

# O autor é o indexador? A prática de indexação na perspectiva das transformações da comunicação científica em ciência aberta

Mariângela Spotti Lopes Fujita

**Como citar:** FUJITA, Mariângela Spotti Lopes. O autor é o indexador? A prática de indexação na perspectiva das transformações da comunicação científica em ciência aberta. In: JORENTE, Maria José Vicentini; PADRÓN, Dunia Llanes; NASCIMENTO, Natália Marinho do; SOUZA, Gabriela de Oliveira (org.). **Contextos Paradigmáticos da Ciéncia da Informação e as transformações em suas práticas.** Marília: Oficina Universitária; São Paulo: Cultura Acadêmica, 2025. p. 177-203. DOI: <https://doi.org/10.36311/2025.978-65-5954-650-3.p177-203>.



All the contents of this work, except where otherwise noted, is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (CC BY-NC-ND 4.0).

Todo o conteúdo deste trabalho, exceto quando houver ressalva, é publicado sob a licença Creative Commons Atribuição-NãoComercial-SemDerivações 4.0 (CC BY-NC-ND 4.0).

Todo el contenido de esta obra, excepto donde se indique lo contrario, está bajo licencia de la licencia Creative Commons Reconocimiento-No comercial-Sin derivados 4.0 (CC BY-NC-ND 4.0).

# CAPÍTULO 5

## O AUTOR É O INDEXADOR? A PRÁTICA DE INDEXAÇÃO NA PERSPECTIVA DAS TRANSFORMAÇÕES DA COMUNICAÇÃO CIENTÍFICA EM CIÊNCIA ABERTA

*Mariângela Spotti Lopes FUJITA*

### 1. INTRODUÇÃO

O ponto de partida de uma transformação nunca é imediato, mas sempre mediado pelo contexto cuja indução é influenciada pelas necessidades de avanços na sociedade humana. Por isso, quando nos referimos a alguma mudança, que mais à frente reconhecemos como paradigma estabelecido, podemos olhar o contexto e verificar que muitas foram as causas e muitos seres humanos estiveram ligados.

Não é diferente quando nos referimos à Ciência Aberta e a participação intensa de pesquisadores na preservação da produção de conhecimento. O conhecimento sempre foi e será um bem a ser cuidado e preservado

porque é a garantia de sobrevivência humana, a razão de sermos e de vivermos. Com o advento da internet e dos documentos digitais, Instituições e pesquisadores tornaram-se mais preocupados com o lugar do conhecimento e da garantia de preservação e acesso sem intermediários. Quando existiam apenas documentos impressos e fisicamente localizados sempre foi realizado por empresas comerciais de bases de dados, bibliotecas e arquivos e os autores precisavam passar por procedimentos que levavam anos para se efetivar e a divulgação científica não era sua competência.

Atualmente o direito do autor em enviar seu texto para publicação sem a mediação de terceiros é possível ser exercido durante o auto arquivamento em sistemas com interface para preenchimento de dados de identificação do documento e de descrição temática do conteúdo. Essa possibilidade surgiu com a criação do ArXiv, originalmente desenvolvido por Paul Ginsparg (Wikipedia, 2023a) e em Computação e Ciência da Informação por meio da Biblioteca CiteSeerx em 1997 criada pelos pesquisadores Lee Giles, Kurt Bollacker e Steven Laurence (Wikipedia, 2023b). As iniciativas ArXiv e CiteSeerx tornaram-se estruturalmente possíveis em 1999, com o movimento de acesso aberto por meio da Open Archives Iniciatiative (OAI) (Van de Sompel; Lagoze, 2000).

O auto arquivamento tem um impacto sociopolítico; é o maior desafio do acesso aberto e a solução que pode gerar maiores impactos para a universalização do acesso à informação e ao conhecimento científico. Em uma reunião realizada em Budapeste em dezembro de 2001, foi discutida a utopia do livre acesso à literatura publicada em revistas científicas. Hoje, isso se tornou realidade graças à combinação de uma tradição e uma nova tecnologia (Rodrigues, 2006, p. 2).

Mais do que parte do movimento de acesso à informação científica, o auto arquivamento pertence à nova configuração do ambiente Web, como uma rede de interação, na qual os próprios internautas interagem e colaboram na descrição dos conteúdos disponíveis, produzindo, classificando e reformulando o que já está disponível.

Na situação de auto arquivamento o autor torna-se um gestor de conteúdo documental e informacional em sistemas de informação que vai

além de simplesmente produzir o conteúdo, mas realizar os procedimentos de descrição para preenchimento de metadados tal como o de autoria, título, instituição responsável, local, data, e os de assunto referentes ao resumo e palavras-chave. Esses dois últimos são representações da informação significativa do conteúdo produzido pelo autor responsáveis pela recuperação por assunto em interfaces de busca. Portanto, o que era da competência dos gestores de bases de dados, os profissionais da informação, passa a ser do próprio autor sem que haja essa mediação informacional mais técnica de preenchimento de metadados durante o auto arquivamento.

A investigação sobre o tema de indexação de autores de trabalhos acadêmicos tem como objeto de análise a atribuição de palavras-chave em sistemas de auto arquivamento em repositórios, periódicos e eventos científicos tendo em vista que o ambiente digital em ciência aberta tornou o autor um indexador e transformou a prática do profissional indexador. O autor e o profissional indexador atuam em diferentes âmbitos do sistema de informação: enquanto os indexadores profissionais, editores e organizadores de eventos atuam no âmbito da organização da informação e são responsáveis por orientarem o uso e funcionamento do sistema de informação, por outro lado, os pesquisadores são autores de produção científica responsáveis pela submissão e uso do sistema. Portanto, as regras e padrões de funcionamento são criadas pelos responsáveis pela organização da informação a serem obedecidas pelos autores.

Nesse sentido, a investigação tem como objetivo discutir com base na literatura publicada sobre a indexação de autores os padrões de funcionamento dos sistemas de informação e as orientações disponíveis para atribuição de palavras-chaves em sistemas de submissão de teses, dissertações, periódicos e de eventos. Essa discussão é importante tendo em vista que a indexação do autor é um procedimento realizado há várias décadas desde o século passado ou quando da criação dos primeiros periódicos científicos publicados. Porém, o autor não é tratado na perspectiva de um indexador que precisa ser orientado quanto aos procedimentos de indexação tal como um indexador profissional que segue regras específicas.

Tal investigação se justifica pelo fato de seu objeto de análise, a atribuição de palavras-chave em sistemas de autoarquivamento em repositó-

rios, periódicos e eventos, científico ser processo de indexação mediador e estratégico para a representação de recursos informacionais em fontes indexadoras definidoras de fator de impacto de produções científicas em todas as áreas de conhecimento. É um projeto prioritário de pesquisa básica cujos resultados contribuirão para o desenvolvimento das áreas de tecnologias estratégicas, tecnologias habilitadoras, tecnologias de produção, tecnologias para o desenvolvimento sustentável e tecnologias para a qualidade de vida.

## **2. O CONTEXTO DE SISTEMAS DE INFORMAÇÃO E SUA EVOLUÇÃO NA CIÊNCIA ABERTA: CONTINUIDADE OU REVOLUÇÃO?**

Pesquisadores em geral que desenvolvem pesquisas têm o desejo e a necessidade de que suas pesquisas sejam publicadas e que sejam citadas. Publicações científicas como artigos, dissertações, teses e trabalhos de eventos científicos tem conteúdos de pesquisas originais e inéditas que precisam ser acessadas por outros pesquisadores com desenvolvimento de pesquisas relacionadas. Quando os resultados de pesquisas publicados não são lidos ou citados os motivos podem ser avaliados de diferentes perspectivas: do veículo de comunicação, dos sistemas de busca, das fontes de informação que armazenam e/ou referenciam a publicação, dos metadados que descrevem a publicação e, também, pela representação do conteúdo elaborada pelos autores das publicações, título, resumo e palavras-chave. Também é válido pensar na grande quantidade de publicações e na falta de estruturação do conhecimento científico legível por máquina (Marcondes; Campos, 2008) que impede a leitura por máquina.

Em que pesem todos os prováveis motivos elencados, é difícil supor que outros pesquisadores ou leitores façam buscas por nomes ou títulos no cenário atual de grandes quantidades de publicações. Quando buscas são realizadas em qualquer interface, o principal e primeiro interesse é o conteúdo da publicação para verificar a relação com a necessidade de conhecimento. A busca por assuntos, invariavelmente, depende das formas

de representação com maior potencial de identificação significativa com o conteúdo representado, o título, resumo e palavras-chave da publicação. Contudo, o acesso por assuntos é resultado da indexação de assuntos, seja pela atribuição de palavras-chave pelo autor ou termos por indexadores profissionais (IFLA, 2011).

O preenchimento dos metadados tem importância para a recuperação dos itens bibliográficos arquivados em sistemas de informação porque o usuário que busca a literatura publicada poderá ter acesso a qualquer desses dados se estiverem corretamente preenchidos, de modo que a recuperação do item bibliográfico poderá ser garantida nas buscas pelo título, autor, editora, ano de publicação, bem como por assunto. Porém, quando o usuário acessa o sistema de informação nem sempre tem ideia dos nomes de autores que, por exemplo, publicaram no ano de 2024 e, assim, a forma de acesso será muito mais importante pelo assunto.

O assunto descrito por palavras-chaves carrega não somente o significado, mas a relevância do conteúdo da área de conhecimento em que se insere e para o domínio de especialidade. Quanto mais precisão cada palavra-chave tiver com relação ao conteúdo da publicação e, principalmente, com o reconhecimento do domínio de especialidade da comunidade de usuários, mais chances de recuperação terá o documento e, consequentemente, de sua leitura e citação em outras publicações. Esse procedimento de atribuição de palavras-chave é um processo de análise de assunto para a indexação de conteúdo que faz parte da comunicação científica.

A comunicação científica é dependente do controle bibliográfico realizado principalmente por empresas editoriais globais que ocuparam um espaço profissional não desenvolvido pelas bibliotecas com relação principalmente ao tratamento temático para representação de assuntos de artigos de periódicos científicos por conta do progressivo aumento de publicações anuais. A citação de Borko e Bernier (1975) explica o aumento da literatura na época:

A imprensa biomédica, por exemplo, foi estimada em publicar 2 milhões de artigos por ano (McCandless, Skweir e Gordon, 1964, p.147). Esses artigos podem ser lidos a uma taxa de dois por hora – assumindo que o leitor esteja atento, saiba ler aproximadamente 70 idiomas e tenha os documentos em mãos. Se a leitura de periódicos for limitada a 1 hora por dia e 365 dias por ano, então levará mais de 27,4 séculos para ler a produção de 1 ano da imprensa biomédica mundial (Borko; Bernier, 1975, p. 6).<sup>1</sup>

A década de 50 do século passado deu início a um acelerado processo de automatização de índices bibliográficos para áreas de conhecimento mais especializados como Física, Química, Medicina entre outras. Essas áreas, devido à urgência de desenvolvimento tecnológico e inovação, necessitam de literatura atualizada publicada principalmente em artigos de periódicos. Surgiram também bibliografias nacionais importantes realizadas por bibliotecas nacionais existentes até a atualidade como, por exemplo, a British National Bibliography (BNB)<sup>2</sup>, “[...] publicada pela primeira vez em 1950 e dedicada a publicação de lista de livros, periódicos e seriados publicados no Reino Unido e na República da Irlanda” (Wikipedia, 2023). Em 1974, a BNB adotou o sistema de indexação PRECIS (PREserved Context Indexing System) para gerar automaticamente o índice de assunto dos itens bibliográficos a partir de um software que admitia a indexação parcialmente automática e parcialmente humana (Fujita, 1989).

Os grandes sistemas online como o DIALOG, por sua vez, tiveram problemas para produzir manualmente os índices de assunto da literatura publicada em artigos de periódicos e passaram a investir em propostas inovadoras de indexação automática para a geração de índices de assunto. O sistema KWIC Index (Key Word In Context) foi criado em algoritmo por Hans Peter Luhn (1960) e aplicado para a geração de índices de assuntos

<sup>1</sup> The biomedical press, for example, has been estimated to publish 2 million papers each year (McCandless, Skweir e Gordon, 1964, p.147). These papers can be read at the rate of two per hour – assuming that the reader is attentive, can read approximately 70 languages, and has the documents at hand. If journal reading is limited to 1 hour per day and 365 days per year, then it will take more than 27.4 centuries to read the output of 1 year of the world's biomedical press (Borko; Bernier, 1975, p.6)

<sup>2</sup> The British National Bibliography · Search national bibliography data for people, organizations, works and subjects ([natbib-lod.org](http://natbib-lod.org))

de literatura técnica do Chemical Titles entre os anos de 1957 e 1959 (Chu, 2010).

O objetivo principal assumido por Luhn era o de disseminação e recuperação da informação tendo em vista o que ele considerava sobre a situação da época em que viveu:

O ritmo acelerado do desenvolvimento científico nos últimos anos acentuou a natureza perecível das novas informações. Como resultado, há uma demanda urgente por comunicação mais rápida nessa área. Parece duvidoso que essa demanda possa ser satisfeita sem romper com alguns dos padrões convencionalmente aplicados à compilação de índices de literatura. Índices especializados para literatura técnica são um meio estabelecido para direcionar engenheiros e cientistas a fontes de informação pertinentes ao seu interesse atual. Qualquer que seja o propósito específico de um índice, uma quantidade substancial de esforço intelectual é necessária para compilá-lo (Luhn, 1960, p.159).

Podemos considerar que o KWIC, antes do PRECIS, teve um significado impactante em sistemas de informação e na Ciência da Informação porque pela primeira vez a tecnologia da área de ciência da computação interagiu com uma proposta inovadora para a indexação de literatura técnico-científica. O impacto maior, ao nosso ver, foi o fato de fazer um índice utilizando o que Luhn (1960, p. 159) denominou de “palavra significativa” ou “chave” extraída automaticamente dos títulos das publicações em linguagem natural. A palavra-chave passou a ter um alto valor pela sua dupla função de representação e recuperação do conteúdo documental assim como os termos extraídos de classificações e descritores de vocabulários controlados. É aplicada em indexação automática e busca por palavras-chave e seu uso atualmente é generalizado em metabuscadores não só de literatura científica. “A abordagem KWIC é indiscutivelmente um marco significativo na IRR<sup>3</sup>” (Chu, 2010, p. 8).

A geração bem-sucedida desses índices trouxe outros problemas relacionados com a falta de controle de vocabulário, mas, por outro lado, demonstrou que era possível realizar indexação automática com a lingua-

<sup>3</sup> Information Representation and Retrieval

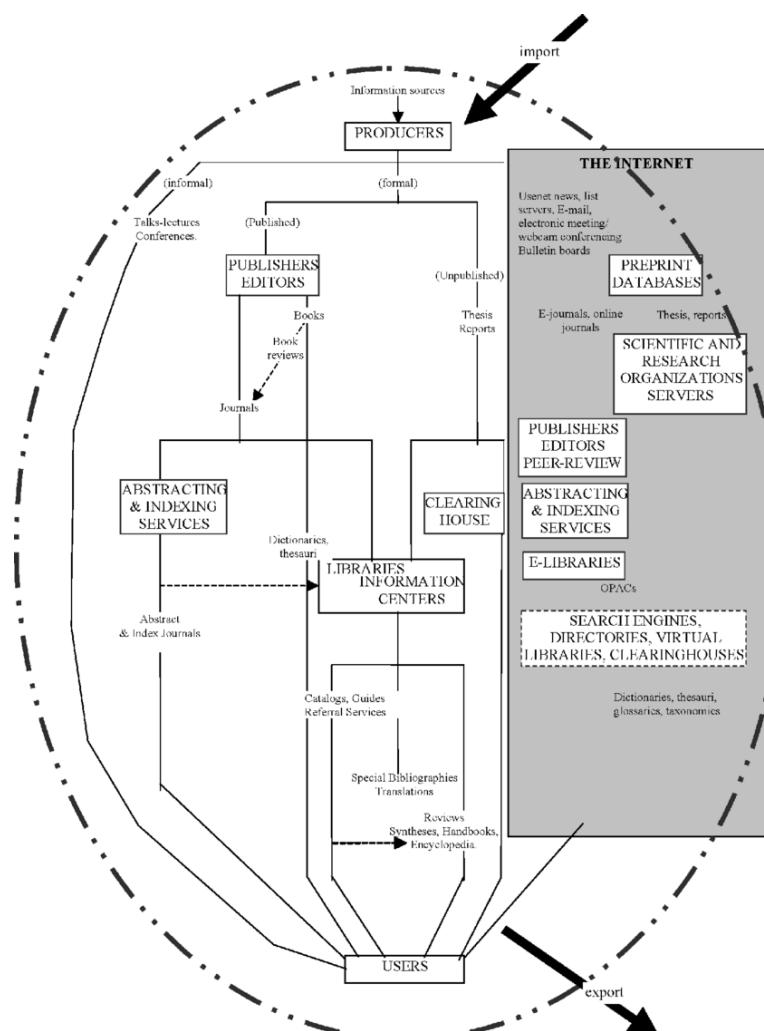
gem natural e o mérito disso está no próprio desenvolvimento posterior da indexação realizada por máquina.

A partir da indexação automática foi necessário tomar conhecimento da interdisciplinaridade com outras áreas de conhecimento além da Ciência da Informação: a terminologia com fundamentos linguísticos, a inteligência artificial ou lógica com a informática, a ciência cognitiva ou a psicologia, a linguística computacional, a semiótica, a estatística e a matemática (Gil Leiva, 2008).

A conexão entre as diferentes áreas de conhecimento para realizar a automação da indexação conduziu a ciência da informação a um patamar dos sistemas de recuperação da informação com desenvolvimento tecnológico e diferentes propostas que aceleraram a comunicação científica atualmente. A indexação automática com palavras-chave atribuídas pelo autor proporcionou as condições ideais para a ciência aberta como abordagem ideológica em contraposição ao comércio de grandes sistemas de informação por assinatura cuja evolução tem contínuo impacto econômico no desenvolvimento científico e são dominantes na demonstração dos resultados de fator de impacto na comunicação científica em todas as áreas do conhecimento.

O ciclo da comunicação científica envolve muitos atores institucionais, comerciais e científicos e, principalmente, os autores e profissionais da informação conforme demonstrado na Figura abaixo com base na abordagem da análise de domínio elaborada por Hjørland (2003, p.95):

Figura 1 - The revised UNISIST-model modified for the domain analytic approach



Fonte: Hjørland, B., 2003, p.95

Nessa representação de 20 anos atrás, Hjørland (2003, p. 95), propõe uma revisão e atualização do modelo de comunicação científica e técnica do UNISIST (1971, p. 96), um Programa Intergovernamental de

Cooperação no Campo da Informação Científica e Tecnológica. Observa-se que no modelo da Figura 1 a Internet passa a ser considerada num amplo contexto de serviços eletrônicos com documentos e informações digitais que na década de 70 ainda não existiam, mas que se consolidaram no mundo atual. Nesse modelo de comunicação científica e tecnológica estão presentes fontes de informação publicadas (periódicos e livros) e não publicados (teses, dissertações e relatórios de pesquisa) que são incluídos tanto em Bibliotecas, Centros de informação quanto em Serviços de indexação e resumo aos quais os usuários têm acesso por meio das interfaces de busca dos catálogos online de bibliotecas e centros de informação bem como de diretórios e bibliotecas virtuais. Em todos esses sistemas de informação a indexação é realizada para atribuição de assuntos representativos do conteúdo de documentos e recursos de informação.

Entretanto, nesse modelo não se faz menção ao acesso aberto e sistemas de informação de acesso aberto, criados por instituições, editores científicos e organizadores de eventos, em que se admite o auto arquivamento pelos próprios autores em repositórios, periódicos e eventos que utilizam softwares de acesso aberto. No modelo de comunicação científica apresentado à presença de sistemas de informação tem as características de editoras comerciais que se dedicam ao comércio de acesso e recuperação da informação em sistemas de acesso fechado e de bibliotecas que tem suas próprias políticas de desenvolvimento de acervo com base na compra de livros publicados e assinatura paga de coleções de periódicos.

O que acontece nesse ciclo de comunicação científica, conforme demonstrado na Figura 1, é que essas palavras-chave serão consideradas em Serviços de indexação e resumo que produzem bases de dados e que adotam a linguagem natural, mas aplicam também o controle de vocabulário com uso de tesouros e vocabulários controlados especializados construídos para aumentar a precisão e diminuir a dispersão da informação durante a recuperação da informação pelo usuário nas interfaces de busca.

A diferença é que na evolução para o acesso aberto o autor passou a ser o indexador do seu próprio conteúdo quando antes a indexação era formalmente realizada por profissionais da informação segundo padrões de procedimentos previamente estabelecidos para a prática profissional e não

para uso de autores. Essa mudança aconteceu não por acaso, mas foi evidenciada pela situação de auto arquivamento em sistemas de acesso aberto.

### **3. INDEXAÇÃO POR PALAVRAS-CHAVE E O CONTROLE DE VOCABULÁRIO EM SISTEMAS DE INFORMAÇÃO**

Antes da ideologia do acesso aberto e do periódico eletrônico o autor sempre foi o indexador de seus artigos na atribuição de palavras-chave conforme orientação aos autores presentes em diretrizes dos periódicos que recomenda a inclusão, após o título, do resumo e das palavras-chave (Fujita, 2004; Fujita; Agustín-Lacruz; Terra, 2018; Terra; Agustín-Lacruz; Fujita, 2022).

Quando o autor de uma tese realiza o autoarquivamento em um repositório de teses e dissertações de sua instituição, o faz com suas próprias palavras-chave porque desconhece orientações sobre indexação e política de indexação (Fujita; Panuto, 2023). Entretanto, não podemos ignorar que o autor científico é um indexador especialista de sua área de conhecimento porque é conhecedor da terminologia e das principais palavras-chave que melhor representam o conteúdo e, principalmente, serão utilizadas pelos demais pesquisadores (Holstrom, 2019).

A indexação sempre teve um nível de complexidade alto tanto nos procedimentos como na aplicação junto à recuperação da informação cuja sintonia deve ser sempre de dependência entre uma e outra porque os resultados são como “causa e efeito”, ou seja, a indexação é causa do efeito qualitativo na recuperação. Assim, entende-se que a indexação de assuntos é uma atividade com impacto social nos ambientes digitais que, ao lado do auto arquivamento, tem a relevância da representação de conteúdos realizada por meio de palavras-chave da linguagem natural. Os benefícios para os pesquisadores que autoarquivam suas produções em sistemas de acesso aberto como os repositórios são: organização da informação e do conhecimento da rede da qual participam; aumentar a visibilidade e o impacto de suas pesquisas; responsabilidade perante a sociedade; ampliação

das possibilidades de diálogo entre os pesquisadores alcançadas pela rede e com a sociedade em geral (Leite, 2009).

A publicação de artigos é realizada com mais frequência e rapidez que livros o que propiciou a necessidade de existirem os Serviços de resumos e indexação fora do contexto de bibliotecas que passaram a elaborar os periódicos internacionais de resumos para o controle bibliográfico internacional. Ao longo do tempo, esses periódicos de resumos antes impressos se tornaram disponíveis eletronicamente por meio de bases de dados digitais que utilizam Hub para comercialização do acesso.

O acesso a essa literatura científica publicada depende da representação de seu conteúdo realizada pelo processo de indexação, seja com linguagem natural ou linguagem controlada. A indexação com linguagem natural permite que palavras sejam extraídas de qualquer parte do texto, título, resumo, palavras-chave e corpo do texto ou mesmo palavras-chave podem ser atribuídas pelo autor do texto em sua linguagem natural com a finalidade de representação do conteúdo. Por outro lado, o processo de indexação com linguagem controlada utiliza, após análise de assunto para identificação e seleção de palavras em linguagem natural extraída do conteúdo textual, vocabulários controlados (tesauro, lista de termos autorizados, lista de cabeçalhos de assunto) para atribuição de descritores. A normalização de forma, função e semântica dos descritores é adotada em vocabulários controlados como controle padronizado de representação da terminologia de especialidade das áreas de conhecimento para eliminar ambiguidade, homônimia e formas ortográficas e criar um padrão de descritor a ser selecionado.

A indexação de assuntos manual ou automatizada, com linguagem natural ou linguagem controlada, sempre esteve presente em serviços de informação especializados principalmente de artigos de periódicos de áreas de conhecimento científicas, tecnológicas e de negócios que utilizam vocabulário mais preciso.

Existe, desde sempre em sistemas de informação, uma convivência híbrida entre linguagem natural e linguagem controlada tendo em vista que a linguagem natural sempre teve maior capacidade de representati-

vidade de ideias e evolução mais rápida de vocabulário, porém com problemas de falta de controle de vocabulário que provocam dispersão, falta de especificidade e de precisão na recuperação de conteúdos em sistemas de informação. A linguagem controlada tem vantagens de representação e recuperação, porém depende de constante atualização para abranger todo o vocabulário significativo. Além disso, o usuário de sistemas de informação realiza suas buscas com sua linguagem natural porque, dificilmente um sistema de informação disponibiliza vocabulários controlados para consulta de descritores.

A diferença está no processo de indexação, que consiste na análise conceitual do documento e na tradução do termo de indexação para o descritor de vocabulário controlado (Lancaster, 2004). Enquanto isso, a palavra-chave pode ser extraída de qualquer parte do documento com ou sem controle de vocabulário e ser atribuída por autores, editores ou mesmo ser gerada automaticamente (Gonçalves, 2008) sem a realização de um processo conceitual padronizado desconhecido dos autores. O processo de indexação realizado por indexadores profissionais é diferente porque eles são treinados para selecionar termos de indexação de acordo com um protocolo específico (Névéol; Dogan; Zhiyong, 2010).

Kipp (2009) relata que poucos estudos sobre palavras-chave de autores em comparação com descritores foram realizados, e que os vocabulários controlados utilizados por indexadores profissionais requerem treinamento para serem utilizados por autores ou usuários. No entanto, a terminologia utilizada pelo autor será sempre diferente do vocabulário controlado utilizado pelo profissional, pois é resultado do conhecimento gerado na área de domínio em evolução, pela própria natureza científica, e é mais específica do que a padronizada que visa obter consistência de termos.

O termo híbrido, segundo Lancaster (2002, p.191) “[...] faz referência aos sistemas de recuperação da informação que utilizam a combinação de termos controlados e a linguagem natural.” Nesse sentido, Lancaster (2002) relembra que quando surgiram os sistemas de indexação automatizados na década de 50 buscava-se uma simplificação na recuperação da informação moldada pelos sistemas de informação que comercializavam a comunicação científica.

Os sistemas híbridos, como denominou Lancaster (2002), utilizariam para o caso de documentos impressos uma combinação de códigos de áreas de assunto e marcação das palavras-chave do próprio título sem que fosse necessário fazer, naquele tempo, fichas de indexação e sim uma gravação direta sobre o documento. Com documentos eletrônicos, Lancaster (2002) indica que o uso do computador ajuda na indexação e o analista examina os termos atribuídos pelo computador de modo a adicionar ou eliminar os termos que considera necessários. Por outro lado, citou outro exemplo de sistema que trabalhava com o mínimo de controle de vocabulário utilizando a combinação de um vocabulário controlado pequeno de descritores mais amplos e uma lista mais exaustiva de descritores em linguagem natural extraídos dos documentos.

Embora Lancaster (2002) se refira a sistemas híbridos não o faz sem justificar a importância do controle de vocabulário na representação e principalmente sua influência na recuperação da informação. Os sistemas híbridos são mais adotados em sistemas de informação mais especializados com delimitação terminológica precisa e de menor quantidade de termos que podem fazer uso da indexação automática e da atribuição de palavras-chaves pelos autores. O que se evidencia é que existem sistemas de informação com controle de vocabulário (Catálogos online de bibliotecas, Biblioteca Virtual de Saúde - BVS), sem controle de vocabulário (periódicos científicos, eventos científicos, bases de dados de textos completos de teses e dissertações) e com a combinação de linguagem controlada e linguagem natural (Repositórios).

Repositório é um sistema de informação de acesso aberto que pode ser um sistema híbrido com linguagem natural e linguagem controlada. Isso porque além de permitir o auto arquivamento pelos autores com atribuição de palavras-chave em linguagem natural, armazena toda a produção científica institucional já publicada, cujos metadados descritivos (autoria, título, dados de publicação) e temáticos (código de áreas de assunto, resumo e palavras-chave em linguagem natural e linguagem controlada) são preenchidos automaticamente durante a operação de interoperabilidade entre bases de dados. Na condição de ser um repositório, foi criado para o armazenamento e preservação institucional mais do que ser um sistema de

informação com objetivos de recuperação da informação em que o processo de indexação para a representação de conteúdos é parte da política de organização da informação.

O controle de vocabulário em sistemas de informação gera contínua padronização que encarece a manutenção realizada por profissionais da informação. Por sua vez, os profissionais da informação não conseguem realizar, ao mesmo tempo, a contínua atualização do controle de vocabulário e a inserção de metadados das publicações em quantidades volumosas. Essa situação favorece os sistemas de informação de acesso aberto no qual é possível que o autor possa realizar o auto depósito de sua publicação e produzir os metadados descritivos e temáticos conforme formato padrão adotado pelo sistema de informação.

Essa é, de fato, uma mudança de paradigma que dá ao autor e ao usuário a responsabilidade de serem atores principais no processo de representação de conteúdo. Com esse protagonismo, tornam-se indexadores especialistas e resolvem grande parte do problema de controle bibliográfico da literatura publicada mundialmente, bem como do acesso de todos à ciência aberta. Desse modo, os sistemas de informação, de modo geral, se beneficiam do envolvimento de todos para realizarem o tratamento padronizado da informação. Porém, nessa condição de indexador o autor é capacitado e orientado quanto aos procedimentos de normalização e de controle de vocabulário?

#### **4. A INDEXAÇÃO DO AUTOR EM SISTEMAS DE INFORMAÇÃO: SÃO NECESSÁRIAS AS ORIENTAÇÕES?**

A política de indexação de qualquer sistema de recuperação da informação tem como principal elemento, conforme Carneiro (1985) e Fujita (2012), o processo de indexação dotado das variáveis: nível de especificidade, nível de exaustividade, escolha da linguagem de indexação, capacidade de revocação e precisão do sistema, consistência e correção. Tanto o processo de indexação quanto suas variáveis são o foco de qualquer orientação

a ser realizada a indexadores profissionais tendo em vista seus objetivos profissionais em função do contexto do sistema de recuperação da informação que habita, conforme estudos realizados por Fujita (2016, 2017) e Gil Leiva e Fujita (2012). Seria necessário, contudo, conhecer objetivos e contextos de outros atores para definir orientações de indexação na determinação de palavras-chave representativas de conteúdos de documentos e recursos informacionais.

O autor de publicações científicas torna-se um indexador quando atribui palavras-chave para representação do conteúdo de textos que redigiu. Essas palavras-chave representam tanto o conteúdo significativo de documentos ou recursos informacionais quanto o domínio de conhecimento do autor que desenvolveu a investigação e a publicou em algum veículo de comunicação científica.

Entretanto, de acordo com Fujita, Terra e Agustin-Lacruz (2018) os periódicos científicos na área de Ciência da Informação e Comunicação não trazem orientações aos autores sobre os procedimentos de indexação por assuntos para determinação de palavras-chave. Em investigação sobre análise das diretrizes para atribuição de palavras-chave aos artigos durante submissão em periódicos brasileiros de Ciência da Informação, Oliveira, Fujita, Dal'Evedove e Martinez Ávila (2020) obtiveram resultados que orientam apenas sobre a quantidade de palavras-chave e seletividade na indexação sem profundidade e sem indicação de uso de vocabulário controlado o que os levam a concluir pela recomendação de elaboração de política de indexação que proporcione assertividade aos autores no momento de atribuição de palavras-chave. Além disso, observa-se que os autores quando realizam autoarquivamento de seus trabalhos acadêmicos em repositórios não são orientados sobre a atribuição de palavras-chave, uso de vocabulários controlados ou expansão de vocabulário (Freitas; Dal'Evedove, 2019) e quando são orientados, as diretrizes de submissão não estão acessíveis publicamente no sistema.

A análise de orientações sobre atribuição de palavras-chave em 5 periódicos das áreas de Educação e Ciência da Informação da coleção Scielo e de trabalhos dos eventos ENANCIB de Ciência da Informação e Reunião Nacional da ANPED da Educação foi analisada por Fujita (2024b). Os

resultados obtidos da análise confirmam que as orientações são formais com relação apenas à quantidade mínima e máxima de palavras-chave e nenhuma orientação sobre análise de assunto foi oferecida aos autores para atribuição de palavras-chave ou uso de vocabulário controlado com exceção do ENANCIB que indica o uso do Tesauro Brasileiro de Ciência da Informação para o controle de vocabulário ou indicação de palavras--chave mais específicas.

No caso do autor, responsável pela atribuição de palavras-chave em sistemas de submissão de trabalhos acadêmicos e publicações, ele é o indexador especialista de domínio que necessita de orientações sobre o processo de indexação e a influência que exerce na visibilidade, citações e fator impacto de suas publicações. Essa situação híbrida de sistemas de informação que utilizam linguagem natural e linguagem controlada é hoje uma realidade que não se pode negar (Fujita, 2020) e a solução, embora conflitante, pode trazer benefícios mútuos se houver uma política de indexação inclusiva dos diferentes atores que realizam a indexação. Para isso, é importante elaborar orientações exclusivamente destinadas ao autor cujos objetivos são diferentes do indexador profissional, tendo em vista que visa a visibilidade de suas publicações científicas para os demais especialistas de seu domínio de conhecimento. As orientações dirigidas ao autor devem incluir esclarecimentos sobre o processo de indexação e as variáveis, exaustividade, especificidade, correção e consistência que influenciam na visibilidade das publicações em fontes indexadoras importante para proporcionar citações e, consequente, fator de impacto.

Strader (2009) que investigou a sobreposição entre palavras-chave atribuídas pelo autor e os cabeçalhos de assunto da Biblioteca do Congresso (LCSH) atribuídos pelo catalogador em teses e dissertações no catálogo online da Ohio State University alerta para o fato de que a maioria dos estudos sobre assunto se concentram em usuários como buscadores de informações e menos no uso de palavras-chave atribuídas por autores de documentos online.

Em estudo sobre coerência semântica na indexação de artigos de periódicos na área de saúde pública indexados no Repositório de Produção Científica da Escola Nacional de Saúde Pública Sergio Arouca da Fundação

Oswaldo Cruz (Fiocruz), Santos, Mollica e Guedes (2019) verificaram baixo grau de coerência na representação temática em amostra de dez artigos. Concluem que é necessário “[...] um olhar mais aprofundado sobre o grau de coerência semântica na representação de artigos publicados na área de saúde pública [...]” e “[...] que na indexação de um documento não há um conjunto correto de termos [...]” (Santos, Mollica; Guedes, 2019, p.229).

Em investigação realizada por Fujita e Tartarotti (2020) sobre análise de palavras-chave da produção científica atribuídas por pesquisadores para submissão de artigos de periódicos indexados na Scopus e no Portal Docentes Unesp, quanto à padronização e controle de vocabulário para diferentes funções em sistemas de armazenagem e recuperação da informação, os resultados obtidos revelaram que o pesquisador, na condição de indexador de sua produção científica, atribuem palavras-chave com alta especificidade, inovação e alta incompatibilidade de termos por problemas ortográficos e semânticos. A principal recomendação do estudo realizado diz respeito aos pesquisadores como indexadores não proficientes que precisam de orientações quanto à padronização de suas palavras-chave e para isso é necessária a elaboração de proposta de política de indexação.

Gollub, Lykke e Tudhope (2014) haviam observado que na maioria dos repositórios, os autores não são treinados em indexação nem recebem quaisquer diretrizes de indexação e advertem que se as sugestões forem derivadas de um vocabulário controlado adequado, a recuperabilidade de seus artigos provavelmente contribuirá para a precisão e a recordação ideais. Os resultados de pesquisa realizada por Tartarotti, Dal’Evedove e Fujita (2018) que analisou a presença de políticas de indexação em ambientes colaborativos mediante aplicação de questionário aos responsáveis por bibliotecas digitais de teses e dissertações no Brasil, revelaram a ausência de política de indexação que oriente e proteja sua prática em ambientes colaborativos. Em estudo mais recente sobre análise de palavras-chave em registro de metadados de assuntos de teses e dissertações, Terra, Agustín-Lacruz, Bernardes, Fujita e Bueno de La Fuente (2021) verificam que embora a atribuição de palavras-chave pelo autor seja uma forma econômica de ampliar o acesso ao conteúdo de teses e dissertações é necessário definir regras para os autores quanto à escolha de palavras-chave.

Em recente análise de 10 repositórios institucionais de diferentes regiões brasileiras realizada por Fujita e Panuto (2024) foi constatado que apenas o Repositório Institucional Unesp possui tutorial para submissão de teses e dissertações com indicação de atribuição de palavras-chave acompanhada de orientações específicas sobre representação de assunto e o uso do Tesauro UNESP para controle de vocabulário. A análise do “Tutorial para uso do Tesauro Unesp” (UNESP, S.d), realizada por Fujita (2024a), demonstrou que o autor de tese ou dissertação é orientado a descrever os assuntos usando um mínimo de 3 descritores do Tesauro Unesp e ao menos um que represente a área de conhecimento. Além disso, o Tutorial esclarece que o Tesauro possui uma rede de termos do mais geral ao mais específico e que é possível navegar por essa rede e escolher outros descritores relacionados ao termo digitado na busca. Nem sempre o termo que o autor busca é o termo preferido ou correto e, nesses casos, o tutorial aconselha o autor a usar o descritor preferido no Tesauro que mais se alinhe à representação da terminologia especializada do domínio para que outros pesquisadores da área de conhecimento reconheçam o assunto tratado no conteúdo da tese ou dissertação. Quando o termo buscado pelo autor não existe no Tesauro Unesp, a orientação do Tutorial indica que o autor solicite a inclusão do termo após análise terminográfica da Comissão Permanente de Tesauro Unesp.

As duas pesquisas (Fujita, 2024a; Fujita; Panuto, 2024) demonstraram que existem orientações quanto à atribuição de assuntos por palavras-chave porém com controle de vocabulário e, assim, o Tutorial se dedica a orientar os autores sobre os procedimentos de uso do Tesauro Unesp quando, antes, é importante a identificação e seleção de termos na análise de assunto do conteúdo pelo autor da tese ou dissertação, ou seja, orientá-lo quanto aos procedimentos de indexação, a análise de assunto e a representação das palavras-chave com uso de vocabulário controlado.

Em orientação específica sobre determinação de palavras-chave para autores de artigos científicos do periódico Revista Brasileira de Educação Especial, Fujita (2004, p. 257) apresenta a metodologia do processo de indexação

[...] para a determinação de palavras chaves utilizando como estratégia de análise de assunto a localização do tema principal pela exploração da estrutura textual combinada com a análise conceitual por questionamento e representação dos termos obtidos pela linguagem documentária do sistema que indexa o periódico.

Nesta proposta está implícita uma visão conceitual e intelectual do processo de análise de assunto que combina a exploração da estrutura textual do artigo com estratégia de identificação de conceitos por questionamento de acordo com proposta metodológica do Modelo de leitura documentária para textos científicos de Fujita (2003). Os termos identificados pela metodologia de análise de assunto serão posteriormente traduzidos com uso de vocabulário controlado, Thesaurus Brasileiro da Educação do INEP, específico da área de Educação.

A análise da literatura publicada sobre a indexação do autor revelou, dessa forma, que os sistemas de informação de periódicos, trabalhos de eventos, teses e dissertações têm atribuição de palavras-chave pelos autores sem orientação quanto aos procedimentos de indexação ou uso de vocabulários controlados. Também não é possível compreender a função dessas palavras-chave na representação e recuperação da informação o que nos leva a considerar que a transformação das práticas de indexação devem ser necessariamente associadas à educação e formação do autor como indexador para que seja consciente do poder de cada palavra-chave que atribui no amplo sistema de comunicação científica.

## **5. CONSIDERAÇÕES FINAIS**

Na análise da trajetória da indexação por palavras-chave durante os anos 50, advento da internet, expansão da web e o aparecimento da ideologia da ciência aberta em contraposição ao aumento exponencial do comércio da comunicação científica por editoras comerciais que souberam fazer a transição da Representação e Recuperação da Informação (IRR) com

uso de tecnologias aplicadas ao processo automático de indexação com o evidente aprimoramento da inteligência artificial.

Tendo em vista a necessidade de elaborar política de indexação que inclua orientação à autores sobre o processo de indexação na representação em sistemas de informação tais como sistemas de autoarquivamento de produção científica em repositórios institucionais, periódicos científicos, eventos e preenchimento de currículos, esta pesquisa propõe a investigação sobre a indexação do autor científico como indexador especialista de domínio.

A proposta de metodologia do processo de indexação com base no Modelo de Leitura Documentária para textos científicos de Fujita (2003) não foi incluída nas orientações aos autores para atribuição de palavras-chaves conforme seção de Forma e preparação dos manuscritos, resumindo as orientações à formalidade de “[...] fazer a indicação após o resumo (mínimo de três e máximo de cinco palavras). Utilizar o site do Thesaurus Brasileiro da Educação do INEP no site [www.inep.gov.br](http://www.inep.gov.br). (Revista Brasileira de Educação Especial, 2024). Entretanto, a indicação do vocabulário controlado especializado em Educação do INEP demonstra que os resultados obtidos com o uso de controle vocabulário como etapa de tradução do processo de indexação pode ter sido um diferencial importante na indexação desse periódico em outros sistemas de informação haja visto estar indexado no Scopus, LILACS, Scielo e outros.

O Modelo de leitura documentária para textos científicos (Fujita, 2003) foi idealizado para aplicação na prática profissional de catalogadores e indexadores e não para indexadores especialistas como autores, editores e organizadores de eventos. Além disso, seria necessário que fosse incorporado à uma política de indexação de assuntos de periódicos, repositórios, e sistemas de gestão de eventos científicos e, necessariamente, descrita operacionalmente em um manual de indexação à semelhança de sistemas de informação bibliográficos como a LILACS da Rede BVS<sup>4</sup> que destinam orientações aos indexadores profissionais com uso do vocabulário controlado DeCS (Descritores em Ciências da Saúde) para controle de vocabulário.

<sup>4</sup> <https://red.bvsalud.org/lildbi-web/>

Diferentes sistemas de informação com requisitos diferenciados de condições tecnológicas, materiais e humanos podem definir diretrizes de política de indexação diferentes entre si, sobretudo se os profissionais que cuidam do sistema de informação tiverem objetivos em prol do aprimoramento da representação e recuperação da informação tendo em vista a visibilidade dos conteúdos do sistema de informação. A representação da informação pelo processo de indexação é a essência da política de indexação e, por isso, deve ser definida mediante orientações metodológicas que visem o conteúdo a ser representado. Nesse sentido, a política de indexação, uma vez definida deve ser registrada em um manual de indexação para que possa ser compreendida, aplicada e aprimorada ao longo do desenvolvimento do sistema de informação.

A metodologia do processo de indexação do Manual de indexação da Unesp é dirigida especificamente aos catalogadores que realizarão a indexação de livros, precisamente porque havia esse interesse e motivação da rede de bibliotecas da Unesp para a representação de assuntos de livros no catálogo online. Esse é um exemplo que pode ser seguido por outros sistemas de informação e dirigidos a diferentes atores, catalogadores, indexadores, autores e leitores com as devidas diferenças metodológicas no processo de indexação conforme objetivos previamente definidos e com base em estudo de estratégias cognitivas e metacognitivas de processos de pensamento para execução de tarefas voltadas para objetivos específicos.

Sugere-se que os procedimentos metodológicos de indexação com as operações de análise de assunto e representação com vocabulário controlado sejam adequados à natureza do processo de atribuição de palavras-chave pelo indexador especialista do domínio de conhecimento tendo em vista a economia de tempo de indexação do profissional e, principalmente, a real possibilidade de existência de termos novos que geram atualizações para os vocabulários controlados. Para isso, será necessário investigar, em sequência, como os autores escolhem suas palavras-chaves e responder questões ainda sem respostas na literatura investigada tais como:

O autor/usuário tem consciência dessa mudança de paradigma com relação à sua atuação como indexador especialista e responsável pela inserção de informações em metadados? Além dessa questão, é preciso saber

ainda: se nessa condição de indexador especialista o autor necessita ser capacitado e orientado quanto aos procedimentos de indexação e de controle de vocabulário? Qual a visão do autor indexador sobre o processo de indexação com procedimentos de análise de assunto e controle de vocabulário? O autor considera primordial a indexação por palavras-chave e por quê?

As respostas a essas questões serão importantes para o aprimoramento da prática de indexação tendo em vista o cenário de transformações da comunicação científica que produziu o contexto paradigmático da Ciência da Informação.

## **REFERÊNCIAS**

ARXIV. *In: WIKIPEDIA: the free encyclopedia*. [San Francisco, CA: Wikimed ia Foundation, 2023]. Disponível em: <https://pt.wikipedia.org/wiki/ArXiv>. Acesso em: 30 ago. 2024.

BRITISH National Bibliography. *In: WIKIPEDIA: the free encyclopedia*. [San Francisco, CA: Wikimed ia Foundation, 2023]. Disponível em: [https://en.wikipedia.org/wiki/British\\_National\\_Bibliography](https://en.wikipedia.org/wiki/British_National_Bibliography). Acesso em: 30 ago. 2024.

BORKO, Harold; BERNIER, Charles L. **Abstracting concepts and methods**. New York: Academy Press, 1975.

CARNEIRO, Marília Vidigal. Diretrizes para uma política de indexação. **Revista da Escola de Biblioteconomia da UFMG**, Belo Horizonte, v. 14, n. 2, p. 221-241, 1985. Disponível em: <https://periodicos.ufmg.br/index.php/reb/article/view/36523/28575>. Acesso em: 10 jul. 2021.

CHU, Heting. **Information representation and retrieval in the digital age**. 2nd. ed. Medford: American Society for Information Science and Technology, Information Today, 2010.

CITESEERX. *In: WIKIPEDIA: the free encyclopedia*. [San Francisco, CA: Wikimed ia Foundation, 2023]. Disponível em: <https://pt.wikipedia.org/wiki/CiteSeerX>. Acesso em: 30 ago. 2024.

FREITAS, Marina Penteado de; DAL'EVEDOVE, Paula Regina. Consistência na indexação por atribuição no repositório institucional da UFSCAR. *In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO*, 20., 2019, Florianópolis. **A Ciência da Informação e a era da Ciência de Dados [...]** Florianópolis: UFSC, 2019. Disponível em: <https://conferencias.ufsc.br/index.php/enancib/2019/paper/view/1203>. Acesso em: 19 dez. 2024.

FUJITA, Mariângela Spotti Lopes. **A leitura documentária do indexador:** aspectos cognitivos e linguísticos influentes na formação do leitor profissional. 2003. 321 f. Tese (Livre-Docência em Análise Documentária e Linguagens Documentárias Alfabeticas) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2003.

FUJITA, Mariângela Spotti Lopes. A representação documentária de artigos científicos em educação especial: orientação aos autores para determinação de palavras chaves. **Revista Brasileira de Educação Especial**, Marília, v. 10, n. 3, p. 257-272, 2004.

FUJITA, Mariângela Spotti Lopes. A política de indexação para representação e recuperação da informação. In: GIL LEIVA, Isidoro; FUJITA, Mariângela Spotti Lopes. **Política de indexação.** São Paulo: Cultura Acadêmica; Marília: Oficina Universitária, 2012. p. 17-30. Disponível em: [http://ebooks.marilia.unesp.br/index.php/lab\\_editorial/catalog/book/32](http://ebooks.marilia.unesp.br/index.php/lab_editorial/catalog/book/32). Acesso em: 19 dez. 2024.

FUJITA, Mariângela Spotti Lopes. **Política de indexação para bibliotecas:** elaboração, avaliação e implantação. Marília; São Paulo: Oficina Universitária; Cultura Acadêmica, 2016. Disponível em: [https://www.biblioteca.unesp.br/portal/arquivos/pdf/politicas-de-indexacao-para-bibliotecas\\_ebook.pdf](https://www.biblioteca.unesp.br/portal/arquivos/pdf/politicas-de-indexacao-para-bibliotecas_ebook.pdf). Acesso em 19 dez. 2024.

FUJITA, Mariângela Spotti Lopes. **Manual de política de indexação para as bibliotecas universitárias da UNESP.** São Paulo: Coordenadoria Geral de Bibliotecas, 2017. Desenvolvimento de material didático ou instrucional. Disponível em: <manual-politica-indexacao-2017.pdf>. Acesso em: 19 dez. 2024.

FUJITA, Mariângela Spotti Lopes. Linguagem natural ou linguagem controlada?: a influência da palavra-chave na representação para indexação e recuperação de informações. **Informacao & Sociedade:** estudos, João Pessoa, v. 30, p. 1-29, 2020. Disponível em: <https://periodicos.ufpb.br/ojs2/index.php/ies/article/view/57041>. Acesso em: 19 dez. 2024.

FUJITA, Mariângela Spotti Lopes. Forms and functions of author keywords in theses and dissertations at the UNESP Institutional Repository (Brazil). **The Canadian Journal of Information and Library Science**, Ontario, v. 47, n. 2, p. 166-175, 1 Oct. 2024. University of Western Ontario, Western Libraries. DOI: 10.5206/cjils-rcsib.v47i2.17628.

FUJITA, Mariângela Spotti Lopes. Análise das funções de palavras- chaves atribuídas por autores em publicações científicas de eventos e periódicos . **RDBCi:** revista digital de Biblioteconomia e Ciência da Informação, Campinas, v. 22, p. e024020, 2024. DOI: 10.20396/rdbcii.v22i00.8676208.

FUJITA, Mariângela Spotti Lopes; PANUTO, Jessica Cristina. Guidelines on assigning the subjects of theses and dissertations in repositories. **IFLA Journal**, The Hague, v. 50, n. 1, p. 160-169, 3 jan. 2024. SAGE Publications. DOI: 10.1177/03400352231217275.

FUJITA, Mariângela Spotti Lopes; AGUSTÍN-LACRUZ, María-Del- Carmen; TERRA, Ana Lúcia. Journals' guidelines about title, abstract and keywords: an overview of information science and communication science areas. **European Science Editing**, United Kingdom, v. 44, n. 4, p. 76-79, 2018. Disponível em: <https://europeanscienceediting.eu/articles/journals-guidelines-about-title-abstract-and-keywords-an-overview-of-information-science-and-communication-science-areas/>. Acesso em: 19 dez. 2024.

GIL LEIVA, Isidoro. **Manual de indización**: teoría y práctica. Gijón: Ediciones TREA, 2008.

GIL LEIVA, Isidoro; FUJITA, Mariângela Spotti Lopes. (org.). **Política de indexação**. São Paulo: Cultura Acadêmica; Oficina Universitária, 2012. v. 1. Disponível em: [https://ebooks.marilia.unesp.br/index.php/lab\\_editorial/catalog/view/32/83/2407-1](https://ebooks.marilia.unesp.br/index.php/lab_editorial/catalog/view/32/83/2407-1). Acesso em: 19 dez. 2024.

GOLUB, Koraljka; TYRKKÖ, Jukka; HANSSON, Joacim; AHLSTRÖM, Ida. Subject indexing in humanities: a comparison between a local university repository and an international bibliographic service. **Journal of Documentation**, London, v. 76, n. 6, p. 1193-1214, 2020. DOI: 10.1108/JD-12-2019-0231.

GONÇALVES, Aline Lima. Uso de resumos e palavras-chave em Ciências Sociais: uma avaliação. **Encontros Bibl**: revista eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 13, n. 26, p. 78–93, 2008. DOI: 10.5007/1518-2924.2008v13n26p78.

HJØRLAND, Birger. Fundamentals of knowledge organization. **Knowledge Organization**, Frankfurt, v. 30, n. 2, p. 87-111, 2003.

HOLSTROM, Chris. Moving towards an actor-based model for subject indexing. **NASKO**, Philadelphia, v. 7, n. 1, p. 120-8, 2019. DOI: 10.7152/nasko.v7i1.15631.

INTERNATIONAL FEDERATION OF LIBRARY ASSOCIATIONS AND INSTITUTIONS. Working Group on Guidelines for Subject Access. National Bibliographic Agencies of the Classification and Indexing Section. **Guidelines for subject access in national bibliographies**. The Hague: IFLA, 2011. Disponível em: <https://www.ifla.org/wp-content/uploads/2019/05/assets/classification-and-indexing/subject-access-by-national-bibliographic-agencies/nba-guidelines-fr-approval.pdf>. Acesso em: 30 ago. 2024.

KIPP, Margaret E. I. User, author and professional indexing in context: an exploration of tagging practices on CiteULike. **Canadian Journal of Information and Library Science**, Ontario, v. 35, n. 1, p. 1-41, 2009.

LANCASTER, Frederick Wilfrid. **Indexação e resumos**: teoria e prática. 2. ed. Brasília, DF: Briquet de Lemos, 2004.

LANCASTER, Frederick Wilfrid. **El control del vocabulário em la recuperación de información**. València: Universitat de Valencia, 2002.

LEITE, Fernando César Lima. **Como gerenciar e ampliar a visibilidade da informação científica brasileira.** Brasília, DF: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2009.

LUHN, H.P. Key word-in-context index for technical literature (kwic index). **American Documentation**, Washington, DC, v. 11, p. 288-295, 1960.

McCANDLESS, R. F. J., SKWEIR, E. A.; GORDON, M. Secondary journals in chemical and biological fields. **Journal of Chemical Documentation**, Washington, DC, v. 4, n. 2, p. 147-53, 1964.

MARCONDES, Carlos Henrique; CAMPOS, Maria Luiza de Almeida. Ontologia e Web Semântica: o espaço da pesquisa em Ciência da Informação. **Ponto de Acesso**, Salvador, v. 2, n. 1, p. 107-136, 2008. Disponível em: <http://www.portalseer.ufba.br/index.php/revistaaci/article/view/2669/1885/>. Acesso em: 30 ago. 2024.

NÉVÉOL, Aurélie; DOĞAN, Rezarta Islamaj; ZHIYONG, Lu. Author keywords in biomedical journal articles. In: AMIASYMPOSIUM, Washington, DC, 2010. **Proceedings** [...] Washington, DC: AMIA, 2010. p. 537-541. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3041277/>. Acesso em: 19 dez. 2024.

OLIVEIRA, Lais Pereira de; FUJITA, Mariângela Spotti Lopes; DAL' EVEDOVE, Paula Regina; MARTINEZ ÀVILA, Daniel. Política de indexação em periódicos da Ciência da Informação: um estudo das diretrizes para atribuição de palavras-chave aos artigos. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 25, n. 4, p. 140-169, 2020. DOI: 10.1590/1981-5344/3876.

RODRIGUES, Eloy. Acesso livre ao conhecimento: a utopia e a realidade. In: ENCONTRO NACIONAL DE BIBLIOTECAS JURÍDICAS, 1., Lisboa, 2004. **Actas** [...] Coimbra: Coimbra Editora, 2006. p. 29-40. Disponível em: <https://hdl.handle.net/1822/4942>. Acesso em: 30 ago. 2024.

SANTOS, Fatima Cristina Lopes dos; MOLLICA, Maria Cecilia Magalhães; GUEDES, Vânia Lisboa Silveira. Coeféncia na representação temática de artigos científicos na área de saúde pública. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 24, n. 3, p. 194-213, out. 2019. Disponível em: <https://www.scielo.br/j/pci/a/FvnB3mwMrJCgqjX5ykKGcmd/>. Acesso em: 15 jul. 2021.

STRADER, C. Rockelle. Author-assigned keywords versus library of congress subject headings: implications for the cataloging of electronic theses and dissertations. **Library Resources & Technical Services**, v. 53, n. 4, p. 243-50, 2009. DOI: <https://doi.org/10.5860/lrts.53n4.243>.

TARTAROTTI, Roberta Cristina Dal'Evedove; DAL' EVEDOVE, Paula Regina; FUJITA, Mariângela Spotti Lopes. Política de indexação em ambiente colaborativo: uma análise em bibliotecas digitais brasileiras de teses e dissertações. **Scire**, Zaragoza, v. 24, p. 45-52, 2018. Disponível em: <https://www.ibersid.eu/ojs/index.php/scire/article/view/4570>. Acesso em: 19 dez. 2024.

TERRA, Ana Lúcia; AGUSTÍN LACRUZ, Carmen; FUJITA, Mariângela Spotti Lopes. Diretrizes dos periódicos de psicologia sobre título, resumo e palavras-chave. **Ciência da Informação**, Brasília, DF, v. 51, n. 2, p. 19-29, maio/ago. 2022.

TERRA, Ana Lúcia, AGUSTÍN LACRUZ, Carmen, BERNARDES, Óscar, FUJITA, Mariângela Spotti Lopes; BUENO DE LA FUENTE, Gema. Subject-access metadata on ETD supplied by authors: a case study about keywords, titles and abstracts in a Brazilian academic repository. **Journal of Academic Librarianship**, Ann Arbor, v. 47, n. 1, e102268, 2021. DOI: 10.1016/j.acalib.2020.102268.

UNIVERSIDADE ESTADUAL PAULISTA. Rede de Bibliotecas da Unesp. Comissão Permanente do Tesauro Unesp. **Tutorial para uso do tesauro Unesp**. São Paulo: Unesp, [2024].

UNISIST. **Study report on the feasibility of a world science information system**. Paris: UNESCO, 1971. Disponível em: <https://unesdoc.unesco.org/ark:/48223/pf0000064862>. Acesso em: 30 ago. 2024.

VAN DE SOMPEL, Herbert; LAGOZE, Carl. The Santa Fe convention of the open archives initiative. **D-Lib Magazine**, v. 6, n. 2, 2000. Disponível em: <https://www.dlib.org/dlib/february00/vandesompel-oai/02vandesompel-oai.html>. Acesso em: 30 ago. 2024