

A PRODUÇÃO DAS MULHERES NA CIÊNCIA DA INFORMAÇÃO A PARTIR DE UMA ANÁLISE DESCRITIVA BASEADA EM MINERAÇÃO DE DADOS

THE PRODUCTION OF WOMEN IN INFORMATION SCIENCE FROM A DESCRIPTIVE ANALYSIS BASED ON DATA MINING

Douglas Farias Cordeiro^a
Kátia Kelvis Cassiano^b

RESUMO

Objetivo: O artigo propõe a realização de um estudo descritivo sobre a participação da mulher na produção científica, em termos de produção bibliográfica e de orientações, no âmbito da Ciência da Informação. **Metodologia:** Utiliza um percurso descritivo, baseado no processo *Knowledge Discovery in Databases*, com o emprego de soluções de mineração de dados para a análise da participação da mulher na produção científica, especificamente na Ciência da Informação. **Resultados:** Os resultados obtidos apresentam indicadores em termos da produtividade no contexto da Ciência da Informação, abordando diferentes visões de acordo com tipos específicos de produção. **Conclusão:** As contribuições alcançadas pelas pesquisadoras na área de Ciência são destacadas, principalmente para produções em periódicos especializados e formação em nível de pós-graduação, demonstrando a representatividade e importância da mulher para o crescimento da área.

Descritores: Gênero. Ciência da Informação. Produção Científica. Informação Científica. Análise Descritiva.

1 INTRODUÇÃO

Por muito tempo a ciência foi considerada um campo de atuação masculina e, embora algumas mulheres tenham participado significativamente

^a Doutor em Ciência da Computação e Matemática Computacional pela Universidade de São Paulo (USP). Docente do Programa de Pós-Graduação em Comunicação da Universidade Federal de Goiás (UFG), Goiânia, Brasil. E-mail: cordeiro@ufg.br

^b Doutora em Engenharia de Sistemas e Computação pela Universidade Federal do Rio de Janeiro (UFRJ). Docente do Programa de Pós-Graduação em Comunicação da Universidade Federal de Goiás (UFG), Goiânia, Brasil. E-mail: katiakelvis@ufg.br

no seu desenvolvimento, nem sempre tiveram o devido reconhecimento. Conforme ressaltado por Valdés e Rubio (2018) o caso de Rosalind Franklin, uma química britânica cuja contribuição não foi reconhecida no anúncio da dupla hélice do DNA, uma das descobertas mais importantes para a biotecnologia nos anos 1950. Cada vez mais as mulheres estudam e praticam ciência. Segundo resultados divulgados pela Elsevier (2017), acerca da produção científica em 12 regiões geográficas e 27 áreas de conhecimento, Brasil e Portugal têm os maiores percentuais (49%) de participação feminina na produção científica dentre os países selecionados. O estudo considerou o quadriênio 2011-2015 e, especificamente para o Brasil, foi observado que 38% das pesquisas e produções acadêmicas do país foram realizadas por mulheres.

Diversos instrumentos em nível mundial têm incentivado a participação feminina na ciência, revelando um entendimento da importância de gerar e apoiar modelos que estimulem a diversidade e incorporem o reconhecimento das mulheres na comunidade científica. Exemplo disso é o programa *For Women in Science da Fundação L'Oréal* e a Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO) que promove, desde 1998, ações de apoio à pesquisa e reconhecimento às mulheres pesquisadoras que, no âmbito de seu trabalho, contribuiram para superar os desafios globais. Nesse sentido, o programa evidencia estratégias de aceleração e incentivos financeiros, por meio da concessão de bolsas de estudos e prêmios.

Nessa mesma vertente, a Fundação Elsevier promove a participação de mulheres na ciência, definindo modelos para desenvolvimento de suas carreiras desde os estágios iniciais. Em 2010, lançou o prêmio *Elsevier Foundation Awards for Early-Career Women Scientists in the Developing World* que reconhece - anualmente e de forma rotativa entre as disciplinas de ciência da vida, química, física e matemática - cinco cientistas, uma de cada das seguintes regiões: América Latina e Caribe, Leste e Sudeste Asiático e Pacífico, Região Árabe, Ásia Central e do Sul, África Subsaariana (OWSD, 2019).

Ainda neste contexto, se destaca o programa *Women in Science*, lançado no Brasil pelo British Council em 2018 e com previsão de duração de três anos. O objetivo do programa é estimular a participação feminina em disciplinas de

ciência, tecnologia, engenharia e matemática desde a educação fundamental, de forma que, na universidade, possam seguir seus interesses nessas áreas. Segundo Morales (2019), o programa promove a geração de capacidades, por meio da agenda científica para mulheres e fortalecimento de redes nacionais e internacionais, apresentando uma forma mais diversificada e representativa de se fazer ciência, que tem como base a troca de experiências e conhecimento entre mulheres de instituições do Brasil e do Reino Unido.

No Brasil, diversas políticas públicas implementadas na última década visando a promoção da equidade de gênero contribuíram para a participação feminina na ciência, conforme relatado por Romeiro e Viola (2018). O Programa Mulher e Ciência, lançado em 2005 pela Secretaria de Promoção da Mulher em parceria com o Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e diversos órgãos relacionados à Educação, Ciência e Tecnologia, incentivou a participação feminina em carreiras acadêmicas e fomentou projetos de pesquisa de estudos de gênero. Por meio dessas ações, o Brasil se tornou pioneiro entre os países com igualdade de gênero no nível de doutorado, com um aumento de 104% no número de matrículas femininas nos programas no biênio 2004-2005 (GROSSI *et al.*, 2016). Apesar disso, cabe destacar que as mulheres seguem maioria nas áreas relacionadas ao cuidado, trato humano e relacionadas à maternagem, enquanto os homens são maioria nas áreas mais exatas, sendo estas áreas estatisticamente mais bem remuneradas no mercado.

O Plano Nacional de Educação, proposto em 2010, definiu ações a serem executadas no período de 2011-2020 para erradicação das discriminações históricas e abolição dos estereótipos de gênero, representando um marco no fomento à participação feminina na comunidade científica brasileira. No que tange à visibilidade das publicações científicas, o estudo conduzido por Carvalho *et al.* (2018) revelou que cerca de metade das publicações brasileiras do quadriênio 2011-2015 foram de autoria de mulheres, um aumento expressivo comparado aos 38% do período 1996-2000. No período de 2011 a 2015, mulheres foram autoras de 50% dos artigos publicados na Scopus (LOPÉZ-BASSOLS *et al.*, 2018) e no período de 2014 a 2017 foram autoras de 72% dos artigos publicados na Web of Science, consolidando o Brasil como o país ibero-

americano com maior porcentagem de artigos científicos que incluem pelo menos uma mulher como autora principal ou coautora (ALBORNOZ *et al.*, 2018).

Apesar desses resultados serem motivadores, ainda se mantêm as diferenças de gênero no que diz respeito à inclusão nos diversos campos profissionais e áreas de conhecimento científico. Lopéz-Bassols *et al.* (2018) demonstraram, por meio de análise de dados do Instituto de Estatística da UNESCO, que na América Latina e Caribe a representatividade das mulheres é significativamente maior nas ciências sociais e humanidades, se comparada às áreas tecnológicas e de engenharia. Ainda assim, conforme relatório divulgado pela Elsevier (2017), apesar de as mulheres conquistarem espaço na comunidade científica, é baixa a participação no que tange à tecnologia e inovação. Em nível mundial, 28% dos pedidos de patente no período de 2011-2015 incluíam pelo menos uma mulher entre os inventores - no Brasil esta proporção era de 19%.

Neste contexto, um dos grandes desafios é a construção de análises e reflexões que estejam focadas em áreas específicas, de modo a construir compreensões que permitam descrever características e particularidades. Diante disso, no âmbito da Ciência da Informação, uma série de trabalhos tem se proposto a investigar a participação da mulher, ou mesmo questões de gênero sob um aspecto mais geral, conforme é descrito ao longo deste trabalho. Em Santo (2008), por exemplo, é realizado um estudo acerca das pesquisas científicas que exploram a temática gênero dentro da Ciência da Informação. Nesta mesma perspectiva, Lima e Dias (2013) apresentam um estudo sobre as produções que promovem debates entre gênero e informação. Por outro lado, em Siciliano, Souza e Meth (2017), é proposto um estudo sobre as contribuições e explorações da temática gênero no âmbito da Ciência da Informação. Em Silva e Romeiro (2018), são apresentados estudos e reflexões que abordam tanto a presença e contribuição da mulher dentro da Ciência da Informação, quanto análises sobre diferentes perspectivas.

Embora existam diversas contribuições e estudos sobre a participação da mulher na Ciência da Informação, ainda existem desafios no que tange a análises baseadas em métodos que possam prover informações embasadas em

dados quantitativos e qualitativos provenientes de registros de produção científica. Diante disso, este artigo tem como objetivo principal apresentar um estudo descritivo sobre a participação das mulheres na área Ciência da Informação no Brasil, por meio da análise de dados referentes aos anos de 1999 a 2018, extraídos de seus currículos disponibilizados na Plataforma Lattes. Foi realizada análise das produções científicas das mulheres nesta área considerando a natureza e a relevância das mesmas. Espera-se que, a partir das análises, fontes de informação sejam identificadas de forma a prover indicadores de visibilidade das mulheres na área de conhecimento Ciência da Informação e fomentar estudos associados.

2 MULHERES E CIÊNCIA DA INFORMAÇÃO

O conceito “Ciência da Informação” tem origem precípua na ação humana de registrar seus conhecimentos, o que se denomina cultura humana, uma vez que ela preconiza a elaboração de documentos, sua salvaguarda e organização (ARAÚJO, 2014). Embora as primeiras instituições dedicadas à salvaguarda das produções humanas tenham surgido durante o Renascimento, somente no século XIX surgiram as disciplinas Arquivologia, Biblioteconomia e Museologia como ciências relacionadas aos seus objetivos. Araújo (2014) defende que a Ciência da Informação tenha surgido em oposição a estas disciplinas, em uma proposta de promover a difusão do conhecimento contido nas obras custodiadas. Nesse sentido, ressalta que nas décadas de 1920 a 1940 diversos cientistas atuavam elaborando índices e resumos, como forma de criar meios democráticos para disseminar a informação e agilizar o trabalho de seus pares.

Assim, como é incontestável a interdisciplinaridade da Ciência da Informação, Queiroz e Moura (2015) destacam que a consolidação desta área de conhecimento foi consequência da demanda informacional da ciência moderna, sendo o avanço tecnológico seu principal catalisador. Ademais, Araújo (2014) defende duas características próprias da Ciência da Informação que refletem sua identidade e autonomia em relação às outras áreas: a preocupação com a disseminação e promoção dos documentos de forma produtiva e o foco no conteúdo destes documentos, o que lhe confere a busca constante pela

informação com valor agregado.

Os primeiros registros da área no Brasil datam da década de 1970, com a criação do Mestrado em Ciência da Informação pelo atual Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) e o lançamento do periódico *Ciência da Informação* em 1972. No que tange à participação feminina nesta área de conhecimento, se percebe um movimento ao longo dos anos na tentativa de ressignificar o protagonismo feminino e sua importante contribuição histórica para o desenvolvimento da área.

Nesse sentido, duas personalidades femininas protagonizam a história da Ciência da Informação no Brasil: Lydia Sambaquy e Celia Zaher. Conforme apresentado por Corrêa e Oliveira (2018), Lydia Sambaquy se destacou na projeção da Biblioteconomia brasileira em nível nacional e internacional por meio da publicação do Esquema de Organização da Biblioteca do Departamento Administrativo do Serviço Público (DASP) e sua atuação visionária na década de 1940 culminou na reformulação do curso de Biblioteconomia, contemplando disciplinas práticas voltadas ao processamento técnico das tarefas de catalogação e classificação, atividades de gestão de bibliotecas e, de forma inovadora, a bibliografia e referência.

Estas ações inovadoras foram determinantes para a criação do atual Ibict como uma instituição de promoção da informação científica, abrindo portas para a emergência da Ciência da Informação no Brasil. Assim, no entendimento de Oddone (2006), a consolidação da área de conhecimento Ciência da Informação no Brasil é consequência da ruptura causada pelo distanciamento entre a biblioteconomia tradicional e as práticas profissionais de Lydia Sambaquy no Ibict.

Igualmente notória, é apontada por Corrêa e Oliveira (2018) a contribuição da cientista Celia Zaher na implantação do Mestrado em Ciência da Informação em 1970, o principal marco do estabelecimento da Ciência da Informação no Brasil. Segundo destacado pelos autores, a atuação desta pesquisadora apresenta aspectos futuristas a respeito das tendências para a área, tendo se destacado na prestação de serviços em organizações renomadas nacionais e internacionais, como o CNPq e UNESCO, e na aplicação da tecnologia aplicada

no âmbito da organização e recuperação da informação, fomentando a informatização dos processos nas bibliotecas.

Sob o olhar da representatividade feminina no âmbito da ciência, Robredo e Vilan Filho (2010) apresentaram a evolução da produção científica nas áreas de informação por sexo e no período de 1980 a 2007, sendo observado um equilíbrio proporcional nas produções de autoria única e maior proporção de mulheres nas autorias múltiplas. O estudo exploratório e descritivo conduzido por Bedin, Feldman e Vianna (2018) apresentaram um mapeamento das publicações disponibilizadas na Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), com objetivo de identificar como a temática Gestão da Informação é tratada sob a perspectiva das mulheres.

Embora tenha se percebido que apenas 8% das autoras têm trajetória de pesquisa em Gestão da Informação, os resultados apresentados por Bedin, Feldman e Vianna (2018) evidenciaram a participação ativa de mulheres na área Ciência da Informação: dos 304 artigos recuperados na base de dados selecionada, 63% foram escritos por mulheres (primeira autoria). Os resultados daquele estudo demonstraram que embora a representatividade das autoras em relação à trajetória de pesquisa em Gestão da Informação seja baixa (apenas 8% das autoras são genuinamente formadas na área), se evidencia qualidade nos estudos conduzidos uma vez que a maioria das publicações de autoria feminina foram realizadas em periódicos de alta relevância na área (Qualis A1 CAPES). Por fim, o estudo ressalta a natureza teórica das contribuições científicas, evidenciando uma carência de pesquisas aplicadas que utilizem a informação para resolver problemas reais da sociedade.

Sena, Cândido e Blattmann (2018) propuseram um estudo exploratório para verificar a atuação de mulheres em pesquisas da área de Ciência da Informação que abordam temas relacionados à tecnologia e inovação. O estudo teve como motivação os números relativos à participação feminina em atividades de pesquisa e desenvolvimento tecnológico no Brasil, que em 2014 já se encontrava entre 45,1% e 55,0%, conforme dados do IBGE (2016) e estudos realizados pela UNESCO (em nível mundial essa representação era de 28,8%). Para o referido estudo, foram selecionados, a partir das palavras-chave

“tecnologia” e “inovação”, 346 artigos científicos brasileiros da base de dados BRAPCI, sendo constatado que 187 (54%) eram de autoria feminina.

Os resultados dos estudos de Sena, Cândido e Blattmann (2018) evidenciaram a contribuição das mulheres na abordagem de temas visionários, tais como a tecnologia como suporte ao patrimônio cultural, construção do conhecimento organizacional para a inovação nas micro e pequenas empresas, análise de redes sociais para inovação, gestão da informação e transferência de tecnologia para a inovação aberta. Ainda, se percebe nas publicações de maior relevância (mensurada no estudo pelo número de citações) a predominância de autoras com formação, em nível de graduação, em Biblioteconomia e Economia, e em relação ao mestrado e doutorado se percebe o domínio da Ciência da Informação, seguido da Economia, Comunicação e Engenharia de Produção. Nesse sentido, o estudo conclui que a temática inovação, no âmbito das publicações analisadas, evidencia o caráter colaborativo da Biblioteconomia e interdisciplinar da Ciência da Informação.

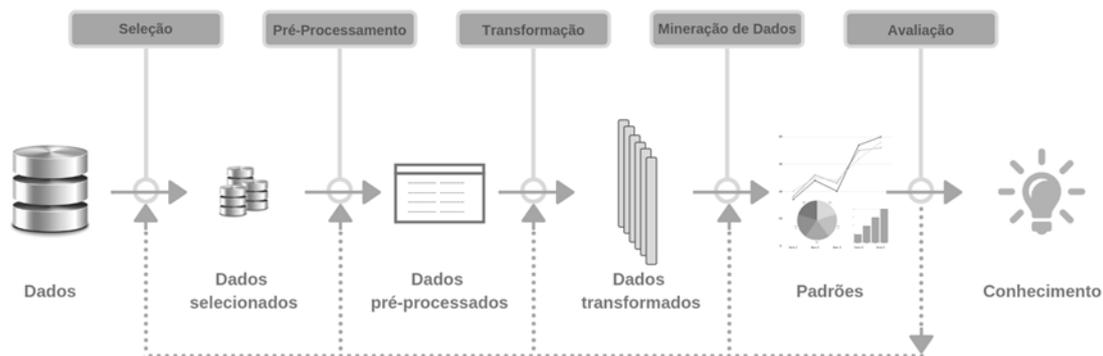
3 METODOLOGIA

O desenvolvimento de análises a partir de conjuntos de dados quantitativos requer a utilização e exploração de estratégias que permitam à obtenção de dados de forma eficiente e assertiva, garantindo a geração de informação útil, e a consequente descoberta de conhecimento. No âmbito do presente trabalho, é proposta a realização de análises de dados de pesquisadores cadastrados na plataforma Lattes, especificamente para a área de Ciência da Informação.

Foi considerado um percurso metodológico baseado no processo de geração de informação proposto por Fayyad Shapiro-Piatetsky e Schmidt (1996), denominado Descoberta de Conhecimento em Bases de Dados - KDD (Knowledge Discovery in Databases), o qual se trata de um modelo clássico que permite o desenvolvimento de análises de forma estratégica através da realização de atividades sequenciais que vão desde a obtenção dos dados até a geração da informação propriamente dita. A Figura 1 apresenta um esquema da aplicação do KDD, com destaque para duas características de grande

importância da metodologia: interatividade, permitindo controle e intervenção nas atividades; e à iteratividade, que possibilita a repetição de atividades anteriores, com o propósito de aumento da acurácia dos resultados obtidos. O KDD é explorado em diversas pesquisas no contexto de estudos em informação e comunicação (CORDEIRO *et al.*, 2022; LIMIRO *et al.*, 2022).

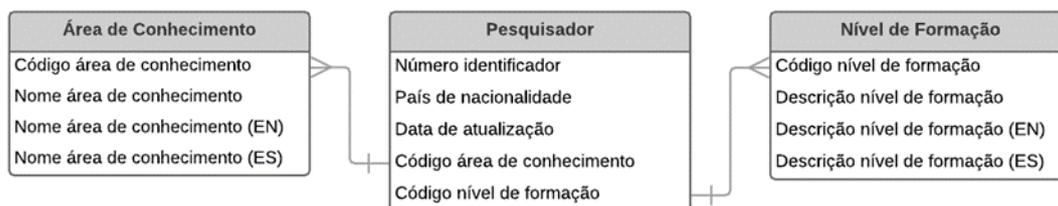
Figura 1 – Processo KDD



Fonte: Adaptado de Fayyad, Shapiro-Piatetsky e Schmidt (1996)

A primeira etapa a ser realizada é a seleção dos dados. Durante esta etapa, são definidos os conjuntos de dados a serem utilizados para resolução do problema, assim como a realização das ações relacionadas à coleta destes dados (GOLDSCHMIDT; PASSOS; BEZERRA, 2015). Para realização dos experimentos, foi inicialmente obtida a lista de todos os pesquisadores cadastrados na base Lattes, disponibilizada no portal do CNPq, contemplando dados referentes aos identificadores únicos dos pesquisadores, nível de formação, e área de conhecimento, conforme apresentado no modelo de dados da Figura 2. Neste sentido, foram obtidos 5.197.727 registros de pesquisadores, dos quais 271.633 possuem titulação igual ou superior a doutorado, representando aproximadamente 5.2% do universo total de pesquisadores.

Figura 2 – Diagrama Entidade-Relacionamento dos currículos cadastrados no Lattes – CNPq



Fonte: Autores

No contexto dos propósitos da pesquisa, foram aplicados filtros sobre o conjunto de dados originais, considerando a área de conhecimento Ciência da Informação, a titulação igual ou superior a doutorado, e o país de nascimento como Brasil. Foram identificadas 16.460 ocorrências para a área em questão, considerando todos os níveis de formação, dos quais 947 registros são de pesquisadores em Ciência da Informação com formação igual ou superior a doutorado, e 845 registros para pesquisadores em Ciência da Informação nascidos no Brasil com formação igual ou superior a doutorado, representando aproximadamente 5,1% da amostra total da área.

A partir da obtenção da amostra de registros a ser considerada, com base no atributo de identificação de cada pesquisador, foram geradas as *urls* (*uniform resource locator*) para coleta dos currículos na Plataforma Lattes. Para tanto, foi gerada uma solução automatizada que consulta a base de registros com a lista de pesquisadores e gera uma saída em formato *csv* (*comma separated values*) com as *urls* associadas, possibilitando a extração dos currículos. A disponibilização dos currículos na plataforma Lattes é feita em formato *html* e em formato *xml*. O formato *xml* apresenta a vantagem de possuir uma estrutura pré-definida, permitindo a extração de dados de forma mais assertiva e direcionada. Diante disso, foi realizada a coleta dos currículos em padrão *xml*.

Uma das análises a ser realizada no âmbito do presente trabalho é a exploração das publicações em periódicos em face das classificações realizadas por meio do programa Qualis da CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). O portal Sucupira disponibiliza, em formato processável, conjuntos de dados com os registros dos periódicos e seus estratos, isto é, a classificação recebida na avaliação dos periódicos. Os dados

contêm os seguintes atributos: *ISSN*, *nome do periódico*, e *estrato*. Para o presente trabalho foram consideradas as classificações referentes à área Comunicação e Informação, levando-se em conta o Qualis Periódicos 2013-2016.

Após a seleção dos dados, deve ser realizada a etapa de pré-processamento, onde são executadas atividades de limpeza dos dados, tratando possíveis registros que apresentem ruídos ou dados faltantes, de forma a garantir a qualidade das análises realizadas (GOLDSCHMIDT; PASSOS; BEZERRA, 2015). No contexto do presente trabalho, durante essa etapa, uma das principais atividades a serem realizadas é a identificação do sexo dos pesquisadores para cada um dos currículos que compõe a base de dados selecionada. É importante destacar que tal informação não se encontra disponível para consulta pública através da plataforma Lattes, o que torna necessária a realização de procedimentos de identificação específicos para tal propósito. Uma estratégia a ser adotada é a identificação manual, através de verificação individual de cada um dos currículos. Entretanto, tal abordagem se torna inviável frente à quantidade de currículos a serem analisados. Diante disso, uma estratégia é o desenvolvimento de soluções computacionais que possibilitem a automatização de tal tarefa.

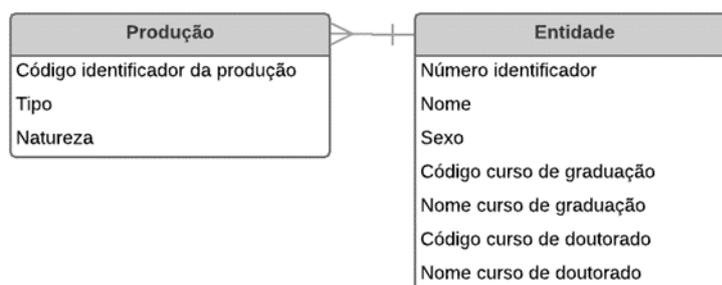
Com o propósito de identificação do sexo dos pesquisadores, foi desenvolvida uma solução utilizando a linguagem R, baseada no pacote *genderBR* (MEIRELES, 2021), que realiza a predição do sexo a partir do primeiro nome com base em dados do Censo de 2010 do Instituto Brasileiro de Geografia e Estatística (IBGE). Através disso, foi possível realizar a inferência de sexo em 98% da amostra. Os elementos da amostra que não puderam ser rotulados de forma automatizada (2% da amostra), tiveram a identificação realizada de forma manual.

Outro procedimento necessário durante a etapa de pré-processamento é a identificação dos tipos de produções de cada um dos pesquisadores. Nos próprios dados obtidos através da plataforma Lattes é possível identificar as produções quanto aos tipos: publicações em periódicos, publicações em eventos, capítulos de livro, participações em bancas, orientações, entre outros.

Entretanto, para as publicações em periódicos é possível ainda associar a classificação proveniente do sistema Qualis da Capes, a qual não está disponível na plataforma Lattes.

No âmbito da classificação das produções em periódicos, os dados referentes a este tipo de registro junto aos currículos dos pesquisadores possuem o atributo de identificação ISSN, o qual foi então cruzado com os dados de classificação disponibilizados pela Capes, para a área Comunicação e Informação, associando para cada uma das ocorrências a classificação devida. Tal procedimento foi realizado através da construção de uma rotina utilizando a linguagem de programação Python. Neste sentido, considerando a identificação do sexo dos pesquisadores e a classificação das produções, os dados pré-processados passaram a seguir o modelo descrito na Figura 3.

Figura 3 – Diagrama Entidade-Relacionamento dos dados pré-processados



Fonte: Autores

A terceira etapa do processo se refere à transformação dos dados para um formato mais adequado para os propósitos de análise. Os dados obtidos através da plataforma Lattes são originalmente disponibilizados em padrão xml, o qual se trata de um padrão de representação de dados através de estruturação baseada em marcadores (tags), semelhante ao padrão html, sendo amplamente utilizado no compartilhamento de dados via Internet (FERNANDES; CORDEIRO, 2016). Apesar disso, no que se refere ao armazenamento e manipulação de dados para propósitos de análises através de processadores de planilhas, o formato csv é mais adequado e interessante, uma vez que possibilita a organização dos dados em formato tabulado. Diante disso, os dados da amostra

considerada foram convertidos para o formato csv, seguindo a estrutura descrita na Figura 3.

No âmbito da metodologia KDD, a mineração de dados é considerada a etapa de maior importância, uma vez que é a principal responsável pela geração de informações. A mineração de dados, de acordo com Amaral (2016), se refere a um conjunto de processos voltados à exploração e análise de grandes conjuntos de dados, com o propósito de levantar padrões, associações, anomalias e previsões. Castro e Ferrari (2016) destacam que a mineração de dados corresponde à aplicação de soluções computacionais que permitam extrair conhecimento em conjuntos de dados pré-processados. De forma geral, as funcionalidades da mineração de dados podem ser exploradas sob duas abordagens: descritiva e preditiva. Os métodos descritivos estão voltados à caracterização, sumarização e discriminação de dados, enquanto os métodos preditivos buscam a inferência ou previsão a partir de análises em conjuntos de dados (HAN; KAMBER; PEI, 2012). No âmbito do presente trabalho serão utilizados métodos descritivos de mineração de dados. Neste sentido, a proposta é gerar análises bivariadas, isto é, baseada em duas variáveis diferentes, a distribuição de frequências das produções científicas dos pesquisadores e o atributo de sexo rotulado junto aos dados obtidos, a fim de identificar possíveis variações e relações entre os atributos considerados.

Após a realização das atividades de mineração de dados, a última etapa é a avaliação, também descrita como interpretação. Nesta etapa, os resultados obtidos são interpretados e analisados através da geração de visualizações. Através da ferramenta de análise Power BI, a qual permite a geração de relatórios dinâmicos e interativos, são gerados gráficos descrevendo os resultados alcançados. É importante destacar que as visualizações preveem a normalização dos dados, com o propósito de garantir a representatividade durante a categorização baseada no sexo dos pesquisadores, promovendo subsídios no que se refere às comparações e reflexões analíticas, também realizadas durante esta etapa, sendo o intuito prover a distribuição em face da produção para cada categoria, e não a comparação numérica bruta dos dados obtidos.

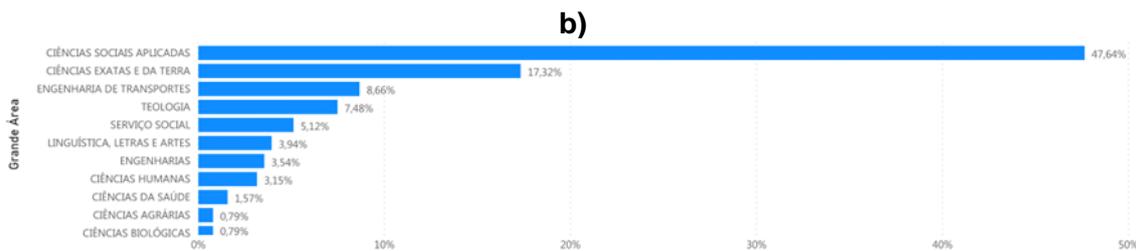
4 RESULTADOS E DISCUSSÕES

A partir da realização das etapas apresentadas nos aspectos metodológicos, contemplando desde a obtenção do conjunto de dados, o pré-processamento, e culminando na geração das visualizações, inicialmente foi gerado um gráfico apresentando as grandes áreas de formação dos pesquisadores que atuam no âmbito da Ciência da Informação. No quantitativo de dados recuperados, estão presentes um total de 547 mulheres (64,7% do total) e 298 homens (35,3%). É importante destacar que a área de formação dos pesquisadores pode ser diferente da sua área de atuação, sendo tal informação contemplada através da Plataforma Lattes. Em todos os gráficos apresentados, a cor rosa é utilizada para representar os dados referentes às mulheres, e a cor azul para representar os dados referentes aos homens (a escolha das cores se deu somente pela diferenciação visual dos dados, não tendo relação com quaisquer outras questões).

A Figura 4 apresenta a distribuição para mulheres e homens em termos percentuais, em que se nota, para as mulheres, que a área de formação possui uma relação maior com a grande área à qual a Ciência da Informação se enquadra, Ciências Sociais Aplicadas, representando um percentual de 72,75%, enquanto, para os homens a presença de pesquisadores provenientes das Ciências Sociais Aplicadas representa 47,64% do total. Se destaca ainda que no caso dos homens, há uma considerável representatividade de pesquisadores provenientes da grande área de Ciências Exatas e da Terra.

Figura 4 – Grande área de formação dos pesquisadores de Ciência da Informação.

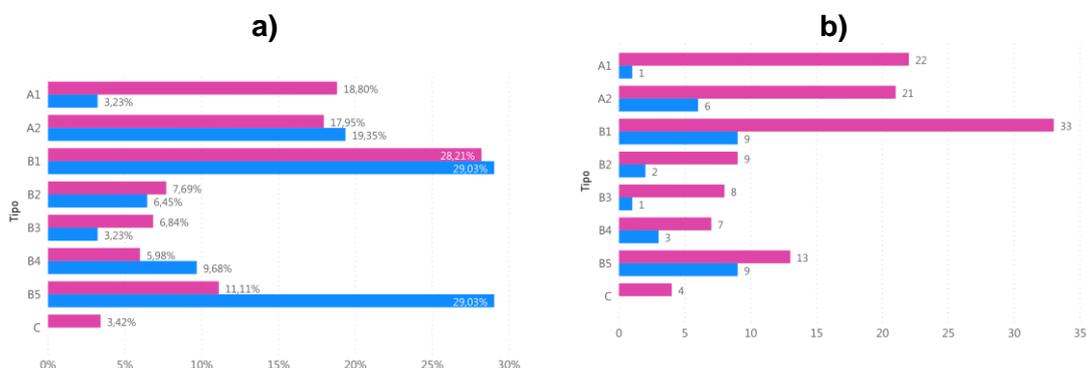




Fonte: Autores

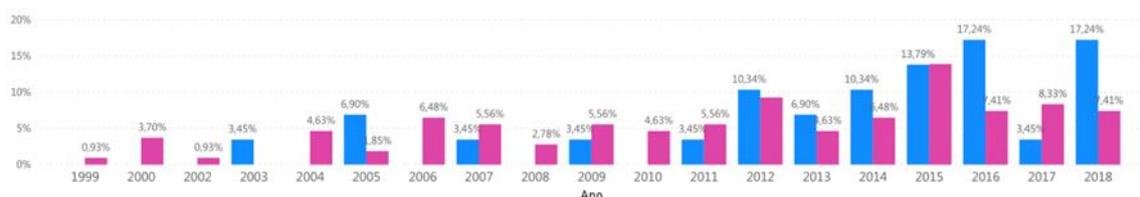
A partir disso, foram analisados também o quantitativo de produção categorizadas com base na classificação disponibilizada pelo Qualis Periódicos 2013-2016. A Figura 5 apresenta os resultados obtidos, em que é possível observar que, em termos proporcionais, as mulheres possuem uma representação consideravelmente maior em periódicos classificados como A1 e A2, sendo que 36,75% da produção em periódicos por parte das pesquisadoras se enquadra nestas categorias, enquanto, para os homens, o percentual é de 22,58%. Além disso, se destaca ainda um considerável número de publicações classificadas como B5 por parte dos homens, totalizando um percentual de 29,03%, o que pode estar associado à grande área de formação dos pesquisadores, uma vez que para as análises foi considerada a classificação Qualis Periódicos para a área de Comunicação e Informação. Outro ponto interessante, destacado na Figura 6, se refere à distribuição temporal percentual das produções em periódicos, a qual demonstra que estas, por parte das mulheres, apresentam uma relativa estabilidade, com pequenas variações percentuais, com exceção para o ano de 2015, que apresentou um maior número de publicações. Por outro lado, as publicações por parte dos homens apresentam uma tendência de crescimento nos últimos anos, o que pode estar associado à um acréscimo no número de pesquisadores do sexo masculino.

Figura 5 – Produção em periódicos com Qualis (a - distribuição relativa por classe, b - valores absolutos)



Fonte: Autores

Figura 6 – Distribuição relativa temporal da produção em periódicos com Qualis

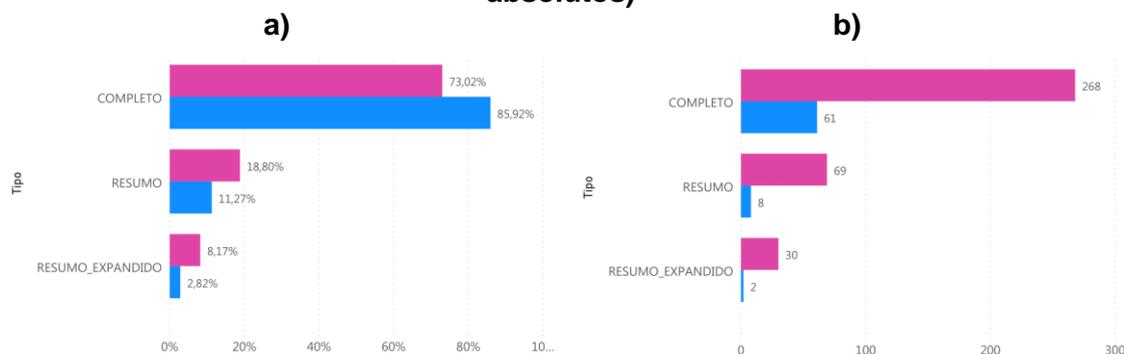


Fonte: Autores

Em relação à produção de trabalhos publicados em anais de eventos, a Figura 7 apresenta os quantitativos percentuais e absolutos, em que neste caso, é possível notar que, em termos relativos, a variação entre a produção das mulheres e dos homens é pequena, tendo os homens uma maior representatividade em trabalhos completos. Entretanto, em termos absolutos, a quantidade de trabalhos completos publicados pelas pesquisadoras é consideravelmente maior que o número de produções, na mesma categoria, pelos homens, constatado pela aferição da média individual por categoria, que, para as mulheres é de aproximadamente um artigo a cada duas pesquisadoras, enquanto para os homens esse valor cai para aproximadamente um artigo completo a cada quatro pesquisadores. Além disso, em termos temporais, assim como observado para outros tipos de produções, verifica-se, na Figura 8, que as produções em eventos para as mulheres apresentam uma maior constância

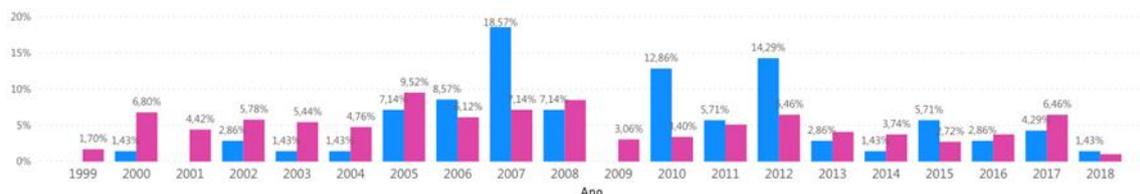
relativa, enquanto para os homens, esse valor é ressaltado por alguns picos de produção (2007, 2010 e 2012), e uma sucessiva diminuição.

Figura 7 –Produção em eventos (a - distribuição relativa por categoria, b - dados absolutos)



Fonte: Autores

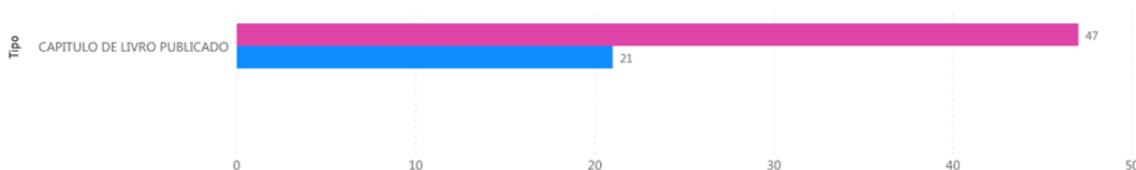
Figura 8 – Distribuição relativa temporal da produção em eventos



Fonte: Autores

Na categoria de produção de capítulos de livros, uma vez que esta foi analisada de forma individual, foram mensurados apenas os valores absolutos de produção, demonstrando que as mulheres possuem um valor total que supera o dobro de produções, nesta categoria, por parte dos homens (Figura 9). Entretanto, realizando a análise sobre um ponto de vista normalizado, tem-se aproximadamente, para as mulheres, a produção de um capítulo de livro a cada 11 pesquisadoras, enquanto para os homens esse valor é de um capítulo de livro a cada 14 pesquisadores.

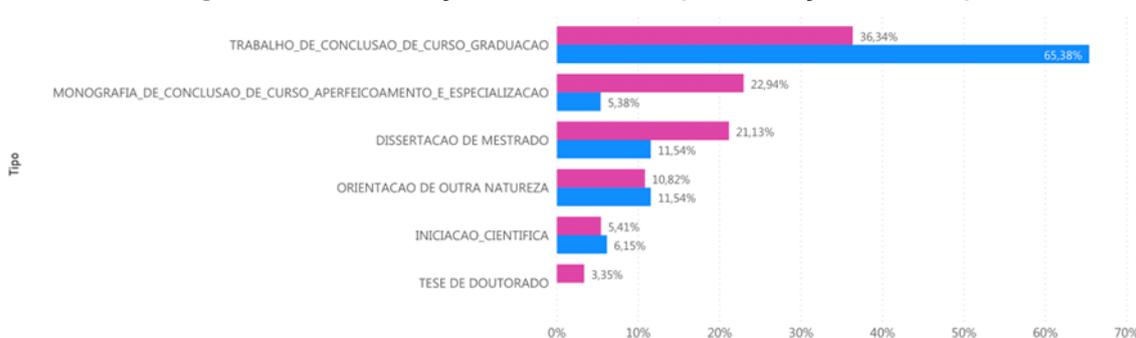
Figura 9 – Produção de capítulos de livro (dados absolutos)



Fonte: Autores

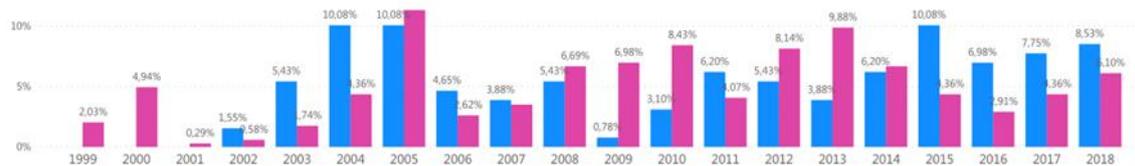
Em relação à formação de alunos, a Figura 10 apresenta a distribuição relativa de orientações por categoria para orientações concluídas. Nos resultados é possível observar que as mulheres possuem uma atuação consideravelmente mais efetiva na pós-graduação, com destaque para orientações de teses de doutorado defendidas. Neste ponto, é fundamental observar que na presente pesquisa foram considerados apenas os currículos Lattes de pesquisadores que declararam sua área de atuação principal como Ciência da Informação, sendo que pesquisadores que, apesar de atuarem em programas de pós-graduação ligados à Ciência da Informação, mas com área de atuação principal no Lattes diferente não foram pontuados. Em termos temporais, foi observado um comportamento numérico semelhante entre mulheres e homens, com alguns desvios pontuais (Figura 11), como no ano de 2004, em que foi constatada uma maior representatividade relativa dos homens, e os anos de 2009, 2010, 2012 e 2013, com uma maior representatividade por parte das mulheres. Os três anos finais apresentaram uma tendência de crescimento semelhante, com maior representatividade relativa numérica por parte dos homens.

Figura 10 – Orientações concluídas (distribuição relativa)



Fonte: Autores

Figura 11 – Distribuição relativa temporal de orientações concluídas



Fonte: autores

5 CONCLUSÃO

Uma série de aspectos culturais, entre os quais se destaca o patriarcado, conferem à mulher uma condição que se configura enquanto um acúmulo de funções, as quais são vista e assumidas como um padrão social, sendo as mulheres denotadas como detentoras de responsabilidades familiares, além de uma cobrança de realização profissional, pessoal e social. Face a isso, diversos estudos, levando-se em conta aspectos interdisciplinares, têm explorado a participação e as significativas contribuições das mulheres na ciência. Entretanto, estes mesmos estudos demonstram que ainda existem obstáculos, principalmente culturais, que perpassam a atuação da mulher na pesquisa científica.

Neste cenário, esta divisão social de gênero que se encontra inserida nas diversas áreas de conhecimento, acaba por tentar impor à mulher um caminho pré-estabelecido, o qual pode ser uma das primeiras barreiras do ingresso da mulher na ciência. Para além disso, a permanência e o desenvolvimento produtivo na pesquisa científica podem ainda sofrer comprometimentos ou particularidades quando mais uma questão é colocada em pauta, a maternidade. Embora existam iniciativas que buscam promover a inserção e participação feminina na ciência, é ainda preciso compreender e prover mecanismos e soluções de questões que perpassam os limites da ciência. Apesar disso, em contraste às imposições e obstáculos existentes, os resultados apresentados neste trabalho trazem indicadores que demonstram, de forma notável, a importância e representatividade da mulher no desenvolvimento da Ciência da Informação, com destaque para as contribuições em termos de produções em periódicos especializados na área, assim como na formação de alunos em nível

de pós-graduação. Acredita-se que este estudo exploratório possa servir como uma fonte de informações para estudos mais especializados na busca pela compreensão de experiências, dificuldades e desafios das mulheres na pesquisa em Ciência da Informação, como também enquanto fonte norteadora para a geração de informações estratégicas para a promoção da igualdade de gênero na área.

REFERÊNCIAS

AMARAL, F. **Aprenda Mineração de Dados: teoria e prática**. Rio de Janeiro: Alta Books, 2016.

ARAÚJO, C. O que é Ciência da Informação? **Informação & Informação**, Londrina, v. 19, n. 1, p. 1-30, 2013. DOI: <https://doi.org/10.5433/1981-8920.2014v19n1p01>.

ALBORNOZ, M.; BARRERE, R.; MATAS, L.; OSORIO, L.; SOKIL, J. Las Brechas de Género en la producción científica Iberoamericana. **Papeles del Observatorio**, Observatorio Iberoamericano de la Ciencia, la Tecnología y la Sociedad de la Organización de Estados Iberoamericanos (OCTS-OEI), n. 9, 2018.

BEDIN, J.; FELDMAN, D.; VIANNA, W. B. Publicações Científicas de autoria de mulheres: um panorama da Gestão da Informação no cenário brasileiro. *In*: SILVA, F. C. G.; ROMEIRO, N. L. (org.). **O protagonismo da mulher na Biblioteconomia e Ciência da Informação**. Florianópolis: ACB, 2018. p. 101-128.

CARVALHO, M. S.; COELI, C. M.; LIMA, L. D. Mulheres no Mundo da Ciência e da Publicação Científica. **Cadernos de Saúde Pública**, Rio de Janeiro, v. 34, n. 3, e00025018, 2018. DOI: <https://doi.org/10.1590/0102-311X00025018>.

CORDEIRO, D. F.; LEAL, M. R. C.; VIEIRA, L. M.; DA SILVA, N. R. Cartografando comentários e sentimentos no perfil de Jair Bolsonaro no Instagram acerca da Covid-19. **Galáxia**, São Paulo, v. 47, e56929, 2022. DOI: <https://doi.org/10.1590/1982-2553202256929>.

CORRÊA, E. C. D.; OLIVEIRA, A. C. D. C. Pelas mãos femininas de Lydia Sambaquy e Celia Zaher. *In*: SILVA, F. C. G.; ROMEIRO, N. L. (org.). **O protagonismo da mulher na Biblioteconomia e Ciência da Informação**. Florianópolis: ACB, 2018. p. 17-44.

DIAS, K. C. O.; LIMA, F. R. A. Levantamento das produções sobre mulheres e relações de gênero nos artigos de periódicos em ciência da informação. *In*: ENCONTRO REGIONAL DE ESTUDANTES DE BIBLIOTECONOMIA,

DOCUMENTAÇÃO E GESTÃO DA INFORMAÇÃO, Juazeiro do Norte, 2012.
Anais [...]. Juazeiro do Norte: EREBD N/NE, 2012.

ELSEVIER. The global research landscape through a gender lens. *In*:
ELSEVIER. **Gender in the Global Research Landscape**. Elsevier, 2017. p.
34-37.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. From data mining to
knowledge discovery in databases. **AI Magazine**, [S. l.], v. 17, n. 3, 1996. DOI:
<https://doi.org/10.1609/aimag.v17i3.1230>.

FERNANDES, J.; CORDEIRO, D. Avaliação de formatos de publicação de
dados abertos governamentais através de indicadores de usabilidade.
Tendências da Pesquisa Brasileira em Ciência da Informação, [S. l.], v. 9,
n. 1, p. 65-83, 2016. Disponível em:
<https://revistas.ancib.org/index.php/tpbci/article/view/377>. Acesso em: 07 de
nov. de 2023.

GOLDSCHMIDT, R.; PASSOS, E.; BEZERRA, E. **Data Mining**: Conceitos,
técnicas, algoritmos, orientações e aplicações. 2. ed. Rio de Janeiro: Elsevier,
2015.

GROSSI, M.; BORJA, S.; LOPES, A.; ANDALÉCIO, A. As mulheres praticando
ciência no Brasil. **Revista Estudos Feministas**, Florianópolis, v. 24, n. 1, p. 11-
30, 2016. DOI: <https://doi.org/10.1590/1805-9584-2016v24n1p11>.

HAN, J.; KAMBER, M.; PEI, J. **Data Mining**: concepts and techniques.
Waltham, MA, USA: Morgan Kaufmann Publishers, 2012.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Pesquisa
de Inovação**: 2014. Coordenação de Indústria, Rio de Janeiro, 2016.
Disponível em: <http://www.pintec.ibge.gov.br/downloads/PUBLICACAO/>.
Acesso em: 08 fev. 2023.

LIMIRO, R. M.; SILVA, N. R.; CORDEIRO, D. F. Mineração de textos para
agrupamento de teses e dissertações por meio de análise de similaridade.
Revista Brasileira e Biblioteconomia e Documentação, São Paulo, v. 18, p.
1-20, 2022. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/1736>.
Acesso em: 08 nov. 2023.

LÓPEZ-BASSOLS, V.; GRAZZI, M.; GUILLARD, C.; SALAZAR, M. **Las
brechas de género en ciencia, tecnología e innovación en América Latina
y el Caribe**: resultados de una recolección piloto y propuesta metodológica
para la medición. Nota técnica no. IDB-TN-1408. Washington, DC: Banco
Interamericano de Desarrollo, 2018.

MEIRELES, F. **genderBR**. Disponível em:
<https://github.com/meirelesff/genderBR>. Acesso em: 16 fev. 2021.

MIKO, I.; LEJEUNE, L. Rosalind Franklin: A Crucial Contribution. *In*: MIKO, I.; LEJEUNE, L. **Essentials of Genetics**, Cambridge, MA: NPG Education, 2009.

MORALES, A. P. Mulheres na Ciência e a Ciência das Mulheres. **Revista Mulheres na Ciência**, [S. l.], n. 1, p. 9-13, 2019. Disponível em: https://www.britishcouncil.org.br/sites/default/files/d1_revista.pdf. Acesso em: 07 nov. 2023.

ODDONE, N. O IBBD e a informação científica: uma perspectiva histórica para a ciência da informação no Brasil. **Ciência da Informação**, Brasília, v. 35, n. 1, p. 45-56, 2006. DOI: <https://doi.org/10.1590/S0100-19652006000100006>.

OWSD - Organização para Mulheres na Ciência para o Mundo em Desenvolvimento. **Elsevier Foundation Awards for Early-Career Women Scientists in the Developing World**. Disponível em: <https://owsd.net/career-development/awards>. Acesso em: 25 jan. 2021.

ROMEIRO, N. L.; VIOLA, C. M. M. Políticas Públicas e o Desenvolvimento da Mulher na Ciência: um estudo sobre as ações no Brasil e em Portugal. *In*: SILVA, F. C. G.; ROMEIRO, N. L. (org.). **O protagonismo da mulher na Biblioteconomia e Ciência da Informação**, Florianópolis: ACB, 2018. p. 249-273.

QUEIROZ, D. G. C.; MOURA, A. M. M. Ciência da Informação: história, conceito e características. **Em Questão**, Porto Alegre, v. 21, n. 3, p. 25-42, 2015. DOI: <https://doi.org/10.19132/1808-5245213.26-42>.

ROBREDO, J.; VILAN FILHO, J. L. Metrias da Informação: histórias e tendências. *In*: ROBREDO, J.; BRASCHER, M. (org.). **Passeios pelo Bosque da Informação: estudos sobre representação e organização da informação e do conhecimento**. Brasília: EROIC, 2010. p. 185-258.

SANTO, P. E. Os estudos de gênero na ciência da informação. **Em Questão**, Porto Alegre, v. 14, n. 2, p. 317-332, 2008. Disponível em: <https://seer.ufrgs.br/index.php/EmQuestao/article/view/6389>. Acesso em: 08 nov. 2023.

SENA, P. M. B., CÂNDIDO, A. C., BLATTMANN, U. Tecnologia e Inovação na Biblioteconomia e Ciência da Informação: Panorama da Produção Científica de Mulheres no Brasil. *In*: SILVA, F. C. G.; ROMEIRO, N. L. (org.). **O protagonismo da mulher na Biblioteconomia e Ciência da Informação**. Florianópolis: ACB, 2018. p. 129-166.

SICILIANO, M.; SOUZA, C. M. S.; METH, C. M. E. S. Sobre o que falamos quando falamos em gênero na ciência da informação? **Informação & Informação**, Londrina, v. 22, n. 2, p. 144-165, 2017. DOI: <https://doi.org/10.5433/1981-8920.2017v22n2p144>.

SILVA, F. C. G.; ROMEIRO, N. L. **O protagonismo da mulher na biblioteconomia e ciência da informação**. Florianópolis: ACB, 2018.

VALDÉS, I.; RUBIO, I. Rosalind Franklin ajudou a desvendar o DNA, mas ficou sem o Nobel. **El País**, 2018. Disponível em:
https://brasil.elpais.com/brasil/2018/02/20/ciencia/1519150633_000516.html.
Acesso em: 08 dez. 2020.

THE PRODUCTION OF WOMEN IN INFORMATION SCIENCE FROM A DESCRIPTIVE ANALYSIS BASED ON DATA MINING

ABSTRACT

Objective: The article proposes a descriptive study on women's participation in scientific production, in terms of bibliographical production and guidelines, within the scope of Information Science. **Methodology:** Uses a descriptive path, based on the Knowledge Discovery in Databases process, using data mining solutions to analyze women's participation in scientific production, specifically in Information Science. **Results:** The results obtained present indicators in terms of productivity in the context of Information Science, addressing different views according to specific types of production. **Conclusion:** The contributions made by researchers in the area of Science are highlighted, mainly for productions in specialized journals and training at postgraduate level, demonstrating the representation and importance of women for the growth of the area.

Descriptors: Gender. Information Science. Scientific production. Scientific Information. Descriptive analysis.

LA PRODUCCIÓN DE MUJERES EN CIENCIAS DE LA INFORMACIÓN A PARTIR DE UN ANÁLISIS DESCRIPTIVO BASADO EN MINERÍA DE DATOS RESUMEN

RESUMEN

Objetivo: El artículo propone realizar un estudio descriptivo sobre la participación de las mujeres en la producción científica, en términos de producción bibliográfica y lineamientos, en el ámbito de las Ciencias de la Información. **Metodología:** Utiliza un camino descriptivo, basado en el proceso de Descubrimiento del Conocimiento en Bases de Datos, con el uso de soluciones de minería de datos para analizar la participación de las mujeres en la producción científica, específicamente en las Ciencias de la Información. **Resultados:** Los resultados obtenidos presentan indicadores en términos de productividad en el contexto de las Ciencias de la Información, abordando diferentes miradas según tipos específicos de producción. **Conclusión:** Se destacan los

aportes realizados por investigadores del área de Ciencias, principalmente para producciones en revistas especializadas y formación a nivel de posgrado, demostrando la representación e importancia de las mujeres para el crecimiento del área.

Descritores: Género. Ciencias de la Información. Producción científica. Información Científica. Análisis descriptivo.

Recebido em: 17.02.2022

Aceito em: 23.11.2023