UNIVERSIDADE DE BRASÍLIA (UNB)
FACULDADE DE CIÊNCIA DA INFORMAÇÃO (FCI)
CURSO DE GRADUAÇÃO EM BIBLIOTECONOMIA

JÉSSICA BILAC GASPARETO

O USO DE TÉCNICAS DE CIÊNCIA DE DADOS PARA ANALISAR A AMBIGUIDADE DE AUTORIA EM PRODUÇÃO CIENTÍFICA DOS PROFESSORES DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO DAS UNIVERSIDADES FEDERAIS BRASILEIRAS

Brasília

2021

JÉSSICA BILAC GASPARETO

O USO DE TÉCNICAS DE CIÊNCIA DE DADOS PARA ANALISAR A AMBIGUIDADE DE AUTORIA EM PRODUÇÃO CIENTÍFICA DOS PROFESSORES DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO DAS UNIVERSIDADES FEDERAIS BRASILEIRAS

Monografia apresentada à banca examinadora como requisito parcial para a obtenção do título de Bacharel em Biblioteconomia pela Faculdade de Ciência da Informação da Universidade de Brasília.

Orientador: Dr. Márcio de Carvalho Victorino

Brasília

2021

Bilac Gaspareto, Jéssica
BG249u O uso de técnicas de ciência de dados para analisar a ambiguidade de autoria em produção científica dos professores dos programas de pós-graduação em ciência da informação das universidades federais brasileiras / Jéssica Bilac Gaspareto; orientador Márcio de Carvalho Victorino. -- Brasília, 2021.
80 p.

Monografia (Graduação - Biblioteconomia) -- Universidade de Brasília, 2021.

1. Comunicação Científica. 2. Ciência de Dados. 3. Ambiguidade entre Autoridades. 4. Colaboração Científica. 5. Comunidade Científica. I. de Carvalho Victorino, Márcio, orient. II. Título.

Universidade de Brasília

FOLHA DE APROVAÇÃO

Título: O USO DE TÉCNICAS DE CIÊNCIA DE DADOS PARA ANALISAR A AMBIGUIDADE DE AUTORIA EM PRODUÇÃO CIENTÍFICA DOS PROFESSORES DOS PROGRAMAS DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO DAS UNIVERSIDADES FEDERAIS BRASILEIRAS

Autor(a): Jéssica Bilac Gaspareto

Monografia apresentada remotamente em **12 de novembro de 2021** à Faculdade de Ciência da Informação da Universidade de Brasília, como parte dos requisitos para obtenção do grau de Bacharel em Biblioteconomia.

Orientador(a) (FCI/UnB): Márcio de Carvalho Victorino

Membro Interno (FCI/UnB): André Luiz Appel

Membro Externo (CEF): José Marcelo Schiessl

Em 12/11/2021.

Documento assinado eletronicamente por **Marcio de Carvalho Victorino**, **Professor(a) de Magistério Superior da Faculdade de Ciência da Informação**, em 12/11/2021, às 23:24, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.

Documento assinado eletronicamente por **José Marcelo Schiessl**, **Usuário Externo**, em 13/11/2021, às 09:24, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.

Documento assinado eletronicamente por **André Luiz Appel**, **Usuário Externo**, em 13/11/2021, às 17:20, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da

Universidade de Brasília.

Documento assinado eletronicamente por **Jessica Bilac Gaspareto**, **Usuário Externo**, em 16/11/2021, às 15:08, conforme horário oficial de Brasília, com fundamento na Instrução da Reitoria 0003/2016 da Universidade de Brasília.

A autenticidade deste documento pode ser conferida no site http://sei.unb.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **7382977** e o código CRC **FB764151**.

Referência: Processo nº 23106.123946/2021-01 SEI nº 7382977

AGRADECIMENTOS

Agradeço, primeiramente, à Deus pela força nessa jornada dos últimos anos que não foram fáceis, mas graças a ele tudo veio a se encaixar da forma correta. A minha mãe, por ser o maior amor da minha vida e representar toda minha existência e história. Ao meu pai, por sempre incentivar o melhor de mim. Aos meus tios Evelyn e Matheus por me escutarem sobre meu trabalho e incentivar o melhor, vocês são essenciais na minha vida e agradeço a oportunidade por poder me reaproximar de vocês. Amo a todos citados aqui de forma descomunal.

Um agradecimento especial a um dos professores que eu mais admiro na Faculdade de Ciência da Informação: Prof. Dr. Márcio de Carvalho Victorino. O senhor foi essencial na minha formação e para o curso que quero seguir na Biblioteconomia. Obrigada imensamente pela oportunidade no PIBIC e por ter se oferecido a realizar este trabalho comigo, espero que seja o primeiro de muitos.

Agradeço a todos que trabalhei no Instituto Brasileiro de Informação em Ciência e Tecnologia, sem dúvidas o local que mais cresci profissionalmente e decidi o que quero seguir. Um agradecimento especial ao Alixandro, você fez total diferença na minha escrita e visão científica. Ao meu chefe no IBICT. Agradeço também tudo que aprendi durante o estágio na Câmara dos Deputados, minha supervisora Terezinha me fez enxergar o papel do bibliotecário de várias formas.

A todos os professores que pude ter contato durante a graduação. Me sinto imensamente grata por ter tido aula com tantos professores incríveis.

Ao Alexandre, meu amigo e irmão de consideração que me acompanhou em todos os momentos difíceis da minha vida e está sempre presente ao meu lado. Obrigada a me ajudar também a limpar os dados das minhas tabelas, foram essenciais para concluir este trabalho.

Ao Victor, uma das maiores amizades que fiz durante meu tempo de graduação. Levarei você pra vida inteira e obrigada por tudo sempre.

A Laura, uma das amizades que eu jamais imaginava que aconteceria. Obrigada pelos conselhos, pelo carinho comigo e com minha família.

As minhas melhores amigas que fiz na graduação, as chamadas “Invertidas”: Pamela, Rebeca, Ana Beatriz, Vivian, Isabela e Letícia. Meninas, vocês sabem o quão importante todas vocês são em minha vida. Sem dúvidas um dos maiores presentes que a UnB me trouxe.

A todos os meus amigos de infância. Ao Cadu, você foi essencial para conseguir finalizar a parte técnica do meu trabalho.

Por último, agradecer a todos que de alguma forma fizeram parte da minha jornada durante a minha graduação. Todos vocês são essenciais e amados por mim, agradeço imensamente pela participação de todos em minha vida.

“Trabalho duro é inútil
para aqueles que não
acreditam em si mesmos.”
– Uzumaki Naruto

RESUMO

A comunicação científica nasce mediante a explosão informacional crescente no mundo e advém da necessidade de organizar novos conhecimentos por meio de produções científicas. Por meio da atribuição de autoridades em trabalhos científicos, é possível identificar produções científicas dentro da comunicação científica, resultando no objeto de estudo deste trabalho. O presente objeto de estudo deste trabalho é salientado na ocorrência de ambiguidades entre nomes de autores. Desta forma, tem como objetivos específicos a contextualização do ciclo da comunicação cientifica e seus pilares e infraestruturas; como a colaboração científica funciona; importância da autoria múltipla; contextualização da bibliometria; além de mostrar como as técnicas de ciência de dados podem atuar na seguinte pesquisa. A metodologia adotada é quantitativa, sendo a coleta de dados proveniente do Portal Sucupira da CAPES, plataforma Lattes e plataforma ORCID. Em seguida, os dados coletados foram utilizados para criação de um esquema no sistema gerenciador de banco de dados MySQL, que possibilitou a organização, manipulação e limpeza dos dados. Os dados provenientes foram utilizados para o desenvolvimento do estudo de caso apresentado neste trabalho e por meio da plataforma Tableau (plataforma utilizada para análise de dados), gerou-se gráficos para discussão do problema apresentado. Em conclusão, foi possível verificar dois tipos diferentes de ambiguidades entre autores dentro dos professores participantes dos programas de pós-graduação em Ciência da Informação nas universidades federais brasileiras.

Palavras-chaves: Comunicação Científica. Ciência de Dados. Ambiguidade entre Autoridades. Colaboração Científica. Comunidade Científica.

ABSTRACT

Scientific communication is born through the growing informational explosion in the world and comes from the need to organize new knowledge through scientific productions. Through the competence of authorities in scientific works, it is possible to identify scientific productions within scientific communication, but there is no object of study in this work. The present object of study of this work is highlighted in the occurrence of ambiguities between authors' names. In this way, its specific objectives are to contextualize the scientific communication cycle and its pillars and infrastructures; how collaboration works; importance of multiple authorship; contextualization of bibliometrics; in addition to showing how data science techniques can work in the following research. The methodology adopted is quantitative, with data collection coming from CAPES' Sucupira Portal, Lattes platform and ORCID platform. Then, the collected data were used to create a schema in the MySQL database management system, which enabled the organization, manipulation and cleaning of the data. The data used were used for the development of the case study presented in this work and through the Tableau platform (a platform used for data analysis), graphs were generated to discuss the problem presented. In conclusion, it was possible to verify the two different types of ambiguities between authors within professors participating in graduate programs in information science in Brazilian universities.

Keywords: Scientific Communication. Data Science. Ambiguity between Authorities. Scientific Collaboration. Scientific community.

LISTA DE QUADROS

LISTA DE FIGURAS

LISTA DE GRÁFICOS

LISTA DE TABELAS

LISTA DE SIGLAS

CI – Ciência da Informação

SGBD – Sistema de gerenciamento de banco de dados

SUMÁRIO

1 INTRODUÇÃO

A ascendência de novas descobertas e avanços científicos foram fundamentais para transformações, de todas as épocas, por meio de mudanças de padrões de comportamento e do acesso à informação na sociedade. Juntamente com o advento da Revolução Industrial e a chegada do século XX – marcado pelo impulso sem precedentes do conhecimento e desenvolvimento tecnológico –, a ciência ganhou mais dimensão juntamente com a produção de novas informações, fazendo com que diversos pesquisadores necessitassem publicar seus trabalhos, nascendo assim a comunicação científica (VALEIRO; PINHEIRO, 2008).

O estudo da comunicação científica se originou por meio de literaturas que estavam correlacionadas com as origens da área da Ciência da Informação. Esta área nasceu por motivações ligadas pela necessidade de garantia ao acesso do crescente volume de documentos científicos produzidos pela comunidade científica presentes no ciclo da comunicação científica.

Dentro dos pontos significativos no ciclo da comunicação científica, destaca-se a comunidade científica que é definida responsável por tratar assuntos envolvendo a escrita científica e o mais importante para este trabalho: a autoridade de autor e suas ocorrências de ambiguidades nominais.

Por meio de produções científicas e a atribuições de autoridades em trabalhos científicos, é possível identificar padrões de citações para exibir a problemática deste trabalho. Também mediante ao reconhecimento do direito legal de um autor sobre um texto, é possibilitado a avaliação da produção cientifica daquele cientista, geralmente, utilizada como parâmetro para concessão de recursos por agências de fomento às pesquisas e como uma quantificação de métricas para indicadores de qualidade na comunicação científica.

Devido ao crescente volume de produções científicas dentro da comunicação científica, a colaboração científica nasce mediante a necessidade de poupar tempo e esforço, dessa forma pesquisadores acaba, potencializando suas pesquisas e gerando reconhecimento científico. Dessa forma, os índices

em tipos de colaborações científicas são utilizados para mensuração de indicadores de qualidade na comunidade científica.

Assim, por meio de produções científicas e a atribuições de autoridades em trabalhos científicos, é possível identificar padrões de citações para exibir a problemática deste trabalho uma vez que ao analisar as produções científicas, sejam elas produtos de colaboração científica ou produções individuais, nota-se ambiguidades entre autoridades dificultando na identificação dos autores para reconhecimento dos créditos acerca de uma produção científica.

A ciência de dados trata-se de uma ciência com foco no tratamento dos dados. As técnicas de ciência de dados tornam-se parte de um papel significativo para a exibição dos problemas de ambiguidades nominais em bases de dados, sendo por meio de suas ferramentas exibidas as possibilidades de apresentação do problema proposto.

O seguinte trabalho inicia-se com uma revisão de literatura que abordará sobre a comunicação científica; comunidade científica e escrita científica; e o papel da Ciência da Informação dentro da comunicação científica. A segunda seção do trabalho dentro da revisão de literatura abordará a colaboração científica, seus tipos de ocorrência e motivações para realizar trabalhos colaborativos. Durante a terceira seção, é caracterizado um dos tipos de colaboração científica: a autoria múltipla e seu uso para indicador de qualidade em produções científicas.

Na quarta seção dentro da revisão de literatura é abordado a autoridade de autores e a importância de valorização do esforço intelectual. Durante a quinta seção abordou-se as ocorrências e tipos de ambiguidades de autorias em produções científicas, sendo um dos pontos mais importantes para o trabalho realizado.

A revisão de literatura é finalizada com as últimas duas seções: bibliometria e ciência de dados que tratam sobre a parte quantitativa do trabalho e são utilizadas para parte prática do estudo de caso presente no trabalho.

O seguinte trabalho propõe exibir por meio de padrões de ambiguidades definidos durante a revisão de literatura envolvendo nomes de autores, sendo

exibidos por meio das técnicas de ciência de dados para gerar uma análise desenvolvida em um estudo de caso.

2 PROBLEMA DA PESQUISA

Como analisar a ambiguidade de autoria considerando a produção científica dos professores de programas de pós-graduação em Ciência da Informação das universidades federais brasileiras?

2.1 Justificativa

A comunicação científica envolve todo um processo de produção e disseminação da informação que, necessita de uma forma de registro para conceder os direitos autoriais a um indivíduo, uma vez que, um pesquisador dedicou esforço e tempo para contribuir para o avanço da ciência.

Uma das formas para mensurar a qualidade de uma produção científica é feito pela atribuição de indicadores de qualidade, ressaltando como um dos principais a autoria múltipla. A autoria múltipla é um indicador que exibe como a contribuição de autores em projetos científicos geram mais reconhecimento e tende a proporcionar um melhor avanço na ciência.

A ambiguidade entre nome de autores, onde um autor pode ter uma enorme gama de variações em seu nome ou diferentes autores com o mesmo nome, podem ser um fator crítico para a análise do indicador de qualidade em autoria múltipla por conta de duplicatas que acarretam na imprecisão dos resultados.

Dessa forma, as técnicas utilizadas em ciência de dados podem ser um fator chave para tratar e limpar os dados que envolvem a autoria múltipla, reduzindo ou eliminando as duplicatas e trazendo precisão para os resultados de busca de autores. As técnicas de ciência de dados também podem ser utilizadas para exibir gráficos e números estatísticos para representar o problema apontado.

3 OBJETIVOS

3.1 Objetivo geral

Analisar a ambiguidade de autoria considerando a produção científica dos professores de programas de pós-graduação em Ciência da Informação das universidades federais brasileiras.

3.2 Objetivos específicos

- Obter dados sobre a produção científica dos professores de programas de pós-graduação em Ciência da Informação das universidades federais brasileiras nas bases de dados abertas.
- Criar uma base de dados relacional para armazenar esses dados de forma organizada.
- Analisar a ambiguidade de autoria utilizando a linguagem SQL.
- Analisar a ambiguidade de autoria utilizando gráficos.

4 REVISÃO DE LITERATURA

4.1 Comunicação Cientifica

A comunicação dispõe de um papel central na ciência pelo fato de que, para ser considerado científico, um determinado conhecimento necessita da aprovação de outros pesquisadores. Essa aprovação se dá por dois momentos – o primeiro ocorre antes da publicação por meio de um teste de qualidade denominado “avaliação prévia[1]” e o segundo ocorre após a publicação quando aprovado na avaliação prévia, sendo publicado como artigo científico e exposto à crítica de todos –, ao ser publicado e acessível aos demais pesquisadores, esse conhecimento pode contribuir para outras pesquisas gerando novos conhecimentos (MUELLER, 2007).

Uma vez que esses conhecimentos são publicados, a comunicação científica nasce mediante a inevitabilidade do registro dos quais os avanços científicos e tecnológicos produzidos pelo ser humano sucedem, isto ocorre para que os precursores de diversas áreas do conhecimento possam prosperar daquela bagagem científica, contribuindo para pesquisas pelo bem da humanidade.

Segundo Targino (2000), a comunicação científica torna-se indispensável para a atividade científica, pois permite a conexão de esforços individuais de membros da comunidade científica, permitindo uma troca contínua de informações e difundindo conhecimento para sucessores ou auferidos de seus predecessores.

O processo da comunicação científica pode ser considerado como um sistema cíclico, pois precisa passar por algumas etapas que devem se repetir para a garantia do avanço científico e retroalimentação deste ciclo. Na Figura 1 é proposto um modelo para representação do ciclo da comunicação científica:

[1] É o processo de julgamento que um manuscrito é submetido antes de uma publicação realizada pelos pares (MUELLER, 2007).

Figura 1 - Ciclo da Comunicação Científica

Fonte: Elaborado pelo autor, com base em Schweitzer, Rodrigues e Varvakis (2011).

4.1.1 Comunidade Científica e Comunicação Científica

De acordo com Mueller (2006), o sistema de comunicação científica pode ser considerado como a infraestrutura da comunidade científica, pois neste sistema há a garantia da qualidade científica por meio de trabalhos que são validados por pares, sendo uma forma de evitar redundâncias e preservar autorias de pesquisas, reforçando o ciclo da comunicação científica (SCHWEITZER; RODRIGUES; VARVAKIS, 2011).

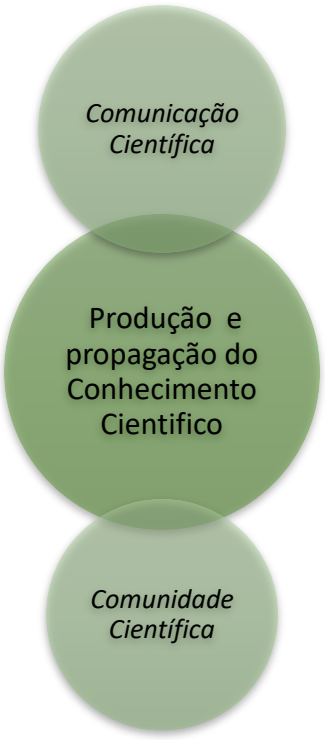
É importante ressaltar que há uma relação íntima entre “comunidade científica” e “comunicação científica”, uma vez que, um funciona como pilar para o outro (MUELLER, 2006). Infere-se que, para haver uma comunidade científica, é imprescindível a ocorrência da comunicação científica a fim de alimentar o ciclo da comunicação científica e reforçando esse pilar científico.

Conforme Leite e Costa (2007), são considerados partes da comunidade científica as universidades – como comunidades acadêmicas – por constituírem

os elementos do sistema científico e também “são consideradas ainda como o cerne da produção do conhecimento, e os processos de comunicação científica permeiam boa parte de suas atividades, o que permite tanto as trocas internas de conhecimento quanto externas, em interação com comunidades científicas” (LEITE; COSTA, 2007, p. 94). Nesta linha de raciocínio, a comunidade científica é constituída por produtores de conhecimento científico que buscam novas bagagem acadêmicas, alimentando a comunicação científica.

Desta forma, é importante lembrar que há uma diferença clara entre a comunicação científica e a comunidade científica, independente da relação íntima entre os dois, a comunicação científica trata-se do veículo e forma de propagação do conhecimento científico; e a comunidade científica designa os tipos de pessoas ou instituições que produzem uma linha de pesquisa. Ressalta-se como o principal objetivo de ambos: produzir e propagar o conhecimento científico assim como mostra a Figura 2.

Figura 2 - Objetivo da Comunidade Científica e da Comunicação Científica



Fonte: Elaborado pelo autor, 2021.

4.1.1.1 Escrita Científica

Conforme Frias (2015), a escrita científica é um dos meios essenciais para a divulgação do conhecimento científico, pois comunica os conteúdos primários – por meio de como essas informações são sintetizadas –, sendo o resultado dessa síntese a manifestação por meio das literaturas brancas[2].

Durante a publicação de um artigo científico é importante ressaltar que é uma publicação com autoria declarada, para apresentar e debater ideais, necessitando como uma das qualidades essenciais na redação de um trabalho científico a objetividade e precisão, para expor um bom trabalho ao público (SILVA, 2019).

Uma boa escrita científica é primordial para que um trabalho tenha seus pontos bem expostos, resultando em uma avaliação positiva por meio dos pares de avaliação, possibilitando a publicação de um artigo de periódico contribuindo para a comunicação científica. De acordo com Silva (2019), ao redigir um trabalho, deve-se levar em consideração os seguintes pontos apresentados no Quadro 1:

Quadro 1 - Pontos para uma Escrita Cientifica de Qualidade

Impessoalidade	Deve-se redigir em terceira pessoa, sem referências pessoais. Utilizam-se expressões como “O presente estudo”. O uso de “nós” e “eu” também devem ser evitados;
Clareza e precisão	Uma expressão clara exige que o autor tenha conhecimento do assunto em seu todo e em suas partes, para que as ideias apresentadas não gerem ambiguidades;
Objetividade	Exige-se que não se use expressões como: “eu penso”, “parece”, “achou-se”; pois indicam subjetividade;
Simplicidade	Ela é sinal de clareza de pensamento. Expressando corretamente as ideias que se deseja transmitir, o texto impõe-se por si mesmo;
Vocabulário técnico	Cada ciência utiliza uma terminologia técnica própria. Deve-se determinar a significação específica dos termos em contexto. Para isso, o autor deve consultar enciclopédias e dicionários especializados;

[2] São documentos convencionais ou formais que apresentam facilidade para identificação, divulgação e obtenção, produzidos dentro dos circuitos comerciais. Os principais utilizados na escrita científica são: periódicos científicos, livros e enciclopédias (GOMES; MENDONÇA; SOUZA, 2007).

Linguagem científica	Escolha frases curtas e simples, que exprimam melhor as ideias. A preferência é pela ordem direta. Evitar inversões desnecessárias. Frases longas e uso de parênteses dificultam a compreensão e tornam a leitura pesada;
Abreviaturas	Não são aconselhadas na redação do texto corrido. São comumente usadas nas notas de rodapé;
Palavras estrangeiras	As palavras estrangeiras devem ser destacadas no texto em forma itálica, em grifo ou negrito. Muito comum o itálico.

Fonte: (SILVA, 2019, p. 1).

4.1.2 Ciência da Informação e a Comunicação Científica

Conforme Miranda (2002), a Ciência da Informação teve seu advento após o fenômeno da “explosão da informação”, situada após a 2°Guerra Mundial, e na necessidade de um controle bibliográfico resultando no tratamento da documentação implícita no processo.

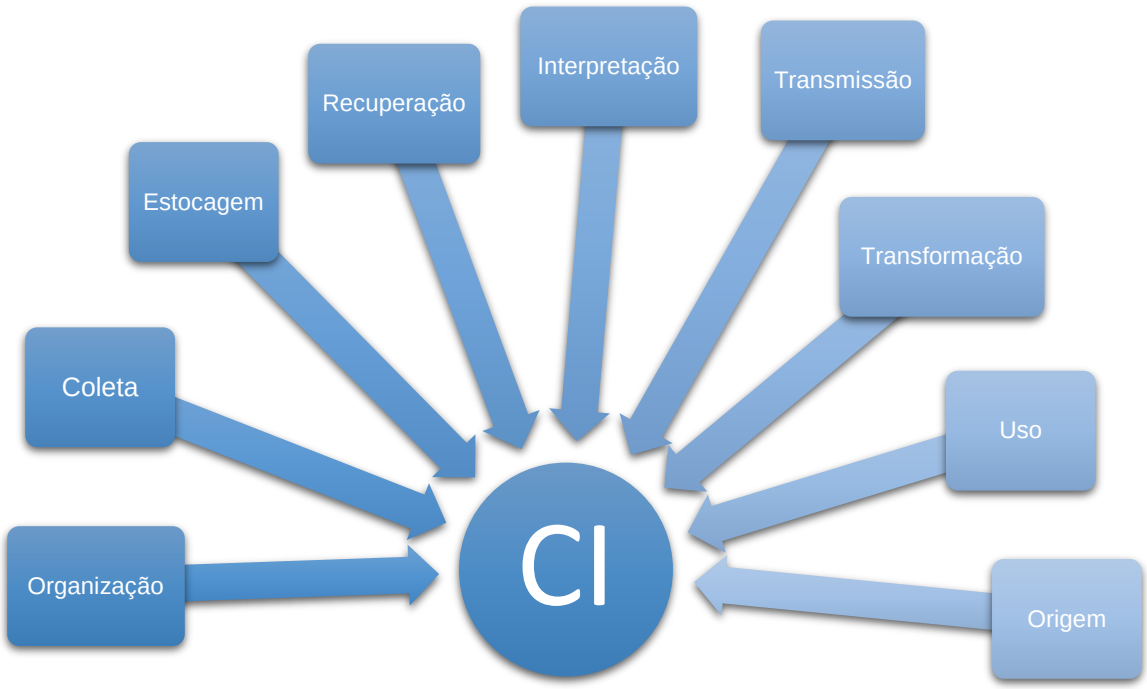
O termo conhecido como “Ciência da Informação” ou (CI), foi criado por volta de 1960, com base de estudos de produção, processamento e uso da informação como atividade predominantemente humana. Todavia, Wellish (1987) assegura que o termo “Ciência da Informação” foi utilizado pela primeira vez em 1959, para caracterizar o estudo do conhecimento registrado (HELPRIN, 1989, *apud* PINHEIRO; LOUREIRO, 1995).

Ao longo de 1962, um grupo de pesquisadores reunidos no *Georgia Institute of Technology* declararam que a Ciência da Informação era a ciência que investigava o comportamento da informação, suas propriedades, forças que regem ao fluxo da informação e por fim seus meios de processamento visando o melhor uso da informação. É uma área que se relaciona e deriva de outros áreas, por isso é considerada uma área emergente de novas disciplinas interdisciplinares (ROBREDO, 2003; VICTORINO 2011).

Na Figura 3 é proposto um modelo com base em Borko (1968), este modelo tem como premissa em suas afirmações acerca de como a Ciência da Informação é definida como uma disciplina que investiga o comportamento da informação por meio de suas propriedades. Este modelo representa alguns dos

processos nos quais as informações sucedem mediante a Ciência da Informação, e a forma como esses procedimentos possuem caráter de ciência pura – por meio de pesquisas de fundamentos – assim como também componente de ciência aplicada – ao desenvolver produtos e serviços –, essas são partes dos comportamentos e propriedades da Ciência da Informação.

Figura 3 - Corpos de Conhecimentos da Informação Ligados a Ciência da Informação



Fonte: Elaborado pelo autor (2021) e Borko (1968).

No decorrer da década de 1990, a Ciência da Informação é definida como um campo dedicado às questões científicas com foco em práticas profissionais voltadas aos problemas em relação à comunicação do conhecimento registrado entre os seres humanos (SARACEVIC, 1992). Em meio a Ciência da Informação e com a emergência de novas disciplinas, destaca-se a bibliometria utilizada por pesquisadores em âmbitos acadêmicos para a criação e parametrização de índices de citações e produções bibliográficas.

A Ciência da Informação apesar de ter seu foco no tratamento da informação e sendo um campo dedicado às questões científicas, muitas das vezes, acaba sendo equiparada e dado por sinônimo de forma equivocada em relação à biblioteconomia.

De acordo com Dias (2007), essa analogia entre a biblioteconomia e a Ciência da Informação se baseia no entendimento de que, a Ciência da Informação é uma área interdisciplinar e acaba abrangendo diversas áreas do conhecimento – biblioteconomia, documentação, arquivologia e outras –, causando a ideia errônea em afirmar que a Ciência da Informação seria um dos nomes da biblioteconomia.

A conclusão que se pode chegar entre essa comparação é que, por se tratar de um campo interdisciplinar, a Ciência da Informação acaba se subdividindo em especialidades e dentro dessas especialidades a biblioteconomia se encontra como é representado na Figura 4.

Figura 4 - Especialidades da Ciência da Informação

Fonte: Elaborado pelo autor, com base em Dias (2007).

A Ciência da Informação por se tratar de uma área do conhecimento com foco nas questões científicas se alinha diretamente com a comunicação científica uma vez que, ao falar dos centros de interesses e de ações da Ciência da Informação, destaca-se ao que se refere a comunicação científica – conhecer os tipos de publicações, características e formas –, mas havendo a necessidade de compreender também as características próprias da informação científica – sua estrutura de processos e seus sistemas de comunicação –, alinhando ambas em prol do avanço científico (MUELLER, 2007).

A comunicação científica é considerada parte essencial referente aos estudos da Ciência da Informação, na qual compõe uma disciplina cujo encargo central é atribuído em questões relacionadas – direta ou indiretamente – com o compartilhamento do conhecimento na sociedade (BAPTISTA *et al.*, 2007).

Dessa forma, ao analisar os pontos de especialidades na qual a Ciência da Informação abrange e a comunicação científica, destaca-se pontos essenciais de convergência entre ambas sendo proposto pelo modelo representado no Quadro 2 abaixo:

Quadro 2 - Convergência entre Ciência da Informação e Comunicação Científica

Ciência da Informação	**Comunicação Científica**
Processos de Comunicação	*Emissor (divulgar avanços científicos) enviar uma mensagem para o receptor (comunidade científica)*
Representação da Informação	*Representar a informação para que usuários possam recuperar artigos de periódicos ou trabalhos acadêmicos, facilitando a comunidade científica*

Fonte: Elaborado pelo autor 2021, com base em Baptista *et al* (2007) e Dias (2007).

4.2 Colaboração Científica

De acordo com Grácio (2018), a coautoria no ramo científico, ainda que em pequenas quantidades, já ocorria no século XVII, tendo o primeiro registro de artigo escrito contando com coautoria entre pesquisadores na data de 1665.

Dessa forma, pensando o produto da ciência como um prol de avanço à sociedade, a colaboração científica é definida por um esforço cooperativo que busca alcançar metas em comum entre os pesquisadores, esforço coordenado e resultados – os trabalhos científicos –, por meio de méritos e responsabilidades compartilhadas (BALANCIERI *et al.*, 2005). Nesta linha de pensamento, cientistas que trabalham em conjunto podem acelerar suas pesquisas por trabalharem em alinhamento com outros pesquisadores ao buscarem o mesmo, resultando no enriquecimento da ciência.

A colaboração científica acaba potencializando o crescimento profissional de um pesquisador, uma vez que, permite o trabalho conjunto de pesquisadores mais renomados com pesquisadores iniciantes (GRÁCIO, 2018). Este crescimento profissional gerado pode ser estipulado com base em novas produções científicas e publicações em revistas que assim, eventualmente, servirão de base para contribuição de novos percussores a fim de renovar e manter atualizado o conhecimento científico.

Um exemplo claro para esta situação pode ser definido como: um trabalho conjunto de um professor ao convidar um de seus alunos para escrever um artigo científico e/ou participar de projetos de iniciação científica em universidades. Uma colaboração dessa não só contribui com o professor, em razão de aumentar sua produção acadêmica, assim como proporcionaria prestigio a um aluno que ainda não dispõe de produções científicas.

Segundo Katz e Martin (1997), a colaboração científica ocorre com diferentes públicos – nações, instituições, grupos de pesquisas –, assim potencializando os resultados quando um grupo põe esforço para chegar no mesmo objetivo. Um dos grandes exemplos para colaboração científica em nações pode ser apresentado durante a pandemia do Covid-19, iniciada no Brasil em 2020, onde cientistas de todo o mundo se mobilizaram para estudar formas de combate à pandemia expondo o potencial da colaboração científica.

De acordo com Hilário, Grácio e Guimarães (2017), as razões que impulsionam a ocorrência de pesquisas colaborativas são diversas e podem variar de acordo com a área do conhecimento e pesquisadores de um mesmo campo. Dentre esses pontos, destaca-se as 17 possíveis motivações,

levantadas por Vanz e Stumpf (2010), fundamentadas por literaturas científicas internacionais e nacionais, apresentadas no Quadro 3:

Quadro 3 - Motivações para Colaboração Científica

1. Desejo de aumentar a popularidade científica, a visibilidade e o reconhecimento pessoal;
2. Aumento da produtividade;
3. Racionalização do uso da mão-de-obra científica e do tempo dispensado à pesquisa;
4. Redução da possibilidade de erro;
5. Obtenção e/ou ampliação de financiamentos, recursos, equipamentos especiais, materiais;
6. Aumento da especialização na Ciência;
7. Possibilidade de "ataque" a grandes problemas de pesquisa;
8. Crescente profissionalização da ciência;
9. Desejo de aumentar a própria experiência através da experiência de outros cientistas;
10. Desejo de realizar pesquisa multidisciplinar;
11. União de forças para evitar a competição;
12. Treinamento de pesquisadores e orientandos;
13. Necessidade de opiniões externas para confirmar ou avaliar um problema;
14. Possibilidade de maior divulgação da pesquisa;
15. Como forma de manter a concentração e a disciplina na pesquisa até a entrega dos resultados ao resto da equipe;
16. Compartilhamento do entusiasmo por uma pesquisa com alguém;
17. Necessidade de trabalhar fisicamente próximo a outros pesquisadores, por amizade e desejo de estar com quem se gosta.

Fonte: VANZ; STUMPF, 2010, p. 50-51.

4.3 Autoria Múltipla

De acordo com *Vilan Filho et al* (2008), a pesquisa em colaboração científica cada vez mais é impulsionada por meio do governo, agências de fomento, universidades e instituições de pesquisa. Este fenômeno ocorre por motivos de que, há uma crença entre a comunidade científica e indivíduos responsáveis por políticas científicas que apoiam e acreditam que as colaborações científicas são um ponto positivo para a ciência e precisam ser encorajadas (KATZ; MARTIN, 1997).

Katz e Martin (1997), afirmam que a colaboração científica pode reduzir custos e acaba aumentando os benefícios da pesquisa, por se tratar de um projeto que contará com pessoas diferentes trabalhando em conjunto em prol de um objetivo comum, mesclando e difundindo conhecimentos para ao avanço de uma pesquisa.

Dentre dos inúmeros tipos de colaboração científica, destaca-se a autoria múltipla. Além de ser frequentemente interpretada como um sinônimo de colaboração, é parametrizado como um indicador pela sua facilidade para mensuração (VILAN FILHO *et al*., 2008).

Para Vilan Filho *et al* (2008), a autoria múltipla pode ser delineada como um texto científico assinado por mais de um autor. Independentemente que durante uma colaboração científica não seja especificado o papel de contribuição de um indivíduo, a autoria múltipla tem servido como um indicador prático e com precisão para apuração da existência de colaboração.

4.3.1 Autoria Múltipla como Indicador de Qualidade

Sobre a autoria múltipla é importante ressaltar que seu conceito e aplicação possuem dois importantes pontos a serem destacados. Levando em consideração o conceito de forma generalizada, a autoria múltipla será todo o esforço de um pesquisador que houver se relacionado ou contribuído de forma significativa com um autor em forma colaborativa (MEADOWS, 1999).

Conforme Martins Filho (1998), no entendimento da lei de direito autoral, o segundo ponto a ser destacado acerca do conceito de autoria múltipla é de que, será considerado coautor aquele cujo o nome – pseudônimo ou sinal convencional – tenha sido utilizado requerendo a devida citação.

Para Vilan Filho (2010), a autoria múltipla é também denominada por diversos outros nomes – coautoria, autoria colaborativa, autoria em parceria ou colaboração – sendo observada na literatura científica sua associação com vários tópicos como colaboração, impacto, visibilidade e produtividade.

De acordo com Meadows (1999), o fator de pesquisadores com mais reconhecimento e produtividade ao trabalharem em colaborações faz com que, esses trabalhos possuam uma maior visibilidade e consequentemente sejam mais citados. Em relação a autoria múltipla, quanto maior for sua evolução, a tendência geral é do crescimento científico (VILAN FILHO, 2010), uma vez que, a colaboração de autores no mesmo trabalho cria uma maior carga cientifica para futuras pesquisas. No Quadro 4 é proposto um modelo com base nas implicações apresentadas por Vilan Filho (2010):

Quadro 4 - Implicações apresentadas pela Múltipla Autoria

Autoria Múltipla e Colaboração	*Pode ser usada somente como indicador parcial de colaboração mesmo com suas vantagens*
Autoria Múltipla e Produtividade	*Quanto maior o percentual de autoria múltipla de uma área, maior será sua produtividade global (maior número médio de publicações)*
Autoria Múltipla e Visibilidade	*A colaboração científica é vista como um indicador de visibilidade em bases de dados internacionais*
Autoria Múltipla e Impacto	*Possui efeito significativo por fazer com que produções com múltipla autoria tivessem maior impacto*

Fonte: Elaborado pelo autor 2021, com base em Vilan Filho (2010) e Meadows (1999).

Dessa forma, a autoria múltipla possui ligação direta com os indicadores de qualidade em produção científica por contar com algumas implicações que influenciam a coautoria. Em síntese, autores que trabalham juntos acabam utilizando o fator de colaboração, no entanto, não é apenas suficiente para ser determinado como um indicador de qualidade.

Ao contar com uma crescente carga científica outros fatores apontados por Vilan Filho (2010) são considerados. O fator de produtividade é um dos essenciais, uma vez que, deriva de uma nova gama de produção científica gerada por meio do número gradual de colaborações científicas fazendo com que, quanto maior for o número de produções de uma área, maior será sua produção em nível global.

Sobre o fator de visibilidade, pode ser considerado um dos mais importantes vistos que, em nível de bases de dados internacionais as colaborações científicas são extremamente bem vistas. No fator de visibilidade, cientistas podem trabalhar com outros cientistas em nível internacional, contribuindo a estas bases de dados.

Por fim, o impacto gerado por meio das autorias múltiplas é solidificado por métricas de citações. Um artigo científico possui seu reconhecimento por meio da quantidade de vezes que o artigo foi citado, aumentando o fator de impacto da revista na qual pertence e trazendo maior visibilidade e reconhecimento aos autores do artigo de periódico. Por meio da Figura 5, é proposto um modelo em quais fatores possuem maior peso nos indicadores de qualidade:

Figura 5 - Peso dos fatores de influência na autoria múltipla

Fonte: Elaborado pelo autor (2021) com base em Vilan Filho (2010).

4.4 Autoridade de Autor

Conforme Garcia *et al* (2010), no âmbito do meio acadêmico, a autoridade ganha extrema relevância, devido ao fato de que ao reconhecer, significa a valorização do esforço intelectual do autor. Nesta linha de pensamento, é de

extrema importância dar os créditos a um pesquisador que dedicou tempo e esforço intelectual para o avanço da ciência.

Por meio do reconhecimento da autoria é permitido a avaliação da produção científica daquele autor, geralmente utilizada como parâmetro para concessão de recursos por meio de agências de fomento a pesquisas, resultando na forma em como a visibilidade daquele autor é afetada no meio acadêmico. De acordo com o Garcia *et al*. (2010):

> Esse reconhecimento age em prol do estabelecimento e da sedimentação da reputação do pesquisador, que passa a ser legitimado no meio. Além disso, garante a continuidade de seus projetos, confere prestígio e reforça a possibilidade de aspirar a posições hierárquicas superiores em sua área de estudo (GARCIA *et al.*, 2010, p. 560).

Na próxima seção será apresentado em como a ambiguidade na forma como autores são citados pode acarretar em uma problemática para o histórico de produções nos quais pesquisadores possuem.

4.5 Ambiguidade de Autoria em Produções Científicas

Os programas de pós-graduação brasileiros inseridos na CAPES necessitam periodicamente que seja feito um levantamento das produções acadêmicas dos pesquisadores, grupos de pesquisa, projetos e entre outros presentes na plataforma de avaliação da CAPES (BRAUNER; ARAÚJO; SANTOS, 2016).

De acordo com Brauner, Araújo e Santos (2016), durante esses levantamentos, recomenda-se que haja um olhar mais crítico acerca das ambiguidades presentes ao alinhamento realizado por softwares. Essas ambiguidades costumam afetar publicações, nomes de eventos e o mais relevante para esta pesquisa, a ambiguidade entre nome de autores.

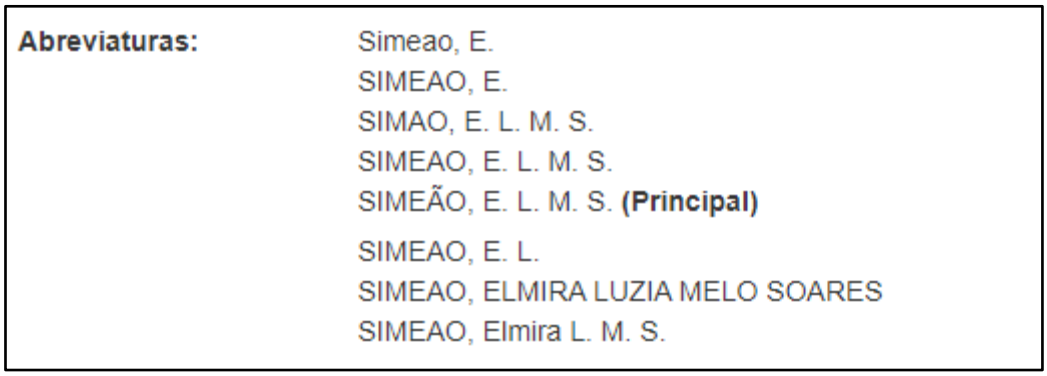
Conforme Mugnaini *et al* (2012), essa problemática acerca da ambiguidade entre nome de autores surge devido ao fato de que, um sistema de Ciência e Tecnologia de qualquer instituição ou país passa por uma avaliação

de sua produção científica. Essa produção necessita estar devidamente inserida e indexada em uma base de dados normalizada, diversificando olhares e indicadores, favorecendo sua apresentação e classificação.

Lee *et* al (2005) afirmam que o problema da ambiguidade de nomes pode haver mais um desdobramento, sendo classificados em dois subproblemas: *split citation* – quando um autor possui diversas variações em seu nome e *mixed citation* – quando autores diferentes possuem nomes iguais –, dessa forma agravando mais a complexidade da desambiguação.

A ocorrência do *split citation* pode ser demonstrada na Figura 6, onde um mesmo autor possui a variação de oito formas para ser citado. Esse tipo de problema é comum ocorrer com autores com sobrenomes incomuns ou quando possuem nomes extensos.

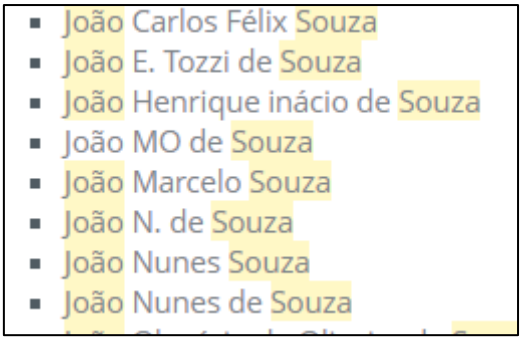
Figura 6 - Exemplo de *Split Citation*



Fonte: Elaborado pelo autor, 2021.

Sobre a outra ocorrência apresentada por Lee *et* al (2005), *mixed citation*, é apresentada na Figura 7. O sobrenome Souza é exibido em diversos resultados da busca por autores com este nome devido ao fato de ser um sobrenome comum. Nesse tipo de ocorrência é bastante comum a ambiguidade acerca da autoria visto que, ao utilizar apenas o sobrenome Souza e um nome, haverá diversos outros autores com o mesmo nome causando imprecisão na desambiguação.

Figura 7 - Exemplo de *Mixed Citation*



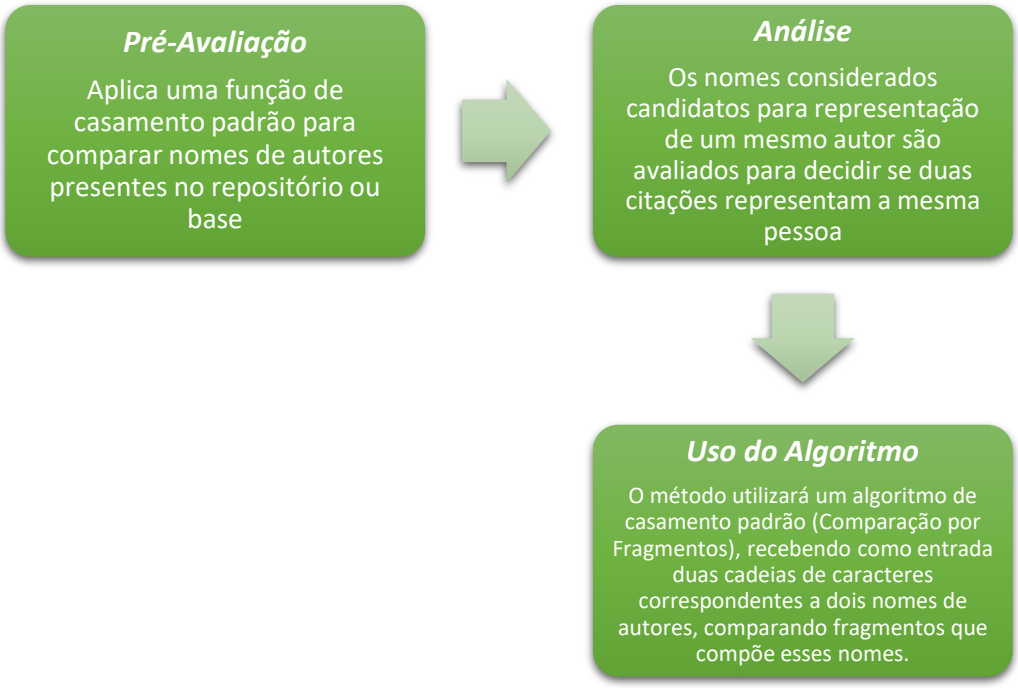
Fonte: Elaborado pelo autor, 2021.

De acordo com Mugnaini *et al* (2012), as causas da ambiguidade entre nome de autores, geralmente, ocorrem no próprio documento quando o autor se autodenomina de variadas formas em diferentes momentos (nome por extenso, abreviado, nome de casado, fantasia ou outras variações) ou quando base de dados geram de forma automática a forma nominal para citação.

A ambiguidade da autoridade torna-se um problema para análises bibliométricas, duplicidades em repositórios digitais e um dos mais críticos, na garantia de créditos em métricas de citações para pessoas errôneas.

Oliveira (2005) propõe como uma das soluções para desambiguação de nomes de autores a elaboração de um índice unificado, esse índice identificaria todas as citações de um autor, independendo na quantidade de variações que esse nome apresenta, durante o método. Na Figura 8 é apresentado como esse método funciona:

Figura 8 - Método Oliveira para Desambiguação de Nomes



Fonte: Elaborado pelo autor (2021) com base em Oliveira (2005).

Na próxima seção será abordado a forma como a bibliometria trabalha com a ambiguidade e desambiguação entre autores, e na forma como a bibliometria pode ser afetada por conta de duplicidades na montagem de indicadores de qualidade.

4.6 Bibliometria

De acordo com Beira *et al* (2020), a bibliometria teve seu advento no início do século XX, por meio de estudos desenvolvidos por Cole e Eales, durante 1917, e foi compreendida naquela época como uma disciplina designada para avaliação de livros.

Conforme Fonseca (1986), a bibliometria é definida como uma técnica quantitativa e estatística, visando mediar os índices de produção e disseminação do conhecimento científico.

Para Araújo (2016), a bibliometria consiste em técnicas estatísticas e matemáticas com finalidade a descrever aspectos da literatura e outros meios de comunicação – análise quantitativa da informação –, originalmente conhecida por Hulme em 1923 como ‘bibliografia estatística’ e por Otlet em 1934 como ‘bibliometria’.

Como dito anteriormente, a bibliometria era uma disciplina voltada para medida de livros – quantidade de edições, exemplares, quantidade de palavras contidas em um livro, espaços ocupados por livros em bibliotecas –, gradualmente foi se voltando para outros estudos em outros formatos de produção bibliográfica (ARAÚJO, 2006).

A bibliometria é utilizada para meios de pesquisas e pode atuar em conjunto com a ciência de dados, facilitando na análise de levantamento de dados a fim de conclusões científicas.

4.6.1 Três Leis Clássicas da Bibliometria

Alvarado (1984, p. 91) afirma que a bibliometria se fundamenta por meio de três leis básicas:

1. **A Lei de Bradford**: para descrever a distribuição da literatura periódica em uma área específica;
2. **A Lei de Lotka**: para descrever a produtividade de autores;
3. **A Lei de Zipf**: para descrever a frequência do uso de palavras em um determinado texto.

A Lei de Lotka, foi formulada em 1926, e construída a partir de um estudo sobre a produtividade dos cientistas, envolvendo a contagem de autores que estavam presentes no *Chemical Abstracts*. Por meio desta analise, Lotka se deu conta que uma grande proporção da literatura científica era produzida por um pequeno número de autores, e um grande número de pequenos produtores se

igualaria, em meios de proporção, ao número reduzido de grandes produtores (ARAÚJO, 2006).

De acordo com Araújo (2006), a segunda lei aborda o conjunto de periódicos – Lei de Bradford –, com objetivo de exibir a extensão na qual artigos científicos com assuntos específicos acabavam aparecendo em periódicos destinados a assuntos não-correlatos, estudando a distribuição de artigos levando em consideração as variáveis de proximidade ou afastamento.

Por fim, Araújo (2006) afirma que a terceira lei clássica da bibliometria – Lei de Zipf – foi formulada em 1949 e descreve a relação entre palavras inseridas em um texto determinado suficientemente grande e a ordem de série destas palavras.

Essas três leis clássicas compõem a bibliometria e suas implicações são usadas em vários contextos acadêmicos e científicos, a fim de auxiliar a análise bibliométrica, provendo valiosos resultados para o avanço científico.

4.7 Ciência de Dados

Os dados possuem uma grande relevância nos meios científicos, tendo como participação direta na ciência e em organizações. Essa importância sé dá por meio do tratamento de insumos essenciais para conduta e andamento de pesquisas – desde o processo de tomada de decisões até o avanço científico. Todavia, apenas nos tempos atuais a ciência de dados tem sido tratada como educação formal, destacando seus profissionais e suas importâncias no mercado de trabalho (CURTY; SERAFIM, 2016).

A ciência de dados *(data science)*, pode ser considerada como um reflexo do ambiente interconectado com sua enorme quantidade de dados disponíveis aos quais conhecemos. A ciência de dados trata-se de uma área interdisciplinar que consegue abordar áreas como as: ciências exatas e engenharias (Comarela *et al*., 2019).

A ciência de dados teve sua ascensão com o advento do desenvolvimento das tecnologias de informação e também das possibilidades de busca, por meio de mecanismos mais aprimorados como buscas avançadas (REIS, 2019).

Reis (2019) afirma que, a ciência de dados é considerada a ciência que reúne múltiplos aspectos de informações por meio de dados e contando como uma equipe multidisciplinar envolvendo profissionais de diversas áreas como: matemáticos, estatísticos, programadores, analistas de dados e bibliotecários.

Apesar do termo “ciência de dados” ser relativamente novo, a busca para compreensão desses dados por meio do trabalho de estatísticos, cientistas e profissionais da informação, já vinham sido abordados em um espaço de discussões antigas (PRESS, 2013; ROLIM, 2018).

Press (2013), afirma que durante 1962, John W. Tukey que era um estatístico norte-americano, defendia a precisão da realização de análises por meio de dados. Em sua obra *The future of data analysis*, é explicado como durante muito tempo o autor acreditava estar interessado apenas em inferências realizadas do particular para o geral, viabilizando apenas métodos provenientes da estatística clássica. Após determinado momento, ele percebeu que seu interesse se destacava na área de análise de dados. Nascendo então a ideia de que partes da estatística deviam passar por uma análise de dados, assumindo características de ciência, e deixando de ser vista apenas como dados brutos no ramo da matemática (ROLIM, 2018).

Analisando uma evolução histórica elaborada por Press (2013), é possível perceber como a ciência de dados teve sua evolução ao longo das últimas décadas. No Quadro 5 é proposto um modelo fundamentado nessa cronologia apresentando os princípios norteadores da atual ciência de dados:

Quadro 5 - Modelo Cronológico dos Princípios Norteadores da Atual Ciência de Dados

Década de 1960	Surgimento de livros, publicações seriadas, artigos, *workshops*, entidades e encontros de especialistas que abordam análise de dados e *big data;*
Em 1966	Lançamento do livro *from data mining to knowledge Discovery in databases* de Usama Fayyad, Gregory Piatetsky-Shapiro e Padhraic Smyth; o periódico *The Journal Data Mining and Knowledge Discovery*, lançado em 1997;
Em 1977	O livro *Exploratory data analysis*, em 1977, de John W. Tukey;
Em 2002	O periódico *Data Science Journal*;
Em 2005	O livro *Competing on analytics* de Thomas H. Davenport, Don Cohen e Al Jacobson;

Em 2009	O artigo *Rise of the data scientist*;
Em 2010	O artigo *What is Data Science?* de Mike Loukides;
Em 2012	O artigo *Data scientist: the sexiest job of the 21st century* de Thomas H. Davenport e D. J. Patil;

Fonte: Elaborado pelo autor (2021), com base em Press (2013) e Rolim (2018, p. 38).

Os autores Maneth e Poulovassilis (2017) afirmam que as principais dificuldades encontradas durante pesquisas em ciência de dados são as seguintes: desenvolver técnicas computacionais capazes de ordenar a gama de variedade e volume dos dados gerados por tecnologias em *web*, móveis e difusas; a proporção de dados produzidos por empresas de grande porte; a aplicação científica e das mídias sociais; a capacidade de desenvolvimento de ferramentas capazes de limpar, transformar, modelar, analisar, e trabalhar com esses dados de forma que cientistas de dados possam mediar com os produtos gerados por meio do *big data*, a fim de garantir segurança, privacidade dos dados de organizações e usuários (ROLIM, 2018).

Conforme Rolim (2018) e mediante a este contexto, nasce a necessidade de um profissional capaz de apresentar soluções e extrair valor dessa quantidade de dados, colaborando com o surgimento do novo campo de estudo chamado de ciência de dados.

O cientista de dados é considerado como a carreira mais "*sexy"* do século 21 (DAVENPORT; PATIL, 2012), essa afirmação tem como base a grande demanda existente no mercado por profissionais capazes de lidar com a *big data* e sua recorrente necessidade de obtenção de resultados a partir dos dados (ROLIM, 2018), que acabam gerando vantagens competitivas para as organizações pois auxiliam na prevenção de cenários e facilitam na tomada de decisões.

De acordo com Curty e Serafim (2016), a formação deste perfil profissional surgiu devido a demanda por profissionais que possuíam capacidade analítica e técnica a fim de lidar com grandes volumes de dados. No mercado atualmente é notável o crescimento desta área e o interesse por diversas instituições em profissionais capacitados na ciência de dados.

Por fim, evidenciando esse contexto de procura por especialistas em ciência de dados, foi realizado uma análise pelo *Google Trends* em 2012, mostrando uma crescente busca de usuários de diversos países, por informações acerca de termos como “*data scientist*” e “*data science*”, essas buscas sempre foram associadas com formação profissional envolvendo: cursos, salários, habilidades requeridas e certificação profissional (CURTY; SERAFIM, 2016; ROLIM, 2018). Desta forma, fica evidente em como organizações e profissionais buscam oportunidades emergentes dessa nova área.

5 METODOLOGIA

A metodologia utilizada neste estudo possui caráter quantitativo dado que, necessita da quantificação dos resultados para montar um padrão de resposta ao problema apresentado. Esta afirmação fundamenta-se no fato de que, para Fonseca (2002), considera-se em como a realidade só pode ser compreendida com base em uma análise de dados brutos, recolhidos com o auxílio de instrumentos neutros e padronizados.

Este trabalho se integra como uma pesquisa exploratória que proporciona o entendimento dos conceitos acerca de comunicação científica, colaboração científica e ciência de dados, além da compreensão em como a autoria múltipla funciona dentro da colaboração científica, destacando suas possíveis problemáticas na mensuração de ambiguidades de autoridades.

De acordo com Gerhardt e Silveira (2009), uma pesquisa com caráter exploratório investiga uma abordagem do fenômeno apresentado na pesquisa por meio do levantamento de informações, este levantamento feito por meio de bibliografias levam o pesquisador a entender melhor e conceituar sua pesquisa.

Sobre o procedimento adotado para este trabalho, utilizou-se como método de pesquisa bibliográfica. O levantamento bibliográfico foi por meio de: livros, bases de dados de acesso restrito e livre. Sendo elas: Brapci, Google Acadêmico, Portal de Periódicos da CAPES, Repositório Institucional de algumas universidades federais, mas majoritariamente foi utilizado de artigos científicos, para o embasamento da pesquisa, encontrados no Google Acadêmico, na Brapci ou no Portal de Periódicos da CAPES.

Os buscadores utilizados na pesquisa foram: “autoria múltipla”, “coautoria”, “métodos para desambiguação de autores”, “bibliometria para desambiguação”, “uso da ciência de dados para desambiguação”, “autoridade de autores”, “comunicação científica”, “comunicação científica AND comunicação científica”, “autoria múltipla como indicador de qualidade científica”.

Concomitantemente ao levantamento de dados, foram definidos por meio do esquema lógico relacional de dados montado no sistema gerenciador de

banco de dados MySQL, quais tabelas e campos seriam necessários para alimentação do banco de dados. Posteriormente a esta definição, foi pensado em quais base de dados seriam levantados estes dados.

Para a coleta de dados, foram extraídos, pela Plataforma Sucupira no Portal da Capes, dados informativos acerca da quantidade de programas de pós-graduação disponíveis sobre Ciência da Informação no Brasil, número de professores inseridos nestes programas de pós-graduação, e a quantidade de universidades integrantes dos cursos de pós-graduação em Ciência da Informação.

Após esse levantamento na Plataforma Sucupira, foram acessados os perfis na plataforma Lattes de cada um dos professores com a finalidade de extrair os números de ORCID. Para levantamento das variações de nomes dos professores (autores), foi utilizado da própria plataforma Sucupira que disponibiliza todas as variações disponíveis utilizadas pelos autores.

Após a extração e organização de todos esses dados apresentados, foi desenvolvido uma base de dados relacional que deu origem ao estudo de caso composto por um esquema relacional criado no sistema gerenciador de banco de dados (SGBD) MySQL.

O MySQL é um *software* que possibilita o uso de um sistema para gerenciamento de banco de dados e apresenta *open source*[3], possuindo licença GPL (*General Public License)*[4]. O MySQL conta com a linguagem SQL e apresenta portabilidade para a maioria das plataformas de linguagens de programação.

O esquema relacional criado tem por objetivo oferecer a organização e ordenação dos dados, para a criação de consultas utilizando a linguagem SQL e visualização dos dados por meio da ferramenta Tableau, a fim de apresentar por meio de gráficos a mensuração da ambiguidade entre autores.

[3] Código aberto.
[4] Licença disponível ao público.

6 ESTUDO DE CASO

Para este estudo de caso foram utilizados os dados provenientes da Plataforma Sucupira do Portal da CAPES, em razão da plataforma conter dados atualizados e correntes acerca dos programas de pós-graduações presentes em universidades federais do Brasil e seus respectivos docentes participantes desses programas.

O presente trabalho tem como objetivo exibir a ambiguidade em nomes de autores vigentes dos programas de pós-graduações presentes na base de dados da Plataforma Sucupira no ano de 2020. Para o desenvolvimento deste estudo de caso foram seguidos os seguintes passos:

- Criação de um modelo relacional de dados para representar o domínio observado;
- Levantamento de fontes de dados que possuem os dados representados no modelo de dados;
- Definição dos Atributos Persistentes para fins de Coleta de Dados;
- Extração dos dados para a carga das tabelas criadas no Banco de Dados Relacional.

6.1 Criação do Modelo Relacional

Nesta primeira etapa foi proposto um modelo relacional utilizado com base na técnica do modelo de entidade-relacionamento (modelo relacional) proposto por Peter Chen (1976).

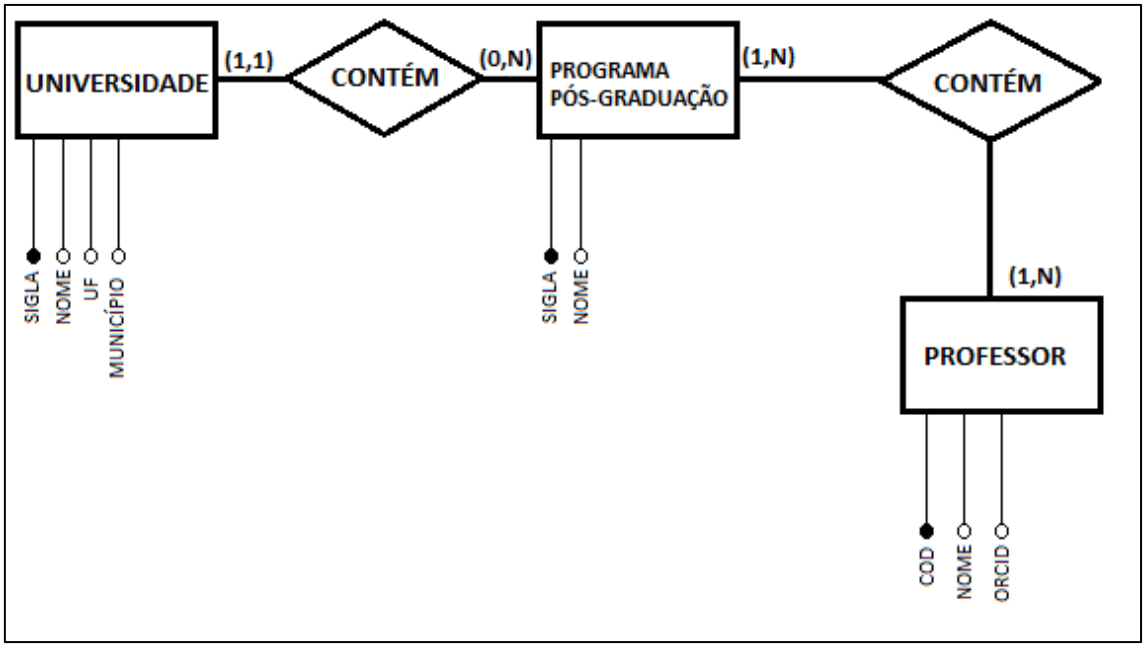
No modelo conceitual do presente estudo de caso, foi pensado em um modelo que possibilitasse a criação de um banco de dados apresentando os seguintes requisitos:

A. O banco de dados deve comportar professores (autores) de cada programa de pós-graduação presentes na Plataforma Sucupira;
B. O banco deve comportar os programas de pós-graduações presentes em cada universidade federal (sendo que algumas universidades possuem mais de um programa de pós-graduação em sua instituição);

C. O banco deve comportar cada universidade que conta com área correlatas acerca de Ciência da Informação;
D. O banco deve comportar uma entidade acerca das variações de nomes de cada professor (autor);

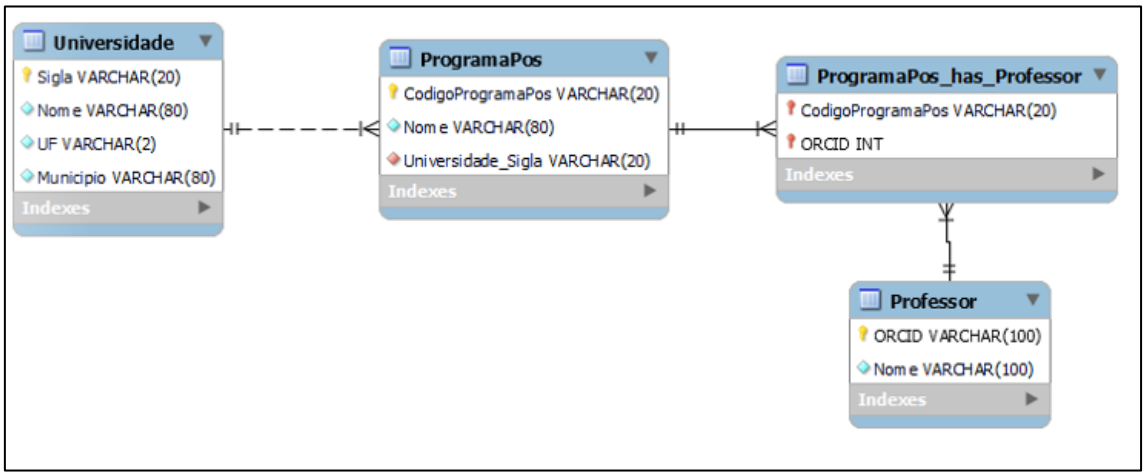
O modelo inicial proposto para o modelo relacional está apresentado na Figura 9, nesse protótipo ainda não havia sido pensado em uma entidade para inserção dos dados acerca das variações de nomes dos professores. A Figura 10 apresenta esse modelo conceitual mapeado para o nível lógico utilizando a ferramenta workbench do sistema gerenciador de banco de dados MySQL.

Figura 9 - Modelo Conceitual Inicial



Fonte: Elaborado pelo autor, 2021.

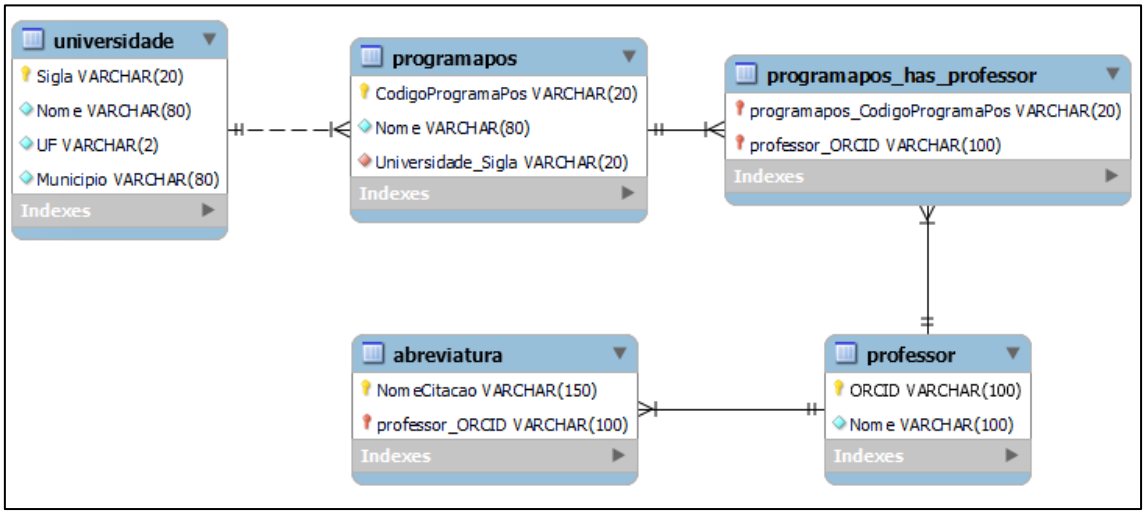
Figura 10 - Modelo Lógico apresentado no MySQL



Fonte: Elaborado pelo autor, 2021.

Após as esquematizações apresentadas, foi pensado em mais uma entidade que conteria os atributos necessários para modelação de uma tabela contendo as variações de ambiguidades dos nomes dos autores. Na Figura 11 é apresentado o modelo final lógico do banco de dados:

Figura 11 - Modelo Final Lógico



Fonte: Elaborado pelo autor, 2021.

6.2 Levantamento de fontes de dados que possuem os dados representados no modelo de dados

Nesta parte do estudo de caso foram selecionadas as fontes de dados que possuem os dados representados no modelo lógico apresentado na seção anterior, levando em consideração os requisitos apresentados.

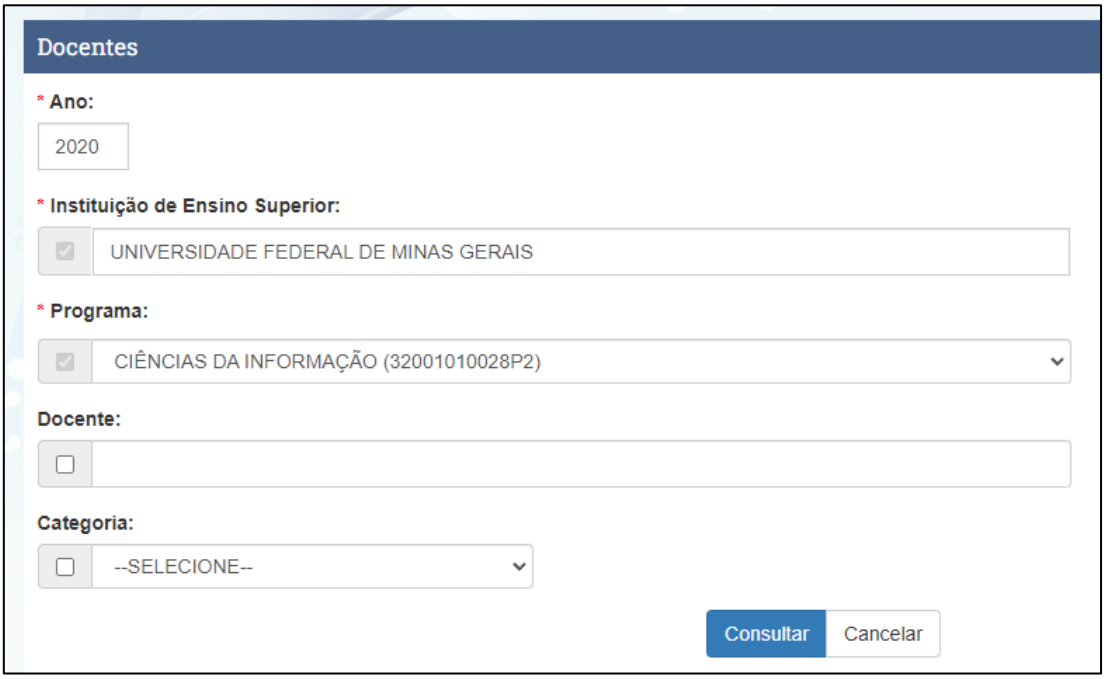
6.2.1 Fonte dos Dados para Professores

Para a entidade professores (autores), os dados vieram da Plataforma Sucupira do Portal da CAPES. Os requisitos para coleta de professores foram os seguintes:

- ✓ Ano: 2020
- ✓ Apenas as instituições de ensino superior presentes na Plataforma Sucupira e presentes em programas de pós-graduações acerca de Ciência da Informação;
- ✓ Categoria de condições de professores permanentes e colaboradores;

Na Figura 12 é apresentado como a busca de cada docente era simulada com base nos refinadores disponíveis pela Plataforma Sucupira:

Figura 12 - Metadados Utilizados para Levantamento dos Dados para Professores



Fonte: Elaborado pelo autor (2021) com base na Plataforma Sucupira do Portal da CAPES.

Após a busca apresentada na Figura 12, os resultados para os professores presentes em cada programa de pós-graduação eram apresentados de forma alfabética. A lupa mostrada na Figura 13 encaminha para maiores detalhes de cada docente participante e apresenta maiores especificidades como: titulação, vínculo com a IES, vínculo com o programa e quantitativo do docente.

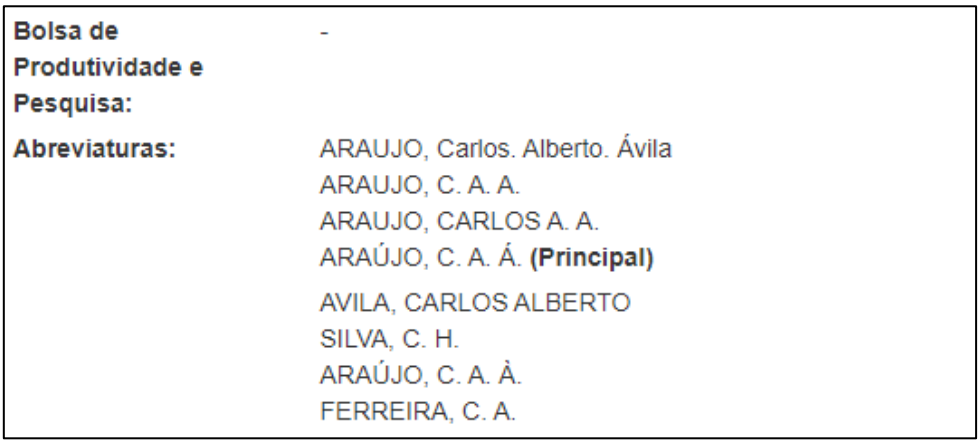
Figura 13 - Resultados das Buscas por Docentes na Plataforma SUCUPIRA

Docente	Categoria	
ADALSON DE OLIVEIRA NASCIMENTO	PERMANENTE	
ALCENIR SOARES DOS REIS	PERMANENTE	
ANA CECILIA NASCIMENTO ROCHA VEIGA	COLABORADOR	
ANA PAULA MENESES ALVES	PERMANENTE	
CARLOS ALBERTO AVILA ARAUJO	PERMANENTE	
CINTIA APARECIDA CHAGAS	PERMANENTE	
CLAUDIO PAIXAO ANASTACIO DE PAULA	PERMANENTE	
CRISTINA DOTTA ORTEGA	PERMANENTE	

Fonte: Elaborado pelo autor (2021) com base na Plataforma Sucupira do Portal da CAPES.

Os dados extraídos e persistente para a montagem do banco relacional foram as abreviaturas. Como mostra a Figura 14, foi por meio das abreviaturas que foi possível analisar as variações de nomes de cada autor e utilização para a alimentação do banco de dados.

Figura 14 - Abreviaturas dos Docentes



Bolsa de Produtividade e Pesquisa:	-
Abreviaturas:	ARAUJO, Carlos. Alberto. Ávila ARAUJO, C. A. A. ARAUJO, CARLOS A. A. ARAÚJO, C. A. Á. **(Principal)** AVILA, CARLOS ALBERTO SILVA, C. H. ARAÚJO, C. A. À. FERREIRA, C. A.

Fonte: Elaborado pelo autor (2021) com base na Plataforma Sucupira do Portal da CAPES.

Acerca da criação de uma chave primária, foi necessário pensar em um campo que seria único para cada docente, ou seja, este número não poderia variar e nem se repetir. Posteriormente, foi utilizado o ORCID de cada professor como sua chave primária. Os professores que não possuem ORCID utilizou-se a contagem numérica como identificador.

Encontrou-se o número do ORCID de cada docente por meio dos perfis disponíveis na Plataforma Lattes, alguns docentes disponibilizavam seu número de ORCID facilitando a modelagem dos dados. Na Figura 15 é apresentado um exemplo de docente com o número de ORCID:

Figura 15 - Exemplificação de ORCID Disponível de Docentes na Plataforma Lattes

Fonte: Elaborado pelo autor (2021) com base na Plataforma Lattes.

6.2.2 Fonte dos Dados para Universidades e Programas de Pós-Graduação

Para a entidade universidades, os dados vieram da Plataforma Sucupira do Portal da CAPES. Os requisitos para coleta de professores foram os seguintes:

- ✓ Cursos avaliados e conhecidos da área de Comunicação e Informação;
- ✓ Dentro da área de Comunicação e Informação foram coletadas apenas os cursos dentro da Ciência da Informação, excluindo os outros cursos (comunicação, desenho industrial, e museologia);
- ✓ Foram separados pelo código disponível de cada programa de pós-graduação para utilizar como sua chave primária;

Na Tabela 1 é apresentado o total de programas de pós-graduação e o total de cursos de pós-graduação inseridos na Plataforma Sucupira do Portal da Capes:

Tabela 1 - Tabela de Programas de Pós-Graduação e Cursos Disponíveis Sobre Ciência da Informação

			TOTAL DE PROGRAMAS DE PÓS-GRADUAÇÃO							TOTAIS DE CURSOS DE PÓS-GRADUAÇÃO				
NOME DA IES	**SIGLA DA IES**	**UF**	**TOTAL**	**ME**	**DO**	**MP**	**DP**	**ME/DO**	**MP/DP**	**TOTAL**	**ME**	**DO**	**MP**	**DP**
FUNDAÇÃO CASA DE RUI BARBOSA	FCRB	RJ	1	0	0	1	0	0	0	1	0	0	1	0
FUNDAÇÃO UNIVERSIDADE FEDERAL DE SERGIPE	FUFSE	SE	1	0	0	1	0	0	0	1	0	0	1	0
UNIVERSIDADE DE BRASÍLIA	UNB	DF	1	0	0	0	0	1	0	2	1	1	0	0
UNIVERSIDADE DE SÃO PAULO	USP	SP	2	0	0	1	0	1	0	3	1	1	1	0
UNIVERSIDADE DO ESTADO DE SANTA CATARINA	UDESC	SC	1	0	0	1	0	0	0	1	0	0	1	0
UNIVERSIDADE ESTADUAL DE LONDRINA	UEL	PR	1	0	0	0	0	1	0	2	1	1	0	0
UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE MESQUITA FILHO, MARÍLIA	UNESP-MAR	SP	1	0	0	0	0	1	0	2	1	1	0	0
UNIVERSIDADE FEDERAL DA BAHIA	UFBA	BA	1	0	0	0	0	1	0	2	1	1	0	0
UNIVERSIDADE FEDERAL DA PARAÍBA, JOÃO PESSOA	UFPB-JP	PB	1	0	0	0	0	1	0	2	1	1	0	0
UNIVERSIDADE FEDERAL DE ALAGOAS	UFAL	AL	1	1	0	0	0	0	0	1	1	0	0	0
UNIVERSIDADE FEDERAL DE MINAS GERAIS	UFMG	MG	2	0	0	0	0	2	0	4	2	2	0	0
UNIVERSIDADE FEDERAL DE PERNAMBUCO	UFPE	PE	1	0	0	0	0	1	0	2	1	1	0	0
UNIVERSIDADE FEDERAL DE SANTA CATARINA	UFSC	SC	1	0	0	0	0	1	0	2	1	1	0	0
UNIVERSIDADE FEDERAL DE SÃO CARLOS	UFSCAR	SP	1	1	0	0	0	0	0	1	1	0	0	0
UNIVERSIDADE FEDERAL DO CARIRI	UFCA	CE	1	0	0	1	0	0	0	1	0	0	1	0
UNIVERSIDADE FEDERAL DO CEARÁ	UFC	CE	1	1	0	0	0	0	0	1	1	0	0	0
UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO	UFES	ES	1	1	0	0	0	0	0	1	1	0	0	0
UNIVERSIDADE FEDERAL DO	UNIRIO	RJ	2	0	0	2	0	0	0	2	0	0	2	0

ESTADO DO RIO DE JANEIRO														
UNIVERSIDADE FEDERAL DO PARÁ	UFPA	PA	1	1	0	0	0	0	0	1	1	0	0	0
UNIVERSIDADE FEDERAL DO RIO DE JANEIRO	UFRJ	RJ	1	0	0	0	0	1	0	2	1	1	0	0
UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE	UFRN	RN	1	0	0	1	0	0	0	1	0	0	1	0
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL	UFRGS	RS	1	1	0	0	0	0	0	1	1	0	0	0
UNIVERSIDADE FEDERAL FLUMINENSE	UFF	RJ	1	0	0	0	0	1	0	2	1	1	0	0
UNIVERSIDADE FUMEC	FUMEC	MG	1	0	0	0	0	1	0	2	1	1	0	0
		Totais	**27**	**6**	**0**	**8**	**0**	**13**	**0**	**40**	**19**	**13**	**8**	**0**

Fonte: Elaborado pelo autor (2021) com base na Plataforma Sucupira do Portal da CAPES.

ME: Mestrado Acadêmico

DO: Doutorado Acadêmico

MP: Mestrado Profissional

DP: Doutorado Profissional

ME/DO: Mestrado Acadêmico e Doutorado Acadêmico

MP/DP: Mestrado Profissional e Doutorado Profissional

Nesta etapa foi possível coletar o nome de cada programa de pós-graduação em conjunto com seu código. As universidades que possuíam mais de um programa de pós-graduação foram cadastradas com seus respectivos programas e quais docentes pertenciam a cada um desses programa de pós-graduação.

6.3 Definição dos Atributos Persistentes para fins de Coleta de Dados

Durante esta etapa foram analisados quais dados seriam relevantes para extrações futuras e quais seriam persistidos no sistema gerenciador de banco de dados MySQL (SGBD).

Para cada entidade foram definidos os seguintes atributos:

- **Entidade Professor**: ORCID, Nome; Universidade
- **Entidade Programa Pós-Graduação**: Código Programa Pós, Nome e Sigla da Universidade;
- **Entidade Universidade**: Sigla, Nome, UF e Município;
- **Entidade Abreviatura**: Nome para Citação e ORCID;

Para os relacionamentos com chaves estrangeiras foram definidos os seguintes atributos:

- **Relacionamento Programa Pós-Graduação e Professor**: Código Programa Pós e ORCID Professor;

Os atributos mais importantes foram aqueles que eventualmente seriam úteis para a identificação da ambiguidade nos docentes inseridos no banco de dados (ORCID e Nome de Citação) como mostra na Figura 16:

Figura 16 - Atributos Pertinentes para Eventuais Consultas

ORCID	NomeCitacao
0000-0001-5709-9388	LYNCH, C. E. C.
0000-0001-5709-9388	LYNCH, C.
0000-0001-5709-9388	LYNCH, CHRISTIAN
0000-0002-4343-8390	SILVA, MARGARETH DA
0000-0002-4343-8390	SILVA, M.
0000-0002-4343-8390	Silva, Margareth da
0000-0002-4343-8390	SILVA., M.
0000-0002-4343-8390	PEREIRA, M. S.
0000-0002-4343-8390	SILVA, MARGARETH
0000-0002-5636-4343	RANGEL, A.
0000-0002-5636-4343	SOUZA, Marina.
0000-0002-5636-4343	RANGEL, A. M. S.

Fonte: Elaborado pelo autor, 2021.

6.4 Extração dos dados para a carga das tabelas criadas no Banco de Dados Relacional

Nesta etapa, os dados extraídos foram inseridos no sistema gerenciador de banco de dados (SGBD) MySQL por meio de tabelas criando uma serie de linhas para cada tabela.

A entidade ‘professor’ é apresentada na Figura 17 com suas respectivas fileiras e atributos alimentados e possuindo no total de 443 professores inseridos no SGBD.

Figura 17 - Tabela Professor e seus Atributos

ORCID	Nome	Universidade
0000-0001-5014-8853	Fabiano Couto Corrêa da Silva	UNIVERSIDADE FEDERAL DO RIO GRANDE DO ...
0000-0001-5057-9936	Thiago Magela Rodrigues Dias	UNIVERSIDADE FEDERAL DE SANTA CATARINA...
0000-0001-5099-112X	Gleice Pereira	UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
0000-0001-5301-1924	MARÍA MANUELA MORO-CABERO	UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE ...
0000-0001-5346-0826	JULIO AFONSO SA DE PINHO NETO	UNIVERSIDADE FEDERAL DA PARAÍBA, JOÃO P...
0000-0001-5361-0644	IVETE PIERUCCINI	UNIVERSIDADE DE SÃO PAULO
0000-0001-5364-3243	IEDA PELÓGIA MARTINS DAMIAN	UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE ...
0000-0001-5383-6723	Rosa da Penha Ferreira da Costa	UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO
0000-0001-5395-683X	Mariana Lousada	UNIVERSIDADE FEDERAL DO ESTADO DO RIO ...

Fonte: Elaborado pelo autor, 2021.

Na Figura 18 é apresentado a entidade ‘universidade’ e seus respectivos atributos e totalizando 23 universidades inseridas no SGBD.

Figura 18 - Tabela Universidade e seus Atributos

Sigla	Nome	UF	Municipio
FCRB	FUNDAÇÃO CASA DE RUI BARBOSA	RJ	Rio de Janeiro
FUFSE	FUNDAÇÃO UNIVERSIDADE FEDERAL DE SERGIPE	SE	São Cristóvão
UDESC	UNIVERSIDADE DO ESTADO DE SANTA CATARINA	SC	Florianópolis
UEL	UNIVERSIDADE ESTADUAL DE LONDRINA	PR	Londrina
UFAL	UNIVERSIDADE FEDERAL DE ALAGOAS	AL	Maceió
UFBA	UNIVERSIDADE FEDERAL DA BAHIA	BA	Salvador
UFC	UNIVERSIDADE FEDERAL DO CEARÁ	CE	Fortaleza
UFCA	UNIVERSIDADE FEDERAL DO CARIRI	CE	Juazeiro do Norte
UFES	UNIVERSIDADE FEDERAL DO ESPÍRITO SANTO	ES	Vitória

Fonte: Elaborado pelo autor, 2021.

Na Figura 19, é apresentado a entidade ‘Programa Pós’ e seus respectivos atributos. Essa entidade totaliza 26 registros, sendo que três universidades possuem dois programas de pós-graduação distintos em uma única instituição.

Figura 19 - Tabela Programa Pós e seus Atributos

CodigoProgramaPos	Nome	Universidade_Sigla
15001016158P5	CIÊNCIA DA INFORMAÇÃO	UFPA
22001018085P8	CIÊNCIA DA INFORMAÇÃO	UFC
22033017002P3	BIBLIOTECONOMIA	UFCA
23001011080P9	GESTÃO DA INFORMAÇÃO E DO CONHECIMENTO	UFRN
24001015049P7	CIÊNCIA DA INFORMAÇÃO	UFPB-JP
25001019077P3	CIÊNCIA DA INFORMAÇÃO	UFPE
26001012171P2	CIÊNCIA DA INFORMAÇÃO	UFAL
27001016175P0	CIÊNCIA DA INFORMAÇÃO	FUFSE
28001010041P0	CIÊNCIA DA INFORMAÇÃO	UFBA
30001013108P0	CIÊNCIA DA INFORMAÇÃO	UFES
31001017138P0	CIÊNCIA DA INFORMAÇÃO - UFRJ - IBICT	UFRJ

Fonte: Elaborado pelo autor, 2021.

A última tabela do SGBD foi a de entidade ‘abreviatura’ como apresentado na Figura 20 e totaliza 1342 registros:

Figura 20 - Tabela Abreviatura e seus Atributos

ORCID	NomeCitacao
0000-0002-3727-1571	ABREU, A. L.
0000-0002-3727-1571	ABREU, A. L. G.
0099	ACHILLES, D.
0000-0002-4859-4544	Adalson Nascimento
0045	Aganette, E.
0045	AGANETTE, E. C.
0045	AGANETTE, ELISANGELA CRISTINA
0032	ALBUQUERQU, M. E. B. C.
0000-0003-3506-0479	ALBUQUERQUE, A. C.
0000-0003-3506-0479	Albuquerque, A.C.
0000-0003-3506-0479	ALBUQUERQUE, ANA CRISTINA

Fonte: Elaborado pelo autor, 2021.

A tabela apresentada na Figura 21, trata-se de um relacionamento com chave estrangeira, essa tabela faz a união das seguintes chaves primárias: ‘CodigoProgramaPos’ e ‘Professor_ORCID’ das entidades ‘Programa Pos’ e ‘Professor’ totalizando 418 registros.

Figura 21 - Relacionamento Estrangeiro entre as Entidades 'Programa Pos' e 'Professor'

CodigoProgramaPos	Professor_ORCID
15001016158P5	0000-0001-5428-2451
15001016158P5	0000-0001-6277-2960
15001016158P5	0000-0001-7439-5779
15001016158P5	0000-0001-9634-1202
15001016158P5	0000-0002-6439-0058
15001016158P5	0000-0002-7314-6487
15001016158P5	0000-0003-0920-992X
15001016158P5	0000-0003-4545-4199

Fonte: Elaborado pelo autor, 2021.

Na Tabela 2 apresentada abaixo é sintetizado cada uma das entidades e relacionamentos com seus totais de registros:

Tabela 2 - Quantidade de Registros em Cada Tabela

Professor	443 registros
Universidade	23 registros
Programas Pós	26 registros
Abreviatura	1342 registros
Relacionamento Estrangeiro (Professor e Programa Pós)	418 registros

Fonte: Elaborado pelo autor, 2021.

7 RESULTADOS

No presente capítulo, apresentam-se os registros de ambiguidades dentro do banco de dados entre os professores inseridos nos programas de pós-graduação em Ciência da Informação do portal da CAPES. A seguinte demonstração será realizada por meio de consultas SQL para simular buscas de resultados específicos no banco de dados e enfim serem sintetizadas por meio de gráficos para apresentar os resultados finais para analises.

7.1 Consultas utilizando a Linguagem SQL

Ao realizar a consulta apresentada na Figura 22, foi possível a recuperação dos dados envolvendo as seguintes tabelas: ‘programapos’, ‘professor’ e o relacionamento entre ambas ‘programa_has_professor’. Esta consulta permitiu que os dados dessas três tabelas fossem combinados com base na relação entre elas - a chave estrangeira - fazendo com que os valores de cada tabela fossem associados entre si por meio da ligação dessas chaves.

Figura 22 - Consulta de Professores Pertencentes a seus Respectivos Programas de Pós-Graduação

```
select universidade.sigla as Universidade, programapos.nome
as ProgramaPos, professor.nome as Professor
from universidade, programapos, programapos_has_professor, professor
where universidade.sigla = programapos.Universidade_Sigla
and programapos.codigoprogramapos = programapos_has_professor.CodigoProgramaPos
and programapos_has_professor.orcid = professor.orcid
order by universidade.sigla
```

Fonte: Elaborado pelo autor, 2021.

Na Figura 23 é exibido parte dos resultados dessa seleção, apresentando como os dados são ordenados por meio das siglas das universidades:

Figura 23 - Apresentação dos Resultados das Tabelas 'CodigoProgramaPos', 'Professor' e 'ProgramaPos_Has_Professor'

Universidade	ProgramaPos	Professor
FCRB	MEMÓRIA E ACERVOS	CHRISTIAN EDWARD CYRIL LYNCH
FCRB	MEMÓRIA E ACERVOS	MARGARETH DA SILVA
FCRB	MEMÓRIA E ACERVOS	APARECIDA MARINA DE SOUZA RANGEL
FCRB	MEMÓRIA E ACERVOS	LUIS FERNANDO SAYÃO
FCRB	MEMÓRIA E ACERVOS	LIA CALABRE DE AZEVEDO
FCRB	MEMÓRIA E ACERVOS	ANA MARIA PESSOA DOS SANTOS
FCRB	MEMÓRIA E ACERVOS	ANTONIO HERCULANO LOPES
FCRB	MEMÓRIA E ACERVOS	ANA LIGIA SILVA MEDEIROS
FCRB	MEMÓRIA E ACERVOS	ISABEL IDELZUITE LUSTOSA DA COSTA
FUFSE	CIÊNCIA DA INFORMAÇÃO	JEFFERSON DAVID ARAUJO SALES
FUFSE	CIÊNCIA DA INFORMAÇÃO	TELMA DE CARVALHO
FUFSE	CIÊNCIA DA INFORMAÇÃO	ALESSANDRA DOS SANTOS ARAUJO
FUFSE	CIÊNCIA DA INFORMAÇÃO	MARTHA SUZANA CABRAL NUNES

Fonte: Elaborado pelo autor, 2021.

A consulta da Figura 24 teve como proposito apresentar os programas de pós-graduação relacionados com cada universidade pertencente.

Figura 24 - Consulta de Programas de Pós-gradução pertencentes a cada Universidade

```
SELECT universidade.sigla, programapos.nome
from universidade, programapos
where universidade.sigla = programapos.Universidade_Sigla
order by universidade.sigla
```

Fonte: Elaborado pelo autor, 2021.

Na Figura 25 é possível a visualização de parte dos resultados apresentando cada universidade federal e seu respectivo programa de pós-graduação.

Figura 25 - Parte dos Resultados de quais Programas de Pós-graduação e suas respectivas Universidades pertencentes

sigla	nome
FCRB	MEMÓRIA E ACERVOS
FUFSE	CIÊNCIA DA INFORMAÇÃO
UDESC	GESTÃO DA INFORMAÇÃO
UEL	CIÊNCIA DA INFORMAÇÃO
UFAL	CIÊNCIA DA INFORMAÇÃO
UFBA	CIÊNCIA DA INFORMAÇÃO
UFC	CIÊNCIA DA INFORMAÇÃO
UFCA	BIBLIOTECONOMIA
UFES	CIÊNCIA DA INFORMAÇÃO
UFF	CIÊNCIA DA INFORMAÇÃO
UFMG	CIÊNCIAS DA INFORMAÇÃO
UFMG	GESTÃO & ORGANIZAÇÃO D...
UFPA	CIÊNCIA DA INFORMAÇÃO
UFPB...	CIÊNCIA DA INFORMAÇÃO
UFPE	CIÊNCIA DA INFORMAÇÃO

Fonte: Elaborado pelo autor, 2021.

Ao realizar a consulta apresentada na Figura 26, foi possível selecionar os dados da tabela ‘universidade’ e ‘programapos’ com proposito de retornar à quantidade de programas de pós graduação em cada universidade.

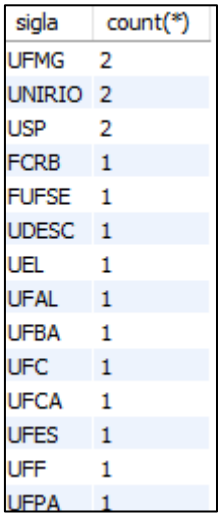
Figura 26 - Consulta de Quantidade de Programas de Pós-graduação em cada Universidade

```
SELECT universidade.sigla, count(*)
from universidade, programapos
where universidade.sigla = programapos.universidade_sigla
group by universidade.sigla
order by count(*) desc
```

Fonte: Elaborado pelo autor, 2021.

Na Figura 27 é apresentado parte dos resultados dessa consulta ordenados de forma crescente:

Figura 27 - Apresentação dos Resultados da Quantidade de Programas de Pós-graduação em cada Universidade



sigla	count(*)
UFMG	2
UNIRIO	2
USP	2
FCRB	1
FUFSE	1
UDESC	1
UEL	1
UFAL	1
UFBA	1
UFC	1
UFCA	1
UFES	1
UFF	1
UFPA	1

Fonte: Elaborado pelo autor, 2021.

Na consulta apresentada da Figura 28 é possível retomar a quantidade de professores participantes de programas de pós-graduações por cada universidade. Isto é possível graças a seleção das tabelas ‘universidade’, ‘programapos’ e o relacionamento entre elas: ‘programapos_has_professor’.

Figura 28 - Consulta da Quantidade de Professores participantes de cada Programa de Pós-Graduação

```
SELECT universidade.sigla, programapos.nome, count(*)
from universidade, programapos, programapos_has_professor
where universidade.sigla = programapos.Universidade_Sigla
and programapos.codigoprogramapos = programapos_has_professor.CodigoProgramaPos
group by universidade.sigla, programapos.nome
order by count(*) desc
```

Fonte: Elaborado pelo autor, 2021.

Na Figura 29 é apresentado parte dos resultados referentes a consulta da Figura 28, nota-se que a função ‘*count(*)*’ foi utilizada para a contagem de professores em cada programa de pós-graduação.

Figura 29 - Parte dos Resultados referentes a Quantidade de Professores por Programas de Pós-Graduação e suas respectivas Universidades

sigla	nome	count(*)
UNESP-MAR	CIÊNCIA DA INFORMAÇÃO	35
UFPB-JP	CIÊNCIA DA INFORMAÇÃO	26
UFSC	CIÊNCIA DA INFORMAÇÃO	25
UNB	CIÊNCIAS DA INFORMAÇÃO	22
UFBA	CIÊNCIA DA INFORMAÇÃO	22
UNIRIO	BIBLIOTECONOMIA	20
FCRB	MEMÓRIA E ACERVOS	19
UDESC	GESTÃO DA INFORMAÇÃO	19
UFF	CIÊNCIA DA INFORMAÇÃO	17
UFPE	CIÊNCIA DA INFORMAÇÃO	16
UFRJ	CIÊNCIA DA INFORMAÇÃO -...	16
FUFSE	CIÊNCIA DA INFORMAÇÃO	15
UFRGS	CIÊNCIA DA INFORMAÇÃO	15

Fonte: Elaborado pelo autor, 2021.

Na Figura 30 é apresentada a consulta utilizada para apresentar as variações de citações de cada professor. Nota-se que o '*order by*' foi utilizado para ordenar os nomes dos professores de forma que cada variação ficasse agrupada com o professor de origem daquele nome.

Figura 30 - Consulta dos Tipos de Variações de Nomes para Citações de cada Professor

```
SELECT professor.nome, abreviatura.nomecitacao
from professor, abreviatura
where professor.orcid = abreviatura.ORCID_Professor
order by professor.nome
```

Fonte: Elaborado pelo autor, 2021.

Na Figura 31 é apresentado parte dos resultados do agrupamento das variações de nomes referentes a cada professor:

Figura 31 - Parte dos Resultados referentes a Variações de Nomes de cada Professor

nome	nomecitacao
ADRIANA ROSECLER ALCARÁ ENGELMANN	ALCARÁ ENGELMANN, A. R. A.
ADRIANA ROSECLER ALCARÁ ENGELMANN	ALCARÁ, A. R.
ADRIANA ROSECLER ALCARÁ ENGELMANN	ALCARA, ADRIANA
ADRIANA ROSECLER ALCARÁ ENGELMANN	ALCARA, ADRIANA R.
ADRIANA ROSECLER ALCARÁ ENGELMANN	ALCARA, ADRIANA ROSECLER
ADRIANA ROSECLER ALCARÁ ENGELMANN	ENGELMANN, Adriana Rosecler ...
ADRIANA ROSECLER ALCARÁ ENGELMANN	ENGUELMANN, A. R. A.
Alberto Calil Elias Junior	CALIL JUNIOR, A.
Alberto Calil Elias Junior	CALIL JUNIOR, ALBERTO
Alcenir Soares dos Reis	REIS, A. S.
Alcenir Soares dos Reis	REIS, ALCENIR SOARES
Alcenir Soares dos Reis	REIS, ALCENIR SOARES DOS
Alegria Celia Benchimol	BENCHIMOL, A.
Alegria Celia Benchimol	BENCHIMOL, A. C.
Alegria Celia Benchimol	BENCHIMOL, ALEGRIA
Alegria Celia Benchimol	BENCHIMOL, ALEGRIA CELIA

Fonte: Elaborado pelo autor, 2021.

Na Figura 32 é apresentado a consulta utilizada para retomar os resultados da quantidade de variações de nomes para cada professor. Essa consulta possibilitou a ordenação dos dados de forma decrescente.

Figura 32 - Consulta da Quantidade de Variações de Nomes para cada Professor

```
SELECT professor.nome, count(*)
from professor, abreviatura
where professor.orcid = abreviatura.ORCID_Professor
group by professor.nome
order by count(*) desc
```

Fonte: Elaborado pelo autor, 2021.

Na Figura 33 é apresentado parte dos resultados referente a quantidade de variações dos nomes de cada professor. Nota-se que os resultados foram ordenados de forma decrescente onde o professor com maior número de variações conta com 12 variações.

Figura 33 - Parte dos Resultados referentes a Quantidade de Variações de Nomes para cada Professor

nome	count(*)
RICARDO CÉSAR GONÇALVES SANT'ANA	12
Gercina Ângela de Lima	11
SILVANA APARECIDA BORSETTI GREGÓRI...	10
Douglas Dyllon Jeronimo De Macedo	10
MARTA LÍGIA POMIM VALENTIM	10
Eliana Silva de Almeida	9
ÂNGELA MARIA GROSSI DE CARVALHO	9
Benildes Coura Moreira dos Santos Maculan	9
Leandro Innocentini Lopes de Faria	9
Maria Teresa Navarro de Britto Matos	8
Renata Maria Abrantes Baracho Porto	8
Maria Giovanna Guedes Farias	8
Frederico Cesar Mafra Pereira	8
ADRIANA ROSECLER ALCARÁ ENGELMANN	7
Carlos Henrique Juvêncio da Silva	7
ROGÉRIO HENRIQUE DE ARAÚJO JÚNIOR	7

Fonte: Elaborado pelo autor, 2021.

Na Figura 34 por meio da consulta apresentada, é possível calcular a média de variações dos nomes de todos os professores presentes na tabela, sendo a média apresentada no valor de 3.38.

Figura 34 - Consulta Referente a média de Variações de Nomes dos Professores

```
SELECT avg(qtd) from professor_citacao_qtd
```

Fonte: Elaborado pelo autor, 2021.

Na Figura 35 é exibido o comando SQL para a criação da visão ‘professor_citacao_qtd’. Essa visão apresenta os nomes dos professores e suas quantidades de variações, sendo criada com propósito para facilitação do cálculo da média de variações dos nomes de todos os professores.

Figura 35 - Comando SQL para Criação da Visão 'Professor_Citacao_Qtd'

```
CREATE
    ALGORITHM = UNDEFINED
    DEFINER = `root`@`localhost`
    SQL SECURITY DEFINER
VIEW `professor_citacao_qtd` AS
    SELECT
        `professor`.`Nome` AS `nome`, COUNT(0) AS `qtd`
    FROM
        (`professor`
        JOIN `abreviatura`)
    WHERE
        (`professor`.`ORCID` = `abreviatura`.`ORCID_Professor`)
    GROUP BY `professor`.`Nome`
```

Fonte: Elaborado pelo autor, 2021.

Na Figura 36 é apresentado parte dos resultados da visão criada, exibindo o nome de cada professor seguido pela quantidade de variações que cada nome representa.

Figura 36 - Parte dos Dados da Tabela Visão ‘Professor_Citacao_Qtd’

nome	qtd
ANA LÚCIA DE ABREU GOMES	6
Daniele Achilles Dutra da Rosa	2
Adalson de Oliveira Nascimento	4
Elisângela Cristina Aganette	4
MARIA ELIZABETH BALTAR CARNEIRO DE ...	6
ANA CRISTINA DE ALBUQUERQUE	3
ADRIANA ROSECLER ALCARÁ ENGELMANN	7
Evelyn Goyannes Dill Orrico	5
JOSE ALMINO DE ALENCAR E SILVA NETO	4
OSWALDO FRANCISCO DE ALMEIDA JÚNIOR	5
MARCO ANTÔNIO ALMEIDA	4
CARLOS CÂNDIDO DE ALMEIDA	4
Carlos Henrique Juvêncio da Silva	7

Fonte: Elaborado pelo autor, 2021.

Na Figura 37 é apresentado uma busca para filtrar professores que possuíam o sobrenome Silva:

Figura 37 - Consulta dos Resultados Apresentando o Sobrenome 'Silva'

```
select professor.nome, abreviatura.NomeCitacao
from professor, abreviatura
where professor.orcid = abreviatura.ORCID_Professor
and NomeCitacao Like '%SILVA%'
group by professor.nome
```

Fonte: Elaborado pelo autor, 2021.

Na Figura 38 é apresentado parte dos resultados envolvendo o sobrenome ‘Silva’ para diferentes professores:

Figura 38 - Variações do sobrenome ‘Silva’ para diferentes Professores

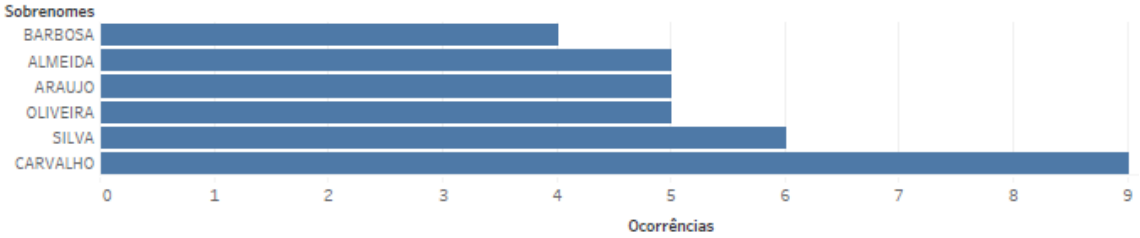
nome	NomeCitacao
JOSE ALMINO DE ALENCAR E SILVA NETO	SILVA NETO, José Almino de Alencar e
ALZIRA KARLA ARAUJO DA SILVA	SILVA, A. K. A.
Carlos Alberto Ávila Araújo	SILVA, C. H.
Carlos Henrique Juvêncio da Silva	SILVA, C. H. J.
Eliezer Pires da Silva	SILVA, E
Elieny Do Nascimento Silva	SILVA, E. DO N.
Eliane Ferreira da Silva	SILVA, E. F.
Fabiano Couto Corrêa da Silva	SILVA, F. C. C.
Fábio Mascarenhas e Silva	SILVA, F. M.
JOSÉ FERNANDO MODESTO DA SILVA	SILVA, J. F. M.
MARGARETH DA SILVA	SILVA, M.
MARIA CLÁUDIA CABRINI GRÁCIO	SILVA, MARIA CLAUDIANE DA
Rubens Alves da Silva	SILVA, R. A.
Rubens Ribeiro Gonçalves da Silva	SILVA, RUBENS R. G. DA
Sérgio Franklin Ribeiro da Silva	SILVA, S. F.
TEREZINHA ELISABETH DA SILVA	SILVA, T. E.

Fonte: Elaborado pelo autor, 2021.

7.2 Apresentação de Gráficos Analíticos

No Gráfico 1, observa-se que o sobrenome com maior número de ocorrências no banco de dados foi ‘Carvalho’ com nove registros de autores diferentes. Este gráfico exclui as duplicatas de sobrenome de um mesmo autor para evidenciar apenas professores diferentes com o mesmo sobrenome.

Gráfico 1 - Sobrenomes com maiores Ocorrências

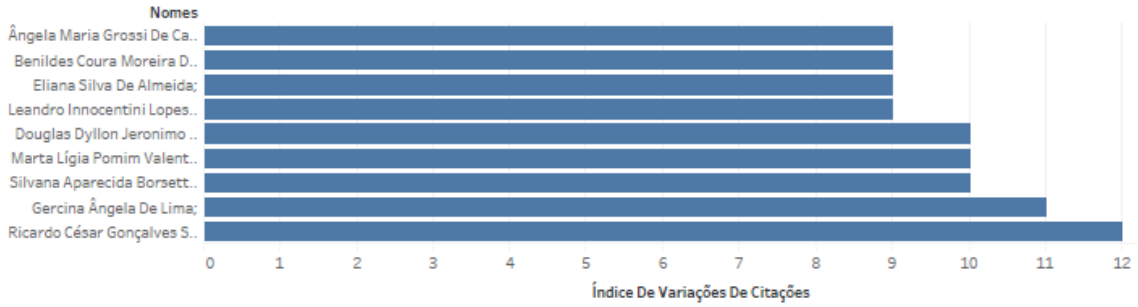


Fonte: Elaborado pelo autor, 2021.

Este caso é um exemplo de *mixed citation* apontado por Lee *et al* (2005), evidenciando que caso um professor opte por utilizar apenas o sobrenome ‘Carvalho’ e ‘ano’ para citações, pode haver ambiguidades com outros professores registrados em bases de dados.

No Gráfico 2, observa-se que este gráfico representa os nove autores com maiores números de variações de nomes para citações. Este é um caso de *split citation* apontado por Lee *et al* (2005), cujo um autor possui diversas formas de ser citado, sendo este problema causado por autores que possuem sobrenomes extensos ou incomuns.

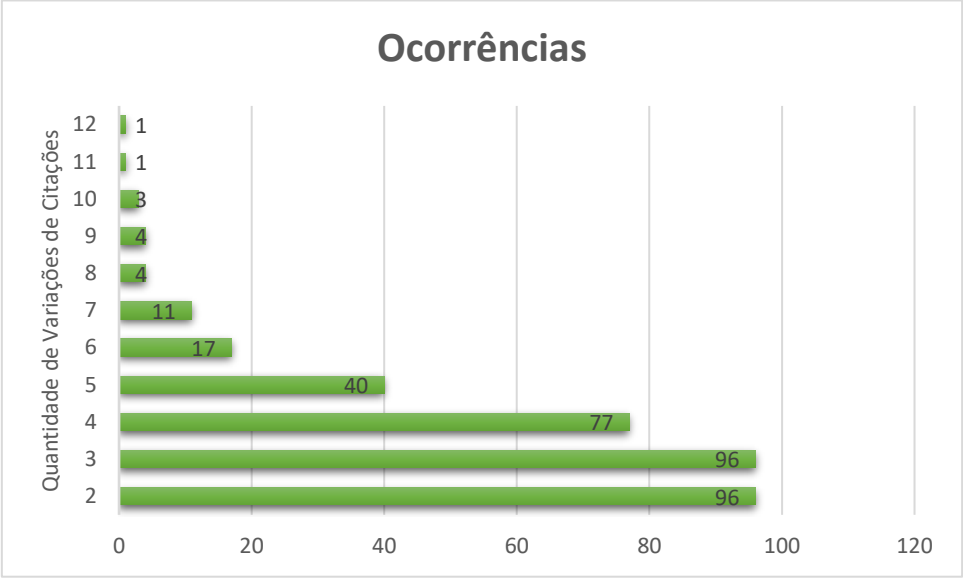
Gráfico 2 - Autores com Maiores Índices de Variações em Citações



Fonte: Elaborado pelo autor, 2021.

No Gráfico 3, observa-se a quantidade de registros dentro do banco de dados (Quantidade de Variações de Citações) em comparação com a quantidade de vezes que foram encontrados esses registros (Ocorrências). Desta forma é possível analisar que os professores com intervalos de 2 até 5 formas de variações foram os mais exibidos no banco de dados totalizando 309 registros.

Gráfico 3 - Índice de Variações de Citações



Fonte: Elaborado pelo autor, 2021.

Em decorrência as análises realizadas anteriormente, verificou-se neste trabalho várias ocorrências de *mixed citation* predominantemente em

sobrenomes considerados comuns nos professores inseridos no banco de dados relacional, sendo os sobrenomes com maiores ocorrências exemplificados no Gráfico 1.

A ocorrência de *split citation* foi exibida na maioria das variações nominais dos professores presentes no banco de dados relacional. No Gráfico 3 exemplifica esta ocorrência durante os intervalos de variações nominais, onde considera-se um valor acima da média, na ocorrência de 5-12 tipos de variações pelo fato de que a maioria dos autores estiveram na margem de 2-4 variações.

8 CONSIDERAÇÕES FINAIS

Neste trabalho foi possível a visualização de dois tipos de ambiguidades que costumam aparecer com frequência em base de dados científicas. Essas duas ambiguidades acabam acarretando em problemas de mensurações em métricas para indicadores de qualidade que acabam afetando os índices de produções científicas, sendo consideradas objetos de estudos significantes para pesquisas dentro da Ciência da Informação.

Por meio das técnicas de ciência de dados, foi possível a visualização das ambiguidades descritas entre os autores presentes na base de dados relacional, provenientes do sistema gerenciador de banco de dados e criado com dados de origem da Plataforma Sucupira do Portal de Periódicos da CAPES e dados captados de outras plataformas. Esta etapa concluiu um dos objetivos específicos proposto no trabalho em obtenção dos dados sobre a produção científica dos professores presentes em bases de dados abertas.

As maiores dificuldades encontradas no desenvolver deste trabalho foram as coletas dos dados, por se tratarem de dados que precisavam passar por uma validação e inserção manual dentro do banco de dados relacional. Esta tarefa exigiu meses de desenvolvimento até a entrega do produto final.

Outro ponto considerado uma dificuldade para a progressão do trabalho se deu por meio de definir quais atributos seriam pertinentes para cada tabela montada, facilitando as eventuais consultas. Desta forma, após definir quais atributos definitivos integrariam o banco de dados relacionais e a criação da base de dados relacional, foram realizadas consultas SQL que possibilitaram gerar gráficos e tabelas para a analise das ambiguidades dentro do banco de dados confirmando a existência do problema previamente apresentado. Esta etapa concluiu os três últimos pontos dos objetivos específicos propostos.

Este trabalho conseguiu alcançar todos os objetos específicos propostos, exibindo como as técnicas de ciência de dados se mostram indispensáveis para expor a existência de padrões de duplicatas, organização dos dados, e por fim para a limpeza dos dados. Este trabalho busca auxiliar em pesquisas futuras

dentro da Ciência da Informação acerca da ambiguidade em autores e suas problemáticas.

9 REFERÊNCIAS

ALVARADO, Rubén Urbizagástegui. A Bibliometria no Brasil. **Ciência da Informação**, v. 13, n. 2, 1984. Disponível em: http://revista.ibict.br/ciinf/article/view/200. Acesso em: 1 set. 2021.

ARAÚJO, Carlos A. A. Bibliometria: evolução histórica e questões atuais. **Em Questão**, v. 12, n. 1, p. 11–32, 10 dez. 2006. Disponível em: https://seer.ufrgs.br/EmQuestao/article/view/16. Acesso em: 1 set. 2021.

BALANCIERI, Renato; BOVO, Alessandro Botelho; KERN, Vinícius Medina; PACHECO, Roberto Carlos dos Santos; BARCIA, Ricardo Miranda. A análise de redes de colaboração científica sob as novas tecnologias de informação e comunicação: um estudo na Plataforma Lattes. **Ciência da Informação**, v. 34, p. 64–77, jan. 2005. Disponível em: https://doi.org/10.1590/S0100-19652005000100008. Acesso em: 1 set. 2021.

BAPTISTA, Ana Alice; COSTA, Sely Maria de Souza; KURAMOTO, Hélio; RODRIGUES, Eloy. Comunicação científica : o papel da Open Archives Initiative no contexto do acesso livre. 2007. Disponível em: https://repositorio.unb.br/handle/10482/635. Acesso em: 17 ago. 2021.

BEIRA, Joana Carlos; GONTIJO, Marília Catarina Andrade; ANNA, Jorge Santa; MACULAN, Benildes Coura Moreira. Indicadores bibliométricos na produção científica em periódicos brasileiros da Ciência da Informação no Estrato A1. **Revista ACB**, v. 25, n. 2, p. 383–408, 22 jul. 2020. Disponível em: https://revista.acbsc.org.br/racb/article/view/1660. Acesso em: 17 ago. 2021.

BORKO, H. Information science: What is it? **American Documentation**, v. 19, n. 1, p. 3–5, 1968. Disponível em: https://doi.org/10.1002/asi.5090190103. Acesso em: 17 ago. 2021.

BRAUNER, Daniela Francisco; ARAÚJO, Ricardo Matsumura de; SANTOS, Glauco Roberto Munsberg dos. Alinhamento de nomes de coautores de publicações científicas : uma abordagem prática. 2016. Disponível em: https://lume.ufrgs.br/handle/10183/169161. Acesso em: 16 ago. 2021.

COMARELA, Giovanni; FRANCO, Gabriel; TROIS, Celio; LIBERATO, Alextian; MARTINELLO, Magnos; CORRÊA, João Henrique; VILLAÇA, Rodolfo. Introdução à Ciência de Dados: Uma Visão Pragmática utilizando Python, Aplicações e Oportunidades em Redes de Computadores. *In*: SCHAEFFER FILHO, Alberto; CORDEIRO, Weverton Luis; CAMPISTA, Miguel Elias (orgs.). **Minicursos do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**. 1. ed. [*S. l.*]: SBC, 2019. p. 246–295. DOI 10.5753/sbc.6555.9.6. Disponível em: https://sol.sbc.org.br/livros/index.php/sbc/catalog/view/65/289/538-1. Acesso em: 18 set. 2021.

CURTY, Renata Gonçalves; SERAFIM, Jucenir Da Silva. A formação em ciência de dados: uma análise preliminar do panorama estadunidense.

Informação & Informação, v. 21, n. 2, p. 307, 20 dez. 2016. Disponível em: https://doi.org/10.5433/1981-8920.2016v21n2p307. Acesso em: 18 set. 2021.

DAVENPORT, Thomas H.; PATIL, D. J. Data Scientist: The Sexiest Job of the 21st Century. **Harvard Business Review**, , seç. Analytics and data science, 1 out. 2012. Disponível em: https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century. Acesso em: 26 set. 2021.

DIAS, Eduardo José Wense. Biblioteconomia e Ciência da Informação: natureza e relações. v. 5, n. Perspectivas em Ciência da Informação, 20 nov. 2007. Disponível em: http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/556/338. Acesso em: 18 set. 2021.

FONSECA, Edson Nery. **Bibliometria. Teoria E Pratica**. 1ª edição. [*S. l.*]: Cultrix, 1986.

FONSECA, J. J. S. **Metodologia da pesquisa científica**. Fortaleza: UEC, 2002. Apostila. Disponível em: http://www.ia.ufrrj.br/ppgea/conteudo/conteudo-2012-1/1SF/Sandra/apostilaMetodologia.pdf. Acesso em: 27 set. 2021.

FRIAS, Ana. A ESCRITA CIENTÍFICA E A DIVULGAÇÃO DO CONHECIMENTO CIENTÍFICO. **Cogitare Enfermagem**, v. 20, n. 2, 28 jun. 2015. DOI 10.5380/ce.v20i2.41922. Disponível em: http://revistas.ufpr.br/cogitare/article/view/41922. Acesso em: 21 ago. 2021.

GARCIA, Carla Costa; MARTRUCELLI, Cristina Ribeiro Nabuco; ROSSILHO, Marilisa de Melo Freire; DENARDIN, Odilon Victor Porto. Autoria em artigos científicos: os novos desafios. **Brazilian Journal of Cardiovascular Surgery**, v. 25, p. 559–567, dez. 2010. Disponível em: https://doi.org/10.1590/S0102-76382010000400021. Aceso em: 18 set. 2021.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo (Org). **Métodos de pesquisa**.

GOMES, S. L. R.; MENDONÇA, M. A. R.; SOUZA, C. M. Literatura cinzenta. *In*: CAMPELLO, B. S.; CENDÓN, B. V.; KREMER, J. M. (orgs.). **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte: Ed. UFMG, 2007. p. 97–114.

GRÁCIO, Maria Claudia Cabrini. Colaboração científica: indicadores relacionais de coautoria. **Brazilian Journal of Information Science: research trends**, v. 12, n. 2, 1 ago. 2018. DOI 10.36311/1981-1640.2018.v12n2.04.p24. Disponível em: https://revistas.marilia.unesp.br/index.php/bjis/article/view/7976. Acesso em: 19 ago. 2021.

HERNÁNDEZ, Mauricio A.; STOLFO, Salvatore J. The merge/purge problem for large databases. **ACM SIGMOD Record**, v. 24, n. 2, p. 127–138, 22 maio 1995. Disponível em: https://doi.org/10.1145/568271.223807. Acesso em: 18 set. 2021.

KATZ, J. Sylvan; MARTIN, Ben R. What is research collaboration? **Research Policy**, v. 26, n. 1, p. 1–18, 1 mar. 1997. Disponível em: https://doi.org/10.1016/S0048-7333(96)00917-1. Acesso em: 18 set. 2021.

KÖHLER, André Fontan; DIGIAMPIETRI, Luciano Antonio. Classificação de autores, instituições e países, por meio de métricas de produção, centralidade e impacto: o campo de turismo no Brasil (periódicos), 1990-2018. **Revista Brasileira de Pesquisa em Turismo**, v. 15, n. 3, p. 2035–2035, 2 jun. 2021. Disponível em: https://doi.org/10.7784/rbtur.v15i3.2035. Acesso em: 18 set. 2021.

KOHLER, Andre Fontan; DIGIAMPIETRI, Luciano Antonio. O campo de turismo no Brasil: caracterização e análise da rede de pesquisadores e sua dinâmica regional. **Perspectivas em Ciência da Informação**, v. 26, n. 2, p. 58–82, 30 jun. 2021. Disponível em: https://periodicos.ufmg.br/index.php/pci/article/view/34989. Acesso em: Acesso em: 27 ago. 2021.

KOSEOGLU, Mehmet Ali; RAHIMI, Roya; OKUMUS, Fevzi; LIU, Jingyan. Bibliometric studies in tourism. **Annals of Tourism Research**, v. 61, p. 180–198, nov. 2016. Disponível em: https://doi.org/10.1016/j.annals.2016.10.006. Acesso em: Acesso em: 27 ago. 2021.

LEE, Dongwon; ON, Byung-Won; KANG, Jaewoo; PARK, Sanghyun. Effective and scalable solutions for mixed and split citation problems in digital libraries. 17 jun. 2005. **Proceedings of the 2nd international workshop on Information quality in information systems** [...]. New York, NY, USA: Association for Computing Machinery, 17 jun. 2005. p. 69–76. DOI 10.1145/1077501.1077514. Disponível em: https://doi.org/10.1145/1077501.1077514. Acesso em: 27 ago. 2021.

LEITE, Fernando César Lima; COSTA, Sely Maria de Souza. Gestão do conhecimento científico: proposta de um modelo conceitual com base em processos de comunicação científica. **Ciência da Informação**, v. 36, p. 92–107, abr. 2007. Disponível em: https://doi.org/10.1590/S0100-19652007000100007. Acesso em: Acesso em: 27 ago. 2021.

MARTINS FILHO, Plínio. Direitos autorais na Internet. **Ciência da Informação**, v. 27, n. 2, 1998. Disponível em: http://revista.ibict.br/ciinf/article/view/800. Acesso em: 24 ago. 2021.

MEADOWS, Arthur Jack. **A comunicação científica**. trad. Antonio Agenor Briquet de Lemos Lemos. Brasília: Briquet de Lemos/livros, 1999.

MIRANDA, Antonio. O Campo da ciência da informação: gênese, conexões e especialidades. *In*: AQUINO, Mirian de Albuquerque (org.). **A ciência da informação e a teoria do conhecimento objetivo: um mal necessário**. João Pessoa: Editora Universitária/UFPB, 2002. p. 9–24.

MUELLER, S. P. M.; PASSOS, E. J. L. As questões da comunicação científica e a ciência da informação. *In*: MUELLER, S. P. M.; PASSOS, E. J. L. (orgs.). **Comunicação científica**. Brasília: Ciência da Informação, 2000. p. 13–22.

MUELLER, Suzana Pinheiro Machado. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, v. 35, p. 27–38, ago. 2006. Disponível em: https://doi.org/10.1590/S0100-19652006000200004. Acesso em: Acesso em: 27 ago. 2021.

MUELLER, Suzana. Literatura científica, comunicação científica e ciência da informação. **Para entender a ciência da informação**. Salvador: EDUFBA, 2007.

MUGNAINI, Rogerio; DIGIAMPIETRI, Luciano Antonio; OLIVEIRA, Laucivaldo Cardoso de; FERREIRA, Sueli Mara Soares Pinto. Normalização de nomes de autores em fontes de informação institucionais: proposta de um método automático de verificação de erros. **Em Questão**, v. 18, n. 3, p. 263–279, 2012. Disponível em: https://seer.ufrgs.br/EmQuestao/article/view/33265. Acesso em: 25 out. 2021.

OLIVEIRA, Jean Wanderlei Alves de. Uma estratégia para remoção de ambiguidades na identificação de autoria de objetos bibliográficos. 1 abr. 2005. Disponível em: https://repositorio.ufmg.br/handle/1843/RVMR-6EAGQK. Acesso em: 27 ago. 2021.

PINHEIRO, Lena Vania Ribeiro; LOUREIRO, José Mauro Matheus. Traçados e limites da ciência da informação. **Ciência da Informação**, v. 24, n. 1, 1995. Disponível em: http://revista.ibict.br/ciinf/article/view/609. Acesso em: 27 set. 2021.

PRESS, Gil. A Very Short History Of Data Science. [*s. d.*]. **Forbes**. Disponível em: https://www.forbes.com/sites/gilpress/2013/05/28/a-very-short-history-of-data-science/. Acesso em: 26 set. 2021.

REIS, Makson de Jesus. Ciência de dados e ciência da informação: guia para alfabetização de dados para bibliotecários. 12 jul. 2019. Disponível em: https://ri.ufs.br/jspui/handle/riufs/12667. Acesso em: 1 set. 2021.

ROBREDO, Jaime. **Da Ciência da Informação. Revisitada aos Sistemas Humanos de Informação**. Brasília: Thesaurus, 2003.

ROLIM, Mesaque Vidal. Análise do perfil do profissional da informação para a atuação como cientista de dados em ambientes de big data : uma perspectiva a partir das disciplinas do curso de Biblioteconomia da UnB. 3 jul. 2018. Disponível em: https://bdm.unb.br/handle/10483/20898. Acesso em: 26 set. 2021.

SARACEVIC, Tefko. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, v. 1, n. 1, 1992. Disponível em: http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235. Acesso em: 27 set. 2021.

SCHWEITZER, Fernanda; RODRIGUES, Rosângela Schwarz; VARVAKIS, Gregório Jean. Comunicação científica e as tecnologias de informação e comunicação. **Comunicação & Sociedade**, v. 32, n. 55, p. 83–104, 27 jun. 2011. Disponível em: https://doi.org/10.15603/2175-7755/cs.v32n55p83-104. Acesso em: 25 out. 2021.

SEMELER, Alexandre Ribas. **Ciência da informação em contextos de e-science: bibliotecários de dados em tempos de Data Science**. 2017. Universidade Federal de Santa Catarina, Tese (doutorado) - Universidade Federal de Santa Catarina, Centro de Ciências da Educação, Programa de Pós-Graduação em Ciência da Informação, Florianópolis, 2017. Disponível em: https://repositorio.ufsc.br/handle/123456789/185593. Acesso em: 12 set. 2021.

SILVA, Antonio Braz de Oliveira e; PARREIRAS, Fernando Silva; MATHEUS, Renato Fabiano; BRANDÃO, Wladmir Cardoso. Redes de co-autoria dos professores da ciência da informação: um retrato da colaboração científica dessa disciplina no brasil. 28 mar. 2013. Disponível em: http://repositorios.questoesemrede.uff.br/repositorios/handle/123456789/704. Acesso em: 19 ago. 2021.

SILVA, José Renato. Escrita Científica. **Revista Ciências em Saúde**, v. 9, n. 2, p. 1–2, 7 jul. 2019. Disponível em: https://doi.org/10.21876/rcshci.v9i2.857. Acesso em: 25 out. 2021.

TARGINO, Maria das Graças. COMUNICAÇÃO CIENTÍFICA: uma revisão de seus elementos básicos. **Informação & Sociedade: Estudos**, 30 jan. 2000. Disponível em: https://periodicos.ufpb.br/ojs/index.php/ies/article/view/326. Acesso em: 17 ago. 2021.

VALEIRO, Palmira Moriconi; PINHEIRO, Lena Vania Ribeiro. Da comunicação científica à divulgação. **Transinformação**, v. 20, p. 159–169, ago. 2008. Disponível em: https://brapci.inf.br/index.php/res/v/116012. Acesso em: 25 out. 2021.

VANTI, Nadia Aurora Peres. Da bibliometria à webometria: uma exploração conceitual dos mecanismos utilizados para medir o registro da informação e a difusão do conhecimento. **Ciência da Informação**, v. 31, p. 369–379, ago. 2002. Disponível em: https://doi.org/10.1590/S0100-19652002000200016. Acesso em: 25 out. 2021.

VANZ, Samile Andrea de Souza; STUMPF, Ida Regina Chittó. Colaboração científica: revisão teórico-conceitual. **Perspectivas em Ciência da Informação**, v. 15, p. 42–55, ago. 2010. Disponível em: https://doi.org/10.1590/S1413-99362010000200004. Acesso em: 25 out. 2021.

VICTORINO, Marcio de Carvalho. **Organização da Informação para dar Suporte à Arquitetura Orientada a Serviços: Reuso da Informação nas Organizações**. 2011. 280 f. Tese de doutorado – Universidade de Brasília, Brasília, 2011.

VILAN FILHO, Jayme Leiro. Autoria múltipla em artigos de periódicos científicos das áreas de informação no Brasil. 24 ago. 2010. Disponível em: https://repositorio.unb.br/handle/10482/7468. Acesso em: 24 ago. 2021.

VILAN FILHO, Jayme Leiro; SOUZA, Held Barbosa de; MUELLER, Suzana. Artigos de periódicos científicos das áreas de informação no Brasil: evolução da produção e da autoria múltipla. **Perspectivas em Ciência da Informação**, v. 13, p. 2–17, ago. 2008. Disponível em: https://doi.org/10.1590/S1413-99362008000200002. Acesso em: 25 out. 2021.