

# A INDEXAÇÃO DE ASSUNTO EM REPOSITÓRIOS INSTITUCIONAIS: MAPEAMENTO DE TENDÊNCIAS NA LITERATURA

## SUBJECT INDEXING IN INSTITUTIONAL REPOSITORIES: MAPPING TRENDS IN THE LITERATURE

Samara Sivirino Marques<sup>a</sup>  
Januario Albino Nhacuongue<sup>b</sup>

### RESUMO

**Objetivo:** analisa os estudos que abordam o processo de indexação em repositórios institucionais na Ciência da Informação. **Metodologia:** utiliza o mapeamento sistemático da literatura da Ciência da Informação dos últimos dez anos. O processo de mapeamento sistemático envolveu as etapas de planejamento, pesquisa, seleção, avaliação da qualidade, extração de dados, classificação de dados e análise dos estudos. Os dados foram extraídos das bases de dados *Web of Science*, BRAPCI e Scopus e analisados nas categorias: autores e instituições, objetivos dos estudos e desafios norteadores, métodos utilizados, contribuições e implicações práticas para a área de estudo. **Resultados:** os resultados revelaram um crescente interesse global na temática, acompanhada por novas abordagens na indexação. No entanto, foram identificadas lacunas significativas, como a necessidade de estudos mais aprofundados sobre a avaliação do impacto dos repositórios, a gestão do ciclo de vida dos dados, a integração com outros sistemas e a utilização de novos modelos de linguagens da Inteligência Artificial (IA). A análise dos métodos utilizados mostrou a diversidade de abordagens na pesquisa, com destaque para a revisão sistemática, análise de conteúdo e estudos de caso. Embora os estudos tenham contribuído significativamente para a melhoria da qualidade da indexação e do acesso à informação, identificaram-se lacunas em áreas como a indexação de dados não textuais e a interoperabilidade entre sistemas. **Conclusões:** o estudo demonstra a importância da indexação para a descoberta e o acesso da informação em repositórios institucionais, ao mesmo tempo em que identifica lacunas e desafios que precisam ser superados. A pesquisa contribui para o avanço do conhecimento na área, oferecendo *insights* valiosos para futuros estudos e práticas profissionais.

---

<sup>a</sup> Mestranda em Ciência da Informação pela Universidade Federal de São Carlos (UFSCar). Assistente de biblioteca do Instituto Federal de São Paulo (IFSP) - Campus Hortolândia. E-mail: samarasmarques28@gmail.com

<sup>b</sup> Doutor em Ciência da Informação pela UNESP, Campus de Marília. Docente do Departamento de Ciência da Informação - UFSCar. E-mail: januario@ufscar.br

**Descritores:** Indexação. Repositórios institucionais. Mapeamento sistemático da literatura. Ciência da Informação.

## 1 INTRODUÇÃO

A explosão de dados gerada pelos avanços das Tecnologias de Informação e Comunicação (TICs) democratizou o acesso à informação. No entanto, a heterogeneidade dos formatos e a ambiguidade das linguagens naturais tornaram os processos de representação e recuperação da informação mais desafiadores, exigindo soluções inovadoras para garantir a eficiência da busca.

Conforme Lancaster (2004) argumenta, além do volume e do tipo de material, a indexação de recursos de informação digitais e em rede traz desafios na decisão sobre a matéria indexável, isto é, sobre os assuntos que precisam ser representados e, conseqüentemente, o vocabulário adequado. Para Fugita (2017), o desafio da análise de assunto para a indexação reside na indissociabilidade entre texto, leitor e, principalmente, o contexto, que molda a interpretação e a atribuição de termos.

Ao analisar alguns acontecimentos que contribuíram para o surgimento e o desenvolvimento da Ciência da Informação, Araujo (2018) destaca novos fenômenos e novos desafios que se colocaram à Ciência da Informação desde o final do século XX. No contexto das teorias contemporâneas que direcionam novos paradigmas da área, o autor cita, por exemplo, o movimento em prol do acesso livre ao conhecimento, a curadoria digital, a representação e organização da informação, entre outros fatores. Um dos principais desafios na organização da informação atualmente reside na gestão de folksonomias, ou seja, de termos e *tags* atribuídos pelos próprios usuários em ambientes de compartilhamento como repositórios digitais, especialmente no contexto do autoarquivamento.

O desafio de preservar os recursos digitais e promover o acesso livre é analisado por Bayeza-Yates e Ribeiro-Neto (2013) na dimensão das necessidades de informação do usuário e seus diferentes níveis de complexidade. Para os autores, a descrição completa da necessidade do usuário não garante a obtenção de bons resultados nos processos de recuperação

porque a busca depende da capacidade de tradução dessa necessidade de informação em uma consulta ou em um conjunto de termos de indexação.

Em um estudo que visava analisar as implicações da indexação por extração no processo de representação na Biblioteca Digital de Monografias da Universidade Federal do Rio Grande do Norte, Santos (2017) identificou vários problemas de recuperação relacionados ao metadado assunto. Esses problemas abrangiam a polissemia, a ocorrência de erros ortográficos, a presença de descritores abrangentes e com multiplicidade de sentidos, abreviações nos termos, uso de siglas, etc., e eram decorrentes das práticas de indexação livre, adotadas pela maioria de repositórios institucionais, associada à ausência ou fragilidade das políticas de indexação.

Considerando que os repositórios institucionais, na sua maioria, adotam práticas de autoarquivamento, permitindo que o próprio usuário registre os metadados do documento, às vezes, comprometendo a qualidade da recuperação, questiona-se como a literatura da Ciência da Informação vem abordando o processo de indexação em repositórios institucionais nos últimos dez anos? Desse modo, objetiva-se analisar os estudos que abordam o processo de indexação em repositórios institucionais na Ciência da Informação.

Embora outros trabalhos tenham tratado sobre o processo de indexação em repositórios institucionais, como por exemplo Alves, Tartarotti e Fugita (2022), Bilhalva Cardoso e Miranda (2020), Assunção *et al.* (2022) e Sales *et al.* (2017), nenhum deles forneceu um mapa desta área de pesquisa na Ciência da Informação. A nossa pesquisa fornece uma visão geral que pode auxiliar a identificação de tópicos mais estudados e de tópicos que precisam de estudos adicionais na Ciência da Informação, contribuindo para o aprimoramento da organização da informação.

A relevância da discussão feita no trabalho pode ser evidenciada por duas vertentes complementares. A primeira internacional, que se manifesta através de iniciativas para conscientização sobre os desafios estratégicos, culturais e tecnológicos envolvendo ativos digitais, bem como para o engajamento das partes. Nesse contexto, a preservação digital é entendida como principal estratégia para garantir o acesso contínuo ou de longo prazo a materiais digitais,

trazendo benefícios sociais, culturais e econômicos, principalmente, no desenvolvimento científico e tecnológico. Nesta vertente, destacamos ações de instituições como *Digital Preservation Coalition*, Federação Internacional de Associações e Instituições Bibliotecárias (IFLA), Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO), entre outras.

A segunda edição das diretrizes da UNESCO (2021) para seleção do patrimônio digital para preservação a longo prazo, fornece uma visão mais profunda sobre os desafios relacionados ao volume, à velocidade, à efemeridade e à variedade de ativos digitais que são criadas e compartilhadas diariamente. Esses desafios não atingem somente instituições patrimoniais como bibliotecas, arquivos e museus, que tradicionalmente têm a responsabilidade de preservar recursos intelectuais e culturais produzidos por toda a sociedade, mas todos os setores públicos e privados.

A segunda vertente é nacional e é evidenciada por estratégias locais para a preservação, o acesso e o uso da informação. Nesta dimensão, destacam-se estratégias da Rede Cariniana de serviços de preservação digital de documentos eletrônicos brasileiros, que tem o objetivo de garantir o acesso contínuo a longo prazo.

O acesso contínuo a longo prazo de ativos digitais, conforme problematizamos, depende da qualidade de representação e recuperação em ambientes como repositórios institucionais e envolve diversas variáveis como usuário e/ou usuário indexador.

O presente trabalho está dividido em cinco seções. Além da introdução, que contextualiza o tema e apresenta o histórico do problema, a segunda seção apresenta o panorama da indexação e dos repositórios institucionais, destacando a sua importância e os desafios enfrentados na atualidade. A terceira seção descreve o processo de mapeamento sistemático e apresenta o protocolo de pesquisa. A quarta seção apresenta os resultados e destaca os principais achados dos estudos e as lacunas para pesquisas futuras. Finalmente, a última seção apresenta as principais conclusões.

## 2 INDEXAÇÃO E REPOSITÓRIOS INSTITUCIONAIS

A explosão informacional e a crescente complexidade da produção científica exigem a implementação de estratégias eficazes para a organização e recuperação da informação. A indexação e os repositórios institucionais, ao fornecerem mecanismos para a descrição e o acesso a objetos digitais, desempenham um papel fundamental nesse contexto, embora enfrentem desafios como a padronização de metadados e a interoperabilidade entre sistemas. A seguir, são analisados alguns dos desafios sobre representação e recuperação da informação.

### 2.1 INDEXAÇÃO

A indexação consiste na atribuição de termos representativos aos assuntos abordados em um documento, visando facilitar a sua localização e recuperação em sistemas de informação.

A indexação e a elaboração de resumos, atividades complementares segundo Lancaster (2004), convergem para um objetivo comum: representar o conteúdo de um documento de forma concisa e precisa, permitindo sua localização e recuperação eficientes. A indexação é uma atividade complexa que envolve a seleção de termos que reflitam o conteúdo semântico e contextual dos documentos. Ao atribuir termos relevantes aos documentos, o indexador cria pontos de acesso que facilitam a localização e a recuperação dos recursos de informação.

O indexador, conforme apontam Silva e Fujita (2004), realiza uma análise aprofundada do documento para identificar com precisão os conceitos-chave que representam sua tematicidade e que serão utilizados para a recuperação da informação, atendendo às necessidades dos usuários. A qualidade da indexação, determinada por sua política e exatidão, é fundamental para garantir a eficácia de um sistema de recuperação da informação, influenciando diretamente a precisão e a relevância dos resultados de busca (Lancaster, 2004).

A literatura sobre indexação em repositórios institucionais converge para a importância de políticas e práticas bem definidas, para garantir a qualidade e

a eficácia na recuperação da informação. Estudos como o de Fujita et al. (2023) evidenciam a relevância da padronização da catalogação de assuntos, enquanto Bandim e Correa (2019) destacam o papel dos vocabulários controlados nesse processo. A qualidade dos metadados, por sua vez, é crucial para a visibilidade e descoberta dos recursos digitais nos repositórios digitais, promovendo o seu acesso e a sua utilização. Yang e Perrin (2017) enfatizam a necessidade e a importância de desenvolvimento de estratégias eficazes para otimizar a indexação e melhorar a capacidade de descoberta do conteúdo.

Para garantir a qualidade da informação recuperada, Tartarotti (2020) defende a necessidade de investir na qualificação dos indexadores, proporcionando-lhes o conhecimento necessário sobre a área do conhecimento, as linguagens documentárias e as técnicas de indexação, além de promover a atualização contínua. As competências e habilidades dos indexadores incluem o domínio da área do conhecimento e a linguagem documentária utilizada; a capacidade de análise textual, síntese e conhecimento das necessidades dos usuários; o treinamento contínuo para garantir a atualização e a padronização dos processos de indexação, entre outras.

Além das abordagens tradicionais, majoritariamente consolidados na Biblioteconomia e Ciência da Informação, modelos colaborativos de indexação social, como os explorados por Santos (2016), apresentam potencial para enriquecer a representação e a recuperação da informação, ao envolver os usuários no processo de atribuição de termos. No entanto, a heterogeneidade dos termos utilizados e a necessidade de mecanismos de controle de qualidade representam desafios para a implementação desses modelos. Esses desafios são intensificados pelos altos custos associados à necessidade de mão de obra especializada de profissionais de informação e pela complexidade de manter a qualidade em contextos com grande volume de informação.

Ademais, o metalinguismo e o multiculturalismo se entrelaçam para criar um cenário complexo na representação e recuperação da informação. A construção de ontologias que capturem nuances culturais e linguísticas, além de lidar com a abstração inerente ao metalinguismo, é um desafio fundamental para garantir a acessibilidade e a relevância da informação em um mundo globalizado.

Os avanços obtidos em vários níveis, apesar de relevantes, revelam a necessidade de pesquisas contínuas para superar desafios como a complexidade dos metadados e a falta de padronização entre os repositórios, conforme destacado por Yang e Perrin (2017). A garantia da sustentabilidade e eficácia dos repositórios institucionais no futuro exige investimentos contínuos em pesquisa e desenvolvimento, visando aproveitar ao máximo as oportunidades oferecidas pelas tecnologias emergentes.

A comparação<sup>1</sup> entre repositórios universitários locais e o Scopus (Golub *et al.*, 2020) revelou uma lacuna significativa na indexação granular de conceitos complexos e interdisciplinares nas humanidades. A ausência de termos de indexação específicos para áreas como estudos culturais, teoria literária e história da arte dificulta a descoberta de pesquisas relevantes e a colaboração entre pesquisadores de diferentes disciplinas. Essa lacuna, que persiste desde a década de 1990, exige o desenvolvimento de tesouros especializados e a implementação de ferramentas de indexação semântica para garantir a acessibilidade e a visibilidade da produção científica em humanidades.

A busca por uma indexação abrangente, associada à necessidade de promover o acesso, impõe aos profissionais da informação a necessidade de desenvolver um repertório de habilidades cada vez mais amplo.

Tendo analisado a indexação e sua importância para a organização do conhecimento, passaremos agora a analisar a sua implementação prática nos repositórios institucionais, explorando as diferentes estratégias de indexação utilizadas nesses ambientes e seus impactos na eficácia da recuperação da informação.

## 2.2 REPOSITÓRIOS INSTITUCIONAIS

Os periódicos científicos foram o canal de informação dominante para a disseminação do conhecimento científico até o início dos anos 1990. No entanto, o alto custo das assinaturas, a dificuldade de publicação e o aumento

---

<sup>1</sup> O estudo analisou 649 artigos revisados por pares, revelando a falta de vocabulários controlados específicos para disciplinas de humanidades.

exponencial da quantidade de títulos limitaram o acesso à informação científica. Como resposta a esses desafios, os repositórios institucionais surgiram como uma alternativa, visando democratizar o acesso à produção científica. Influenciados por um crescente movimento em prol do acesso aberto, esses repositórios foram concebidos para garantir a livre disponibilização dos resultados de pesquisas financiadas com recursos públicos, promovendo assim a disseminação e a visibilidade do conhecimento científico (FREITAS, 2019).

O avanço das TICs impulsionou a produção científica e, conseqüentemente, a necessidade de ferramentas eficazes para organizar e disseminar esse conhecimento. Os repositórios institucionais surgiram como uma resposta a essa demanda, oferecendo um ambiente digital para armazenar, preservar e disponibilizar a produção acadêmica de forma aberta e acessível. A implementação bem-sucedida de um repositório depende de uma política institucional clara e abrangente, que defina os objetivos, as responsabilidades e os procedimentos a serem seguidos. Essa política deve contemplar aspectos como o acesso aberto, a gestão de coleções, a definição de metadados e o estabelecimento de um fluxo de trabalho eficiente. Além disso, a participação ativa da comunidade acadêmica, especialmente dos pesquisadores, o acompanhamento e a atualização constantes, são fundamentais para o sucesso do repositório, garantindo o seu povoamento e a disseminação da produção científica (Torino, 2017).

Para Sayão (2017), a crescente produção de dados científicos, caracterizada por sua heterogeneidade e complexidade, exige a implementação de infraestruturas robustas para sua gestão e preservação. A curadoria de dados, um conjunto de práticas que visam garantir a acessibilidade, reutilização e longevidade dos dados, emerge como um elemento fundamental nesse contexto, exigindo a integração de tecnologias, políticas e a colaboração de diversos atores.

Os repositórios institucionais desempenham um papel crucial na preservação e disseminação dos dados de pesquisa, mas a implementação eficaz desses sistemas enfrenta desafios relacionados à definição de políticas de autoarquivamento, à garantia da interoperabilidade dos metadados e à

proteção da privacidade dos dados em todo o seu ciclo de vida, o que exige uma abordagem multidisciplinar e a colaboração entre diferentes atores.

O aumento da produção científica e a necessidade de ampliar a visibilidade das instituições de ensino superior têm impulsionado o desenvolvimento de repositórios institucionais como ferramentas eficazes para a divulgação da produção acadêmica (ARAKAKI, 2019).

Para a UNESCO (2022) “a ideia por trás da Ciência Aberta é permitir que as informações científicas, dados e resultados sejam mais amplamente acessíveis (acesso aberto) e aproveitados de forma mais confiável (dados abertos)”. A ciência aberta pode ajudar a resolver grandes desafios locais e globais ao tornar resultados de pesquisa e tecnologia acessíveis a todos, mas pesquisadores alertam que o Brasil ainda não tem uma política específica para ciência aberta.

Reconhecendo que a ciência aberta não deve apenas fomentar um maior compartilhamento do conhecimento científico exclusivamente entre as comunidades científicas, mas também promover a inclusão e o intercâmbio de conhecimento acadêmico originário de grupos tradicionalmente sub-representados ou excluídos (como mulheres, minorias, indígenas, acadêmicos de países menos favorecidos e línguas com poucos recursos) e contribuir para reduzir as desigualdades no acesso ao desenvolvimento científico, às infraestruturas e às capacidades entre diferentes países e regiões (UNESCO, 2022, p. 5)

Os Repositórios Institucionais exibem novas estratégias para o compartilhamento da informação científica sendo um notável recurso para a visibilidade dessa informação. A implementação de repositórios institucionais é uma estratégia eficaz para aumentar a visibilidade da produção científica das universidades, fortalecer suas metas de desenvolvimento institucional e contribuir para a sua avaliação institucional.

As novas formas de colaboração entre pesquisadores e a sociedade civil estão reconfigurando as dinâmicas de produção e validação do conhecimento, desafiando os modelos tradicionais de ciência e promovendo a construção de um conhecimento mais relevante e socialmente engajado.

Conforme demonstrado por Aguiar *et al.* (2024), o monitoramento da Política Científica e Tecnológica é fundamental para compreender e aprimorar a interação da sociedade com a produção científica, indo além da mera

comunicação e abrangendo indicadores como engajamento e participação. O monitoramento contínuo da Política Científica e Tecnológica é essencial para acompanhar as mudanças nas demandas da sociedade e ajustar as estratégias de comunicação da ciência, e isso inclui políticas e estratégias de organização, disseminação, acesso e uso da informação armazenada em repositórios digitais.

Para os autores Fachin, Blatmann e Caldin (2019, p. 89) “[...] acesso aberto é um meio de disseminação do conhecimento sem restrições de fronteiras, taxas, com o conhecimento do autor na disponibilização do conteúdo, de forma a socializar rapidamente as pesquisas e os pesquisadores”.

A modelagem conceitual FRBR, em conjunto com as TICs, revolucionou os processos de catalogação, possibilitando a criação de catálogos mais flexíveis e capazes de atender às necessidades dos usuários contemporâneos, conforme descrito por Arakaki, Simionato e Santos (2017):

[...] a inserção das tecnologias no campo da informação significa além de um grande valor intrínseco, uma repercussão favorável, não só naqueles que trabalham diretamente com a área de informação, mas para todos aqueles que, no dia a dia, necessitam de informação. A dimensão tecnológica assumiu um papel preponderante nos processos sociais e econômicos, afetando inclusive a atuação profissional. Para o bibliotecário é imprescindível que desempenhe um papel ativo direcionado à orientação de seus usuários. Entre as suas funções, estão relacionadas a seleção das fontes de informação, acesso, uso e reuso informacional.

Entre as características dos repositórios institucionais estão a disponibilização de conteúdos em acesso aberto e a realização do autoarquivamento. A indexação livre, frequentemente utilizada no autoarquivamento, embora facilite a disponibilização de documentos, apresenta limitações em termos de abrangência e precisão quando comparada à indexação realizada por especialistas. Conforme aponta Freitas (2019), a invisibilidade de um documento na busca eletrônica compromete sua utilidade e, conseqüentemente, seu impacto.

Diante da importância da indexação para a eficácia dos repositórios institucionais, a próxima seção detalhará a metodologia empregada neste estudo para analisar a literatura da Ciência da Informação.

### 3 METODOLOGIA

Para responder à questão norteadora, realizou-se um estudo exploratório e descritivo, por meio do mapeamento sistemático da literatura. O mapeamento sistemático da literatura é um método de pesquisa usado para analisar um domínio específico de forma abrangente, categorizando a literatura relevante para identificar padrões, lacunas e áreas para pesquisas futuras. É particularmente útil em áreas com um escopo amplo como Ciência da Informação, onde percepções exploratórias, em parte relacionadas ao vasto e variado volume de publicações, se mostram necessárias.

Para Felderer e Carver (2018), ao contrário das Revisões Sistemáticas da Literatura, que buscam responder a uma questão de pesquisa específica com base em todas as evidências disponíveis, os mapeamentos sistemáticos fornecem um mapa dos principais tópicos que podem servir de base para pesquisas futuras.

Embora as revisões de escopo e de mapeamento sejam variações<sup>2</sup> do mapeamento sistemático da literatura, conforme Christou, Parmaxi e Zaphiris (2024), ambas auxiliam os pesquisadores a entenderem o escopo e a amplitude da literatura em um determinado campo, identificar lacunas na pesquisa e fornecer uma visão geral abrangente das evidências disponíveis.

O processo de mapeamento sistemático desenvolvido neste trabalho envolveu as etapas descritas por Felderer e Carver (2018): planejamento do estudo, pesquisa de estudos, seleção de estudos, avaliação da qualidade do estudo, extração de dados, classificação de dados, análise e relatórios.

O planejamento do estudo inclui a justificativa da necessidade de uma revisão, o gerenciamento do protocolo de revisão, a especificação das questões

---

<sup>2</sup> A revisão de escopo é um tipo de síntese de conhecimento que usa um processo sistemático para mapear as evidências sobre um assunto e identificar ideias-chave, teorias, fontes e lacunas de conhecimento. O objetivo de uma revisão de escopo é incluir todas as informações relevantes disponíveis, incluindo literatura cinzenta. A revisão de mapeamento é orientada por uma questão e inclui uma pesquisa completa e sistemática de um campo amplo. Ele identifica o corpo da literatura atualmente disponível sobre um assunto e aponta quaisquer lacunas nas evidências. A revisão do mapeamento pode ser conclusiva na descrição das evidências disponíveis e na identificação de lacunas, enquanto a revisão do escopo é de natureza exploratória (Christou; Parmaxi; Zaphiris, 2024, p. 4).

de pesquisa, o desenvolvimento do protocolo de revisão e a validação do protocolo de revisão. Os fatores que motivam a questão norteadora deste estudo (como a literatura da Ciência da Informação vem abordando o processo de indexação em repositórios institucionais?) foram descritos na seção da introdução.

A identificação de estudos primários relevantes para inclusão no mapeamento pode ser feita por meio de diferentes abordagens. Para atingir um nível aceitável de integridade dentro das restrições deste estudo, combinamos a pesquisa manual em períodos selecionados com o método *Backwards Snowballing*<sup>3</sup>, examinando trabalhos citados por estudos incluídos na revisão (Felderer; Carver, 2018). As bases selecionadas para a pesquisa de estudos são Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (BRAPCI), Biblioteca Digital Brasileira de Teses e Dissertações (BDTD) e Web of Science (WoS).

O processo de seleção dos estudos em um mapeamento é uma etapa crítica que garante a validade e relevância dos resultados obtidos. Esse processo envolve a definição de critérios de inclusão e exclusão, baseados nos objetivos do estudo; a realização do processo de seleção e a determinação da relação entre artigos e estudos.

O planejamento do nosso estudo foi baseado na abordagem PICOC: População ou problema, Intervenção, Comparação ou controle, Resultados e Contexto (Felderer; Carver, 2018). O protocolo do mapeamento sistemático está sintetizado no Quadro 1.

**Quadro 1** - Protocolo do mapeamento sistemático da literatura

| <b>Critérios</b>  | <b>Descrição</b>  |
|-------------------|---|
| Título            | A indexação de assunto em repositórios institucionais: uma análise a partir do mapeamento sistemático da literatura                   |
| Objetivo          | Analisar os estudos que abordam o processo de indexação em repositórios institucionais na Ciência da Informação                       |
| Questão principal | Como a literatura da Ciência da Informação vem abordando o processo de indexação em repositórios institucionais nos últimos dez anos? |

---

<sup>3</sup> O método *Forward Snowballing* examina trabalhos que citam artigos incluídos na revisão (Felderer; Carver, 2018, p. 51).

|                               |  |  |
|-------------------------------|--|--|
| Questões secundárias          | Q1: Qual a participação de diferentes tipos de atores (autores, instituições e países) nas pesquisas sobre indexação em repositórios institucionais nos últimos dez anos?  |  |
|                               | Q2: Quais os principais objetivos das pesquisas sobre indexação em repositórios institucionais e como se relacionam com os desafios enfrentados na prática da indexação?   |  |
|                               | Q3: Quais metodologias de pesquisa são mais utilizadas nos estudos?  |  |
|                               | Q4: Quais as principais contribuições teóricas e práticas das pesquisas para a área da Ciência da Informação e para a gestão de repositórios?  |  |
| População                     | Propostas (estudos de caso e discussões) de autores da Ciência da Informação sobre a indexação de assuntos em repositórios institucionais  |  |
| Intervenção                   | Métodos e abordagens para enfrentar os desafios atuais da indexação de assuntos em repositórios institucionais   |  |
| Comparação ou controle        | Identificação dos principais insights ou avanços, em comparação com os trabalhos citados por estudos incluídos na revisão  |  |
| Resultado                     | Mapeamento de métodos, técnicas, cenários e aplicações para aprimorar a organização e o acesso da informação em repositórios institucionais  |  |
| Contexto de Aplicação         | Aplicações voltadas para a área de organização da informação, em particular a indexação de assunto em repositórios institucionais  |  |
| Termos                        | Indexação de assunto ou metadado assunto   | <i>Subject index or subject metadata</i> |
|                               | Repositórios institucionais  | <i>Institutional repositories</i>        |
| Idioma                        | Português e inglês   |  |
| Pesquisa de estudos           | Artigos publicados nos principais periódicos e nas principais bases que cobrem tópicos da Ciência da Informação  |  |
| Métodos de seleção            | Síntese do resumo, dos objetivos, dos principais achados e das principais contribuições teóricas e práticas  |  |
| Bases de dados                | <i>Web of Science</i> , BRAPCI e Scopus  |  |
| Inclusão                      | Estudos diretamente relacionados ao tema de estudo; estudos publicados em conferências ou periódicos revisados por pares; estudos publicados nos últimos dez anos; estudos da área da Ciência da Informação      |  |
| Exclusão                      | Estudos duplicados ou versões mais antigas do trabalho atual; documentos não acessíveis online; descarte de artigos menos relevantes para manter o tamanho gerenciável do mapa.                                  |  |
| Qualidade                     | Agrupamento de trabalhos semelhantes de conjuntos de autores; descrição dos procedimentos de análise de dados; especificação das limitações dos resultados e da aplicação dos resultados em situações concretas. |  |
| Campos de extração dos dados  | Principais percepções, objetivos, metodologia, contribuições e aplicações  |  |
| Síntese narrativa dos estudos | Organização e descrição dos estudos em categorias lógicas; análise dos principais achados dentro de cada uma das categorias; e síntese dos resultados.   |  |

**Fonte:** elaboração própria

O rigor aplicado nas etapas de planejamento, seleção, e análise dos estudos permite garantir a confiabilidade e a validade dos resultados obtidos. A aplicação de critérios estruturados e sistemáticos assegurou a inclusão de estudos relevantes e alinhados aos objetivos do mapeamento, permitindo uma visão abrangente e consistente do domínio de pesquisa. A seguir, na seção de resultados, são apresentados os achados do mapeamento sistemático da literatura, incluindo as principais categorias identificadas, as tendências observadas no campo de estudo e as lacunas que emergiram a partir da análise dos dados coletados. Esse panorama oferecerá *insights* fundamentais que podem contribuir para o avanço da área e para a formulação de pesquisas futuras.

#### 4 RESULTADOS E DISCUSSÃO

A presente seção apresenta os resultados obtidos a partir do mapeamento sistemático da literatura, com o objetivo de organizar e interpretar os dados coletados de forma clara e objetiva.

Nas etapas de busca e seleção dos estudos, foi realizada uma pesquisa exaustiva nas bases de dados *Web of Science*, *Google Scholar* e BRAPCI, combinando os descritores “indexação de assunto OR metadado assunto AND repositórios institucionais” ou “*Subject index OR subject metadata AND Institutional repositories*”. Essa busca resultou em um conjunto inicial de 693 artigos na base *Web of Science*, 32 artigos na base *Scopus* e 116 artigos na BRAPCI, totalizando 841 artigos, o que atesta a importância e a amplitude dos estudos sobre indexação em repositórios institucionais na comunidade científica nacional e internacional.

**Tabela 1** – Estudos primários recuperados de cada base

| <b>Base de dados</b> | <b>Quantidade</b> |
|----------------------|-------------------|
| Web of Science       | 693               |
| Scopus               | 32                |
| BRAPCI               | 116               |
| <b>Total</b>         | <b>841</b>        |

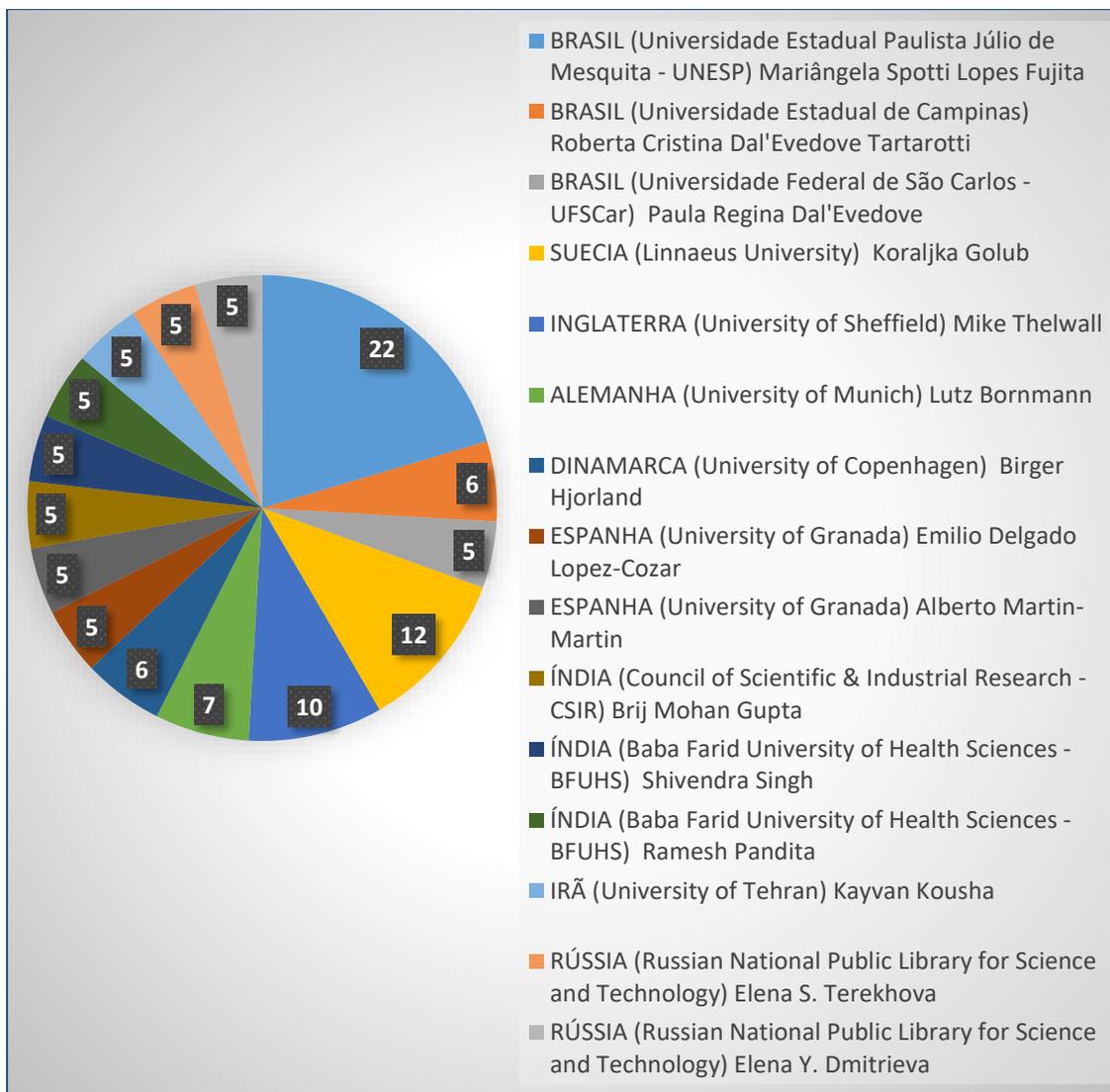
**Fonte:** Dados de pesquisa

Todos os estudos primários recuperados foram submetidos a um rigoroso processo de seleção e análise, a fim de garantir a inclusão apenas dos que atendiam aos critérios de inclusão estabelecidos e manter o tamanho gerenciável do mapa (Felderer; Carver, 2018). Dessa forma, 150 estudos selecionados foram analisados e categorizados com base nos critérios estabelecidos, permitindo identificar padrões, tendências, e lacunas no campo investigado.

A síntese narrativa dos estudos foi estruturada em torno de quatro categorias analíticas, a fim de proporcionar uma visão holística e comparativa das pesquisas. As categorias "autores e instituições", "objetivos dos estudos e desafios norteadores", "métodos utilizados", "contribuições e implicações práticas para a área de estudo" permitiram identificar as principais questões investigadas, as dificuldades enfrentadas pelos pesquisadores, os instrumentos e procedimentos adotados, bem como as contribuições teóricas e práticas de cada estudo para o campo da Ciência da Informação. A categoria "autores e instituições" permitiu analisar a relação entre os autores e as instituições produtoras de pesquisa na área de indexação em repositórios institucionais. A categoria "objetivos dos estudos e desafios norteadores" concentrou-se em identificar os propósitos de cada pesquisa e compreender as questões problematizadas pelos autores. A categoria "métodos utilizados" detalhou os instrumentos e procedimentos metodológicos empregados em cada estudo, permitindo avaliar a rigorosidade e a validade das pesquisas. Por fim, a categoria "contribuições e implicações práticas para a área de estudo" analisou o impacto de cada pesquisa para o avanço do conhecimento na área e suas possíveis aplicações práticas em diferentes contextos.

A análise detalhada dos estudos nessas categorias, apresentada a seguir, é fundamental para compreender o estado da arte da pesquisa e identificar as principais lacunas de conhecimento na área.

**Gráfico 1** – Categoria de autores e instituições de referência em indexação de assunto em repositórios institucionais



**Fonte:** Dados de pesquisa

A análise da categoria "autores e instituições" oferece uma visão geral da produção científica na área, permitindo identificar, principalmente, os autores e instituições que mais se dedicam aos estudos relacionados à indexação de assunto em repositórios institucionais. Quanto à diversidade geográfica, embora não explicitamente demonstrado no Gráfico 1, a presença de autores de diferentes países e instituições indica a existência de colaborações internacionais, o que enriquece a pesquisa e amplia o alcance dos resultados.

O gráfico 1 lista autores como Mariângela Spotti Lopes Fujita, Koraljka Golub, Mike Thelwall, Lutz Bornmann, Birger Hjørland e Roberta Cristina

Dal'Evedove Tartarotti, cuja trajetória consistente de pesquisa na área se reflete no número significativo de publicações. Contudo, a presença de outros autores com menos publicações sugere a entrada de novos pesquisadores no campo, o que favorece novas abordagens.

Os dados apresentados no Gráfico 1, obtidos da *Web of Science*, oferecem uma visão geral da produção científica sobre indexação de assunto em repositórios institucionais nos últimos dez anos, considerando a quantidade de cinco ou mais publicações dos autores nesta temática. A *Web of Science* foi escolhida por sua ampla cobertura, mas é importante ressaltar que a base de dados pode apresentar um viés de publicação e de língua, subestimando a produção em outras fontes e idiomas. Portanto, os resultados apresentados devem ser interpretados com cautela, considerando as limitações da base de dados utilizada e do contexto analisado.

Na síntese narrativa, a categoria "objetivos dos estudos e desafios norteadores" foi crucial para a compreensão abrangente das propostas de investigação e dos problemas motivadores dos estudos analisados. A análise dos objetivos e problemas de pesquisa apresentados revela uma diversidade de abordagens e interesses dentro da temática da indexação e dos repositórios institucionais, que foram agrupadas nas seguintes dimensões:

1. Melhoria da indexação e recuperação da informação. Nesta dimensão, destacam-se propostas para o desenvolvimento de novas técnicas de indexação, por exemplo, explorando técnicas como indexação semântica, indexação baseada em proximidade e indexação de dados não estruturados. Também se destacam propostas para avaliação de métodos de indexação em termos de precisão, revocação e outros indicadores; melhoria da eficiência da pesquisa em repositórios, por meio de desenvolvimento de algoritmos e modelos de indexação mais eficientes; desenvolvimento de *softwares* e ferramentas para automatizar e melhorar os processos de indexação. O artigo "*Semantic indexing of open scientific literature to help users discover and navigate through publications networks*", de Michel, Toulet e Bobasheva

- (2022), por exemplo, propõe um índice semântico para indexação semântica e visualização da literatura científica, facilitando a exploração de regras de associação temática e redes de coautoria.
2. Análise de repositórios institucionais. A maioria das propostas nesta dimensão concentra-se na avaliação do estado da arte ou da situação atual dos repositórios institucionais nos contextos nacional e internacional; na identificação dos principais desafios enfrentados pelos repositórios e proposição de soluções; e na análise de políticas de indexação e das práticas de gestão de repositórios. O artigo *“The indexing policy in the practices of brazilian institutional repositories: a diagnostic study from the perception of managers and indexers”*, de Fugita *et al.* (2023), por exemplo, analisa percepções de gestores e indexadores sobre a política de indexação, avaliando elementos e variáveis das práticas profissionais. Como principais contribuições, destaca a necessidade de melhorias nas práticas dos repositórios institucionais analisados em relação à catalogação de assuntos, padronização de metadados de assuntos, controle de vocabulário e recuperação de informações.
  3. Promoção do acesso aberto e da visibilidade da pesquisa. Nesta dimensão, destacam-se iniciativas e estratégias para incentivar pesquisadores a depositarem seus trabalhos em repositórios institucionais; para aumentar a visibilidade de ativos digitais em repositórios; e para promover o acesso aberto.
  4. Uso de tecnologias e ferramentas. Nesta dimensão, algumas propostas propõe a utilização de técnicas de mineração de dados para analisar grandes volumes de dados e extrair informações relevantes. O artigo *“Data mining as a tool for information retrieval in digital institutional repositories”*, de Fusco e Leticia (2014), por exemplo, sugere a utilização de técnicas de mineração de dados para melhoria dos resultados de busca.

Resumidamente, os objetivos de pesquisa da maioria dos estudos na

temática de indexação e repositórios institucionais buscam aprimorar os processos de indexação, aumentar a visibilidade da pesquisa, facilitar o acesso à informação e otimizar o uso de repositórios institucionais.

A partir da análise dos objetivos e dos desafios de pesquisa apresentados, é possível identificar algumas lacunas e oportunidades para futuras investigações na área de indexação e repositórios institucionais. Embora haja estudos sobre o uso de técnicas de Inteligência Artificial (IA) na indexação, ainda há um grande potencial para explorar o uso de aprendizado de máquina para aprimorar a recuperação da informação. A integração entre repositórios institucionais e sistemas de gestão de dados de pesquisa ainda é um desafio, especialmente no que diz respeito à interoperabilidade e à garantia de qualidade dos dados. Alguns dos desafios de integração se estendem ao *Linked Data*, campo com grande potencial para melhorar a descoberta e a reutilização da informação.

No que tange à avaliação do impacto dos repositórios, embora haja estudos sobre a visibilidade da pesquisa em repositórios, são necessários mais estudos para quantificar o impacto real desses ambientes na citação dos artigos e na carreira dos pesquisadores. A gestão do ciclo de vida dos dados é um tema complexo e sensível que exige mais investigação, principalmente, por abarcar aspectos éticos e de privacidade. A necessidade de desenvolver repositórios que atendam a diferentes linguagens e culturas (multiculturalismo) é um desafio crescente no âmbito da ciência aberta e do paradigma de integração de diferentes sistemas de conhecimento.

A aplicação de tendências tecnológicas atuais como *blockchain*, para garantir a autenticidade e a integridade dos dados depositados em repositórios, ou Inteligência artificial generativa, para gerar resumos e metadados automatizados, pode revolucionar os processos de indexação.

A categoria “métodos utilizados” revela uma diversidade de abordagens, refletindo a complexidade e a multidisciplinaridade do campo da indexação e dos repositórios institucionais. As principais estratégias utilizadas são: revisão sistemática, pesquisa exploratória, análise de conteúdo, estudos comparativos, estudos de caso, webometria e criação de protótipos. Embora os métodos

utilizados permitam a investigação e compreensão dos respectivos objetos de pesquisa, algumas lacunas podem ser identificadas. Por exemplo, no contexto de estudos longitudinais, que possibilitariam o acompanhamento da evolução dos repositórios institucionais ao longo do tempo e avaliação do impacto das mudanças nas políticas e tecnologias. Do mesmo modo, o uso de experimentos controlados para testar a eficácia de intervenções em políticas e práticas de indexação, integrando métodos consolidados de estudos de usuários na Biblioteconomia e Ciência da Informação como *card sorting*, grupos focais, protocolo verbal, entre outros, pode abrir espaço para novas abordagens.

Conforme mencionado anteriormente, a categoria “contribuições e implicações práticas para a área de estudo” analisou o impacto de cada pesquisa para o avanço do conhecimento na área e suas possíveis aplicações práticas em diferentes contextos. As principais contribuições dos estudos sobre indexação de assunto em repositórios institucionais são: a) melhoria da qualidade da indexação - necessidade de aprimorar a catalogação de assuntos, padronização de metadados e controle de vocabulário para garantir a recuperação precisa da informação; b) desenvolvimento de novas ferramentas e técnicas – a criação de índices semânticos, a utilização de técnicas de processamento de linguagem natural e o desenvolvimento de ontologias são exemplos de avanços tecnológicos que contribuem para a melhoria da indexação; c) combinação de diferentes abordagens - os estudos abordam a indexação sob diversas perspectivas, como a percepção dos usuários, a análise de políticas institucionais, o desenvolvimento de ferramentas e a avaliação de resultados; e d) alinhamento com o acesso aberto – a maioria dos estudos está alinhada aos princípios do acesso aberto, buscando garantir a visibilidade e a disseminação da produção científica (Bandim, 2017; Fujita *et al.* 2023; Tartarotti, 2020; Torino, 2017; Santos, 2016; Yang; Perrin, 2017).

O trabalho “*From necessary maintenance to intentional management: a workflow for remediating and enhancing institutional repository metadata*”, de Clark, Smith e Walker (2024), por exemplo, avaliou a qualidade dos metadados utilizados no processo de representação em repositórios e demonstrou como a utilização do *OpenRefine*, do *R* e do *Python* podem corrigir, complementar ou

substituir a descrição por metadados exportados de outras bases como *Web of Science*. A pesquisa “*Searching for Swedish LGBTQI fiction: the librarians’ perspective*”, de Golub, Bergenmar e Humelsjo (2023), destaca uma lacuna importante na busca por ficção sueca LGBTQI+. Os autores argumentam que os serviços de busca *online* atuais, focados em categorias amplas como gênero, tempo e lugar, não conseguem atender às necessidades dos usuários que buscam características mais específicas e subjetivas, como ritmo e estilo, presentes nas obras de ficção.

As pesquisas analisadas demonstram a relevância da indexação de assuntos para a descoberta da informação em repositórios institucionais. No entanto, ainda há desafios a serem superados, como a necessidade de desenvolver métodos para indexar dados não textuais, garantir a interoperabilidade entre diferentes sistemas, garantir a ética e privacidade na indexação e avaliar o impacto da indexação na visibilidade da pesquisa. Para avançar nessas subáreas, é fundamental a colaboração interdisciplinar entre bibliotecários, cientistas da informação, engenheiros da computação e outros profissionais, visando desenvolver soluções inovadoras e eficazes.

## 5 CONSIDERAÇÕES

A questão de como a literatura da Ciência da Informação tem abordado o processo de indexação em repositórios institucionais nos últimos dez anos é central para o desenvolvimento de estratégias eficazes de organização e recuperação da informação nesses ambientes. O mapeamento sistemático realizado neste estudo permitiu identificar as principais tendências, os principais desafios e as principais oportunidades nessa área, contribuindo para o avanço do conhecimento e a melhoria das práticas de indexação em repositórios institucionais.

As categorias “autores e instituições” (Q1) e “objetivos de estudo e desafios norteadores” (Q2) revelam uma notável diversidade geográfica e um crescente número de colaborações, evidenciando o interesse global na temática da indexação em repositórios institucionais. Os resultados indicam avanços significativos na compreensão dos desafios e oportunidades da indexação, com

destaque para o desenvolvimento de novas técnicas e a exploração de diferentes abordagens metodológicas. No entanto, a pesquisa identificou lacunas importantes, como a necessidade de estudos mais aprofundados sobre a avaliação do impacto dos repositórios, a gestão do ciclo de vida dos dados, a integração com outros sistemas e a aplicação de novas tecnologias como *blockchain* e IA. A emergência de novas tecnologias oferece um grande potencial para revolucionar os processos de indexação e gestão de repositórios, abrindo novas perspectivas para futuras pesquisas.

As categorias “métodos utilizados” (Q3) e “contribuições e implicações práticas para a área de estudo” (Q4) do mapeamento sistemático revelaram uma diversidade de métodos e abordagens na pesquisa sobre indexação em repositórios institucionais, com destaque para a revisão sistemática, análise de conteúdo e estudos de caso. Embora os estudos tenham contribuído significativamente para a melhoria da qualidade da indexação e do acesso à informação, identificaram-se lacunas em áreas como a indexação de dados não textuais, a interoperabilidade entre sistemas e a avaliação do impacto da indexação. A necessidade de abordagens interdisciplinares e o uso de novas tecnologias, como IA generativa, se apresentam como caminhos promissores para o avanço da pesquisa nessa área.

A presente pesquisa, ao focar na indexação de assunto em repositórios institucionais nos últimos dez anos e utilizar principalmente a *Web of Science*, apresenta limitações quanto à abrangência da análise e generalização dos resultados. No entanto, apesar dessas limitações, os achados contribuem para a compreensão dos desafios e oportunidades na área, especialmente para futuras pesquisas que buscam análises mais abrangentes e aprofundadas do tema.

## REFERÊNCIAS

AGUIAR, C. M. G. D.; SALLES FILHO, S. L. M.; PEREIRA, S. P.; COLUNAGNATI, F. A. B. Are we on the right path? Insights from Brazilian universities on monitoring and evaluation of Public Communication of Science and Technology in the digital environment. *Journal of Science Communication (JCOM)*, V. 23, n. 6, 2024.

Disponível em: [https://jcom.sissa.it/article/pubid/JCOM\\_2306\\_2024\\_A01/](https://jcom.sissa.it/article/pubid/JCOM_2306_2024_A01/). Acesso em 3 dez. 2024.

ALVES, L. S. S.; TARTAROTTI, R. C. D. E.; FUJITA, M. S. L. Avaliação de vocabulário controlado para a representação e recuperação de teses e dissertações em repositório institucional. **Revista Ibero-Americana de Ciência da Informação**, [S. l.], v. 15, n. 1, p. 283–294, 2022. DOI: 10.26512/rici.v15.n1.2022.42496. Disponível em: <https://periodicos.unb.br/index.php/RICI/article/view/42496>. Acesso em: 10 dez. 2024.

ARAKAKI, F. A.; SIMIONATO, A. C.; SANTOS, P. L. V. A. C. Catalogação e tecnologia: interseções com a Web Semântica. **Inf. Prof.**, Londrina, v. 6, n. 2, p. 03 – 19, jul./dez. 2017. Disponível em: <http://dx.doi.org/10.5433/2317-4390.2017v6n2p03>. Acesso em: 06 jun. 2021.

ARAKAKI, F. A.; CONEGLIAN, C. S.; SANTOS, P. L. V. A. da C.; SANTAREM SEGUNDO, J. E. Disponibilidad de metadatos en linked data para repositorios digitales. **Anales de Documentación**, v. 22, n. 2. 2019. Disponível em: <https://revistas.um.es/analesdoc/article/view/327441>. Acesso em: 05 jun. 2021.

ARAUJO, Carlos Alberto Ávila. **O que é ciência da informação**. Belo Horizonte: KMA, 2018.

ASSUNÇÃO, S. S.; DINIZ, J. V. C.; FUJITA, M. S. L.; REDIGOLO, F. M. Representação temática de material audiovisual em repositórios multimídia de instituições federais brasileiras: um estudo de caso com conceito paradigmático. **Scire: representación y organización del conocimiento**, v.28, n.1, p. 23-32, 2022. DOI: 10.54886/scire.v28i1.4755\_ Acesso em: 11 jan. 2023.

BANDIM, A. S. M.; CORREA, F. R. Indexação automática por atribuição de artigos científicos em português da área de Ciência da Informação. **Transinformação**, [S. l.], v. 31, 2019. Disponível em: <https://periodicos.puc-campinas.edu.br/transinfo/article/view/5921>. Acesso em: 28 jun. 2022.

BILHALVA CARDOSO, C. I.; MIRANDA, A. C. D. Modelo de Controle da Autoridade de nomes pessoais: uma proposta para o Repositório Institucional da Universidade Federal do Rio Grande. **ÁGORA: Arquivologia em debate**, [S. l.], v. 30, n. 60, p. 130–147, 2019. Disponível em: <https://agora.emnuvens.com.br/ra/article/view/894>. Acesso em: 20 set. 2023.

FACHIN, J.; BLATMANN, U.; CALDIN, C. F. Tendências e uso de repositórios de acesso aberto. **Ponto de Acesso**, [S. l.], v. 13, n. 2, p. 86–115, 2019. Disponível em: <https://periodicos.ufba.br/index.php/revistaici/article/view/32440>. Acesso em: 20 dez. 2022.

CHRISTOU, E.; PARMAXI, A.; ZAPHIRIS, P. A systematic exploration of scoping and mapping literature reviews. **Universal Access in the Information Society (UAIS)**, mai. 2024. <https://doi.org/10.1007/s10209-024-01120-3>. Disponível em: <https://link.springer.com/article/10.1007/s10209-024-01120-3>. Acesso em: 3 dez. 2024.

FELDERER, M.; CARVER, C. J. Guidelines for systematic mapping studies in Security Engineering. *In*: OTHMANE, L. B.; JAATUN, M. G.; WEIPPL, E. Empirical research for software security: foundations and experience. Nova Iorque: CRC Press Taylor & Francis Group, 2018.

FREITAS, M. P. **Autoarquivamento e representação de assunto**: estudo analítico de teses e dissertações do Repositório Institucional da UFSCar. Dissertação (Mestrado em Programa de Pós-Graduação em Ciência da Informação) - Universidade Federal de São Carlos, 2019.

FUJITA, M. S. L. A identificação de conceitos no processo de análise de assunto para indexação. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 1, n. 1, p. 60–90, 2004. DOI: 10.20396/rdbci.v1i1.2089. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/2089>. Acesso em: 13 mar. 2022.

FUJITA, M. S. L. A leitura documentária e o processo de compreensão do indexador: memorial de investigação científica. *In*: FUJITA, M. S. L.; NEVES, D. A. B.; DAL'EVEDOVE, P. R. (org.). Leitura documentária: estudos avançados para a indexação [online]. Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2017, p. 15-49. Disponível em: <http://books.scielo.org/id/3pk5m>. Acesso em: 3 fev. 2022.

FUJITA, M. S. L.; TARTAROTTI, R. C. D.; DAL'EVEDOVE, P. R.; PANUTO, J. C. The indexing policy in the practices of Brazilian institutional repositories: A diagnostic study from the perception of managers and indexers. **Journal of Librarianship and Information Science**, v. 56, n.2, jan. 2023. P. 424-442. <https://doi.org/10.1177/09610006221145544>. Disponível em: <https://journals.sagepub.com/doi/abs/10.1177/09610006221145544>. Acesso em: 4 nov. 2023.

GOLUB, K.; TYRKKO, J.; HANSSON, J.; AHLSTROM, I. Subject indexing in humanities: a comparison between a local university repository and an international bibliographic service. **Journal of Documentation**, v.76, n.6, p. 1193-1214, out. 2020. DOI 10.1108/JD-12-2019-0231. Acesso em: 13 out. 2024.

LANCASTER, F. W. **Indexação e resumos**: teoria e prática. 2. ed. Brasília, DF: Briquet de Lemos/Livros, 2004.

SALES, L. F.; ROCHA, L. L.; CAVALCANTI, M. T.; SALES, L. F.; ROCHA, L. L. Desenvolvimento de um vocabulário controlado para o repositório institucional carpedien. **Revista Eletrônica de Comunicação, Informação e Inovação em Saúde**, v. 11, 2017. DOI: [10.29397/reciis.v11i10.1398](https://doi.org/10.29397/reciis.v11i10.1398) Acesso em: 11 jan. 2023.

SANTOS, R. F. dos. Indexação em repositórios digitais: uma abordagem sobre o metadado assunto da Biblioteca Digital de Monografias da UFRN. **Revista Informação na Sociedade Contemporânea, [S. l.]**, v. 1, p. 1–22, 2017. DOI: 10.21680/2447-0198.2017v1n0ID12279. Disponível em: <https://periodicos.ufrn.br/informacao/article/view/12279>. Acesso em: 9 dez. 2024.

SAYÃO, L. F. o papel dos repositórios digitais na curadoria de dados de pesquisa. *In: VECHIATO, F. L. et al. (Org.). Repositórios digitais: teoria e prática.* Curitiba: EDUTFPR, 2017.

TARTAROTTI, R. C. D. Representação temática em repositórios institucionais de bibliotecas universitárias: a percepção de gestores e bibliotecários catalogadores-indexadores da USP, UNESP e UNICAMP. **RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação**, Campinas, SP, v. 18, n.00, 2020. DOI: 10.20396/rdbci.v18i00.8660803. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/8660803>. Acesso em: 18 dez. 2024.

TORINO, E. Políticas em repositórios digitais: das diretrizes à implementação. *In: VECHIATO, F. L. et al. (Org.). Repositórios digitais: teoria e prática.* Curitiba: EDUTFPR, 2017.

Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO): The UNESCO/PERSIST - Guidelines for the Selection of Digital Heritage for Long Term Preservation. 2. ed. UNESCO, 2021. Disponível em: <https://repository.ifla.org/items/9662f8de-5bfc-4e4a-8035-1d2a6add8913>. Acesso em: 14 fev. 2024.

Organização das Nações Unidas para a Educação, a Ciência e a Cultura (UNESCO): Recomendação da UNESCO sobre Ciência Aberta. UNESCO, 2022. Disponível em: [https://unesdoc.unesco.org/ark:/48223/pf0000379949\\_por](https://unesdoc.unesco.org/ark:/48223/pf0000379949_por). Acesso em: 14 mar. 2023.

YANG, L.; PERRIN, J. M. Introduction to the special issue of “digital collection metadata & Internet discovery”. *Journal of Web Librarianship*, v.11, p. 153-156, 2017. doi: 10.1080/19322909.2017.1383136. Acesso em: 9 nov. 2023.

## SUBJECT INDEXING IN INSTITUTIONAL REPOSITORIES: MAPPING TRENDS IN THE LITERATURE

### ABSTRACT

**Objective:** to analyze the studies that address the indexing process in institutional repositories in Information Science. **Methodology:** uses systematic mapping of Information Science literature from the last ten years. The systematic mapping process involved the stages of planning, searching, selection, quality assessment, data extraction, data classification and analysis of the studies. The data was extracted from the *Web of Science*, BRAPCI and Scopus databases and analyzed in the following categories: authors and institutions, study objectives and guiding challenges, methods used, contributions and practical implications for the field of study. **Results:** The findings revealed a growing global interest in the subject, accompanied by new approaches to indexing. However, significant gaps were identified, such as the need for more in-depth

studies on assessing the impact of repositories, data lifecycle management, integration with other systems and the use of new Artificial Intelligence (AI) language models. The analysis of the methods used showed the diversity of approaches in the research, with an emphasis on systematic review, content analysis and case studies. Although the studies have made a significant contribution to improving the quality of indexing and access to information, gaps were identified in areas such as indexing non-textual data and interoperability between systems. **Conclusions:** the study demonstrates the importance of indexing for the discovery and access of information in institutional repositories, while identifying gaps and challenges that need to be overcome. The research contributes to the advancement of knowledge in the area, offering valuable *insights* for future studies and professional practices.

**Keywords:** Indexing. Institutional repositories. Systematic literature mapping. Information Science.

## INDEXACIÓN TEMÁTICA EN REPOSITARIOS INSTITUCIONALES: TENDENCIAS BIBLIOGRÁFICAS

### RESUMEN

**Objetivo:** analizar los estudios que abordan el proceso de indización en repositorios institucionales en Ciencias de la Información. **Metodología:** Utiliza mapeo sistemático de la literatura en Ciencia de la Información de los últimos diez años. El proceso de mapeo sistemático involucró las etapas de planificación, búsqueda, selección, evaluación de la calidad, extracción de datos, clasificación de datos y análisis de los estudios. Los datos se extrajeron de las bases de datos *Web of Science*, BRAPCI y Scopus y se analizaron en las siguientes categorías: autores e instituciones, objetivos del estudio y retos orientadores, métodos utilizados, contribuciones e implicaciones prácticas para el campo de estudio. **Resultados:** Los resultados revelaron un creciente interés mundial por el tema, acompañado de nuevos enfoques de indización. Sin embargo, se detectaron importantes lagunas, como la necesidad de estudios más profundos sobre la evaluación del impacto de los repositorios, la gestión del ciclo de vida de los datos, la integración con otros sistemas y el uso de nuevos modelos lingüísticos de Inteligencia Artificial (IA). El análisis de los métodos utilizados mostró la diversidad de enfoques de la investigación, con énfasis en la revisión sistemática, el análisis de contenido y los estudios de casos. Aunque los estudios contribuyeron significativamente a mejorar la calidad de la indización y el acceso a la información, se detectaron lagunas en ámbitos como la indización de datos no textuales y la interoperabilidad entre sistemas. **Conclusiones:** El estudio demuestra la importancia de la indización para el descubrimiento y el acceso a la información en repositorios institucionales, al tiempo que identifica lagunas y retos que deben superarse. La investigación contribuye al avance del conocimiento en el área, ofreciendo valiosas *perspectivas* para futuros estudios y prácticas profesionales.

**Palabras clave:** Indexación. Repositorios institucionales. Mapeo sistemático de la literatura. Ciencia de la información.

**Recebido em:** 20.12.2024

**Aceito em:** 23.12.202