

Estado da publicação: O preprint não foi submetido para publicação

Modelo para o desenvolvimento de estratégia de busca para a recuperação da informação científica e tecnológica: o caso da impressão 3D

Andréia Cristina Galina, Jacqueline Leta

<https://doi.org/10.1590/SciELOPreprints.8141>

Submetido em: 2024-02-26

Postado em: 2024-05-23 (versão 2)

(AAAA-MM-DD)

Justificativa da versão: Melhorias na explicação e visual do modelo.

Modelo para o desenvolvimento de estratégia de busca para a recuperação da informação científica e tecnológica: o caso da impressão 3D

Model for developing a search strategy for retrieving scientific and technological information: the case of 3D printing

Andréia Cristina Galina¹, ORCID <http://orcid.org/0000-0002-3670-9172>

Jacqueline Leta², ORCID <http://orcid.org/0000-0002-3271-7749>

¹ Doutoranda do Programa de Gestão e Educação em Biociências, Instituto de Bioquímica Médica, Universidade Federal do Rio de Janeiro - UFRJ

² Docente do Programa de Gestão e Educação em Biociências, Instituto de Bioquímica Médica, Universidade Federal do Rio de Janeiro - UFRJ

RESUMO

Após a segunda guerra mundial, a ciência passa a ser compreendida como um setor vital para o desenvolvimento e progresso de um país, e, conseqüentemente, as nações passaram a investir mais recursos para as atividades de pesquisa. A necessidade de gerenciar esses recursos e de compreender a dinâmica científica e tecnológica estimulou o desenvolvimento de áreas relacionadas a métricas de desempenho e de impacto da ciência, como bibliometria, cientometria e patentometria. Estudos dessas áreas sobre um determinado tema têm como etapa inicial a recuperação da informação, a qual requer o uso de uma estratégia de busca eficiente para recuperar o máximo de documentos de interesse com o mínimo de falsos positivos. OBJETIVO O presente estudo propõe um modelo para o desenvolvimento de estratégia de busca para a recuperação da informação científica e tecnológica em um tema emergente, a impressão 3D, em estudos métricos quantitativos. RESULTADOS Para o caso da impressão 3D, comparamos a quantidade de resultados recuperados com publicações sobre o tema, e obtivemos um quantitativo 70% superior ao identificado na literatura para documentos científicos e 44% superior para patentes. CONCLUSÃO A utilização do modelo pode ampliar consideravelmente a quantidade de informação científica e tecnológica recuperada, possibilitar a identificação de termos para tecnologias emergentes antes que esses se estabeleçam e proporcionar melhoria na qualidade dos dados, evitando conclusões equivocadas.

KEYWORDS: estratégia de busca, recuperação da informação, impressão 3D

ABSTRACT

After the Second World War, science began to be understood as a vital sector for the development and progress of a country and, consequently, nations invested more resources for research activities. The need to manage these resources and to understanding scientific and technological dynamics stimulated the development of fields related to metrics on science performance and impact, such as bibliometrics, scientometrics and patentometrics. Studies in these fields on a given topic have information retrieval as their initial stage, which requires the use of an efficient search strategy to retrieve the maximum number of documents of interest with the minimum number of false positives. **OBJECTIVE** To propose a model for developing a search strategy for retrieval of scientific and technological information on an emerging topic, 3D printing, in quantitative metric studies. **RESULTS** For the case of 3D printing, we compared the quantity of retrieved results with publications on the subject, and we obtained a quantity 70% higher than that identified in the literature for scientific documents and 44% higher for patents. **CONCLUSION** The use of the model can significantly increase the amount of retrieved scientific and technological information, enable the identification of terms for emerging technologies before they become established, and enhance data quality by avoiding misleading conclusions.

KEYWORDS: search strategy, information retrieval, 3D printing

Introdução

Os avanços da ciência e da tecnologia a partir do século XX, em especial após a Segunda Guerra Mundial, que levaram, por exemplo, ao aprimoramento do radar, à produção em escala industrial da penicilina e o desenvolvimento da bomba atômica entre outros, não deixam dúvidas que o desenvolvimento científico e tecnológico foi crucial para o fim da segunda grande guerra.

A política em prol o avanço da atividade científica e tecnológica foi impulsionada pela publicação do relatório “*Science the endless frontier*”, organizado por Vannevar Bush, a pedido do Presidente dos Estados Unidos, Franklin D. Roosevelt, que foi publicado em 1945. Nele, se propõe a elaboração de uma

política de Estado voltada para a ciência, pois esta passou a ser identificada como uma atividade vital para o progresso e desenvolvimento de um país (BUSH, 1945). O relatório orientava que o governo dos EUA deveria financiar a pesquisa básica realizada por universidades e difundia a ideia de que a oferta do conhecimento, gerado por esse tipo de pesquisa, era a chave para que empresas privadas desenvolvessem pesquisa aplicada para a solução de problemas, promovendo, assim, o desenvolvimento tecnológico, com a produção de produtos e de serviços. Essa sequência de passos que se inicia com pesquisa básica, seguida da pesquisa aplicada, que acarreta o desenvolvimento de produtos, serviu de base para o modelo de inovação linear que foi amplamente utilizado em duas abordagens: “science push” e “demand pull”, nas quais a ciência assume um papel central para se atingir a inovação (IACONO; DE ALMEIDA; NAGANO, 2011).

A aceitação dessa nova perspectiva sobre a relação entre ciência e tecnologia definiu, inicialmente, o curso da política de C&T dos EUA, sendo rapidamente disseminada para a maioria dos países industrializados (IACONO; DE ALMEIDA; NAGANO, 2011), levando à ampliação dos incentivos ao setor e à expansão de organizações, de infraestrutura e de recursos humanos (GODINHO, 2007). Neste contexto, a política científica passa a ser de identificação de prioridades, o que gera a necessidade de indicadores específicos para uma compreensão aprofundada da dinâmica Científica e Tecnológica, que envolve aspectos políticos, econômicos, sociais e/ou culturais (VELHO, 2001). Assim, a partir dos anos de 1960, atividades de mensuração, acompanhamento e avaliação de C&T de países desenvolvidos passaram a ser rotineiras (VELHO, 2001) por três principais motivos: (a) garantir que a ciência participasse dos objetivos econômicos e sociais dos países, (b) a limitação de recursos financeiros e (c) as evidências de que a alocação de recursos realizada por cientistas vinha se mostrando ineficiente em vários aspectos, como por exemplo na falta de equidade para diferentes regiões.

Ainda na década de 1960, a coleta de informações de insumos (*input*) foi realizada por órgãos setoriais responsáveis pelas atividades de C&T, como a Organização das Nações Unidas para a Educação, a Ciência e a Cultura¹

¹ United Nations Educational, Scientific and Cultural Organization

(UNESCO) e a Organização para a Cooperação e Desenvolvimento Econômico (OCDE) (VELHO, 2001). Como resultado dessas iniciativas temos o Manual de Frascati, publicado em 1963 pela OCDE, onde se propõe a normalização de metodologias para a uniformidade na produção de estatísticas internacionais, com o olhar em recursos financeiros e humanos (VELHO, 2001).

Na década de 1970 surgiu a necessidade de novos indicadores de resultados (*output*) que fossem capazes de identificar as tendências e resultados de políticas já implantadas no setor de C&T (VELHO, 2001). Essa nova necessidade impulsionou o desenvolvimento de disciplinas que dão suporte a métricas da ciência como a bibliometria, cientometria e patentometria.

Os indicadores de C&T obtidos a partir de técnicas da bibliometria e da cientometria têm três grupos alvo de aplicação: (a) os especialistas destas duas disciplinas que buscam ampliar e consolidar o conhecimento específico delas; (b) os especialistas de outras disciplinas que se apropriam desses indicadores para compreender seu próprio campo; e (c) os cientistas políticos e/ou gestores de ciência, que se apropriam desses indicadores para uso na gestão e política de ciência (GLÄNZEL, 2003). Também se discute a interação desses campos de pesquisas com outros, como biblioteconomia, recuperação da informação, cientometria, informetria, tecnometria e sociologia da ciência.

No presente artigo tratamos, portanto, de temática relacionada à Recuperação da Informação (RI), que segundo Pielke (2010) é um campo originalmente ligado à Ciência da Computação e que, com o passar do tempo, foi sendo incorporado à Ciência da Informação (CI), e, nas últimas décadas, tornou-se indispensável a diversos outros campos do conhecimento (PIELKE, 2010).

Nos estudos cientométricos², a RI inclui uma etapa inicial e decisiva que é a elaboração da estratégia de busca. Considerando, portanto, a relevância dessa etapa em estudos métricos e a diversidade do público-alvo, o presente estudo propõe um modelo para o desenvolvimento de estratégia de busca para a recuperação da informação científica e tecnológica, usando como exemplo uma temática emergente, a impressão 3D.

² Com base no estudo de Glanzel, a partir deste ponto do trabalho, o termo cientometria (ou estudos cientométricos) será utilizado como sinônimo ao termo bibliometria (ou estudo bibliométrico)

A seguir, apresentamos os principais desenvolvimentos ocorridos na RI e estratégia de busca, assim como modelos.

Principais marcos, desenvolvimentos e modelo para RI

Segundo Araújo (2009), a RI representa a aplicação mais significativa no campo da CI e que faz parte de uma das seis correntes teóricas da CI, dentro dos estudos de natureza matemática, juntamente com a bibliometria (ÁVILA, 2009). Para Targino (1995), a RI se configura como o conjunto de conhecimentos relativos à origem, coleta, organização, armazenamento, recuperação, interpretação, transferência, transformação e utilização da informação, ou seja, refere-se a todo o ciclo informacional. Para a autora, essa corrente emergiu como decorrência natural do processo de evolução da Biblioteconomia e Documentação (TARGINO, 1995), que tem como marco o primeiro catálogo de biblioteca realizado no século II a.C por Calímaco (SANDERSON; CROFT, 2012).

Com o acelerado desenvolvimento tecnológico é observado que a partir do início do século XX já existiam depósitos de patentes sobre dispositivos para auxiliar a recuperação da informação. Um exemplo foi o pedido de patente feito por Soper, em 1918, sobre um dispositivo que tinha como objetivo agilizar a recuperação de documentos que possuíssem categorias combinadas; outros depósitos de patentes incluem os realizados por Goldberg entre 1920 e 1930 para a recuperação por entrada (SANDERSON; CROFT, 2012).

A explosão da informação ocorreu após a segunda guerra mundial e alguns autores entendem que o desenvolvimento da CI teve como motivação as indagações de Bush (1945) em seu artigo "*As we may think*", onde é definido o problema crítico da explosão de informação, que pode ser descrito: "Como tornar acessível e disponível o conhecimento, considerando a explosão de informação?" Bush propõe uma solução para essa questão: o dispositivo mecânico denominado de Memex (BUSH, 1996), cuja ideia principal era permitir o armazenamento e a recuperação de grandes volumes de informação.

Poucos anos depois, Cavin Mooers, cientista da computação, propõe outra solução: um sistema de codificação tipo cartão perfurado e, ao explicar o funcionamento de seu dispositivo, ele propõe a primeira definição para o conceito RI, como sendo: um método que viabiliza que um usuário transforme sua

necessidade de informação em uma lista de referências de documentos com informações relevantes³ (MOOERS, 1951).

Ainda na década de 1950, pesquisas apontavam a preocupação sobre a indexação e a recuperação de documentos. E na década de 1960, há um movimento nas bibliotecas no sentido de indexar itens de suas coleções através de palavras-chave. Computacionalmente se tem a formalização de algoritmos para classificar documentos, a introdução de feedback de relevância e a associação estatística de termos com significado semântico semelhante (FERNEDA, 2003).

O avanço das tecnologias de informática alterou a natureza do RI fazendo com que, cada vez mais, a recuperação da informação deixasse de ser um processo manual, a partir de material físico e/ou impresso, para um processo computacional a partir de um material que podia ser armazenado e manipulado através de uma máquina. Assim, na década de 1970, houve o desenvolvimento de pesos de frequência de termo (ocorrência de palavras-chave dentro de um documento) e de frequência inversa de ocorrência (palavras menos comuns tendiam a se referir a conceitos específicos importantes na recuperação), assim como o desenvolvimento do modelo de espaço vetorial (SANDERSON; CROFT, 2012) que trouxeram ganhos significativos para a RI. Outros desenvolvimentos importantes foram a derivação do modelo probabilista no princípio de classificação de probabilidades e em outras modelagens e a incorporação da dependência de termos na recuperação classificada (SANDERSON; CROFT, 2012).

A partir da década de 1980, a popularização do computador e o avanço nas Tecnologias da Informação e Comunicação (TICs) permitiram o aumento da capacidade de armazenamento e processamento. Desde então, com a expansão da rede mundial de computadores e a viabilização de seu uso, a internet passou a ser a principal fonte de busca de informação de qualquer natureza e para viabilizar o processo de busca é muito comum o uso de ferramentas, motores, mecanismos ou interfaces de busca dentro de sites,

³ "Information retrieval is the name for the process or method whereby a prospective user of information is able to convert his need for information into an actual list of citations to documents in storage containing information useful to him."

portais (IACONO; DE ALMEIDA; NAGANO, 2011) e Sistemas de Recuperação da Informação (SRI).

A RI se faz hoje digitalmente através dos SRI que, para Godinho são “um conjunto de operações consecutivas executadas para localizar, dentro da totalidade de informações disponíveis, aquelas realmente relevantes”, ou seja, para o autor, as funções dos SRI incluem a seleção, análise, indexação e busca de informações (GODINHO, 2007).

Sobre SRI, dois aspectos, mais especificamente, influenciam o presente trabalho e serão explorados mais à frente. O primeiro é a indexação de documentos, onde é relevante destacar o papel dos descritores, que são a representação do assunto temático nos documentos e envolvem a descrição do conteúdo através de termos, os quais são normalmente retirados de vocabulários controlados denominados de tesouro, para a sintetização de seu conteúdo. O segundo aspecto, elencado por Lancaster (1966), como indispensável a qualquer SRI, é a taxa de revocação⁴ (*recall* é o termo equivalente na língua inglesa) que é o número de documentos procurados recuperados dividido pelo número total de documentos procurados existentes na coleção, e a taxa de precisão⁵, que se refere ao número de documentos procurados recuperados dividido pelo número total de documentos recuperados.

Os modelos de RI são utilizados dentro dos SRI e possuem diferentes categorizações, como a apresentada por Baeza-Yates e Ribeiro Neto (2013) que indica três categorias baseados em: texto, links e em objetos multimídia. Como o objeto do presente trabalho é a RI em texto nos atentaremos apenas a essa categoria, cujas subdivisões são apresentadas na **Figura 1**.

Para os documentos tipo texto, observamos duas subcategorias: os modelos clássicos de RI para texto não estruturado e o texto semiestruturado (BAEZA-YATES; RIBEIRO-NETO, 2013).

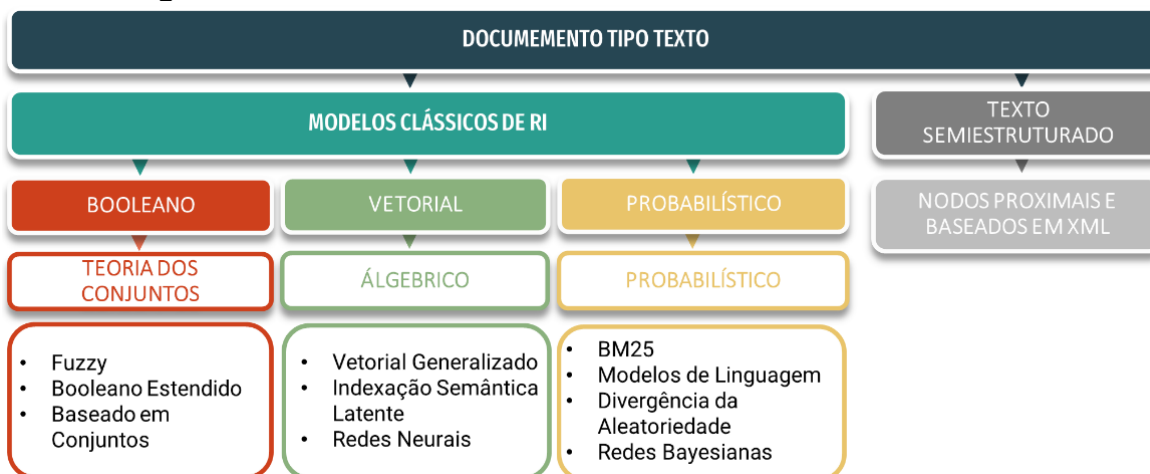
Os modelos clássicos consideram que o documento é composto por palavras-chave que os representa (termos de indexação) e essas são utilizadas na

⁴ Taxa de revocação = número de documentos desejados recuperados dividido pelo número total de documentos desejados existentes na coleção

⁵ Taxa de precisão = número de documentos desejados recuperados dividido pelo número total de documentos recuperados

recuperação da informação e pode ser feita através dos métodos: booleano, vetorial ou probabilístico (VELHO, 2001).

Figura 1: Taxonomia de modelos de RI baseados em texto



Fonte: adaptado de Baeza-Yates (BAEZA-YATES; RIBEIRO-NETO, 2013).

O método booleano usa a teoria dos conjuntos para representar os documentos e as buscas através de termos de indexação com o uso de operadores “and”, “or” e “not”, entre outros. No método vetorial os documentos e buscas são representados como vetores em um espaço de n dimensões e, por isso, se diz que é um modelo algébrico, uma vez que apresenta pesos associados aos termos de indexação para calcular o grau de similaridade, o que auxilia o ordenamento dos resultados. Por fim, no método probabilístico, as representações dos documentos e buscas são baseadas na teoria das probabilidades que utiliza pesos para melhoria da ordenação.

Em um texto não estruturado as informações estão em forma livre, o oposto do que ocorre nos textos estruturados que possuem uma estrutura rígida como banco de dados e planilhas eletrônicas. Já os textos semiestruturados são o meio termos entre as opções anteriores. Eles possuem uma estrutura bem definida, como título, seções e parágrafos. No entanto, dentro dessa estrutura, ele é composto por texto não estruturado. As técnicas de indexação para textos semiestruturados são nodos proximais e baseados em XML (VELHO, 2001).

O modelo de RI implementado em um SRI é normalmente uma informação sigilosa, podendo fazer parte da estratégia de negócios de uma empresa. Um exemplo de modelo de RI de sucesso é o algoritmo de PageRank da Google.

Esse exemplo ilustra um caso de sucesso da resolução do problema enunciado por Bush ainda nos anos de 1940, que tem como aspecto central a RI e que assume grande protagonismo no século XXI, em que a população do planeta, incluindo os cientistas, precisam fazer escolhas e decidir sobre como acessar e recuperar a informação de interesse frente ao volume cada vez maior de informação circulando, em especial, nos ambientes virtuais.

Essa nova demanda justifica o presente trabalho que propõe um modelo para auxiliar o desenvolvimento da estratégia de busca, com peculiaridades distintas de trabalhos anteriores, os quais são apresentados mais à frente.

O papel da estratégia de busca na RI

Bates distingue o processo de busca em base de dados em duas etapas: uma realizada pelos indivíduos e outra pelo SRI (BATES, 1990). Muito embora os SRIs tenham o papel de facilitar a recuperação da informação, um desafio sempre presente está no desenvolvimento da estratégia de busca, por ser um processo complexo e que, segundo Kremer, é tão importante quanto a qualidade das bases de dados, onde se faz necessário, além do conhecimento do tema, a compreensão e familiaridade dos SRI que serão utilizados (KREMER, 1985).

A estratégia de busca é definida por Lopes como “Uma técnica ou conjunto de regras para tornar possível o encontro entre uma pergunta formulada e a informação armazenada em uma base de dados” (LOPES, 2002), ou seja, é o meio que o usuário se comunica com o SRI e pode ser realizada de diferentes formas com resultados muito distintos.

Após expor diversos pontos que autores consideram importantes para o desenvolvimento de uma estratégia de busca, Lopes (2002), em seu artigo de revisão, propõe um modelo para desenvolvimento da estratégia de busca em sete etapas. Vale explicar que, naquele momento, a autora não tinha acesso direto às bases de dados, mas somente através da biblioteca de sua instituição, que prestava assistência aos pesquisadores para a recuperação da informação. Considerando esse contexto, em que as buscas eram conduzidas por bibliotecários, ou seja, profissionais com treinamento e acesso à informação e com conhecimento sobre o processo de indexação, características e particularidades de bases de dados, a autora propõe um modelo pautado nas seguintes etapas:

- 1- discussão do tópico de pesquisa, cujo objetivo é deixar claro ao intermediário a ideia da pesquisa e como sugestão a essa compreensão se tem a pergunta - Como os resultados da busca irão ser aplicados?;
- 2- conhecimentos básicos sobre os instrumentos de busca, se refere as possibilidades existentes que ajudarão no desenvolvimento e seleção dos termos como por exemplo o uso de tesauro e uma listagem de palavras-chave;
- 3- formulação provisória, se refere a estratégia inicial feita pelo intermediário onde se agrupa termos similares com ajuda do pesquisador;
- 4- compreensão lógica dos conjuntos, o intermediário precisa o conhecimento da teoria dos conjuntos e desenvolvimento de estratégia para computador e auxiliar o pesquisador nessa compreensão;
- 5- interdisciplinaridade, entendimento dos campos correlatos a pesquisa;
- 6- eliminação de termos indesejados, e por último;
- 7- especificação de parâmetros relevantes, aqui a autora traz os seguintes questionamentos ao pesquisador: A busca deve ser limitada nos anos mais recentes? Quais as bases de dados que provavelmente irão fornecer as mais relevantes citações? O pesquisador quer todas as citações que mencionam uma autoridade particular ou somente as que são autorizadas por uma pessoa particular? (LOPES, 2002).

Anos antes, Rogers (1980) propôs um modelo com 6 passos, a saber:

- 1- esclarecer a questão do estudo (entrevista);
- 2- estabelecer os parâmetros da busca baseados na entrevista;
- 3- identificar o(s) sistema(s) onde deverá ser feita a busca;
- 4- traduzir (indexar) a questão para a linguagem do sistema;
- 5- realizar a busca; e
- 6- fornecer a informação (TUCKER; MARSHALL, A. P.; TUCKER, 1980).

Os modelos propostos por Lopes e Rogers têm o mesmo objetivo da presente proposta assim como todas as suas etapas são abrangidas em nosso modelo,

porém, com forma e sequência diferentes visto que não há, em nossa proposta, o papel do intermediário (o profissional da biblioteca), mas sim a figura do pesquisador que é estimulado a desenvolver uma sequência de raciocínio para amadurecimento da estratégia que deve ser focada em sua pergunta de pesquisa. O modelo proposto tem o diferencial da recuperação da informação tecnológica representa pelas patentes e agrega informações de bases de patentes existentes.

Esses dois exemplos de modelos podem ser aplicados a qualquer tema ou área, no entanto, existem algumas áreas que possuem modelos específicos para sanar suas particularidades, como exemplo a área de saúde, que possui os modelos de estratégia como: PICO, PICO, PICOT, PICOD, SPICE, SPIDER, PCC, ECLIPSE entre outros. Esses modelos consideram aspectos intimamente associados à área da saúde e, por isso, muitas vezes, devem ser incorporados nas estratégias de busca, como: achados clínicos, etiologia, manifestações clínicas, diagnóstico, testes diagnóstico, prognóstico, terapia, prevenção, experiência e significado, melhoria, tempo, população, instrumento de coleta entre outros (CÁSSIO; ARAÚJO, 2020).

A literatura sobre modelos de estratégia de busca desenvolvidos por autores brasileiros nos últimos 10 anos, com foco em recuperação da informação em bases de documentos científicos e/ou de patentes, voltados para estudos métricos, indica três artigos, os quais têm objetivos diferentes da presente proposta. Os três estudos são apresentados brevemente a seguir.

No artigo “Proposta de metodologia para a RI da produção científica em CI na base BRAPCI” de 2010, os autores propõem uma metodologia de RI para uma base de dados de documentos acadêmicos e científicos em CI, que se baseia na atribuição de pesos aos campos textuais de busca de acordo com sua importância, aprimorando a relevância do conteúdo na RI (FREITAS *et al.*, 2010). Em “Information retrieval in institutional repositories using the summarization technique derived from the selection of Cassiopeia attributes” de 2020, os autores exploram técnicas que otimizam a busca e RI em bases de textos acadêmicos, avaliando se o uso da técnica de sumarização, baseada na seleção de atributos (palavras) do modelo Cassiopeia, pode auxiliar na recuperação de informações, com o objetivo de reduzir a sobrecarga de informações e aumentar a precisão dos resultados fornecidos ao usuário (CRUZ; GUELPELI, 2020). Já no artigo

“Recuperação da informação científica sobre doenças tropicais negligenciadas: análise comparativa da Scopus, Pubmed e *Web of Science*” de 2017, os autores desenvolvem um método para recuperar informações científicas sobre as Doenças Tropicais Negligenciadas a partir de bases de dados internacionais através de um modelo que busca informações científicas usando os termos padronizados definidos nos Descritores em Ciências da Saúde da Biblioteca Virtual de Saúde (DeCS) (SOBRAL, 2017).

Em relação aos três artigos apresentados brevemente acima, os dois primeiros visam aumentar a revocação e precisão, proposta que difere do objetivo do presente estudo que visa, principalmente, aumentar a revocação da RI para aplicação em estudos quantitativos. O terceiro artigo tem um foco específico para uma área cujo diferencial, o uso de descritores, é abrangido pelo modelo proposto no presente estudo.

Não foram identificados na literatura estudos, com os recortes feitos em autores brasileiros para os últimos dez anos, com propostas de modelo de recuperação da informação em base de patentes, outra característica de nosso modelo. E, além das comparações realizadas anteriormente, o modelo proposto no presente estudo tem o diferencial da multiplicidade de fontes para recuperar o máximo de termos relevantes para serem incluídos na estratégia de busca, além de permitir, através dos grafos, a identificação de áreas correlatas e termos novos ainda não estabelecidos.

Por fim, sobre a importância da escolha da base de dados, suas características e qualidade do SRI que será utilizado, explanamos no Apêndice 1, informações sobre os recursos importantes a serem verificados antes da escolha do SRI a ser utilizar.

A seguir, a proposta de modelo para auxiliar o desenvolvimento de estratégia de busca, focada em estudos cientométricos e patentométricos e que pode ser aplicado a temas emergentes, é apresentada.

O Modelo para Desenvolvimento de Estratégia de Busca

O modelo é apresentado na Figura 2 e está dividido em nove etapas: pré-requisito, termos semente, SRI científica, tesouro, grafo, especialista, estratégia científica, SRI tecnológica e estratégia tecnológica. As setas da figura mostram o caminho a ser seguido.

Figura 2: *Esquema do modelo para desenvolvimento da estratégia de busca em estudos quantitativos.*



Fonte: Elaboração própria

O modelo se inicia com a etapa **Pré-requisito** que parte do entendimento que, para o melhor desenvolvimento de qualquer estratégia de recuperação de informação, é necessário que a pergunta de pesquisa, os objetivos do estudo e a escolha do SRI sejam claros e estejam bem definidos, o que levará à seleção de termos semente e à escolha dos SRI mais apropriados. Para a seleção do SRI, é essencial considerar os principais SRI utilizados em estudos cientométricos, que incluem a *Web of Science (WoS)*, *Scopus*, *Dimensions*, *Lens* e outras plataformas de pesquisa de patentes, como *Espacenet*, *USPTO*, *Derwent Innovations Index (DII)*, *Lens* e *Orbit Intelligence*. Esses serão denominados de forma genérica de SRI. Essa consideração se deve pois elas não se resumem apenas a um banco de dados, mas possuem um sistema complexo de recuperação da informação além de outras funcionalidades.

Para a definição do SRI a ser utilizado, recomenda-se que sejam observados os seguintes critérios: especialidade, acesso, tipos de documentos, campos exportados, formatos de arquivos, filtros, operadores (de proximidade, exclusão,

booleanos e de truncamento) e quantidade de dados exportada por vez (esses critérios são trabalhados no apêndice 1). Também é necessário decidir pela utilização de um ou mais SRI e, para isso, deve-se avaliar o custo-benefício entre o tempo previsto para a retirada de duplicidades, padronização de dados e a quantidade de dados que será agregada. Caso se opte pela utilização de mais de um SRI, é importante utilizar estratégias de busca semelhantes com adaptações na escrita conforme especificação de cada SRI.

O pacote em R *biblioverlap* (VIEIRA; LETA, 2023), desenhado para analisar metadados⁶ extraídos dos SRI usados em estudos de bibliometria, é uma ferramenta que, entre outras funcionalidades, auxilia a tomada de decisão sobre a escolha de usar um ou mais SRI, pois ele permite comparar diferentes conjuntos de dados, marcando aqueles que estão em duplicidade e ainda cria um diagrama de Venn-Euler para representar os diferentes conjuntos de dados e suas sobreposições.

Devido à crescente quantidade de dados, que atingiu 59 ZB em 2020 e é estimado que irá alcançar 175 ZB até 2025 (REINSEL; GANTZ; RYDNING, 2017), empresas estão aproveitando essa disponibilidade para lançar novos negócios com base na visão que esses dados fornecem (REINSEL; GANTZ; RYDNING, 2017). Com uma previsão de mercado para a análise de dados de alto desempenho que vai de US\$ 87,50 bilhões em 2023 para US\$ 236,43 bilhões até o final de 2028 (MORDOR INTELLIGENCE, 2022), empresas que oferecem SRI e análise de dados tecnológicos estão implementando melhorias contínuas e agregando novos dados. Por exemplo, a *Questel* complementa seus dados de proteção intelectual com informações de literatura científica (QUESTEL, 2024).

Da mesma forma empresas que há décadas trabalham com documentos científicos, como a *Clarivate Analytics*, proprietária da *Web of Science*, estão incorporando outros tipos de dados, como os de propriedade intelectual, e expandindo os serviços oferecidos. Segundo a empresa elas oferecem dados, *insights* e análises para inovação (CLARIVATE, 2023). Com isso, melhorias em

⁶ Chamaremos aqui de metadados os dados que são disponibilizados pelos SRI, que facilitam a recuperação e organização da informação. O conceito de metadados (dado sobre dados) tem como origem a catalogação utilizada em bibliotecas e são utilizados para auxiliar a gestão e organização dos dados (MARICATO; JOSÉ MACÊDO, 2017).

serviços, análise de dados e inclusão de uma variedade de dados estão sendo disponibilizadas rapidamente por SRI.

Após finalização das decisões iniciais, temos a etapa denominada **Termos Semente**, que é uma lista de termos extraídos das leituras realizadas em: documentos, mídia, jornais, normas, relatórios técnicos, teses, dissertações, patentes, anais de eventos, além de palavras-chave utilizadas em artigos e artigos de revisão que o pesquisador buscou para delimitar, inicialmente, o tema. Com a lista inicial de termos semente, segue-se para a etapa **SRI Científica**, que consiste na realização de buscas individualizadas no/s SRI a partir de cada um dos termos semente, recuperação e testes. Nesse momento de exploração, o pesquisador pode usar mais de um SRI, mesmo que este não seja usado para a recuperação dos metadados após a finalização da estratégia.

Nessa fase, se deve fazer testes com cada um dos termos com o objetivo de validá-los, identificar e eliminar erros da estratégia que está sendo desenvolvida. Ainda para a validação dos dados se deve fazer a leitura de resumos aleatoriamente para checar se a estratégia recuperou algum falso positivo, ou seja, quando um termo corresponde a busca realizada, mas entre os documentos recuperados há algum que não se refere ao interesse do estudo (um exemplo: um estudo cujo interesse é a produção científica referente ao Zika vírus, o uso do termo de busca “zika” recuperará documentos sobre a raça de coelhos zika que não é de interesse do estudo, mas vem corretamente através do termo utilizado).

A próxima etapa do modelo que é a recuperação de outros termos em potenciais é chamada de **Tesouro**, pois aqui o pesquisador consultará listas de vocabulário controlado e estruturado de termos utilizados tanto para análise de assuntos quanto para a recuperação de documentos (UNESCO, 2024). A ideia é, portanto, identificar a existência de um tesouro da área de interesse de seu estudo, pesquisar pelos termos já conhecidos e buscar a existência de outros. Teste cada um deles na base de dados e, se confirmada a importância desses termos, acrescente-os na nova estratégia de busca.

Volte para o **SRI Científica** e recupere os metadados para serem utilizados na próxima etapa que é a do **Grafo**, esses serão utilizados para a criação de um grafo de co-ocorrência de palavras-chave.

Há vários programas disponíveis que permitem a criação de grafos desta natureza, como o VOSviewer,⁷ que é gratuito e com interface e comandos de uso relativamente simples. Através deste programa é possível evidenciar os termos que mais se destacam e exportar a lista de palavras-chave com a quantidade de ocorrências e força dos nós.

A lista de palavras-chave auxilia a identificação de novos termos relacionados ao estudo que devem ser testadas no SRI Científica e as identificadas como de possíveis contribuições devem ser acrescentadas na estratégia de busca. Caso não seja identificado nenhum termo, passe para a fase seguinte.

Para temas emergentes é provável que se leve mais tempo até que se consolide a denominação que se tornará mais usual e a rede de palavras-chave ajudará nessa identificação.

Se achar pertinente verifique a existência de novos termos no Tesouro para os novos termos. Com os novos termos obtidos através do Grafo, o pesquisador procederá uma nova busca (etapa SRI científico) e, com base nos documentos recuperados, se achar pertinente, poderá verificar a ocorrência de sinônimos (etapa Tesouro), repetindo esse processo até considerar que a ocorrência de novos termos foi esgotada.

De posse de uma lista de termos relacionados ao tema de interesse, a próxima etapa é a consulta ao **Especialista** para validar a estratégia de busca ou fazer os ajustes. Participações em eventos científicos sobre o tema de interesse são iniciativas importantes para encontrar especialistas que podem, então, atuar como consultores e, assim, avaliar a lista de termos e sanar dúvidas pontuais. Caso novos termos sejam indicados pelo especialista, retorne à etapa do SRI, teste os termos e acrescente-os, se pertinentes, à lista de inclusão.

Na etapa **Estratégia Científica**, que ocorre após a consulta a um especialista, há a definição dos termos de interesse à pesquisa, os quais são usados na recuperação da informação no SRI científica escolhido na etapa pré-requisito. Caso o estudo englobe temas relacionados a tecnologia, desenvolvimento de produtos ou a análise de patentes, passe para a etapa **SRI Tecnológica**. Nesta etapa, recomenda-se que o pesquisador adapte a estratégia científica para as especificações do SRI Tecnológica, acrescentando aos termos já existentes os

⁷ <https://www.vosviewer.com>

códigos de classificação de patentes, como a Classificação internacional de patentes⁸ (termo em inglês *International Patent Classification (IPC)*) utilizada nas bases de dados com sua sigla em inglês e a Classificação Cooperativa de Patentes (termo em inglês *Cooperative Patent Classification (CPC)*⁹) e, se conveniente, use como termos a descrição desses códigos. Cabe ao pesquisador avaliar o nível (classe, subclasse ou grupo) da classificação que deve ser utilizado, conforme o tema e pergunta do estudo. Identificar os termos de pesquisa dentro das classificações internacionais é essencial para uma busca eficiente nas plataformas de patentes. No entanto, o nível que se deve utilizar (classe, subclasse, grupo) deve ser avaliado conforme o tema de estudo.

A etapa final é a **Estratégia Tecnológica** e ocorre após a identificação e definição dos termos obtidos na/s base/s tecnológica/s e o acréscimo destes termos à estratégia científica (se for o caso). Cabe ao pesquisador decidir se é necessário ou não uma nova validação com um especialista. Sugerimos que isso se faça caso exista a descoberta de muitas alterações nas classificações e testes no SRI tecnológica.

Feito os passos anteriores, a estratégia final está pronta para ser utilizada no SRI Tecnológica.

Observe que as etapas Tesouro, Grafo e Especialista têm como retorno a alimentação da lista de termos e execução de testes no SRI Científica ou Tecnológica, o que é indicado pela seta em dois sentidos. Cada vez que a entrada nas etapas gera uma mudança na lista de termos, novos testes devem ser realizados. Cabe ao pesquisador avaliar o momento de finalização dessa realimentação e passagem para a etapa seguinte. Essa sequência ficará mais clara no exemplo de aplicação do modelo no próximo tópico.

A etapa Especialistas é independente das demais, sendo uma etapa coringa, podendo ocorrer sempre que for necessário. A consulta a um especialista no início do estudo para um brainstorming inicial facilita o amadurecimento dos termos, no entanto se isso não for possível é sugerido a consulta na fase final para validação ou ajuste da estratégia. No modelo essa etapa foi colocada no momento crucial da consulta ao especialista.

⁸ Criada em 1971 com o objetivo de padronizar de forma hierárquica as patentes e modelos de utilidade de acordo com a área tecnológica (WIPO, 2021).

⁹ Estabelecida em 2013 é baseada no IPC e tem o objetivo de ser mais detalhada e específica.

A seguir demonstraremos a aplicação do modelo para o tema impressão 3D.

Aplicação do modelo para o tema Impressão 3D

Manufatura Aditiva (MA) é um termo guarda-chuva sinônimo de impressão 3D que envolve sete processos e é definido, segundo a ASTM52900:2015, como o “processo de união de materiais para criar peças a partir de modelos em 3D, geralmente camada por camada, em oposição à fabricação subtrativa e à metodologia de fabricação formativa” (SANTOS, 2020).

Embora a impressão 3D não seja uma tecnologia recente, ela continua em franco crescimento, conforme indicado no relatório Wohlers de 2023, que mostra que a indústria de impressão 3D teve um crescimento em dez setores distintos e com receita global que ultrapassou 10 bilhões de dólares, enquanto os serviços contribuíram com mais de 7 bilhões de dólares. Essa indústria registrou uma taxa de crescimento médio notória de 18,3%, mantendo um crescimento de dois dígitos em 25 dos últimos 34 anos (WOHLER ASSOCIATES, 2023).

Outro aspecto de destaque da impressão 3D que evidencia sua importância na atualidade é sua contribuição na indústria 4.0, sendo um de seus pilares, à medida que a Indústria 4.0 se afasta da produção em massa e entra na personalização em massa (KHORASANI *et al.*, 2022), uma das características da impressão 3D.

O resumo dos passos da aplicação do modelo é apresentado na Figura 3, já sua descrição mais completa com a estratégia de busca utilizada nos SRI com os operadores e quantidade é apresentada no Apêndice 2.

Com base na pergunta de partida (a impressão 3D tem um domínio de pesquisa multifacetado?), foi definido o objetivo do estudo que foi identificar as áreas do conhecimento que se apropriam da temática impressão 3D.

Para a escolha do **SRI Científico** foram selecionadas inicialmente as fontes *Scopus*, *WoS* e *Dimensions*. Todas atendem as demandas de nossa pergunta de partida, com filtros, operadores booleanos, de truncamento, e permitem o *download* dos metadados com os campos necessários ao estudo. Como a quantidade de documentos e, portanto, metadados a se recuperar era grande,

optou-se por utilizar a *Scopus* uma vez que esse SRI permite a exportação da maior quantidade de metadados por vez¹⁰.

Os testes para patentes foram realizados na *DII* e *Questel Intelligence*. A escolha pela *Questel* se deu pela quantidade de campos exportados que é superior aos campos exportados no metadados da *DII*, o que permite uma análise mais detalhada.

Figura 3: Aplicação do modelo para MA.



Na etapa **Pré-requisito**, foi definida a pergunta de pesquisa, objetivos e seleção das bases a se utilizar.

Nosso ponto de partida foi o **Termo Semente** 3D print, que é o termo mais amplo para as tecnologias aditivas. Esses termos possuem diferentes possibilidades de escrita na língua inglesa como *3d print* e *three dimensional print*; essas variações foram utilizadas como termos semente.

Com os termos semente, procedemos para a etapa **SRI Científico**, realizando testes com cada termo para identificar a contribuição deles na recuperação de

¹⁰ Apenas na fase de revisão deste artigo e após o estudo que utilizou essa estratégia de busca já ter sido publicado que tivemos o conhecimento da plataforma aberta Lens (lens.org) que agrega patentes e trabalhos acadêmicos, possui todas as funcionalidades das demais, tem uma grande cobertura e seu grande diferencial é ser gratuita para acesso público quando a utilização é sem fins lucrativos.

documentos de interesse, além de identificar possíveis variações desses termos e a existência de falso positivo. O resultado dos testes foi acrescentado na lista de inclusão e exclusão da estratégia, conforme indicado no Apêndice 2.

Durante os testes, identificamos o termo “additive manufacturing” que foi acrescentado à estratégia. No entanto, com o desenvolvimento do estudo, foi possível compreender sua centralidade no tema, passando a ser o principal termo do estudo. Como resultado dessa etapa, foram recuperados 88.152 documentos.

Continuando a sequência da Figura 3, para a etapa **Tesouro**, considerando o caráter transdisciplinar do tema MA, buscamos por possíveis sinônimos em diferentes tesouros, como o de ciência da informação no “Tesouro Brasileiro da Ciência da Informação”, saúde no *Medical Subject Headings*¹¹ (*MeSH*), engenharia no *Engineering Village*, entre outros. Nessa etapa, foi encontrado, apenas no *MeSH*, o termo *Stereolithography* que é um processo de MA.

Com o acréscimo de *Stereolithography* na estratégia de busca, o número de documentos recuperados aumentou para 89.374. No entanto, com a descoberta desse processo, foi necessário rever a literatura sobre impressão 3D e, desta vez, buscando por outros processos e tecnologias ligadas a MA. Nesse processo, reconhecemos alguns, os quais, após os testes individuais, foram acrescentados a nossa lista, aumentando para 92.980 documentos recuperados. Nessa etapa, algumas dúvidas surgiram em relação a adequação ou não de alguns termos como por exemplo *Electron Beam Melting and Polyjet inkjet printing* (WONG; HERNANDEZ, 2012) (HUANG *et al.*, 2013), que aparecem em alguns artigos de revisão como sinônimo de MA, mas que não aparecem em outros.

As dúvidas foram anotadas e apresentadas na consulta a um **Especialista**, que ocorreu em um *workshop* sobre impressão 3D. Neste evento que contou com especialistas que atuam na indústria, foi possível conhecer outras tecnologias além das já identificadas. Sobre a dúvida mencionada anteriormente, o especialista informou sobre a existência da *ISO/ASTM 52900:2015* (antiga *ASTM-F2792*), elaborada pelo *Technical Committee Additive manufacturing*

¹¹ O *MeSH* é definido como um “vocabulário controlado e hierarquicamente organizado, utilizado para indexar, catalogar e pesquisar informações biomédicas relacionadas a saúde” (NATIONAL LIBRARY OF MEDICINE (US), 2023).

(ISO/TC 261), em cooperação com o *Committee F42 on Additive Manufacturing Technologies*, conforme *ISO/IEC Guide 21-1:2005* (STANDARD, 2015), norma essa que tinha sido identificada na literatura estudada sobre o tema.

Na *ASTM 52900* são estabelecidos os princípios gerais e a terminologia para a MA. No documento também é descrita a dificuldade de estabelecer definições claras para a nomenclatura associada a MA, pois durante o desenvolvimento dessa tecnologia houve diversos termos e definições diferentes em uso. Esta característica é comum às tecnologias emergentes, que precisam de um período para que se estabeleça as nomenclaturas mais adequadas, o que pode ser uma dificuldade na recuperação da informação científica.

Todos os processos de MA existentes na *ASTM 52900* foram, então, incluídos na estratégia de busca. Consultamos ainda todas as normas publicadas pela *ASTM* sobre MA na procura de outros termos, no entanto, não identificamos outras contribuições. Realizamos, assim, a primeira recuperação dos documentos e seus metadados no SRI. Nessa etapa, a estratégia de busca recuperou 132.043 documentos.

Seguindo para o desenvolvimento do **Grafo**, a partir dos metadados da Scopus foi elaborada a rede de coocorrência de palavras-chave para todas as palavras-chave utilizando o *software* VOSviewer, onde foi possível identificar aproximadamente o período em que os termos *bioprint* e *bioplotter* começam a serem utilizados, o que permitiu evidenciar a aplicação de impressão 3D na área biomédica e da saúde. Após testes no SRI, esses termos foram incluídos na estratégia de busca, o que resultou em 133.900 documentos recuperados.

Durante um evento científico, onde a estratégia de busca foi apresentada em detalhe, um **Especialista** questionou sobre o uso do termo “rapid prototyping”, que no português é “prototipagem rápida”. Na nota brasileira da ABNT NBR ISO/ASTM 52900:2018, que é a tradução da ISO/ASTM 52900, consta que prototipagem rápida ou impressão 3D são sinônimos de MA. No entanto, segundo Peter Foy e Dale Moody (FOY; MOODY, 2020), a prototipagem rápida é uma tecnologia utilizada tanto na impressão 3D quanto na manufatura

subtrativa¹², não sendo, portanto, uma tecnologia exclusiva para MA¹³. Desta forma, esse termo foi excluído da estratégia de busca para garantir maior precisão da estratégia (retirada de falsos positivos).

Com a exclusão do termo realizada no passo anterior, finalizamos a estratégia de busca para a informação científica. Na etapa **Estratégia Científica** recuperamos no SRI 116.199 registros, utilizando a estratégia nos campos: *title*, *abstract* e *keywords*.

Os dados da etapa anterior alimentam a etapa de **SRI Tecnológico** e, após consultas aos códigos internacionais de patentes IPC e CPC, verificação e testes de cada tecnologia listada na estratégia de busca científica, foi constatado que nenhum código relacionado às tecnologias era exclusivo. O código do IPC B33Y referente a MA foi utilizado finalizando a estratégia de busca para dados de tecnologia.

Na **Estratégia Tecnológica**, recuperamos a informação tecnológica aplicando a estratégia de busca nos campos: *title*, *abstract*, *description* e *object of invention* na plataforma da *Questel Intelligence*, com um retorno de 139.609 documentos de patentes e 98.014 famílias de patentes.

Comparando resultados

Para avaliação do resultado do modelo proposto no presente estudo, foi realizado um levantamento na literatura a fim de identificar estudos sobre a produção científica ou tecnológica no tema MA/impressão 3D. Foram identificados doze estudos, cujas estratégias são apresentadas no Quadro 1 com a especificação do tipo de documento, período, quantidade de documentos recuperados, base de dados e estratégia de busca utilizada.

Em relação a diversidade de termos usados nas estratégias de buscas, cinco estudos utilizam variações do termo *3D print*, sete utilizam MA, dois incluem algum processo relacionado a impressão 3D, dois utilizam *bioprint* (variação

¹² A manufatura subtrativa é um conjunto de processos embasados na mudança da forma de um material através da remoção de materiais, sendo a usinagem o processo mais importante (SANTOS, 2020) em contraposição a “Fabricação aditiva, i.e. fabricação de objetos tridimensionais [3D] por fabricação aditiva, deposição, aglomeração aditiva ou camadas aditivas, p. ex. por impressão 3D, estereolitografia ou sinterização seletiva por laser [2015.01] (Fonte: <http://ipc.inpi.gov.br/classifications/ipc/ipcpub/?notion=scheme&version=20240101&symbol=B33Y&menulang=pt&lang=pt&viewmode=f&fipopc=no&showdeleted=yes&indexes=no&headings=yes¬es=yes&direction=o2n&initial=A&cwid=none&tree=no&searchmode=smart>)

utilizada na área de saúde) e, por fim, três deles utilizam *rapid prototyping*, termo já discutido anteriormente.

No presente trabalho, utilizamos variações de impressão 3D, manufatura aditiva, os sete processos de manufatura aditiva e variações utilizadas na área da saúde como *bioprint* e *bioplotter*. O termo *rapid prototyping* não foi utilizado por não ser uma tecnologia exclusiva de processo aditivo, como explicado anteriormente. Já sobre o número de termos usados nas estratégias, observamos que apenas três estudos usaram mais de 10 termos, no entanto vale lembrar que existem estratégias que podem ser otimizadas com o uso de operadores coringas que utilizados no final de uma palavra, e a partir de seu uso aceita quaisquer terminações. Nossa estratégia utilizou 49 termos, que são: (((("Three Dimensional" OR "3D" OR "3 dimensional") AND ((print*) AND NOT (printex OR printo OR printmaking OR printgrammetry OR printboard OR printsipa OR printsip)) OR "additive manufactur*" OR Stereolithography OR "fused deposition model*" OR "Laminat* Object Manufactur*" OR "Selective Laser Sintering" OR "Laser Engineer* Net Shap*" OR "Electron Beam Melt*" OR polyjet OR "inkjet print*" OR "binder Jet*" OR "exone" OR "voxeljet" OR "direct* energy deposit*" OR "laser metal deposit*" OR "laser engineer* net shap*" OR "direct* metal deposit*" OR "direct* light fabricat*" OR "laser-based metal deposit*" OR "laser freeform fabricat*" OR "laser direct casting" OR lasform OR "laser additive manufactur*" OR "electron beam freeform fabricat*" OR "material extrusion" OR "fused filament fabricat*" OR "material jet*" OR "multi-jet model*" OR "powder bed fusion" OR "selectiv* laser sinter*" OR "direct metal laser sinter*" OR "selective laser melt*" OR "electron beam melt*" OR "selectiv* heat sinter*" OR "multi-jet fusion" OR "direct metal laser melt*" OR "sheet laminat*" OR "selective deposition laminat*" OR "ultrasonic additive manufactur*" OR "vat photopolymerizat*" OR "digital light process*" OR "continuous liquid interface production" OR bioprint* OR bio-print* OR bioplotter OR bio-plotter).

Quadro 1: Comparação de estratégias de busca em estudos cientométricos e patentométricos sobre impressão 3D.

Autores	Tipo de documento	Período (qtd. de anos)	Num. Docs Base de dados	Estratégia
Jin et al	Artigo e artigo de revisão	1995-2016 (22)	2769 Web of Science	"3D print*" or "three-dimensional print*" or "additive manufacturing"
Gupta; Dhawan	Publicações	2007-2016 (10)	7309 Scopus	"3D Printing" or "Three Dimensional Printing"
Jin et al	Artigo e artigo de revisão	2005-2016 (12)	2679 Web of Science	"3D print*" or "three-dimensional print*" or "additive manufacturing"
Ca vigglioli; Ughetto	Artigo, congresso e capítulos de livro	1982-2017 (36)	14926 Scopus	"Additive manufacturing" or "3D printing"
Repanovici; Nedelcu	Não informado	1983-2018 (36)	11529 Web of Science	3D printing in Engineering Fabrication field
Zarrabaitia-Bilbao et al	Artigo	2010-2019 (10)	6692 Web of Science	(3D or 3 D or 3-D or 3 dimension* or 3-dimension* three* dimensions* or desktop* or additive* or freeform) and (print* or manufactur* or fabricat*) plus (rapid* prototyp* or layer by layer or layer-by-layer)
Jemghili; Ait Taleb; Khalifa	Artigo, congresso e capítulos de livro	2010-2020 (11)	5482 Web of Science	"additive manufacturing" or "3 D printing" or "freeform manufacturing" or "layered manufacturing" or "rapid prototyping" - considerando todas as diferentes formas de escrita
Muhammad et al	Artigo	1999-2019 (21)	5697 Web of Science	"3D printing methods"
Bai et al	Artigo, artigo de revisão e patente	1990-2020 (31)	17505 Web of Science e Derwent Innovations	"3D printing" or "3 D printing" or "Three-dimensional printing" or "3 dimensional printing" or "3D Printable" or "3 D Printable" or "Three-dimensional Printable" or "3 dimensional Printable" or "3D print" or "3 D print" or "Three-dimensional Print" or "3 dimensional print" or "3D Printed" or "3 D Printed" or "Three-dimensional Printed" or "3 dimensional Printed" or "3D bioprinting" or "3 D bioprinting" or "Three-dimensional bioprinting" or "3 D bioprinted" or "3D bioprint" or "3 D bioprint" or "3D bioprint" or "3 D bioprinted" or "3 dimensional bioprinted" or "Three-dimensional bioprinted" or "3D bioprintable" or "Three-dimensional bioprintable" or "3 dimensional bioprintable" or "bioprinting" or "bioprint" or "bioprinted" or "bioprintable"
Miao et al	Patente	1990-2017 (38)	34090 Derwent Innovation	3D printing, Additive manufact, rapid prototype, Stereo Lithography Apparatus, Bioprinting, 3D mosaic, Digital brick lay*, Layered manufact*, Solid freeform fabrication, Laminated object Manufact, Digital Light Process, Selective Laser Sinter, Fused Deposition Model*, Direct Laser Fabrication, Direct Metal Deposition, Laser-clad-forming technology, Electron beam selective melt
Garcia-Leon; Gomez-Camperos; Jaramillo Zheng et al	Publicações	2008-2021 (14)	1271 Scopus	"additive manufacturing" or "3D printing" and "mechanical properties"
	Artigos	2000-2021 (22)	4767 Scopus	("additive manufacturing*" OR "print*") AND ("polymer*") AND ("multi-material*" OR "composite*" OR "dissimilar*")

Considerando a taxa de recuperação, ou seja, a quantidade de documentos recuperados dividida pelo total de anos de publicações dos 12 estudos, observamos que o trabalho com a maior taxa de recuperação foi de 730,9, no

artigo de Gupta e Dhawan (2018), onde foram recuperadas 7.309 publicações em dez anos. Ao utilizarmos a estratégia de busca do modelo proposto em nosso estudo e usando as mesmas especificações (tipo de documento, período, base, campos) utilizadas por Gupta e Dhawan, chegamos ao resultado de 12.409 artigos recuperados, o que representa um aumento de 69,8% em relação ao que os autores recuperaram.

Para os documentos tecnológicos, a maior recuperação identificada foi de 34.090 na DII para um período de 28 anos, que corresponde a uma taxa de recuperação de 1.217,5 patentes por ano (MIAO *et al.*, 2020). Da mesma forma, ao utilizarmos a estratégia desenvolvida em nosso estudo na base de patentes DII, foi possível recuperar 98.014 documentos em 56 anos, o que corresponde a uma taxa de recuperação de 1.750,2 patentes/ano, que corresponde a um aumento de 43,7% em relação à taxa que os autores encontraram.

Discussão e Considerações Finais

O objetivo desse trabalho foi propor um modelo de desenvolvimento de estratégia de busca para a recuperação da informação científica e tecnológica para auxiliar estudos métricos da ciência e tecnologia. Ele tem como público-alvo interessados na recuperação da informação científica e tecnológica.

O modelo propõe o desenvolvimento em vários passos, sempre que necessário, retornando a etapas anteriores e fazendo testes dentro do SRI para identificar as variações de diferentes termos, identificando os termos apropriados (melhoria de especificidade) e inapropriados (melhoria na precisão). Recomenda a utilização de tesouro para identificação de estrutura hierárquica e ligações entre termos, e ainda utiliza a coocorrência de palavras-chave para identificar variações de termos e sua utilização em diferentes áreas de pesquisa. Cabe ao pesquisador a constante reflexão para avaliar quando ele deve passar para os passos seguintes, tendo uma participação ativa e amadurecimento sequencial do tema, das características dos SRI e dos softwares utilizados.

No exemplo de aplicação para MA, o resultado na recuperação de documentos obtido a partir do modelo proposto é superior aos encontrados na literatura, isso porque ele inclui termos pouco utilizados, ou que estão em desuso, e ainda identifica variações de termos utilizados em diferentes áreas de estudo.

Um levantamento de dados incorreto ou incompleto pode levar a resultados imprecisos e conclusões equivocadas acerca do perfil ou tendência de pesquisa em um campo ou temas. Além da quantidade maior de documentos retornados, o modelo proposto também se diferencia pela possibilidade de identificação de particularidades de temas emergentes, multidisciplinares e/ou transversais, como é exatamente o caso usado neste trabalho: impressão 3D.

Por fim, vale mencionar que o modelo aqui proposto não necessita de desenvolvimento de *scripts*, programas, cálculos complexos ou computacionalmente pesados para se chegar ao resultado, podendo ser aplicado a qualquer tema e formas de documentos, em especial, as patentes.

Agradecimentos

A CAPES pela concessão da bolsa de doutorado à Andreia Galina (Proc. 88882.331955/2019-01) e financiamento do projeto de pesquisa coordenado por Jacqueline Leta (Proc. 434146/2018-8). Agradecemos ainda a contribuição dos pesquisadores Fabio Castro Gouveia, Rita de Cássia P. Machado e Célia R.S. Barbalho.

Contribuição de autoria: Andréia Cristina Galina: Conceitualização, curadoria de dados, análise formal, investigação, metodologia, validação, escrita, visualização, escrita-rascunho original.

Jacqueline Leta: Conceitualização, aquisição de financiamento, administração do projeto, supervisão, escrita-análise e edição.

Declaração de Conflito de Interesses: As autoras declaram não terem qualquer conflito de interesse.

Referências

ÁVILA, C. A. A. Correntes teóricas da ciência da informação. **Ciência da Informação**, v. 38, n. 3, p. 192–204, 2009. .

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca**. [S. l.: s. n.], 2013. v. 2, .

BATES, Mar. J. Where should ther person stop and the information search iterface start? **Information Processing & Management**, v. 26, n. 5, p. 575–591, 1990. <https://doi.org/0306-4573/90>.

BUSH, V. As we may think. **Interactions**, v. 3, n. 2, p. 35–46, 1996. <https://doi.org/10.1145/227181.227186>.

BUSH, V. **Science The Endless Frontier A Report to the President by Vannevar Bush , Director of the President Roosevelt’s Letter Science - The Endless Frontier**, n. July. [S. l.: s. n.], 1945. Disponível em: <https://www.nsf.gov/od/lpa/nsf50/vbush1945.htm>, visited in 07/05/2021.

CÁSSIO, W.; ARAÚJO, O. Recuperação da informação em saúde: construção, modelos e estratégias Health information retrieval: construction, models and strategies Recuperación de información de salud: construcción, modelos y estrategias. **Convergências em Ciência da Informação**, v. 3, n. 2, p. 100–134, 2020.

CLARIVATE. Clarivate - Let our intelligence move you. 2023. Disponível em: <https://clarivate.com/>. Acesso em: 10 fev. 2024.

CRUZ, L. A.; GUELPELI, M. V. C. Information Retrieval in Institutional Repositories Using The Summarization Technique Derived from the Selection of Cassiopeia Attributes. **Brazilian Journal of Development**, v. 6, n. 11, p. 88022–88041, 2020. <https://doi.org/10.34117/bjdv6n11-286>.

FERNEDA, E. **Recuperação de Informação: análise sobre a contribuição da ciência da computação para a ciência da informação**. 2003. 1–147 f. Universidade de São paulo, 2003.

FOY, P.; MOODY, D. Thermal Spray, Putting a Face on 3D Printing. 2020. **Plasma Powder and Systems, Inc.** Disponível em: <https://plasmapowders.com/technology/thermal-spray-putting-a-face-on-3d-printing/>. Acesso em: 14 set. 2020.

FREITAS, J. L.; BUFREM, L. S.; FAUSTINO, R.; JUNIOR, G. **Proposta de Metodologia para a Recuperação da Produção Científica em Ciência da Informação na base BRAPCI**. [S. l.: s. n.], 2010.

GLÄNZEL, W. **Bibliometrics as a research field: A course on Theory and Application of Bibliometric Indicators**. [S. l.: s. n.], 2003. Disponível em: https://www.researchgate.net/publication/242406991_Bibliometrics_as_a_research_field_A_course_on_theory_and_application_of_bibliometric_indicators.

GODINHO, M. M. Indicadores de C&T, inovação e conhecimento: onde estamos? Para onde vamos? **Análise Social**, v. 42, n. 182, p. 239–274, 2007.

HUANG, S. H.; LIU, P.; MOKASDAR, A.; HOU, L. Additive manufacturing and its societal impact: A literature review. **International Journal of Advanced Manufacturing Technology**, v. 67, n. 5–8, p. 1191–1203, 2013. <https://doi.org/10.1007/s00170-012-4558-5>.

IACONO, A.; DE ALMEIDA, C. A. S.; NAGANO, M. S. Interação e cooperação de empresas incubadas de base tecnológica: Uma análise diante do novo paradigma de inovação. **Revista de Administracao Publica**, v. 45, n. 5, p. 1485–1516, 2011a. <https://doi.org/10.1590/S0034-76122011000500011>.

KHORASANI, M.; LOY, J.; GHASEMI, A. H.; SHARABIAN, E.; LEARY, M.; MIRAFZAL, H.; COCHRANE, P.; ROLFE, B.; GIBSON, I. A review of Industry 4.0 and additive

manufacturing synergy. **Rapid Prototyping Journal**, v. 28, n. 8, p. 1462–1475, 2 ago. 2022. <https://doi.org/10.1108/RPJ-08-2021-0194>.

KREMER, J. M. Estratégia de Busca Search Strategy. **Revista da Escola de Biblioteconomia da UFMG**, v. 14, n. 2, p. 187–220, 1985. .

LANCASTER, F. W. Evaluating the Small Information Retrieval System*. **Journal of chemical documentation**, v. 6, n. 3, p. 158–160, 1966. Disponível em: <https://pubs.acs.org/sharingguidelines>.

LOPES, I. L. Estratégia de busca na recuperação da informação : revisão da literatura. **Ciência da Informação**, v. 31, p. 30–71, 2002. .

MARICATO, J. de M.; JOSÉ MACÊDO, D. Influência dos manuais da OCDE e da RICYT na literatura científica: contribuições para a construção de indicadores de CTel? 2017. **XVIII Encontro Nacional de pesquisa em Ciência da Informação - ENANCIB 2017** [...]. Marília: [s. n.], 2017.

MIAO, Z.; DU, J.; DONG, F.; LIU, Y.; WANG, X. Identifying technology evolution pathways using topic variation detection based on patent data: A case study of 3D printing. **Futures**, v. 118, n. March 2019, p. 102530, 2020. DOI 10.1016/j.futures.2020.102530. Disponível em: <https://doi.org/10.1016/j.futures.2020.102530>.

MOOERS, C. N. Zatocoding applied to mechanical organization of knowledge. **American Documentation**, v. 2, n. 1, p. 20–32, 1951. <https://doi.org/10.1002/asi.5090020107>.

MORDOR INTELLIGENCE. **High-performance data analytics market (2023-2028)**. Industry Report. [S. l.]: Mordor Intelligence, 2022. Disponível em: <https://www.mordorintelligence.com/pt/industry-reports/high-performance-data-analytics-market>.

NATIONAL LIBRARY OF MEDICINE (US). Medical Subject Headings (MeSH). 2023. **Mesh Terms**. Disponível em: <https://www.nlm.nih.gov/mesh/meshhome.html>. Acesso em: 14 jun. 2021. (container-title: Mesh Terms).

PIELKE, R. In retrospect: Science — the endless frontier. **Nature**, v. 466, n. 7309, p. 922–923, 2010. <https://doi.org/10.1038/466922a>.

QUESTEL. Questel Orbit Intelligence - IP Intelligence Software. 11 fev. 2024. **Questel**. Disponível em: <https://www.questel.com/patent/ip-intelligence-software/orbit-intelligence/>.

REINSEL, D.; GANTZ, J.; RYDNING, J. **Data Age 2025: The Evolution of Data to Life-Critical**. [S. l.]: Seagate, abr. 2017. Disponível em: <https://www.seagate.com/files/www-content/our-story/trends/files/Seagate-WP-DataAge2025-March-2017.pdf>.

SANDERSON, M.; CROFT, W. B. The history of information retrieval research. **Proceedings of the IEEE**, v. 100, n. SPL CONTENT, p. 1444–1451, 2012a. <https://doi.org/10.1109/JPROC.2012.2189916>.

SANTOS, G. A. dos. **Tecnologias mecânicas: materiais, processos e manufatura avançada**. [S. l.]: Érica, 2020. v. 1, . . Acesso em: 10 fev. 2024.

SOBRAL, N. V. Recuperação da Informação Científica sobre doenças Tropicais Negligenciadas: Análise Comparativa da Scopus, Pubmed e Web of Science. *In: XVIII Encontro Nacional de Pesquisa em Ciência da Informação-ENANCIB 2017* [...]. Marília-SP: [s. n.], 2017.

STANDARD, I. S. O. **ISO/ASTM 52900: 2015 Additive manufacturing-General principles-terminology**. [S. l.]: Standard, A. S. T. M., 2015. Disponível em: <https://www.iso.org/standard/69669.html>.

TARGINO, M. das G. A INTERDISCIPLINARIDADE DA CIÊNCIA DA INFORMAÇÃO COMO ÁREA DE PESQUISA. **Informação & Sociedade: Estudos**, v. 5, n. 1, p. 12–17, 1995. .

TUCKER, J. M.; MARSHALL, A. P., issue editor; TUCKER, J. M. **Current Library Use Instruction**. [S. l.: s. n.], 1980. v. 29, .

UNESCO. Vocabulary information. 12 fev. 2024. **UNESCO Thesaurus**. Disponível em: <http://vocabularies.unesco.org/browser/thesaurus/en/>. Acesso em: 14 jun. 2021.

VELHO, L. M. L. S. Estratégias para um sistema de indicadores de C & T no Brasil. **Parcerias Estratégicas**, n. 13, p. 109–121, 2001.

VIEIRA, G.; LETA, J. **Biblioverlap**. [S. l.: s. n.], 15 nov. 2023. Disponível em: <https://cran.r-project.org/web/packages/overlapping/index.html>.

WIPO. International Patent Classification (IPC). 2021. **World Intellectual Property Organization**. Disponível em: <https://www.wipo.int/classifications/ipc/en/>.

WOHLER ASSOCIATES. Wohlers Report 2023. 2023. **<https://wohlersassociates.com/product/wr2023/>**. Disponível em: <https://wohlersassociates.com/product/wr2023/>. Acesso em: 22 set. 2023.

WONG, K. V.; HERNANDEZ, A. A Review of Additive Manufacturing. **ISRN Mechanical Engineering**, v. 2012, p. 1–10, 2012. <https://doi.org/10.5402/2012/208760>.

Apêndice 1: Critérios para escolha do SRI.

Critérios para a escolha do Sistema de Recuperação da Informação Científica

Acesso
A Scopus e a WoS são pagas, já a Dimensions possui algumas funcionalidades gratuitas limitadas para download e outras funcionalidades pagas e a Lens é totalmente gratuita.

Campos Exportados
Nem todos os campos visualizados estão disponibilizados para download. Faça uma conferência prévia dos campos que são exportados.

Filtros e Operadores
Quanto maior a complexidade e recorte do tema mais os filtros e operadores ganham importância. Verifique se os filtros existentes atendem a necessidade da pesquisa. Os operadores de proximidade, exclusão e booleanos são disponibilizados por todas as plataformas com pequenas variações em sua escrita.



Especificidade

Existem SRI multidisciplinares como *Dimensions*, *Scopus* e *Web of Science* e *Lens*, cuja cobertura varia significativamente. I, plataformas especializadas por área do conhecimento (IEEE Xplore para engenharias) ou por região de interesse como a Scielo (12 países da América Latina, Caribe, África do Sul e Portugal).

Tipos de documentos

Diversos são os tipos de documentos: artigos, artigos de revisão, conferências, cartas, livros, capítulos de livros, material editorial, datasets, grants, ensaios clínicos, etc. Usualmente se utiliza artigos, no entanto é necessário avaliar quais contribuem com o problema da pesquisa. Para um tema emergente é interessante analisar conferências, pois o processo de publicação é mais dinâmico e rápido.

Formatos de arquivo

Se deve verificar se o tipo de arquivo exportado é compatível com os softwares de análise que serão utilizados. As três bases científicas de exemplo permitem exportação no formato BibTex e csv, além de outras possibilidades.

Quantidade de dados

A quantidade de dados exportada por vez pode ser um limitador para grandes quantitativos se executada diretamente na plataforma (sem APIs). A WoS permite o download de 1000 documentos, a Scopus 20000, Dimensions 500 e Lens 1000 documentos por vez.

Versão 2.0

¹ Singh VK, Singh P, Karmakar M, Leta J, Mayr P. The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*. 2021;126(6):5113.

Critérios para a escolha do Sistema de Recuperação da Informação Tecnológica

Tipos de documentos

Alguns escritórios disponibilizam outros tipos de propriedade intelectual além de patentes como modelo de utilidade, mas a maioria dos escritórios disponibilizam depósitos de patentes/família de patentes via Tratado de Operação de Patentes - PCT). As bases comerciais possuem dados de patentes e outras informações como bolsa de valores, informações de empresas entre outros, essas informações variam muito da base utilizada e do tipo de plano adquirido.

Formatos de arquivo

Todos os SRI que permitem a exportação de metadados testada permite exportação em formato de planilha eletrônica, além de outras opções. Os SRI comerciais e a Lens permitem a exportação em diversos formatos, normalmente entre as opções se tem csv e txt entre outros, já o Espacenet, USPTO, DII e Google Patents o download é no formato pdf.

Quantidade de dados

A quantidade de metadados permitida depende muito do plano assinado, mas normalmente é um volume muito superior do que a quantidade permitida nas bases científicas. A DII permite 1000 documentos por vez e a Dimensions Analytics e Dimensions Pluscom (pagas) permite o download de arquivos com até 50 mil linhas por vez.



Especificidade e Acesso

Escritórios, institutos de patentes ou de propriedade intelectual/industrial, disponibilizam dados de forma on-line e gratuita, para consulta e download individual, as mais abrangentes são: Espacenet2 (do Instituto Europeu de Patentes), PATENTSCOPE3 (da Organização Mundial da Propriedade Intelectual), USPTO4 (escritório de patentes e marcas dos Estados Unidos). Outros exemplos gratuitos são a Google Patents5 e Lens6. Existem ainda bases nacionais como o INPI7 do Brasil e LATIPAT pra busca em documentos em português e espanhol8. Ao e cadastrar nessas plataformas existem vantagens para o download de dados. Quanto as plataformas pagas temos a Derwent Innovations Index (DII) da Clarivate Analytics e muitas outras opções de plataformas comerciais com análises inclusas que permitem a exportação de grandes volumes de dados.

Campos Exportados

Existe uma variação de campos exportados no metadados, que devem ser verificados por plataforma, mas normalmente as pagas possuem a maior quantidade de campos para exportação.

Filtros e Operadores

São grandes as opções de filtros encontradas, com exceção na Derwent e Google patentes. Quando aos operadores é comum terem os operadores básicos booleanos e de aproximação a maioria dos SRI os possuem.

2<https://worldwide.espacenet.com>
6[Lens.org](https://lens.org)
Versão 2.0

3<https://patentscope.wipo.int/search/pt/search.jsf>
7inpi.org.br

4<https://pubs.uspto.gov/pubwebapp>
8<https://ip.espacenet.com/>

5www.google.com.br/patents

Apêndice 2: Aplicação do modelo com a estratégia de busca em etapas e quantidade de documentos recuperados.

Passo Etapa	Estratégia	Resultados
1	SRI	
2	Temos Semente	
3	SRI Científico	88.152
4	Thesaurus	89.374
5	SRI Científico	92.980
6	Especialista	132.043
7	Grato	133.900
8	Especialista	116.199
9	SRI Tecnológico	
10	Estratégia Tecnológica	139.609 pacientes 96.014 famílias

Passo Etapa	Estratégia Científica	Resultados
1	SRI	
2	Temos Semente	
3	SRI Científico	88.152
4	Thesaurus	89.374
5	SRI Científico	92.980
6	Especialista	132.043
7	Grato	133.900
8	Especialista	116.199
9	SRI Tecnológico	
10	Estratégia Tecnológica	139.609 pacientes 96.014 famílias

Passo Etapa	Estratégia Tecnológica	Resultados
1	SRI	
2	Temos Semente	
3	SRI Científico	88.152
4	Thesaurus	89.374
5	SRI Científico	92.980
6	Especialista	132.043
7	Grato	133.900
8	Especialista	116.199
9	SRI Tecnológico	
10	Estratégia Tecnológica	139.609 pacientes 96.014 famílias

This preprint was submitted under the following conditions: The authors declare that they are aware that they are solely responsible for the content of the preprint and that the deposit in SciELO Preprints does not mean any commitment on the part of SciELO, except its preservation and dissemination. The authors declare that the necessary Terms of Free and Informed Consent of participants or patients in the research were obtained and are described in the manuscript, when applicable. The authors declare that the preparation of the manuscript followed the ethical norms of scientific communication. The authors declare that the data, applications, and other content underlying the manuscript are referenced. The deposited manuscript is in PDF format. The authors declare that the research that originated the manuscript followed good ethical practices and that the necessary approvals from research ethics committees, when applicable, are described in the manuscript. The authors declare that once a manuscript is posted on the SciELO Preprints server, it can only be taken down on request to the SciELO Preprints server Editorial Secretariat, who will post a retraction notice in its place. The authors agree that the approved manuscript will be made available under a Creative Commons CC-BY license. The submitting author declares that the contributions of all authors and conflict of interest statement are included explicitly and in specific sections of the manuscript. The authors declare that the manuscript was not deposited and/or previously made available on another preprint server or published by a journal. If the manuscript is being reviewed or being prepared for publishing but not yet published by a journal, the authors declare that they have received authorization from the journal to make this deposit. The submitting author declares that all authors of the manuscript agree with the submission to SciELO Preprints.

Este preprint foi submetido sob as seguintes condições:

- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que os dados, aplicativos e outros conteúdos subjacentes ao manuscrito estão referenciados.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa, quando aplicável, estão descritas no manuscrito.
- Os autores declaram que uma vez que um manuscrito é postado no servidor SciELO Preprints, o mesmo só poderá ser retirado mediante pedido à Secretaria Editorial do SciELO Preprints, que afixará um aviso de retratação no seu lugar.
- Os autores concordam que o manuscrito aprovado será disponibilizado sob licença [Creative Commons CC-BY](#).
- O autor submissor declara que as contribuições de todos os autores e declaração de conflito de interesses estão incluídas de maneira explícita e em seções específicas do manuscrito.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.
- O autor submissor declara que todos os autores do manuscrito concordam com a submissão ao SciELO Preprints.