto práticos, como por exemplo, profissionais da computação.

O processo de construção do conhecimento e da ciência está diretamente vinculado ao potencial de divulgação e publicação de resultados. Apesar da existência de diversos estudos sobre elaboração de indicadores bibliométricos, há a necessidade de se desenvolver soluções tecnológicas que possibilitem a visualização desses indicadores, que são essenciais para a compreensão dos impactos das políticas e das dinâmicas sociais envolvendo ciência e tecnologia.

A Plataforma Lattes apresenta-se como uma relevante fonte de dados para a elaboração de indicadores e este trabalho mostrou a viabilidade da implementação de uma interface interativa para a visualização de indicadores de produção científica das ICTs brasileiras a partir de dados dessa plataforma extraídos pela ferramenta automatizada syncclattes. Um desafio encontrado em softwares bibliométricos diz respeito a preparação dos dados que alimentarão a ferramenta para gerar a visualização, pois cada sistema aceita um padrão de entrada específico. O protótipo que foi desenvolvido não tem esse problema, uma vez que trabalha com a notação JSON. Além disso, ele poderá ser utilizado por usuários sem conhecimento prévio a respeito de softwares bibliométricos, o que constitui mais uma contribuição deste trabalho.

A Visualização de Informação é uma área multidisciplinar que pode ser utilizada em conjunto com outras técnicas de extração de conhecimento, porém, somente com a visualização é possível prover mecanismos suficientes para obter o conhecimento necessário. Para tanto, um gráfico não deve ter por objetivo simplesmente representar números de forma visual, ele precisa possibilitar análise e interpretação adequadas da informação que está sendo disponibilizada. Por sua vez, a familiaridade com os dados facilita o entendimento dos indicadores bibliométricos, possibilitando ao usuário formular e testar novas hipóteses que podem ser verificadas por meio das visualizações.

Como resultados, foi possível estabelecer um breve panorama sobre a produção científica da unidade-caso, sendo que a colaboração entre os 50 pesquisadores da UFSCar que mais produzem artigos em eventos e em periódicos é um indicador que permite a elaboração de diversos estudos. O uso de componentes gráficos como o grafo permitiu aplicar regras da visão humana, como a forma, o que traz mais sentido aos dados que estão sendo analisados, facilitando o entendimento da informação por parte do usuário da interface interativa. A interface interativa foi desenvolvida com base em análises de trabalhos acadêmicos e de portais que abordam os elementos essenciais de bons gráficos de informação, constituídos por conteúdo rico, execução sofisticada e visualização convidativa.

Por fim, considera-se que este projeto poderá colaborar para os estudos sobre otimização da visualização de indicadores de produção científica, que, por sua vez, são utilizados para a avaliação da ciência. Futuras pesquisas poderiam ampliar a amostra de pesquisadores e comparar os dados bibliométricos que foram obtidos no presente trabalho com resultados de publicações e também com outros softwares que trabalham com a Plataforma Lattes para a geração de indicadores científicos, indo ao encontro das boas práticas para construção de gráficos, mais especificamente à etapa de revisão de dados. Há também a necessidade de melhorias tanto na integração da interface interativa com o synclattes, quanto a respeito da ampliação de metadados extraídos dos currículos Lattes, o que permitiria a implementação de gráficos com conteúdos mais ricos.

REFERÊNCIAS

ALVES, A. D.; YANASSE, H. H.; SOMA, N. Y. SUCUPIRA: a System for Information Extraction of the Lattes Platform to Identify Academic Social Networks. **6th Iberian Conference on Information Systems and Technologies**, Chaves - Portugal, 2011. Disponível em: <https://www.researchgate.net/publication/224251537_SUCUPIRA_A_system_for_Information_extraction_of_the_Lattes_Platform_to_identify_academic_social_networks>. Acesso em: 10 jul. 2019.

ALVES, M. C. **Visualização de informação para simplificar o entendimento de indicadores sobre avaliação da ciência e tecnologia**. 2015. 120 f. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) - Universidade Federal de São Carlos, São Carlos, 2015.

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

ARAKAKI, F. A.; SIMIONATO, A. C.; SANTOS, P. L. V. A. da C. Catalogação e tecnologia: interseções com a Web Semântica. **Informação@Profissões**, Londrina, v. 6, n. 2, p. 03 - 19, jul./dez. 2017.

AUTRAN, M. M. M. et al. Perfil de produção acadêmica dos programas brasileiros de pós-graduação em Ciência da Informação 2008-2012. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 20, n. 4, p. 57-78, out./dez. 2015.

BAPTISTA, A. A.; MACHADO, A. B. Um gato preto num quarto escuro: falando sobre metadados. **Revista de Biblioteconomia de Brasília**, v. 25, n. 1, p. 77-90, 2001. Disponível em: <<http://www.brapci.inf.br/v/a/8514>>. Acesso em: 25 maio 2018.

BAPTISTA, D. M. O impacto dos metadados na representação descritiva. **Revista ACB: Biblioteconomia em Santa Catarina**, v. 12, n. 2, p. 177-190, 2007. Disponível em: <<http://www.brapci.inf.br/v/a/8026>>. Acesso em: 25 maio 2018.

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.

BORKO, H. **Information science: what is it?** American Documentation, v. 19, n. 1, p. 3-5, 1968.

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

BOUTIN, E. et al. Les reseaux comme outils d'analyse en bibliométrie. Un cas d'application: les reseaux d'auteurs. **Cahier de la Documentation**, n. 1, p. 3-13, mars. 1996.

BRINTON, W. C. **Graphic Presentation**. 1939. From the collection of the Prelinger Library, San Francisco, California, 2008.

BUCKLAND, M. **Information as thing**. JASIS, v. 42, n. 5, p. 351-260, 1991.

BUSH, V. **As we may think**. Atlantic Monthly, v. 176, n. 1, p.101-108, 1945. Disponível em: <<http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm>>. Acesso em: 26 nov. 2017.

CARD, S. K.; MACKINLAY, J. D.; SHNEIDERMAN, B. **Readings in information visualization: using vision to think**. São Francisco: Morgan Kaufmann, 1999. 686 p.

CARVALHO, L. A.; CRIPPA, G. Ciência da informação: histórico, delimitação do campo e a sua perspectiva sobre a área da comunicação. **Perspectivas em Ciência da Informação**, v. 18, n. 4, 2013. Disponível em: <<http://www.brapci.inf.br/v/a/14034>>. Acesso em: 02 dez. 2017.

CASTRO, R. C. F. Impacto da Internet no fluxo da comunicação científica em saúde. **Revista de Saúde Pública**, São Paulo, v. 40, n esp., p. 57-63, 2006. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-89102006000400009&lng=en&nrm=iso>. Acesso em: 31 maio 2019.

CONSELHO NACIONAL DE DESENVOLVIMENTO CIENTÍFICO E TECNOLÓGICO (CNPq). **Plataforma Lattes**. Brasília: CNPq, 2017. Disponível em: <<http://lattes.cnpq.br>>. Acesso em: 20 nov. 2017.

CTIT UFMG - COORDENADORIA DE TRANSFERÊNCIA E INOVAÇÃO TECNOLÓGICA DA UFMG. **Somos UFSCar**. 2019. Disponível em: <<http://somos.ufscar.br>>. Acesso em: 13 jul. 2019.

DATAVIVA. **Data Viva beta**. 2019. Disponível em: <<http://dataviva.info>>. Acesso em: 01 maio 2019.

DAVYT, A.; VELHO, L. **A avaliação da ciência e a revisão por pares: passado e presente. Como será o futuro?** História, ciência, saúde - Manguinhos, Rio de Janeiro, v. 7, n. 1, p. 93-116, mar./jun. 2000.

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 nov. 2017.

ENGWALL, L.; BLOCKMANS, W.; WEAIRE, D. Bibliometrics: issues and context. **Portland Press Limited**. 2014. Disponível em: <http://www.portlandpresspublishing.com/sites/default/files/Editorial/Wenner/WG_87/WG_87_chapter%201.pdf>. Acesso em: 10 jul. 2019.

ERDMANN, A. L. et al. A avaliação de periódicos científicos qualis e a produção brasileira de artigos da área de enfermagem. **Revista Latino-Americana de Enfermagem**, São Paulo, USP, v. 17, n. 3, jun. 2009.

FARIA, L. I. L. et al. Análise da produção científica a partir de publicações em periódicos especializados. In: BRENTANI, R. R.; CRUZ, C. H. B.; SUZIGAN, W.; FURTADO, J. E. M. P.; GARCIA, R. C. (Org.). **Indicadores de Ciência, Tecnologia e Inovação em São Paulo - 2010**. 1 ed. São Paulo: FAPESP, 2011, v. 1, p. 1-71.

FEW, S. **Information Dashboard Design: The Effective Visual Communication of Data**. O'Reilly, 2006.

FEW, S. **Simple Visualization Techniques for Quantitative Analysis**. California: Analytics Press, 2009.

FONTES, B. A. S. M.; STELZIG, S. Sobre trajetórias de sociabilidade: a idéia de ralé social enquanto mecanismo criador de novas redes sociais. **Política & Sociedade**, Florianópolis, v. 3, n. 5, 2004, p. 57-77.

FRANCO, N. M. G.; SARVO, D. O.; FARIA, L. I. L. ANÁLISE DE REDES POR COOCORRÊNCIA DE PALAVRAS-CHAVE: identificação de possíveis colaborações científicas. **6º Encontro Brasileiro de Bibliometria e Cientometria**, Rio de Janeiro, 2018. Disponível em: <https://ebbc.inf.br/ebbc6/docs/6EBBC2018v2018_07_27.pdf>. Acesso em: 10 jul. 2019.

FREIRE, G. H. A. Ciência da Informação: temática, histórias e fundamentos. **Perspectivas em Ciência da Informação**, v. 11, n.1, p. 6-19, 2006.

FREITAS, C. M. D. S. et al. Introdução à visualização de informações. **RITA**, Porto Alegre, v. 8, n. 2, p. 143-158, 2001.

GEOCAPES. **Sistema de Informações Georreferenciadas**. 2019. Disponível em: <<https://geocapes.capes.gov.br>>. Acesso em: 01 maio 2019.

GIL, A. C. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Atlas, 2008.

GREGOLIN, J. A. R. et al. Análise da produção científica a partir de indicadores bibliométricos. In: LANDI, F. R.; GUSMÃO, R. (Org.). **Indicadores de ciência, Tecnologia e Inovação em São Paulo-2004**. Fapesp, São Paulo, v. 1, p. 1-44, 2005.

HERMAN, I.; MELANÇON, G.; MARSHALL, M. S. Graph Visualization and Navigation in Information Visualization: A Survey. **IEEE Transactions on Visualization and Computer Graphics**, v. 06, n.1, p. 24-43, 2000.

HICKS, D. et al. The Leiden Manifesto for research metrics. **Nature**, v. 520, p. 429-431, 2015.

HILLMANN, D. **Using Dublin Core**. 2005. Disponível em: <<http://dublincore.org/documents/usageguide>>. Acesso em: 25 maio 2018.

JQUERY. **Category: Ajax.** 2018. Disponível em: <<http://api.jquery.com/category/ajax>>. Acesso em: 15 maio 2018.

KEIM, D. A. Information Visualization and Visual Data Mining. **IEEE Transactions on Visualization and Computer Graphics**, v. 7, n. 1, p. 100-107, 2002.

LANE, J. Let's make science metrics more scientific. **Nature**, v. 464, p. 488-489, 2010.

LIMA, R. A. **Análise bibliométrica da atividade científica em bioprospecção (1986 - 2006)**. 2007. Dissertação (Mestrado) - Instituto de Geociências, Universidade Estadual de Campinas, Campinas, 2007.

MACIEL, R. S. **A Plataforma Lattes como recurso estratégico para a gestão dos Programas de Pós-Graduação: uma análise baseada na produção de artigos científicos**. 2018. 183 p. Dissertação (Mestrado em Ciência da Informação) - Universidade Federal de São Carlos, São Carlos, 2018.

MAPA DO TCU. **Auditoria do Tribunal de Contas da União**. 2019. Disponível em: <<https://meapffc.apps.tcu.gov.br>>. Acesso em: 01 maio 2019.

MASCHIETTO, L. G. **Interface interativa para visualização de indicadores de produção científica a partir de dados da Plataforma Lattes: procedimento para tratamento bibliométrico dos dados**. 2019. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) - Universidade Federal de São Carlos, São Carlos, 2019.

MATIAS, M. S. O. **Base referencial para o povoamento de repositórios institucionais: coleta automatizada de metadados da Plataforma Lattes**. 2015. 86 p. Dissertação (Mestrado em Gestão de Organizações e Sistemas Públicos) - Universidade Federal de São Carlos, São Carlos, 2015.

MEADOWS, A. J. **A comunicação científica**. Brasília: Briquet de Lemos/Livros, 1999.

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

MUGABE, D. A.; FERNANDES, J. A.; CORREIA, P. F. **Avaliação da associação estatística num diagrama de dispersão por estudantes universitários**. 2012. Disponível em: <<http://repositorium.sdum.uminho.pt/bitstream/1822/20671/1/2012%20SIEMb.pdf>>. Acesso em: 10 jul. 2019.

NARIN, F.; OLIVASTRO, D.; STEVENS, K. S. Bibliometric theory, practice and problem. **Evaluation Review**, Thousand Oaks, California, v. 18, n. 1, p. 65-76, 1994.

NASCIMENTO, H. A.; FERREIRA, C. B. R. Visualização de Informações - Uma Abordagem Prática. In: UNISINOS, 2005. Rio Grande do Sul. XXV Congresso da

Sociedade Brasileira de Computação. **Anais...** Rio Grande do Sul: São Leopoldo, 2005, p. 1262-1312.

OKUBO, Y. Bibliometric indicators and analysis of research systems: methods and examples. **OECD Science, Technology and Industry Working Papers**, Paris, v. 97, n. 41, 1997/1, OECD, 1997.

PAINEL LATTES. **Estatísticas da Base de Currículos da Plataforma Lattes**. 2019. Disponível em: <<http://estatico.cnpq.br/painelLattes>>. Acesso em: 01 maio 2019.

PORTAL BRASILEIRO DE DADOS ABERTOS. **dados.gov.br**. 2019. Disponível em: <<http://dados.gov.br>>. Acesso em: 01 maio 2019.

REIS, J. E. **Incipiência da disponibilidade de indicadores bibliométricos e altmétricos nos repositórios institucionais brasileiros**. 2016. 120 p. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) - Universidade Federal de São Carlos, São Carlos, 2016.

ROEMER, R. C.; BORCHARDT, R. **Meaningful Metrics: A 21st-Century Librarian's Guide to Bibliometrics, Altmetrics, and Research Impact**. Association of College and Research Libraries. Chicago, Illinois, 2015. 251 p.

ROSSI, L.; FREIRE, I. L.; MENA-CHALCO, J. P. Genealogical index: A metric to analyze advisor-advisee relationships. **Journal of Informetrics**, v. 11, n. 2, p. 564-582, 2017. ISSN 1751-1577.

SANTOS, S. M.; MUGNAINI, R. Comunicação científica em periódicos e a internacionalização das ciências brasileiras. In: CARNEIRO, F. F. B.; FERREIRA NETO, A.; SANTOS, W. (Org). **A comunicação científica em periódicos**. Appris, p. 73-93, 2019.

SARACEVIC, T. Ciência da Informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996. Semestral. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235/22>>. Acesso em: 30 nov. 2017.

SARVO, D. O. **Avaliação da Plataforma Lattes como fonte de coleta de metadados para povoamento automatizado em repositórios institucionais**. 2018. 96 p. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) - Universidade Federal de São Carlos, São Carlos, 2018.

SHNEIDERMAN, B. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In: **Proceedings of IEEE Symposium on Visual Languages**, Boulder, CO, p. 336-343, 1996.

SILVA, C. G. **Exploração de bases de dados de ambientes de Educação a distância por meio de ferramentas de consulta apoiadas por Visualização de Informação**. 2006. Tese (Doutorado) - Instituto de Computação, Universidade Estadual de Campinas, Campinas, 2006.

SILVA, J. L. C.; FREIRE, G. H. A. Um olhar sobre a origem da ciência da informação: indícios embrionários para sua caracterização identitária. **Encontros Bibli**, v. 17, n. 33, p. 1-29, 2012. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2012v17n33p1>>. Acesso em: 10 nov. 2017.

SMIT, J. W.; BARRETO, A. A. **Ciência da informação**: base conceitual para a formação do profissional. In: VALENTIM, M.L. (Org.). Formação do profissional da informação. São Paulo: Polis, 2002.

SPINAK, E. Indicadores cientiométricos. **Ciencia da Informação**, Brasília, v. 27, n. 2, p. 141-148, 1998. Disponível em: <<http://revista.ibict.br/ciinf/article/view/795/826>>. Acesso em: 28 nov. 2017.

TARGINO, M. G. Comunicação científica: uma revisão de seus elementos básicos. **Informação & Sociedade**, v. 10, n. 2, p. 1-27, 2000.

UNWIN, A. Good Graphics? In: CHEN, C.; HÄRDLE, W.; UNWIN, A. **Handbook of Data Visualization**, Springer, 2008, p. 57-78.

VAN ECK, N. J. et al. A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. **Journal of the American Society for Information Science and Technology**, p. 2405-2416. 2010.

VAN RAAN, A. F. J. Advances in bibliometric analysis: research performance assessment and science mapping. In: **Bibliometrics: Use and Abuse in the Review of Research Performance**, London: Portland Press Ltd, 2014, v. 87.

VANZ, S. A. S.; SILVA FILHO, R. C. O protagonismo das revistas na comunicação científica: histórico e evolução. In: CARNEIRO, F. F. B.; FERREIRA NETO, A.; SANTOS, W. (Org). **A comunicação científica em periódicos**. Appris, p. 19-44, 2019.

VIDOTTI, M. V. **Caracterização e discussão da produção científica de uma instituição de ensino superior privada**. 2016. 94 p. Dissertação (Mestrado em Ciência, Tecnologia e Sociedade) - Universidade Federal de São Carlos, São Carlos, 2016.

VIEIRA, J. M. L.; CORREA, R. F. Visualização da informação na construção de interfaces amigáveis para sistemas de recuperação de informação. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 16, n. 32, p. 73-93, 2011. ISSN 1518-2924. DOI: 10.5007/15182924.2011v16n32p73.

VISPUBLICA. **Modelo de visualização de dados públicos**. 2019. Disponível em: <<http://vispublica.gov.br>>. Acesso em: 01 maio 2019.

WONG, D. M. THE WALL STREET JOURNAL. **Guide to Information Graphics: THE DOS AND DON'TS OF PRESENTING DATA, FACTS, AND FIGURES**. New York; London: W.W. Norton & Company, 2010.

YAMAGUCHI, J. K. **Diretrizes para a escolha de técnicas de visualização aplicadas no processo de extração do conhecimento**. 2010. 182 p. Dissertação (Mestrado em Ciência da Computação) - Universidade Estadual de Maringá, Maringá, 2010.

ANEXO 1 - JSON DE NÚMERO DE PUBLICAÇÕES POR PESQUISADOR.

Arquivo JSON criado por meio de procedimento para tratamento bibliométrico dos dados dos currículos Lattes extraídos pela ferramenta synclattes. É utilizado nas seguintes visualizações: gráfico de barras, barras empilhadas e nuvem de etiquetas.

```
[{
  "nome": "Jose Antonio Eiras",
  "artigos_eventos": 581,
  "artigos_periodicos": 369,
  "idlattes": "6653223814715498"
}, {
  "nome": "Walter José Botta Filho",
  "artigos_eventos": 545,
  "artigos_periodicos": 371,
  "idlattes": "8956458007749112"
}, {
  "nome": "João Batista Fernandes",
  "artigos_eventos": 604,
  "artigos_periodicos": 298,
  "idlattes": "2691336077459602"
}, {
  "nome": "Elson Longo da Silva",
  "artigos_eventos": 214,
  "artigos_periodicos": 683,
  "idlattes": "9848311210578810"
}, {
  "nome": "Orlando Fatibello Filho",
  "artigos_eventos": 529,
  "artigos_periodicos": 343,
  "idlattes": "9859737944357808"
}, {
  "nome": "Paulo Cezar Vieira",
  "artigos_eventos": 493,
  "artigos_periodicos": 292,
  "idlattes": "8923844563320377"
}, {
  "nome": "Maria Fatima das Graças Fernandes da Silva",
  "artigos_eventos": 480,
  "artigos_periodicos": 268,
  "idlattes": "0457632122660653"
}, {
  "nome": "Claudio Shyinti Kiminami",
  "artigos_eventos": 418,
  "artigos_periodicos": 310,
  "idlattes": "5443002033733395"
}, {
  "nome": "Jose Teixeira Freire",
  "artigos_eventos": 519,
  "artigos_periodicos": 151,
  "idlattes": "2270677706718167"
}, {
  "nome": "Anselmo Ortega Boschi",
  "artigos_eventos": 455,
  "artigos_periodicos": 213,
  "idlattes": "3259846931278607"
}
```

```

}, {
  "nome": "Ruth Herta Goldschmidt Aliaga Kiminami",
  "artigos_eventos": 450,
  "artigos_periodicos": 193,
  "idlattes": "7643216999108223"
}, {
  "nome": "Aparecida Maria Catai",
  "artigos_eventos": 434,
  "artigos_periodicos": 195,
  "idlattes": "5801652590531684"
}, {
  "nome": "Joaquim de Araújo Nóbrega",
  "artigos_eventos": 333,
  "artigos_periodicos": 291,
  "idlattes": "8833989058164529"
}, {
  "nome": "Edson Roberto Leite",
  "artigos_eventos": 135,
  "artigos_periodicos": 475,
  "idlattes": "1025598529469393"
}, {
  "nome": "Julio Zukerman Schpector",
  "artigos_eventos": 220,
  "artigos_periodicos": 378,
  "idlattes": "4252331837170383"
}, {
  "nome": "Claudemiro Bolfarini",
  "artigos_eventos": 331,
  "artigos_periodicos": 260,
  "idlattes": "9231627080617037"
}, {
  "nome": "Ducinei Garcia",
  "artigos_eventos": 391,
  "artigos_periodicos": 193,
  "idlattes": "5236758856687873"
}, {
  "nome": "Audrey Borghi e Silva",
  "artigos_eventos": 387,
  "artigos_periodicos": 190,
  "idlattes": "4855616925791895"
}, {
  "nome": "Marisa Narciso Fernandes",
  "artigos_eventos": 446,
  "artigos_periodicos": 126,
  "idlattes": "3021357484136942"
}, {
  "nome": "Pedro Manoel Galetti Junior",
  "artigos_eventos": 371,
  "artigos_periodicos": 194,
  "idlattes": "7398754661670478"
}, {
  "nome": "Orlando Moreira Filho",
  "artigos_eventos": 352,
  "artigos_periodicos": 193,
  "idlattes": "3927076022470557"
}, {
  "nome": "Zilda Aparecida Pereira Del Prette",
  "artigos_eventos": 409,
  "artigos_periodicos": 122,
  "idlattes": "8113238388739093"
}, {
  "nome": "Antonio Gilberto Ferreira",

```

```

"artigos_eventos": 299,
"artigos_periodicos": 205,
"idlattes": "3676462220401452"
}, {
  "nome": "Eniceia Gonçalves Mendes",
  "artigos_eventos": 380,
  "artigos_periodicos": 110,
  "idlattes": "3897627554738983"
}, {
  "nome": "Maria Cristina P Innocentini Hayashi",
  "artigos_eventos": 380,
  "artigos_periodicos": 104,
  "idlattes": "7263318849588556"
}, {
  "nome": "Ernesto Chaves Pereira de Souza",
  "artigos_eventos": 281,
  "artigos_periodicos": 199,
  "idlattes": "1505400360366643"
}, {
  "nome": "Edson Rodrigues Filho",
  "artigos_eventos": 267,
  "artigos_periodicos": 203,
  "idlattes": "3667941735597178"
}, {
  "nome": "Angélica Maria Penteado Martins Dias",
  "artigos_eventos": 241,
  "artigos_periodicos": 229,
  "idlattes": "8637344058859147"
}, {
  "nome": "Márcio Raymundo Morelli",
  "artigos_eventos": 318,
  "artigos_periodicos": 148,
  "idlattes": "0172837599844991"
}, {
  "nome": "Raquel de Lima Camargo Giordano",
  "artigos_eventos": 349,
  "artigos_periodicos": 114,
  "idlattes": "9695542424889786"
}, {
  "nome": "Jose Renato Coury",
  "artigos_eventos": 354,
  "artigos_periodicos": 91,
  "idlattes": "8156854796793004"
}, {
  "nome": "Roberto de Campos Giordano",
  "artigos_eventos": 331,
  "artigos_periodicos": 111,
  "idlattes": "0834668419587001"
}, {
  "nome": "Maria Amelia Almeida",
  "artigos_eventos": 371,
  "artigos_periodicos": 64,
  "idlattes": "0316944549621648"
}, {
  "nome": "Luiz Antonio Carlos Bertollo",
  "artigos_eventos": 196,
  "artigos_periodicos": 238,
  "idlattes": "7069073110550804"
}, {
  "nome": "Lucia Cavalcanti de Albuquerque Williams",
  "artigos_eventos": 349,
  "artigos_periodicos": 84,

```

```

    "idlattes": "6962646759651464"
  }, {
    "nome": "Sofia Cristina Iost Pavarini",
    "artigos_eventos": 308,
    "artigos_periodicos": 123,
    "idlattes": "1983620301963081"
  }, {
    "nome": "Jose Carlos de Toledo",
    "artigos_eventos": 307,
    "artigos_periodicos": 120,
    "idlattes": "9421627326330417"
  }, {
    "nome": "Pedro Augusto de Paula Nascente",
    "artigos_eventos": 277,
    "artigos_periodicos": 147,
    "idlattes": "1387043320265189"
  }, {
    "nome": "Romeu Cardozo Rocha Filho",
    "artigos_eventos": 216,
    "artigos_periodicos": 199,
    "idlattes": "5200717413563550"
  }, {
    "nome": "Francisco Tadeu Rantin",
    "artigos_eventos": 295,
    "artigos_periodicos": 115,
    "idlattes": "3546242299713690"
  }, {
    "nome": "Reinaldo Morabito Neto",
    "artigos_eventos": 219,
    "artigos_periodicos": 187,
    "idlattes": "4194801952934254"
  }, {
    "nome": "Victor Carlos Pandolfelli",
    "artigos_eventos": 163,
    "artigos_periodicos": 243,
    "idlattes": "7369376873984839"
  }, {
    "nome": "Dilson Cardoso",
    "artigos_eventos": 299,
    "artigos_periodicos": 101,
    "idlattes": "2462847535959232"
  }, {
    "nome": "Sonia Cristina Juliano Gualtieri",
    "artigos_eventos": 264,
    "artigos_periodicos": 136,
    "idlattes": "9973185407856723"
  }, {
    "nome": "Eloisa Tudella",
    "artigos_eventos": 246,
    "artigos_periodicos": 153,
    "idlattes": "3207258034860219"
  }, {
    "nome": "Roseli Esquerdo Lopes",
    "artigos_eventos": 300,
    "artigos_periodicos": 95,
    "idlattes": "1507752191797249"
  }, {
    "nome": "Mauricio Jamami",
    "artigos_eventos": 263,
    "artigos_periodicos": 132,
    "idlattes": "4274129560250050"
  }, {

```



```
"nome": "Irineu Bianchini Junior",
"artigos_eventos": 269,
"artigos_periodicos": 121,
"idlattes": "0659731944736389"
}, {
  "nome": "Aline Maria de Medeiros Rodrigues Reali",
  "artigos_eventos": 313,
  "artigos_periodicos": 64,
  "idlattes": "3250195451890332"
}, {
  "nome": "Alzir Azevedo Batista",
  "artigos_eventos": 166,
  "artigos_periodicos": 207,
  "idlattes": "6469642481998660"
}]
```

ANEXO 2 - JSON DE COAUTORIA ENTRE PESQUISADORES.

Arquivo JSON criado por meio de procedimento para tratamento bibliométrico dos dados dos currículos Lattes extraídos pela ferramenta synclattes. É utilizado nas seguintes visualizações: grafo e matriz de colaboração.

```
{
  "nodes": [{
    "nome": "Jose Antonio Eiras",
    "qtde_artigos": 950,
    "idlattes": "6653223814715498",
    "depto": "Departamento de Física"
  }, {
    "nome": "Walter José Botta Filho",
    "qtde_artigos": 916,
    "idlattes": "8956458007749112",
    "depto": "Departamento de Engenharia de Materiais"
  }, {
    "nome": "João Batista Fernandes",
    "qtde_artigos": 902,
    "idlattes": "2691336077459602",
    "depto": "Departamento de Química"
  }, {
    "nome": "Elson Longo da Silva",
    "qtde_artigos": 897,
    "idlattes": "9848311210578810",
    "depto": "Departamento de Química"
  }, {
    "nome": "Orlando Fatibello Filho",
    "qtde_artigos": 872,
    "idlattes": "9859737944357808",
    "depto": "Departamento de Química"
  }, {
    "nome": "Paulo Cezar Vieira",
    "qtde_artigos": 785,
    "idlattes": "8923844563320377",
    "depto": "Departamento de Química"
  }, {
    "nome": "Maria Fatima das Graças Fernandes da Silva",
    "qtde_artigos": 748,
    "idlattes": "0457632122660653",
    "depto": "Departamento de Química"
  }, {
    "nome": "Claudio Shyinti Kiminami",
    "qtde_artigos": 728,
    "idlattes": "5443002033733395",
    "depto": "Departamento de Engenharia de Materiais"
  }, {
    "nome": "Jose Teixeira Freire",
    "qtde_artigos": 670,
    "idlattes": "2270677706718167",
    "depto": "Departamento de Engenharia Química"
  }, {
    "nome": "Anselmo Ortega Boschi",
    "qtde_artigos": 668,
    "idlattes": "3259846931278607",
```

```

    "depto": "Departamento de Engenharia de Materiais"
  }, {
    "nome": "Ruth Herta Goldschmidt Aliaga Kiminami",
    "qtde_artigos": 643,
    "idlattes": "7643216999108223",
    "depto": "Departamento de Engenharia de Materiais"
  }, {
    "nome": "Aparecida Maria Catai",
    "qtde_artigos": 629,
    "idlattes": "5801652590531684",
    "depto": "Departamento de Fisioterapia"
  }, {
    "nome": "Joaquim de Araújo Nóbrega",
    "qtde_artigos": 624,
    "idlattes": "8833989058164529",
    "depto": "Departamento de Química"
  }, {
    "nome": "Edson Roberto Leite",
    "qtde_artigos": 610,
    "idlattes": "1025598529469393",
    "depto": "Departamento de Química"
  }, {
    "nome": "Julio Zukerman Schpector",
    "qtde_artigos": 598,
    "idlattes": "4252331837170383",
    "depto": "Departamento de Química"
  }, {
    "nome": "Claudemiro Bolfarini",
    "qtde_artigos": 591,
    "idlattes": "9231627080617037",
    "depto": "Departamento de Engenharia de Materiais"
  }, {
    "nome": "Ducinei Garcia",
    "qtde_artigos": 584,
    "idlattes": "5236758856687873",
    "depto": "Departamento de Física"
  }, {
    "nome": "Audrey Borghi e Silva",
    "qtde_artigos": 577,
    "idlattes": "4855616925791895",
    "depto": "Departamento de Fisioterapia"
  }, {
    "nome": "Marisa Narciso Fernandes",
    "qtde_artigos": 572,
    "idlattes": "3021357484136942",
    "depto": "Departamento de Ciências Fisiológicas"
  }, {
    "nome": "Pedro Manoel Galetti Junior",
    "qtde_artigos": 565,
    "idlattes": "7398754661670478",
    "depto": "Departamento de Genética e Evolução"
  }, {
    "nome": "Orlando Moreira Filho",
    "qtde_artigos": 545,
    "idlattes": "3927076022470557",
    "depto": "Departamento de Genética e Evolução"
  }, {
    "nome": "Zilda Aparecida Pereira Del Prette",
    "qtde_artigos": 531,
    "idlattes": "8113238388739093",
    "depto": "Departamento de Psicologia"
  }, {

```

```

"nome": "Antonio Gilberto Ferreira",
"qtde_artigos": 504,
"idlattes": "3676462220401452",
"depto": "Departamento de Química"
}, {
"nome": "Eniceia Gonçalves Mendes",
"qtde_artigos": 490,
"idlattes": "3897627554738983",
"depto": "Departamento de Psicologia"
}, {
"nome": "Maria Cristina P Innocentini Hayashi",
"qtde_artigos": 484,
"idlattes": "7263318849588556",
"depto": "Departamento de Ciência da Informação"
}, {
"nome": "Ernesto Chaves Pereira de Souza",
"qtde_artigos": 480,
"idlattes": "1505400360366643",
"depto": "Departamento de Química"
}, {
"nome": "Edson Rodrigues Filho",
"qtde_artigos": 470,
"idlattes": "3667941735597178",
"depto": "Departamento de Química"
}, {
"nome": "Angélica Maria Penteado Martins Dias",
"qtde_artigos": 470,
"idlattes": "8637344058859147",
"depto": "Departamento de Ecologia e Biologia Evolutiva"
}, {
"nome": "Márcio Raymundo Morelli",
"qtde_artigos": 466,
"idlattes": "0172837599844991",
"depto": "Departamento de Engenharia de Materiais"
}, {
"nome": "Raquel de Lima Camargo Giordano",
"qtde_artigos": 463,
"idlattes": "9695542424889786",
"depto": "Departamento de Engenharia Química"
}, {
"nome": "Jose Renato Coury",
"qtde_artigos": 445,
"idlattes": "8156854796793004",
"depto": "Departamento de Engenharia Química"
}, {
"nome": "Roberto de Campos Giordano",
"qtde_artigos": 442,
"idlattes": "0834668419587001",
"depto": "Departamento de Engenharia Química"
}, {
"nome": "Maria Amelia Almeida",
"qtde_artigos": 435,
"idlattes": "0316944549621648",
"depto": "Departamento de Psicologia"
}, {
"nome": "Luiz Antonio Carlos Bertollo",
"qtde_artigos": 434,
"idlattes": "7069073110550804",
"depto": "Departamento de Genética e Evolução"
}, {
"nome": "Lucia Cavalcanti de Albuquerque Williams",
"qtde_artigos": 433,

```

```

        "idlattes": "6962646759651464",
        "depto": "Departamento de Psicologia"
    }, {
        "nome": "Sofia Cristina Iost Pavarini",
        "qtde_artigos": 431,
        "idlattes": "1983620301963081",
        "depto": "Departamento de Gerontologia"
    }, {
        "nome": "Jose Carlos de Toledo",
        "qtde_artigos": 427,
        "idlattes": "9421627326330417",
        "depto": "Departamento de Engenharia de Produção"
    }, {
        "nome": "Pedro Augusto de Paula Nascente",
        "qtde_artigos": 424,
        "idlattes": "1387043320265189",
        "depto": "Departamento de Engenharia de Materiais"
    }, {
        "nome": "Romeu Cardozo Rocha Filho",
        "qtde_artigos": 415,
        "idlattes": "5200717413563550",
        "depto": "Departamento de Química"
    }, {
        "nome": "Francisco Tadeu Rantin",
        "qtde_artigos": 410,
        "idlattes": "3546242299713690",
        "depto": "Departamento de Ciências Fisiológicas"
    }, {
        "nome": "Reinaldo Morabito Neto",
        "qtde_artigos": 406,
        "idlattes": "4194801952934254",
        "depto": "Departamento de Engenharia de Produção"
    }, {
        "nome": "Victor Carlos Pandolfelli",
        "qtde_artigos": 406,
        "idlattes": "7369376873984839",
        "depto": "Departamento de Engenharia de Materiais"
    }, {
        "nome": "Dilson Cardoso",
        "qtde_artigos": 400,
        "idlattes": "2462847535959232",
        "depto": "Departamento de Engenharia Química"
    }, {
        "nome": "Sonia Cristina Juliano Gualtieri",
        "qtde_artigos": 400,
        "idlattes": "9973185407856723",
        "depto": "Departamento de Botânica"
    }, {
        "nome": "Eloisa Tudella",
        "qtde_artigos": 399,
        "idlattes": "3207258034860219",
        "depto": "Departamento de Fisioterapia"
    }, {
        "nome": "Roseli Esquerdo Lopes",
        "qtde_artigos": 395,
        "idlattes": "1507752191797249",
        "depto": "Departamento de Terapia Ocupacional"
    }, {
        "nome": "Mauricio Jamami",
        "qtde_artigos": 395,
        "idlattes": "4274129560250050",
        "depto": "Departamento de Fisioterapia"
    }

```

```

}, {
  "nome": "Irineu Bianchini Junior",
  "qtde_artigos": 390,
  "idlattes": "0659731944736389",
  "depto": "Departamento de Hidrobiologia"
}, {
  "nome": "Aline Maria de Medeiros Rodrigues Reali",
  "qtde_artigos": 377,
  "idlattes": "3250195451890332",
  "depto": "Departamento de Teorias e Práticas Pedagógicas"
}, {
  "nome": "Alzir Azevedo Batista",
  "qtde_artigos": 373,
  "idlattes": "6469642481998660",
  "depto": "Departamento de Química"
}],
"links": [{
  "source": 0,
  "target": 16,
  "value": 506
}, {
  "source": 1,
  "target": 7,
  "value": 453
}, {
  "source": 1,
  "target": 15,
  "value": 344
}, {
  "source": 2,
  "target": 6,
  "value": 671
}, {
  "source": 2,
  "target": 5,
  "value": 650
}, {
  "source": 2,
  "target": 26,
  "value": 197
}, {
  "source": 2,
  "target": 22,
  "value": 77
}, {
  "source": 3,
  "target": 13,
  "value": 420
}, {
  "source": 4,
  "target": 38,
  "value": 74
}, {
  "source": 5,
  "target": 2,
  "value": 650
}, {
  "source": 5,
  "target": 6,
  "value": 614
}, {
  "source": 5,

```

```
    "target": 26,  
    "value": 173  
  }, {  
    "source": 5,  
    "target": 22,  
    "value": 61  
  }, {  
    "source": 6,  
    "target": 2,  
    "value": 671  
  }, {  
    "source": 6,  
    "target": 5,  
    "value": 614  
  }, {  
    "source": 6,  
    "target": 26,  
    "value": 191  
  }, {  
    "source": 6,  
    "target": 22,  
    "value": 58  
  }, {  
    "source": 7,  
    "target": 1,  
    "value": 453  
  }, {  
    "source": 7,  
    "target": 15,  
    "value": 450  
  }, {  
    "source": 10,  
    "target": 28,  
    "value": 156  
  }, {  
    "source": 10,  
    "target": 16,  
    "value": 47  
  }, {  
    "source": 11,  
    "target": 17,  
    "value": 176  
  }, {  
    "source": 13,  
    "target": 3,  
    "value": 420  
  }, {  
    "source": 15,  
    "target": 7,  
    "value": 450  
  }, {  
    "source": 15,  
    "target": 1,  
    "value": 344  
  }, {  
    "source": 16,  
    "target": 0,  
    "value": 506  
  }, {  
    "source": 16,  
    "target": 10,  
    "value": 47  
  }
```

```
}, {
  "source": 17,
  "target": 11,
  "value": 176
}, {
  "source": 18,
  "target": 39,
  "value": 78
}, {
  "source": 18,
  "target": 2,
  "value": 32
}, {
  "source": 19,
  "target": 20,
  "value": 51
}, {
  "source": 19,
  "target": 33,
  "value": 41
}, {
  "source": 20,
  "target": 33,
  "value": 168
}, {
  "source": 20,
  "target": 19,
  "value": 51
}, {
  "source": 22,
  "target": 2,
  "value": 77
}, {
  "source": 22,
  "target": 5,
  "value": 61
}, {
  "source": 22,
  "target": 6,
  "value": 58
}, {
  "source": 22,
  "target": 26,
  "value": 40
}, {
  "source": 23,
  "target": 32,
  "value": 22
}, {
  "source": 26,
  "target": 2,
  "value": 197
}, {
  "source": 26,
  "target": 6,
  "value": 191
}, {
  "source": 26,
  "target": 5,
  "value": 173
}, {
  "source": 26,
```



```
    "target": 22,  
    "value": 40  
  }, {  
    "source": 28,  
    "target": 10,  
    "value": 156  
  }, {  
    "source": 29,  
    "target": 31,  
    "value": 212  
  }, {  
    "source": 31,  
    "target": 29,  
    "value": 212  
  }, {  
    "source": 32,  
    "target": 23,  
    "value": 22  
  }, {  
    "source": 33,  
    "target": 20,  
    "value": 168  
  }, {  
    "source": 33,  
    "target": 19,  
    "value": 41  
  }, {  
    "source": 38,  
    "target": 4,  
    "value": 74  
  }, {  
    "source": 39,  
    "target": 18,  
    "value": 78  
  }, {  
    "source": 41,  
    "target": 1,  
    "value": 37  
  }  
}]  
}
```