

Sistematização de modelo de avaliação do controle de vocabulários em repositórios: relato de pesquisa com o Repositório Institucional Unesp

Mariângela Spotti Lopes Fujita ¹ 

RESUMO

Introdução: O controle de vocabulário em sistemas de armazenamento, tratamento e recuperação de recursos de informação é necessário para se obter consistência entre a indexação e a recuperação de modo a evitar a dispersão informacional. Repositórios digitais em universidades são, atualmente, fundamentais na organização e gestão do conhecimento gerado pela produção científica, tecnológica, artística e administrativa, entretanto, é preciso verificar a disponibilização de vocabulário controlado e como se realiza o controle de vocabulário. **Objetivo:** Com o objetivo de sistematizar proposta para controle de vocabulário e uso de vocabulários controlados em repositórios universitários administrados por bibliotecas foi elaborado modelo de avaliação que se propõe a sistematizar métodos, procedimentos, recursos e técnicas. **Metodologia:** Para isso, o desenvolvimento da investigação realizou pesquisa exploratória com pesquisa bibliográfica e documental e pesquisa aplicada no Repositório Institucional Unesp. **Resultados:** Os resultados obtidos constituíram-se em um Plano de Ação, discutido e elaborado por Grupo de Estudos, composto de seis ações e nove estudos para avaliação e controle de vocabulário em repositórios universitários acerca de: controle de vocabulário na indexação por profissionais e não profissionais; controle de vocabulário na recuperação; uso de metadados de assuntos de trabalhos acadêmicos; compatibilização de palavras-chave; análise de variações terminológicas em nível semântico, sintático e pragmático; análise de logs de transação de buscas por assuntos. **Conclusão:** Conclui-se que a sistematização das ações em modelo de avaliação é relevante para que repositórios universitários incorporem os avanços oferecidos pelo controle de vocabulário em suas rotinas e, principalmente, pela contribuição de novos termos oriundos da evolução científica e tecnológica.

PALAVRAS-CHAVE

Controle de vocabulário. Linguagens documentárias. Repositórios institucionais. Avaliação.

Correspondência do autor

¹Universidade Estadual Paulista
Marília, SP, Brasil /
e-mail: mariangela.fujita@unesp.br

Systematization of the evaluation model of vocabulary control in repositories: research report with the Unesp Institutional Repository

ABSTRACT

Introduction: Vocabulary control in information resource storage, treatment and retrieval systems is necessary to obtain consistency

between indexing and retrieval in order to avoid informational dispersion. Digital repositories in universities are currently fundamental in the organization and management of knowledge generated by scientific, technological, artistic, and administrative production, however, it is necessary to verify the availability of controlled vocabulary and how vocabulary control is carried out. **Objective:** With the objective of systematizing a proposal for vocabulary control and the use of controlled vocabularies in university repositories managed by libraries, an evaluation model was developed that proposes to systematize methods, procedures, resources, and techniques. **Methodology:** For this, the development of the investigation carried out exploratory research with bibliographic and documentary research and applied research in the Unesp Repository. **Results:** The results obtained constituted an Action Plan, discussed and elaborated by the Study Group, composed of six actions and nine studies to evaluate and control vocabulary in university repositories about: vocabulary control in indexing by professionals and non-professionals; vocabulary control in retrieval; use of subject metadata from academic papers; keyword matching; analysis of terminological variations at semantic, syntactic and pragmatic levels; analysis of transaction logs for searches by subject. **Conclusion:** It is concluded that the systematization of actions in an evaluation model is relevant for university repositories to incorporate the advances offered by vocabulary control in their routines and, mainly, for the contribution of new terms arising from scientific and technological evolution.

KEYWORDS

Vocabulary control. Documentary languages. Institutional repositories. Evaluation.

| 2

CRediT

- **Reconhecimentos:** Agradecimentos à Coordenadoria Geral de Bibliotecas da Unesp.
- **Financiamento:** Este estudo foi financiado pela agência brasileira do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) para a bolsa produtividade em pesquisa. Processo: 304760/2017-0
- **Conflitos de interesse:** Os autores certificam que não têm interesse comercial ou associativo que represente um conflito de interesses em relação ao manuscrito.
- **Aprovação ética:** Sim.
- **Disponibilidade de dados e material:** Os conjuntos de dados gerados e/ou analisados durante o presente estudo estão disponíveis no <https://dmptool.org/plans/60792>
- **Contribuições dos autores:** A autoria é única que realizou conceituação, análise formal, investigação, metodologia, supervisão, validação, redação -rascunho original: FUJITA, M. S. L.



JITA: HS. Repositories.



Artigo submetido ao sistema de similaridade

Submetido em: 25/03/2022 – Aceito em: 26/05/2022 – Publicado em: 31/05/2022

1 INTRODUÇÃO

Com a produção acelerada dos documentos nato-digitais (criados em ambiente digital) e a preocupação com a preservação, o repositório aparece como uma alternativa para o depósito seguro dos objetos digitais. A grande vantagem, porém, é que podem ser utilizados em instituições públicas e privadas para disseminação de toda a pesquisa que é produzida, além de proporcionarem o autoarquivamento pelos autores, o que confere melhor divulgação das pesquisas. Por isso, existem muitos repositórios e, de forma global, o *Ranking Web of Repositories*¹ registra 3885 repositórios em todo o mundo, em sua mais recente edição de fevereiro de 2022, entre os quais 3751 são repositórios institucionais.

É um serviço de informação científica (em ambiente digital e interoperável) que gerencia a produção intelectual de uma instituição de ensino e pesquisa. É responsável pela reunião, armazenamento, organização, preservação, recuperação e, principalmente, pela disseminação da informação científica que é produzida na instituição. Segundo Lynch (2003, p. 2) é “[...] um conjunto de serviços que a universidade oferece aos membros de sua comunidade, visando o gerenciamento e disseminação dos recursos digitais criados pela instituição e pelos seus membros.”

Ademais, contém sistema de recuperação da informação por pontos de acesso do documento e dos metadados o que possibilita acesso ao documento digital contido no repositório. No caso do assunto, assim como em outras bases de dados é possível verificar se houve controle do vocabulário utilizado para representação na indexação ou na busca.

Com a possibilidade de autoarquivamento pelo autor, o repositório torna-se um sistema de informação mais dinâmico e amigável porque permite a interatividade com seus usuários que passam a construí-lo socialmente e, com isso, querem ser visíveis além de precisarem garantir a preservação digital em um sistema institucional que proverá a confiabilidade probatória a agência de fomentos, demais instituições e comunidade científica. Essa interação proporcionada pelo autoarquivamento exige do Repositório um compromisso com padrões descritivos e temáticos continuamente avaliados e aplicados para a garantia da visibilidade.

Por outro lado, a interação do autor e do usuário do repositório traz o benefício de sua terminologia de domínio de conhecimento aplicada para atribuição de palavras-chave em metadados de assuntos durante autoarquivamento e durante a estratégia de busca. É uma terminologia especializada em domínios de conhecimentos utilizada entre pares na comunidade científica que acompanha a evolução e a inovação cujos conteúdos documentais advêm de pesquisas científicas para geração de novos conhecimentos. Por outro lado, essa terminologia científica possui variações terminológicas, principalmente em nível sintático e semântico, que precisam de controle de vocabulário com uso de vocabulários controlados para a garantia de visibilidade almejada.

Tendo em vista que o repositório pratica o uso combinado das linguagens natural e controlada de forma híbrida, essa situação pode beneficiar o repositório caso haja uma política de indexação para autores e bibliotecários que proporcione as diretrizes necessárias para o controle de vocabulário. Para isso, é necessário estudo de avaliação de controle de vocabulário em repositórios de modo a elaborar proposta adequada para o uso de vocabulário controlado durante a atribuição de assuntos bem como do uso da linguagem natural para contínua atualização do vocabulário controlado.

Com o objetivo de elaborar proposta para controle de vocabulário e uso de vocabulários controlados em repositórios universitários administrados por bibliotecas foi elaborado modelo de avaliação que se propõe a sistematizar métodos, procedimentos, recursos e técnicas em prol da elaboração de uma política de indexação para repositórios.

¹ [TRANSPARENT RANKING: All Repositories \(February 2022\) | Ranking Web of Repositories \(webometrics.info\)](https://webometrics.info/)

2 FUNDAMENTAÇÃO TEÓRICA

O controle de vocabulário é exercido com auxílio de vocabulário controlado à exemplo de tesouros, lista alfabética autorizada de termos, listas de cabeçalhos de assunto, entre outros.

A Norma ISO 25964-2 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p.16), sobre tesouros para a recuperação da informação, considera o controle de vocabulário essencial porque no discurso comum um termo pode ter mais de um significado e a escolha de um termo preferido para representar um conceito específico nunca é direta porque conceitos podem ser expressos de várias maneiras. Por isso, o tesouro tem função importante na mediação entre os termos utilizados no discurso e aqueles que funcionam efetivamente para a recuperação da informação o que implica o usuário aceitar um grau de artificialidade no vocabulário controlado para o alcance de benefícios na recuperação.

O conceito de controle de vocabulário para a Norma Z39.19-2005 “*Guidelines for the construction, format, and management of monolingual controlled vocabularies*” (AMERICAN NATIONAL STANDARDS INSTITUTE/ NATIONAL INFORMATION STANDARDS ORGANIZATION, 2005, p.10) significa que é

[...] o processo de organização de uma lista de termos (a) para indicar qual de dois ou mais termos sinônimos está autorizado a ser utilizado; (b) para distinguir entre homógrafos; e (c) para indicar relações hierárquicas e associativas entre os termos no contexto de um vocabulário controlado ou lista de cabeçalho de assunto. (tradução nossa)

A norma, desse modo, considera a organização do vocabulário controlado como o próprio controle de vocabulário, mas, ao mesmo tempo, ressalta as funções que o vocabulário controlado realiza, tais como a indicação do termo sinônimo autorizado e quais relações hierárquicas e associativas existem entre termos. Portanto, o controle de vocabulário está presente em um vocabulário controlado que, por sua vez, é usado para realizar o controle de vocabulário. | 4

O controle de vocabulário, dessa forma, está ligado ao uso de um vocabulário controlado que “[...] é essencialmente uma lista de termos autorizados.” (LANCASTER, 2004, p.19), porém, vai além de uma lista porque os termos autorizados estão organizados em uma estrutura semântica que controla termos sinônimos, homógrafos e relacionados entre si, seja por relação hierárquica ou relação associativa.

Os vocabulários controlados têm, caracteristicamente, dupla função com objetivo de reciprocidade no controle de vocabulário, porque são usados durante os processos de representação e busca. Lancaster (2002, p.22) demonstra a dupla função pelos objetivos do controle de vocabulário:

1. Facilitar a representação consistente dos assuntos por indexadores e usuários que recuperam, evitando a dispersão dos elementos relacionados. Isto se consegue com o controle (agrupação) dos sinônimos e quasisinônimos e a distinção de homógrafos;
2. Facilitar a realização de uma busca ampla sobre uma matéria vinculando os termos com relações paradigmáticas e sintagmáticas

Hjorland considera que o princípio do vocabulário controlado segue a regra de Cutter de que “[...] é sempre a expressão mais específica e mais adequada que deve ser pesquisada no vocabulário e atribuída a documentos.” e que, “Desta forma, as expressões para os tópicos a serem recuperados são mais previsíveis.” (HJORLAND, 2008, p.89)

Os avanços da indexação no que tange à sua avaliação são medidas válidas para assegurar a vantagem do uso de vocabulários controlados em assegurar consistência, tanto na representação quanto na busca, de modo que “[...] um conceito ou tema apareça expresso sempre da mesma forma” (MOREIRO GONZALEZ, 2004, p. 51).

Outra vantagem é que a evolução dos vocabulários controlados, provavelmente influenciada pela indexação social na internet, tem oferecido modos de visualização cada vez mais intuitivos e projetados para usuários não profissionais que necessitam de suporte terminológico para o alcance de especificidade ou exaustividade em suas buscas. Na evolução de vocabulários controlados, o tesouro é, sem dúvida, utilizado por instituições nacionais e internacionais pela forma de apresentação facilitadora da compreensão de conceitos e de sua contextualização em uma determinada área de conhecimento. Sobre isso, a Norma ISO 25964 (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p.vi) considera “[...] que, no passado, os tesouros foram projetados para profissionais de informação treinados em indexação e pesquisa, hoje há uma demanda por Vocabulários que os usuários não treinados acharão intuitivos [...]”.

É baseado nessa premissa que a Norma ISO 25964 sustenta a aplicação do tesouro como vocabulário controlado, também, em situações em que os computadores façam as escolhas, ou seja, “Se o indexador e o pesquisador forem orientados para escolher o mesmo termo para o mesmo conceito, os documentos relevantes serão recuperados” (INTERNATIONAL ORGANIZATION FOR STANDARDIZATION, 2011, p.vi).

Os vocabulários controlados são fundamentais no momento da indexação. É um instrumento de tradução da linguagem natural de modo a diminuir os problemas de variação terminológica além de inconsistências da forma escrita. Vocabulário controlado é uma lista de termos autorizados que auxilia na indexação, melhorando os aspectos relacionados a recuperação com o uso de termos consistentes ao conteúdo dos documentos; controlar sinônimos, optando-se por uma única forma padronizada, com remissivas de todas as outras formas; diferenciar homógrafos; reunir ou ligar termos cujos significados apresentem uma relação mais estreita entre si. No entanto, é mais que uma lista, uma vez que, os termos autorizados estão organizados em estrutura semântica que possibilita o controle de termos sinônimos, homógrafos e relacionados entre si, seja por relação hierárquica ou relação associativa (LANCASTER, 2004, p.14-19).

Porém, o controle de vocabulário em repositórios institucionais não é uma tarefa fácil por uma série de fatores e aspectos peculiares não presentes em outros sistemas de informação como, por exemplo, as modalidades de autoarquivamento pelo autor e o povoamento automático. Com os resultados obtidos da análise do critério de uso de vocabulários controlados ou ferramentas terminológicas em uma amostra de 35 repositórios universitários espanhóis, Barrionuevo Almuzara, Alvite Díez, Rodríguez Bravo (2012, p.98) observaram que “[...] a função principal foi manipulada por listas de cabeçalhos de assunto e palavras-chave e, em menor grau, classificações, tesouros e listas de descritores.” e que, com a opção de autoarquivamento os autores podem determinar suas próprias palavras-chaves sem consulta aos assuntos que o sistema oferece. Concluem que “[...] o volume de termos sem controle que podem ser incluídos em repositórios não é limitado, uma circunstância que parece exigir alguma forma de padronização.”

Em investigação sobre vocabulários controlados, Fujita e Tolare (2019) realizaram análise de recursos de interface de 86 repositórios brasileiros para identificação de tipos de vocabulários controlados. Os resultados identificam que 81% utilizam listas de termos em ordem alfabética sem controle de vocabulário e em 65% dos repositórios incluem em seus metadados palavras-chave da linguagem natural e termos de vocabulários controlados. Consideram que a lista de termos em ordem alfabética, derivadas de palavras-chave da linguagem natural, são vocabulários controlados de menor complexidade comparados aos tesouros e que poderiam ser aprimoradas com a aplicação de controle de vocabulário. Sobre a integração de termos de vocabulários controlados com palavras-chave da linguagem natural atribuídas pelos autores, Fujita e Tolare (2019) consideram necessária para a atualização dos vocabulários por dois motivos: as listas de termos incorporam termos e palavras-chaves e o quantitativo de termos e palavras-chave em metadados ampliam a visibilidade da produção científica arquivada no repositório.

3 METODOLOGIA

A metodologia adotada para a realização desta investigação sobre avaliação do controle de vocabulário e uso de vocabulários controlados em repositórios institucionais possui um caráter descritivo exploratório de cunho etnográfico em razão da necessidade de extração de dados e informações diretamente da realidade.

A pesquisa etnográfica tem como objetivo a descoberta de novas relações e de novas formas de entendimento da realidade a partir da observação e visão dos participantes sobre os significados dos resultados obtidos em sua prática cotidiana. Para isso utilizou as técnicas de observação participante, entrevista e análise de entrevista pelo pesquisador com base nos aspectos observados e resultados obtidos de estudos teóricos, bem como análise da documentação, que se organizaram em três etapas: 1) exploração, que consiste na seleção dos problemas, do local e os primeiros contatos com o campo de estudo; 2) decisão, ou a busca dos dados para compreensão e interpretação do fenômeno; e 3) explicação da realidade, pela análise de todo o processo vivenciado pelo pesquisador, através dos relatos dos profissionais sobre a atividade desenvolvida (MAIA, 2007).

O desenvolvimento da investigação etnográfica foi realizado com duas orientações metodológicas: a primeira que discutiu, observou, interagiu sobre uso e avaliação do controle de vocabulário com Grupo de Estudos formado com pesquisadores da temática, catalogadores de bibliotecas universitárias, gestor e profissionais de apoio de repositórios universitários; e, a segunda que realizou e discutiu propostas viáveis para avaliação do controle de vocabulário em um repositório universitário com acompanhamento do grupo.

A primeira orientação, de natureza etnográfica, constituiu o Grupo de Estudos para o estudo de avaliação do controle de vocabulário e uso de vocabulários controlados em repositórios institucionais, mediante reuniões quinzenais durante dois anos, de modo a promover a reflexão crítica com a finalidade de identificar problemas ou avaliar mudanças durante a realização de seminários de vivências profissionais ou compartilhamento de experiências. Nestes seminários foi apresentada a sistematização teórica e metodológica sobre o controle de vocabulário na representação e recuperação da informação e avaliação do controle de vocabulário em repositórios institucionais administrados por bibliotecas a partir das considerações analíticas e comparativas do desenvolvimento da pesquisa, bem como a proposta de modelo metodológico para avaliação e controle de vocabulário em repositórios universitários.

A pesquisadora e o grupo interagiram com o objeto da pesquisa e analisaram todo o processo que permitiu o aprimoramento do modelo metodológico com a proposta de um Plano de Ação para implantação, manutenção e avaliação cuja sistematização resultou no modelo de avaliação do controle de vocabulário em repositórios institucionais, objeto de análise da segunda orientação metodológica da pesquisa etnográfica. A segunda orientação realiza a sistematização do modelo metodológico para avaliação e controle de vocabulário em um repositório universitário com acompanhamento do grupo. O repositório universitário utilizado para a execução do plano foi o Repositório Institucional Unesp, coordenado por Grupo Gestor e desenvolvido por Coordenação Executiva que acompanha atividades da Equipe Técnica. A segunda orientação se desdobrou em duas fases: a fase de conhecimento sobre o estudo de caso com o Repositório Institucional da Unesp e a fase de discussão do Plano de Ação para elaboração do modelo metodológico de avaliação e controle de vocabulário com acompanhamento do Grupo de Estudos.

Para o andamento da primeira fase, denominada estudo analítico-descritivo do controle de vocabulário no Repositório Institucional Unesp, foram realizadas reuniões conjuntas do Grupo de Estudos com a Coordenação Executiva e Equipe Técnica como forma de obtermos um grau de interação com a realidade profissional. Realizamos entrevistas com membros da Coordenação Executiva e Equipe Técnica para obtermos a visão sobre os procedimentos a respeito das atividades do Repositório, bem como sobre a necessidade e importância do controle

de vocabulário e uso de vocabulários controlados. A análise da documentação foi realizada para contextualizar e complementar as informações coletadas nas entrevistas.

No estudo analítico-descritivo do controle de vocabulário no Repositório Institucional Unesp buscou-se analisar a relevância do controle de vocabulário e seu uso no Repositório Institucional Unesp a partir das reuniões com o Grupo de Estudos como também mediante entrevista realizada com a Coordenação Executiva e Equipe Técnica. Para a entrevista, analisou-se a documentação para contextualizar as questões e complementar as informações coletadas nas entrevistas.

A entrevista com a Coordenação Executiva e Equipe Técnica teve como objetivo obter a visão sobre os procedimentos a respeito das atividades do Repositório, bem como sobre a necessidade e importância do controle de vocabulário e uso de vocabulários controlados. As questões foram elaboradas com base na análise: a) da literatura e da documentação sobre o Repositório Institucional Unesp (RI-Unesp); e, das respostas de questionário sobre política de indexação² preenchido pela Equipe Executiva do RI- Unesp. O questionário elaborado pelo Subgrupo de Trabalho da Rede Sudeste tinha a temática de política de indexação em repositórios. A finalidade era obter uma visão geral sobre os procedimentos a respeito das atividades do Repositório, bem como sobre a necessidade e importância do controle de vocabulário e uso de vocabulários controlados. Buscou-se compreender para posterior análise o contexto institucional, a realidade de atuação de profissionais quanto, aos elementos, variáveis, processos e instrumentos que envolve o manejo do controle de vocabulário para elaborar um diagnóstico da política de indexação nos repositórios.

Na segunda fase de discussão da viabilidade de aplicação do Plano de Ação no Repositório Institucional Unesp a pesquisadora e o Grupo de Estudos realizaram, com a participação da Coordenação Executiva e Equipe Técnica, reuniões de análise e discussão do Plano de Ação para sistematização do modelo metodológico de avaliação do controle de vocabulário. Dessa forma, os resultados do estudo analítico-descritivo do controle de vocabulário no Repositório Institucional Unesp da primeira fase e a discussão do Plano de Ação para elaboração do modelo metodológico de avaliação e controle de vocabulário com acompanhamento do Grupo de Estudos da segunda fase, serão apresentados, respectivamente, nas duas seções seguintes.

4 ANÁLISE DO REPOSITÓRIO INSTITUCIONAL UNESP NA PERSPECTIVA DO CONTROLE DE VOCABULÁRIO

Nesta pesquisa, o estudo do Repositório Institucional da Unesp teve especial enfoque e foi o ambiente para o desenvolvimento da investigação. Com a finalidade de elaboração de Plano de Ação com o Grupo de Estudos para sistematização de modelo metodológico para avaliação de controle de vocabulário em repositórios universitários, foi inicialmente elaborada análise do Repositório Institucional Unesp na perspectiva do controle de vocabulário por meio da literatura sobre sua criação e funcionamento mediante análise de questionário e entrevista com Equipe Executiva.

Para entender o histórico de criação e implementação do repositório utilizou-se de revisão de marcos legais e de publicações da própria equipe de implantação para contextualização. O Repositório Institucional Unesp, criado em outubro de 2013 em conjunto com os repositórios da Universidade de São Paulo (USP) e da Universidade Estadual de Campinas (UNICAMP), somam-se ao Repositório da Produção Científica do CRUESP (Conselho de Reitores das Universidades Estaduais Paulistas).

O repositório surgiu a partir da Portaria Unesp número 88, de 28 de fevereiro de 2013, que instituiu o Grupo Gestor da Política do Repositório Institucional Unesp (GRI- Unesp),

² Questionário disponibilizado pela Rede Sudeste de Repositórios Institucionais

responsável pelo desenvolvimento, implantação e manutenção do repositório da universidade com o objetivo de “armazenar, preservar, disseminar e possibilitar o acesso aberto, como bem público global, à produção científica, acadêmica, artística, técnica e administrativa da Universidade.” (UNIVERSIDADE ESTADUAL PAULISTA, 2013, p. 47). A Coordenadoria Geral de Bibliotecas da Unesp (CGB) faz parte do Grupo Gestor e é responsável pela coordenação executiva do projeto. Tem como incumbência garantir a inclusão da produção no Repositório junto à Equipe Técnica, formada por bibliotecários e profissionais analistas de sistemas.

Para implementação do Repositório Institucional Unesp, foram definidas quatro metas que refletem seu objetivo (ASSUMPCÃO; SILVA; FERREIRA; BASTOS, 2014, p. 4):

1. inclusão da produção científica institucional publicada no período de 2008 a 2012 e indexada na Web of Science;
2. inclusão da produção científica institucional publicada em periódicos da SciELO;
3. inclusão da produção científica institucional publicada no período de 1976 a 2007 e indexada na Web of Science;
4. inclusão da produção científica institucional indexada na Scopus.

A inauguração do Repositório CRUESP fez a Unesp ter como objetivo incluir toda a produção científica da universidade no repositório. Desse modo, a meta inicial era incluir a produção de pesquisadores da universidade indexada nas bases de dados *Web of Science* e *Scopus* e publicada em periódicos da *Scientific Electronic Library Online* (SciELO). Para alcançar essa meta de maneira eficiente, mesmo com o prazo para a inauguração do Repositório CRUESP e do próprio Repositório Institucional Unesp, foram usados processos de coleta, de conversão e de importação automática dos registros referentes a essa produção científica.

O Repositório está organizado em comunidades que representam as unidades universitárias da Unesp e divididas em subcomunidades que representam os departamentos e Programas de Pós-Graduação, onde há coleções para os diferentes tipos de documentos de autoria dos docentes e/ou discentes associados ao departamento ou ao Programa de Pós-Graduação. Da mesma forma, os artigos científicos e os outros materiais são inseridos no Repositório apenas pela Equipe Técnica responsável por essa atividade. Não é possível publicar no Repositório Institucional Unesp, pois ele não edita e não publica qualquer documento, apenas os que já foram publicados em revistas científicas, anais de eventos etc.

O Repositório faz parte do movimento de acesso aberto da produção científica e qualquer pessoa pode consultar e fazer *download* dos documentos do Repositório. Há apenas uma restrição em algumas dissertações e teses com acesso ao texto completo quando está restrito durante um período escolhido pelo autor (período de embargo). Qualquer pessoa pode se cadastrar no Repositório, com ele o autor consegue se inscrever para receber notificações sobre os novos documentos que forem adicionados ao repositório. Para realizá-lo é preciso cadastrar-se no site do repositório com o e-mail e seguir as instruções disponíveis na página.

As atividades que envolvem a implantação, manutenção e aprimoramento do Repositório são desenvolvidas pela Coordenadoria Geral de Bibliotecas (CGB) por meio de suas parcerias com o Núcleo de Educação a Distância (NEaD), com as Pró-Reitorias de Pós-Graduação, de Graduação, de Pesquisa, de Extensão e de Administração, com a Fundação Editora Unesp e com a Agência Unesp de Inovação (AUIN).

Desde que foi criado, o Repositório Institucional Unesp conseguiu resultados satisfatórios que atendem plenamente a comunidade acadêmica e obteve em julho de 2016 a sexta posição no Ranking do *Web of Institutional Repositories*, com repositórios nacionais, e a 233ª posição no ranking mundial de repositórios. Em 2022, é um dos cinco maiores repositórios no Brasil, de acordo com a edição de fevereiro de 2022 do *Ranking Web of Institutional Repositories*, e está entre os 30 maiores repositórios institucionais mundiais na 24ª posição (WEBOMETRICS, 2022).

Em março de 2022, o repositório dispõe de 173848 registros dos quais 55% são artigos (95420), 17% são dissertações de mestrado (30248), 9% são teses de doutorado (16278), 6%

são trabalhos de conclusão de curso (10383) e outros materiais como trabalhos apresentados em evento (8902), resumo (6579), resenha (1984), editoriais (754), cartas (643), capítulos de livro (550), livros (487), patentes (408), podcasts (275), erratas (254), notas (209), teses de livre-docência (167) boletim (78), revista (74), dados de pesquisa (56), jornal (34), relatório (26), data paper (13), plano de gestão de dados (9), biografia (7), objeto educacional (6), partitura (2), regulamento (1) e vídeo (1). Em suma, o Repositório Institucional Unesp atende satisfatoriamente sua comunidade permitindo a disseminação e o acesso à produção desenvolvida na universidade.

Para incluir a produção de pesquisadores da universidade indexada nas bases de dados foram usados recursos de coleta, de conversão e importação automática dos registros das produções. Para isso, as coletas automáticas auxiliaram na inclusão de registros no repositório a partir de fontes de dados como a *Scopus*, *Web of Science*, *SciELO*, *PubMed*, Currículo Lattes e do catálogo Athena que são integrados ao perfil *Open Research and Contributors Identification* (ORCID) dos docentes e pesquisadores da Unesp. O repositório utiliza o ORCID para integrar a produção científica de pesquisadores nas bases de dados. A Unesp foi a primeira universidade a utilizá-lo no Brasil, assim toda a produção acadêmica dos pesquisadores faz parte do registro dos pesquisadores e evita o retrabalho de preencher e atualizar seus dados em outros locais.

Está organizado em comunidades que representam as unidades universitárias da Unesp e divididas em subcomunidades que representam os departamentos e Programas de Pós-Graduação, possui coleções para os todos os tipos de documentos da comunidade universitária. As informações gerais sobre o Repositório Institucional Unesp contida no site esclarecem que:

- Utiliza o software DSpace;
- Na barra lateral esquerda do repositório encontra-se um menu de buscas com tipo de produção, data do documento, autor, título, palavra-chave. Em palavras-chave apresenta uma lista alfabética de 329.103 termos em português, inglês, francês e italiano, que estão presentes nos metadados dos documentos.
- Formato de Metadados: Dublin Core.
- Tipo de produção: produção acadêmica e científica, produção administrativa, produção artística, produção comemorativa - Unesp 40 anos, produção cultural, produção técnica.
- Tipos de materiais: possui 28 tipos diferentes de materiais que somam 173821; os três tipos com maiores quantidades são: artigos (95420), dissertação de mestrado (30237), tese de doutorado (16270) e Trabalho de Conclusão de Curso (10375)³.

| 9

Para entender o contexto atual do repositório e de como os profissionais desenvolvem as funções e operações necessárias ao funcionamento foi realizada entrevista com a gestora e duas bibliotecárias da Equipe Executiva do Repositório Institucional Unesp em outubro de 2021 que responderam as questões durante a entrevista gravada. De acordo com as explicações da metodologia foram formuladas 12 questões, conforme Apêndice 1. A primeira parte com cinco questões foi formulada a partir de literatura publicada sobre o Repositório Institucional Unesp e de sua documentação e a segunda parte, com mais sete questões, tomou como base as respostas às questões do “Questionário sobre Política de Indexação em Repositórios” (analisadas acima pela autora).

Conforme respostas do “**Questionário sobre Política de Indexação em Repositórios**” analisamos que:

- **A formação o e desenvolvimento da coleção digital do Repositório Institucional Unesp** envolve o povoamento (coleta, captura ou colheita feita de forma automática ou semiautomática), os depósitos (submissão de documento) e autoarquivamento;

³ Dados de 24 de fevereiro de 2022 disponíveis em: [Repositório Institucional UNESP](https://repositorio.inunesp.br/)

- A **equipe do repositório** é formada por nove servidores da universidade, sendo dois analistas, três bolsistas do curso de Biblioteconomia, um assistente e três bibliotecárias dentre elas a coordenadora que também compõe o Comitê gestor do Repositório. Atualmente conta com 36 bibliotecários da rede de bibliotecas da Unesp dedicados às tarefas de indexação e catalogação.
- A **Plataforma utilizada no repositório é DSpace** e utiliza os softwares auxiliares para tratamento da informação: Duplicate Checker; Oxygen; Adobe; MarcEdit; Libre office. O padrão de metadados utilizado é o Dublin Core;
- O repositório possui **padronização de metadados** para definição de campos obrigatórios, repetitivos e de descrição com o intuito de melhorar a qualidade do armazenamento dos documentos.
- O **perfil dos usuários do repositório** é a própria comunidade acadêmica e abrange também a comunidade externa em geral;
- No quesito **prática da indexação**, o repositório não realiza um controle de autoridade para os registros migrados de fontes externas ao repositório, como também não dispõe de uma política de indexação escrita e formalizada. No entanto, apresenta um manual para procedimentos práticos do processo de indexação de assuntos e durante o processo é utilizado como auxílio automático para facilitar a operação o tesouro Unesp;
- Em relação a **qualidade da indexação** o nível de especificidade dos termos de indexação não está estabelecido na determinação de assuntos dos documentos, o repositório possui uma indicação de mínimo três termos ou assuntos por documento;
- No que tange as **ferramentas para indexação**, o Repositório Institucional Unesp não utiliza ferramentas de validação/correção automática de termos/assuntos para garantir a correção e consistência de assuntos e de nomes (geográficos, nomes de pessoas, identificadores, séries e títulos), inclusive utiliza termos/assuntos sem controle de vocabulário, em linguagem natural (palavras-chave) combinada com a linguagem de indexação para a representação temática;
- O repositório utiliza mais de uma **linguagem de indexação para representação temática** como lista de cabeçalhos de assuntos, tesouros, cabeçalhos de assunto e de autoridades da instituição, tesouro da instituição, autoridades da *Library of Congress* e da Biblioteca Nacional. Da mesma forma, não realiza enriquecimento e manutenção da linguagem de indexação que consiga abranger a interoperabilidade/compatibilidade semântica de vocabulários controlados e não oferece sistema de marcação (*tagging*) para indexação de textos pelos usuários;
- No que se refere a **avaliação da indexação** o repositório no momento não realiza testes ou ensaios para a avaliação periódica da prática de indexação pela recuperação e não dispõe de relatórios publicados desta avaliação.
- A entrevistada relata que a universidade possui um grupo de bibliotecários que estuda a linguagem Unesp e a indexação que, entre suas atividades, realiza estudos sobre estes procedimentos, no entanto o grupo não está vinculado a equipe do repositório.

Em continuidade, na análise das respostas das questões da **primeira parte da entrevista** com a Equipe Executiva do repositório a gestora do Repositório Institucional Unesp (Gestora) foi esclarecido que:

a) **Política de indexação (para validação de assuntos):**

A política de indexação já estabelecida para o catálogo Athena pode ser utilizada futuramente também para o repositório institucional. Atualmente, o repositório não passa

compartilha os descritores atribuídos na validação de assuntos com base no tesauro, uma vez que não existe uma interoperabilidade do catálogo com o repositório. A rede de bibliotecas possui várias bases de dados, uma delas é o repositório e outra é o catálogo Unesp.

Acreditamos que poderia ser aproveitada a validação de assuntos do catálogo bibliográfico, com a devida atualização/inclusão de novos termos. Mas não temos certeza em relação às questões técnicas, pois necessitamos de um estudo específico para realizar testes para esse reaproveitamento. (Gestora)

Por outro lado, a política de indexação do catálogo precisa reaproveitar e rever as palavras-chave escolhidas pelo autor, pois é o produto de indexação do autor durante autoarquivamento no Repositório Institucional Unesp e não deveria ser eliminado do catálogo online Athena, mas sim passar por uma validação de novos termos para o tesauro, até porque o catálogo não permite que o próprio autor faça o autoarquivamento e a determinação de palavras-chave.

Então, [...] eu acredito que agora ... a política está estabelecida, entendo que esta política de indexação, ela é viva, ativa [...] e aí eu acredito que é muito de ... acionar o grupo e fazer, trazer estas reflexões para esse grupo, para que possam compreender também essa importância da linguagem utilizada pelo próprio usuário e para enriquecer na verdade, eu acredito, isso eu acho que sempre leva a um enriquecimento e uma melhoria na recuperação da informação, então eu na verdade, vejo desta forma, que seria algo é no sentido de acionar o grupo e para que a gente pudesse trazer está reflexão e atualização da própria política. (Gestora)

b) Metas do Repositório Institucional Unesp de 2014:

| 11

As bibliotecas e a CGB não vão conseguir atingir as metas sem é preciso a vontade política e institucional da universidade. Para que o repositório, realmente, possa ter um bom desenvolvimento, é muito importante o engajamento do gestor universitário neste projeto pois é um benefício não só para as bibliotecas, mas para toda a universidade e comunidade externa.

[...] o repositório institucional tem uma portaria que o regulamenta recentemente a gente teve uma política de acesso aberto aprovada inclusive por todos, [...], eu percebo hoje que nós somos procurado há por exemplo fazer um estudo do "impacto da pandemia " então a universidade reconhece que o ambiente do repositório possuem dados e esses dados falam, recentemente [...] as pessoas estão começando a entender principalmente porque por conta de fato que veio lá da agência de fomento, por exemplo, que está aqui, lança uma política de acesso aberto e torna obrigatório ter a produção armazenada no repositório, não só a produção mas os dados de pesquisa, entregar um plano de gestão, então assim eu entendo também que [...] a Unesp antes de 2013 ela já havia passado por três iniciativas de repositório e nenhuma deu certo a quarta tentativa que é a justamente quando a Fapesp diz para as três universidades estaduais "vocês implantem o repositório institucional porque nós vamos soltar um portaria informando aos pesquisadores que se eles não depositarem, que se a produção não estiver no repositório eles não iram receber o financiamento, o recurso financeiro [...] (Gestora)

[...] no ambiente do repositório a gente é, por conta da infraestrutura. Isso é um pouco maior extrapola até a equipe gestora na parte de execução e entra um pouco mais na coordenadoria de tecnologia da informação que agora tem comprado espaço em nuvem pra que a gente possa conseguir aproveitar e armazenar esse conteúdo que é de um tamanho maior. Então os vídeos geralmente são armazenados e depositados no Youtube e o que a gente faz, na verdade, a proposta é de se dar link, fazer a descrição do vídeo e dar link da onde este objeto está hospedado. (Gestora)

c) **Produção científica dos pesquisadores da Unesp no Repositório Institucional Unesp:**

Foi citado o exemplo do Instituto de Artes, o qual não costumava ser contemplado com algumas oportunidades da Reitoria devido a sua diferente produção. Entretanto, atualmente, existem mais repositórios que possuem obras artísticas e museológicas e, tendo em vista, a diversidade de produção científica que a universidade tem e considerando a relevância da área de humanas, é importante entender, segundo a gestora, que vai existir uma diversidade de tipos documentais maior ainda do que existe hoje.

[...] então em relação ao Instituto de Artes [...] a gente até começou a conversar porque muito dos materiais que eles produzem a gente também vai precisar de todo um trabalho de checagem da parte de direito autoral e eles falam [...] que regulamentam a produção deles porque geralmente eles produzem essas partituras, enfim, e já comercializam então precisamos ver aí toda essa parte no caso aí das partituras que nós temos aqui no repositórios [...] elas estão, se eu não estou enganada, dentro de uma comunidade que é da comemoração de 40 anos Unesp onde o Instituto, o próprio Instituto de Artes ... é desenvolveu e criou a partitura pro hino da Unesp, então muito provavelmente esteja ligado a essa comunidade, é isso. Os vídeos, por exemplo, nós não armazenamos... (Gestora)

d) **Teses e Dissertações**

As teses e dissertações representam quantidade significativa da produção científica da Unesp no Repositório e estamos cientes de que é um tipo de material que também está no catálogo online cujo software sofreu atualização e estamos trabalhando em questões técnicas para garantir a migração de dados entre ambas as bases de dados.

12

Neste momento, porque nós não mudamos do software Aleph para a plataforma Alma, então antes no software Aleph ainda, a gente carregava é ... de um para o outro, então se as teses e dissertações estão inseridas no repositório ... isso é migrado para o catálogo. Agora, com a plataforma ALMA, a gente tá trabalhando nisso [...] relação, então para que a gente tenha que ter no repositório o que se tenha também no Athena [...] nós acreditamos que a avaliação de assuntos do que existe no catálogo bibliográfico, lógico que ... com a devida atualização e inclusão de novos termos, nós acreditamos sim que nós podemos aproveitá-la para o repositório... mas ainda e eu acho que [...] nós, na verdade, ainda não temos certeza dessas questões técnicas, do que é que se, o que envolveria se nós teríamos condições, como é que seria esse aproveitamento, esse reaproveitamento, então eu entendo que nós precisamos fazer um estudo específico, para conseguir ter esse reaproveitamento da validação de assuntos que já existe no catálogo bibliográfico. (Gestora)

e) **O Tesouro Unesp nos tutoriais de autoarquivamento:**

Os bibliotecários atuam diretamente no processo de revisão dos registros de diferentes tipologias documentais que entram no repositório. Para realizar tal atividade seguem as recomendações do Tutorial de Verificação disponível no site do repositório quanto ao uso do Tesouro Unesp.

[...] então como nós temos esse tutorial ele tem essa recomendação para que se use para as seguintes tipologias artigos, capítulos de livros, dados de pesquisa, plano de gestão, TCC, tese e dissertação que está passando por revisão neste momento porque, enfim, a gente vai incluir uma parte de justificativa a própria ... pediu que a gente fizesse até mesmo uma revisão da própria portaria que regulamenta o autoarquivamento então assim os documentos de teses e dissertações eles estão passando por revisão e com certeza vão ter esse tipo de recomendação [...] recentemente a gente abriu o autoarquivamento de TCC por exemplo [...] o repositório está ganhando uma dimensão cada vez maior de inserção dentro dos

nossos trabalhos [...] não está dando pra ficar somente com um profissional pra fazer essa validação, então elas já estão querendo colocar mais pessoas da equipe pra ajudar na validação desses registros [...] elas tem uma recomendação pra se utilizar o tesouro Unesp é isso que elas tem no nosso tutorial de verificação e validação de registros dentro do repositório [...] (Gestora)

Na **segunda parte da entrevista**, a Gestora respondeu as questões derivadas das respostas ao questionário de Política de Indexação em Repositórios com a finalidade de esclarecer os seguintes aspectos:

a) Quantidade de profissionais bibliotecários dedicados à tarefa de indexação e catalogação:

São 36 profissionais bibliotecários da rede de bibliotecas da Unesp que realizam a revisão dos registros que entram no repositório decorrentes de autoarquivamento, porém, não indexam esses registros.

[...] esses 36 profissionais, na verdade, estão inseridos no processo de revisão do registro que entra no repositório e ali eles têm, no manual de revisão, a instrução de revisão de registros, eles têm sim uma recomendação de que eles devem consultar o ... tesouro Unesp. (Gestora)

Os [...] profissionais são responsáveis pela validação de registros dentro do repositório institucional e utilizam para essa atividade as recomendações descritas no tutorial para verificação (artigos, capítulo de livro, dados de pesquisa, plano de gestão e TCC). (Gestora)

b) Função dos softwares auxiliares utilizados no tratamento da informação

| 13

Cada um dos softwares possui uma função específica e importante. É possível, por exemplo, verificar a duplicação em outros procedimentos que não estão apenas no repositório e importar os registros de metadados de uma base de e-books para dentro do próprio catálogo bibliográfico:

***Duplicate Checker:** é uma das ferramentas do módulo de qualidade de metadados [...] ele mostra possíveis registros duplicados dentro do repositório. Atualmente é utilizado para outros projetos.*

***Oxygen:** ferramenta para desenvolvimento e aplicação de folhas de estilo para transformação de registros das bases de dados no formato adotado pelo DSpace.*

***Adobe Acrobat:** programa para edição de arquivos em formato PDF.*

***MarcEdit:** faz transformação de metadados do MARC21 para MARC XML. Atualmente não utilizamos mais essa ferramenta para o R.I..*

***Libreoffice:** é um pacote Office. Utilizamos o Libreoffice Calc porque ele consegue [...] de maneiras mais efetivas que o Microsoft Excel. (Gestora)*

c) Padronização de metadados

O repositório utiliza o Dublin Core para padronização dos metadados e a ISO 639 para definição de idiomas.

d) Indexação e política de indexação no Repositório Institucional Unesp

Não é realizada atividade de indexação para nenhum tipo de documento e os profissionais seguem o tutorial de verificação de registros de autoarquivamento que recomenda o uso do tesouro Unesp para seleção dos descritores. Não existe uma política de indexação definida para o Repositório e se adota o Tesouro Unesp para o controle de vocabulário na revisão de registros de algumas tipologias documentais por profissionais e durante o

autoarquivamento por autores. Ainda não é possível incluir o Tesouro Unesp em auxílio automático ou semiautomático incorporado ao Repositório Institucional Unesp.

[...], mas nós não temos indexação dentro do repositório não temos ninguém que trabalhe [...] a única coisa que nós temos [...] eles têm um tutorial para verificação e validação de registros nesse tutorial existe a recomendação para que se use o tesouro Unesp. (Gestora)

A análise do Repositório Institucional Unesp na perspectiva do controle de vocabulário obtida mediante resultados da literatura, documentação, questionário e entrevista revela que o controle de vocabulário é realizado com a recomendação do uso do Tesouro Unesp especificamente para as tarefas de revisão de registros pelos bibliotecários e durante autoarquivamento pelos autores de algumas tipologias documentais para o que são disponibilizados tutoriais específicos. A modalidade que mais arquiva documentos no Repositório Institucional Unesp é a de povoamento automático, sem possibilidade de indexação pelo profissional e a segunda modalidade é o autoarquivamento pelos autores cujos registros são validados pelos profissionais. Além disso, não se aplica a modalidade de arquivamento mediado em que o profissional elabora o registro e faz a indexação com uso de vocabulário controlado. O Repositório Institucional Unesp não possui política de indexação descrita em manual, mas, recentemente, a Equipe Executiva investiu significativo esforço na elaboração de tutoriais especialmente orientados à revisão dos registros para os profissionais e tutoriais orientados ao autoarquivamento de diferentes tipologias documentais em que se recomenda o uso do Tesouro Unesp para o controle de vocabulário.

Ressalte-se que o envolvimento com a Equipe Executiva durante o desenvolvimento desta investigação foi inspirador para o desenvolvimento dos tutoriais. Entretanto, será necessária a elaboração de diretrizes para o tratamento temático e a padronização dos metadados de assuntos que, certamente, ajudarão na formalização de uma política de indexação orientada para as especificidades de gestão e funcionamento do Repositório Institucional Unesp. Outro ponto a ser observado é a ausência de realização de testes ou ensaios para avaliação de controle de vocabulário cujos resultados poderiam servir de parâmetro para calibrar os indicadores válidos na definição da política de indexação. Com essa visão, detalhamos a seguir a sistematização de modelo de avaliação do controle de vocabulário em Repositórios Institucionais com base nos estudos desenvolvidos pelo Grupo de Estudos com a colaboração e participação da Equipe Executiva do Repositório Institucional Unesp.

| 14

5 SISTEMATIZAÇÃO DE MODELO DE AVALIAÇÃO DO CONTROLE DE VOCABULÁRIO EM REPOSITÓRIOS INSTITUCIONAIS (RI)

O modelo metodológico de avaliação do controle de vocabulário para Repositórios Institucionais foi desenvolvido com base nas contribuições dos estudos realizados em Plano de Ação pelos pesquisadores do Grupo de Estudos conforme primeira e segunda orientações metodológicas de natureza etnográfica relatadas na seção de metodologia. A sistematização dos resultados dos estudos propõe metodologias de avaliação de controle de vocabulário a serem aplicadas com objetivos definidos em processos de organização e representação da informação realizados em repositórios institucionais. Para tanto, foi necessário indicar as metodologias propostas em tais estudos e identificar os processos e sistemas de organização e representação da informação compatíveis.

O Plano de Ação, composto de seis ações e 9 estudos, foi inicialmente discutido com a Equipe Executiva após seminário de apresentação do projeto de pesquisa seguido de reuniões de discussão dos principais problemas e demandas do Repositório Institucional. As ações foram definidas em função do resultado dessas reuniões cuja interação entre pesquisadores, catalogadores, profissionais de TI, bibliotecários e gestora demandaram nove propostas de estudos incluídas em seis ações conforme demonstrado no Quadro 1 a seguir:

Quadro 1. Plano de Ação: ações e estudos correspondentes

AÇÕES	ESTUDOS
AÇÃO 1: Avaliação do controle de vocabulário na indexação	ESTUDO 1: Avaliação da análise de assunto por autores no autoarquivamento de teses e dissertações em repositórios: estudo de observação com Protocolo Verbal;
	ESTUDO 2: Proposta de padronização de palavras-chave atribuídas por pesquisadores na submissão da produção científica em diferentes sistemas de informação que realizam a gestão e divulgação: uma proposta de política de indexação para pesquisadores e autores
	ESTUDO 3: A política de indexação em repositórios institucionais do Brasil: estudo diagnóstico na percepção de gestores e indexadores;
AÇÃO 2: Avaliação do controle de vocabulário na recuperação	ESTUDO 4: Avaliação do processo de indexação de assuntos pela recuperação da informação: uma análise nos repositórios institucionais do CRUESP;
AÇÃO 3: Análise de metadados de teses, dissertações e TCCs no Repositório;	ESTUDO 5: Metadados de assuntos de teses e dissertações no catálogo das bibliotecas da Unesp e no Repositório Institucional Unesp: estudo exploratório sobre controle de vocabulário
AÇÃO 4: Estudo da compatibilização de palavras chaves (linguagem natural) e tesouro Unesp conforme Santos (2020):	ESTUDO 6: Controle de vocabulário Repositório Institucional Unesp: proposta de mapeamento e compatibilização com o Tesouro Unesp
AÇÃO 5: Análise das variações terminológicas em nível semântico, sintático e pragmático com uso de processadores de linguagem natural.	ESTUDO 7: Análise de variações terminológicas em nível sintático, semântico e pragmático no vocabulário do Repositório Institucional Unesp
AÇÃO 6: Estudo da análise de logs	ESTUDO 8: Revisão bibliográfica atualizada sobre análise de logs e metodologia e verificar qual a influência dos estudos de análise de logs para avaliação da recuperação da informação em repositórios
	ESTUDO 9: Análise experimental da base de logs de busca por assuntos do Repositório Institucional Unesp

Fonte: Elaborado pela autora (2022).

Cada estudo foi realizado por grupos de pesquisadores do Grupo de Estudos com a colaboração da Equipe Executiva que acompanhou e participou do desenvolvimento durante os anos de 2020 e 2021. O Repositório Institucional Unesp serviu como universo de pesquisa para todos os estudos conforme orientação metodológica de natureza etnográfica.

As ações 1 e 2 são dedicadas à avaliação do controle de vocabulário cujos estudos desenvolvem metodologias específicas tendo em vista os processos de indexação e de recuperação como objeto de pesquisa. Na ação 3 o objeto de pesquisa é o metadados de assunto que proporciona avaliação da padronização do controle de vocabulário. As ações 4, 5 e 6 são voltadas para a linguagem natural como objeto de pesquisa tendo em vista a padronização e compatibilização para contínua atualização de vocabulários controlados. As metodologias de avaliação do controle de vocabulário em repositórios desenvolvidas por cada estudo são identificadas por seus respectivos objetivos e funções conforme Quadro 2 a seguir:

Quadro 2. Metodologias de avaliação do controle de vocabulário em repositórios: objetivos e funções

METODOLOGIAS	OBJETIVO	FUNÇÃO
Estudo 1: Observação de padrões e estratégias utilizadas por autores de teses e dissertações durante a indexação para atribuição de palavras-chave no autoarquivamento em repositório institucional	Avaliação do controle de vocabulário na indexação	Controle de vocabulário na indexação

Estudo 2: Análise de palavras-chave atribuídas por pesquisadores para submissão de artigos de periódicos indexados na base de dados Scopus e no Portal Docentes Unesp, quanto à padronização e controle de vocabulário para diferentes funções em sistemas de armazenagem e recuperação da informação	Avaliação do controle de vocabulário na indexação	Controle de vocabulário na indexação
Estudo 3: Estudo diagnóstico da política de indexação em repositórios institucionais brasileiros pela percepção de gestores e bibliotecários catalogadores-indexadores	Avaliação da política de indexação	Elaboração da política de indexação
Estudo 4: Avaliação da indexação de assuntos pela abordagem da recuperação com usuários no Repositório Institucional por meio de buscas por assuntos no Repositório Institucional da Unesp utilizando linguagem natural (palavras-chave das pesquisas em andamento dos usuários) e em um segundo momento utilizando linguagem controlada (Tesouro Unesp)	Avaliação do controle de vocabulário na recuperação	Controle de vocabulário na recuperação
Estudo 5: Análise comparada dos procedimentos de tratamento de teses e dissertações existentes no catálogo e no Repositório para avaliação do controle de vocabulário	Análise de metadados	Padronização da indexação de assuntos para controle de vocabulário
Estudo 6: Mapeamento e compatibilização das palavras-chave dos registros do Repositório Institucional da Unesp (RIU) com o Tesouro Unesp, por meio da compatibilização sintática, considerando-se os processos de igualdades e similaridades entre os termos.	Mapeamento e compatibilização de palavras-chave com o Tesouro Unesp	Atualização de vocabulários controlados
Estudo 7: Identificar e descrever o conceito e os tipos de variações terminológicas em nível sintático, semântico e pragmático que ocorrem ou que são passíveis de ocorrência no vocabulário controlado do Repositório Institucional Unesp e elaborar estratégias para seu tratamento terminológico.	Análise de variações terminológicas	Atualização de vocabulários controlados
Estudos 8 e 9: Análise de logs de buscas dos usuários para atualização de vocabulários controlados	Análise de logs	Atualização de vocabulários controlados

Fonte: Elaborado pela autora.

Com a finalidade de estabelecer uma sequência de aplicação das metodologias de avaliação do controle de vocabulário em repositórios institucionais foram analisadas as contribuições de cada estudo com vistas à aplicabilidade no Repositório Institucional Unesp conforme transcritas abaixo:

Estudo 1: Recomendar que os sistemas de autoarquivamento incluam tutoriais sobre atribuição de palavras-chave com controle de vocabulário sem a imposição de que são obrigados a utilizar somente termos controlados. A característica mais atual da palavra-chave tende a representar assuntos mais específicos dentro das ciências e, em comparação, os termos de indexação de um vocabulário controlado tendem a ser mais estáveis e se conectar a assuntos mais amplos, o que

determina uma complementaridade entre ambos e não permite a exclusão, mas sim a convivência em um sistema híbrido de representação e recuperação da informação.

Estudo 2: Elaborar proposta de política de indexação para padronização de palavras-chave atribuídas por autores e pesquisadores, na submissão da produção científica em diferentes sistemas de informação que realizam a gestão e divulgação científica. Discutir a elaboração de uma política de organização e representação da informação a ser seguida, que possa ser continuamente avaliada e atualizada e que forneça diretrizes aos docentes/pesquisadores na atribuição padronizada das palavras-chave em suas produções bibliográficas. Recomenda-se dar ciência aos docentes dos resultados desta pesquisa, para que possam corrigir suas palavras-chave de artigos, à luz de orientações de padronização e consistência.

Estudo 3: Elaborar manual de política de indexação e recomendar elementos e variáveis de política de indexação no Repositório Institucional da Unesp para melhorar o tratamento temático com a padronização de condutas quanto aos metadados;

Estudo 4: Adaptar e aplicar metodologia de avaliação da indexação pela recuperação da informação com usuários em repositórios institucionais; realizar comparação dos resultados obtidos entre repositórios institucionais a fim de elaborar um panorama da situação da recuperação por assuntos; recomendar elementos de política de indexação no Repositório Institucional da Unesp;

Estudo 5: Elaboração de orientações à indexação de teses e dissertações para autores e catalogadores com uso do Tesouro Unesp realizarem o controle de vocabulário;

Estudo 6: Gerar o mapeamento das palavras-chave com o Tesouro Unesp, por meio da compatibilização sintática, considerando-se os processos de igualdades e similaridades entre os termos;

Estudo 7: Identificação de um conjunto de elementos que permitam orientar ou reorientar as políticas de indexação empregadas no vocabulário do Repositório Institucional Unesp e, como benefícios indiretos, a serem verificados a médio prazo, o incremento na representatividade do referido vocabulário pela perspectiva de sua comunidade usuária.

Estudos 8 e 9: Extração pela Análise de Transações de Logs (TLA) da linguagem natural empregada pelos usuários para realizar buscas em três fases sequenciais: coleta, preparação e análise de dados.

A análise das contribuições de cada estudo, realizada em 3 reuniões conjuntas do Grupo de Estudo com a gestora e a Equipe Executiva, foi determinante para a elaboração de proposta de sistematização e aplicação das metodologias em três eixos abaixo descritos:

EIXO 1 - Estudo diagnóstico da política de indexação no repositório: terá início pelo *estudo diagnóstico da política de indexação no repositório* (Estudo 3) com a finalidade de obter as informações necessárias sobre requisitos de organização usuários e recursos financeiros, bem como elementos e variáveis de organização e representação da informação. Para completar o estudo diagnóstico indica-se a *análise de metadados de assunto* (Estudo 5) dos documentos que entram no repositório pela modalidade de autoarquivamento de modo a assegurar uma padronização da indexação e definir diretrizes para elaboração de orientações aos autores. Ainda no estudo diagnóstico, recomenda-se o *estudo de avaliação da indexação pela abordagem de recuperação por assuntos com usuários* (Estudo 4) para análise comparada da linguagem natural com a linguagem controlada do Tesouro Unesp que permita obter resultados quanto à correção, especificidade e exaustividade no sistema de busca e no Tesouro Unesp.

EIXO 2 - Elaboração da política de indexação do repositório: Os três estudos do Eixo 1 fornecerão resultados para a elaboração da Política de Indexação do Repositório, fundamental para o autoarquivamento dos autores e validação dos metadados pelos bibliotecários. A Política de indexação, dessa forma, precisará levar em consideração duas etapas de indexação: da análise de assunto para atribuição de palavras-chave e da representação das palavras-chave por vocabulário controlado. Para a primeira etapa recomenda-se a aplicação das metodologias de

avaliação dos estudos de análise de atribuição de palavras-chaves (Estudo 1) e observação de padrões e estratégias utilizadas por autores durante a indexação para atribuição de palavras-chave no autoarquivamento (Estudo 2) para adequação das orientações na política de indexação.

EIXO 3 - Elaboração e atualização de vocabulário controlado: O Tesouro Unesp é o vocabulário controlado a ser utilizado para a etapa de representação das palavras-chave por autores e bibliotecários e, por isso, deverá ser continuamente atualizado para que atenda aos critérios de qualidade da indexação, correção, especificidade e exaustividade. Nesse sentido, os estudos que visam a atualização de vocabulários controlados são vitais para manter a adesão ao uso de controle de vocabulários durante a indexação e a validação da indexação na modalidade de autoarquivamento. O estudo de mapeamento e compatibilização de palavras-chaves presentes nos metadados (Estudo 6) com o Tesouro Unesp é realizado por meio de processos de igualdades e similaridades da compatibilização sintática que poderá aumentar a quantidade de relações com termos autorizados, em especial a relação de equivalência. A análise de variações terminológicas em nível sintático, semântico e pragmático (Estudo 7) que ocorrem no Tesouro Unesp é um estudo que poderá auxiliar na elaboração de estratégias de tratamento terminológico para incluir contínuas atualizações. Por outro lado, a linguagem natural utilizada por usuários do Repositório durante a estratégia de busca poderá ser estudada pela análise de logs de buscas dos usuários (Estudos 8 e 9) para atualização contínua do Tesouro Unesp com a inserção de novos termos que serão selecionados a partir de resultados de procedimento específico de análise de logs.

6 CONSIDERAÇÕES FINAIS

| 18

A proposta de sistematização e aplicação das metodologias de avaliação do controle de vocabulário em repositórios consiste, fundamentalmente, no desenvolvimento de uma política de indexação cuja elaboração se realiza com base nos 3 eixos: Diagnóstico, Elaboração da Política de indexação e Atualização do Vocabulário Controlado.

Com essa proposta, a gestora e a Equipe Executiva avaliaram a sistematização das metodologias e, em função, das rotinas de trabalho com o Repositório Institucional Unesp e do engajamento do Grupo de Estudos, definiram que inicialmente será realizado, a curto prazo, o estudo de mapeamento e compatibilização de palavras-chaves presentes nos metadados (Estudo 6) tendo em vista ser viável a execução e, principalmente, que a atividade de autoarquivamento de publicações e trabalhos acadêmicos por autores pesquisadores da universidade tem demandado um vocabulário cada vez mais especializado em cada domínio de especialidade. Tal demanda é cada vez maior, o que acarreta a geração de novas palavras-chave sem controle de vocabulário. Além disso, a lista de palavras-chave disponibilizada no Repositório Institucional da Unesp apresenta problemas de controle de vocabulário que dificultam seu uso por usuários do repositório. A partir do estudo de mapeamento e compatibilização de palavras-chave presentes nos metadados, novos termos poderão ser avaliados pela Comissão Permanente do Tesouro Unesp. A urgência desse estudo tem respaldo no fato de que existe um vocabulário controlado de uso dos bibliotecários que se encontra disponível para usuários e autores, cuja utilização tem sido recomendada nos tutoriais de autoarquivamento, o que aumenta o interesse em atribuir termos controlados tanto na indexação como na recuperação.

Essa decisão, embora seja importante e justificada para o momento de falta de controle de vocabulário do repositório, não exclui a necessidade de realizar o diagnóstico e a política de indexação a ser realizada posteriormente. A equipe executiva é reduzida e conta com apenas dois bibliotecários e um profissional da área de Ciência da Computação e, por isso, o Grupo de Estudos decidiu continuar a desenvolver os estudos até a elaboração final da política de indexação com apoio da Comissão Permanente do Tesouro Unesp.

A proposta de controle de vocabulário e uso de vocabulários controlados em repositórios universitários administrados por bibliotecas exige que estudos de avaliação sejam realizados para sistematizar métodos, procedimentos, recursos e técnicas adequados e exequíveis para o contexto do repositório. A estrutura, o funcionamento e a gestão de um repositório apresentam diferenças significativas em relação a outros sistemas de recuperação da informação ao permitir a interação dos autores no autoarquivamento, a atribuição de assuntos pelos autores, o armazenamento de diferentes tipologias documentais e de recursos de informação e, mais notadamente, da possibilidade de manter um sistema híbrido de representação da informação com palavras-chave da linguagem natural e de vocabulários controlados. Esse contexto necessita da elaboração de uma política de indexação adequada para repositórios que considere todos os atores e fatores em prol do controle de vocabulário e da riqueza terminológica de especialidade da linguagem natural.

REFERÊNCIAS

AMERICAN NATIONAL STANDARDS INSTITUTE/NATIONAL INFORMATION STANDARDS ORGANIZATION. Z39.19-2005. **Guidelines for the construction, format, and management of monolingual controlled vocabularies**. Bethesda, Maryland: NISO Press, 2005. Disponível em: <http://www.niso.org/standards/resources/Z39-19-2005.pdf> . Acesso em: 06 de março de 2022.

ASSUMPÇÃO, Fabrício Silva *et al.* A conversão de registros na implantação de repositórios institucionais: o caso do Repositório Institucional UNESP. *In: SEMINÁRIO NACIONAL DE BIBLIOTECAS UNIVERSITÁRIAS*, 18, 2014, Belo Horizonte. **Anais [...]** Belo Horizonte: UFMG, 2014. p. 1-16. Disponível em: <http://repositorio.Unesp.br/handle/11449/123645> Acesso em: 06 de março de 2022.

| 19

BARRUONUEVO ALMUZARA, Leticia *et al.* A study of authority control in Spanish university repositories. **Knowledge Organization**, v.39, n.2, p. 95-103, 2012. Disponível em: http://www.ergon-verlag.de/isko_ko/downloads/ko_39_2012_2_e.pdf . Acesso em: 06 de março de 2022.

HJØRLAND, Birger. What is knowledge organization (KO)? **Knowledge Organization**, v.35, n.2/3, p.86-101, 2008. Disponível em: <https://bit.ly/3a8YgCm>. Acesso em: 30 de maio de 2022.

FUJITA, Mariângela Spotti Lopes; TOLARE, Jéssica Beatriz. Vocabulários controlados na representação e recuperação da informação em repositórios brasileiros. **Informação & Informação (Online)**, v.24, p. 93 - 125, 2019. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/view/37985>. Acesso em: 30 de maio de 2022.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 25964-1:2011 **Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 1: Thesauri for information retrieval**. Geneva: International Organization for Standardization, 2011.

INTERNATIONAL ORGANIZATION FOR STANDARDIZATION. ISO 25964-1:2011 **Information and documentation -- Thesauri and interoperability with other vocabularies -- Part 2: Interoperability with other vocabularies**. Geneva: International Organization for Standardization, 2013.

LANCASTER, Frederick Wilfrid. **Indexação e resumos: teoria e prática**. 2.ed.rev.atual. Trad. de Antonio Agenor de Briquet de Lemos. Brasília: Briquet de Lemos/Livros, 2004. 452p. (Título original: Indexing and abstracting in theory and practice)

LANCASTER, Frederick Wilfrid. **El control del vocabulario en la recuperación de información**. 2.ed. rev. Trad. de Alejandro de la Cueva Martín. València: Universitat de València, 2002. (Título original: Vocabulary control for information retrieval; Educació. Materials, 12)

LYNCH, Clifford A. Institutional repositories: essential infrastructure for scholarship in the digital age. **Association of Research Libraries**, Washington, DC., n.226, p. 1-7, fev. 2003. Disponível em: <https://bit.ly/3NKVBx9>. Acesso em: 06 de março de 2022.

MAIA, Graziela Zambão Abdian. Pesquisa etnográfica e estudo de caso. *In*: MACHADO, Lourdes Maria Machado; MAIA, Graziela Zambão Abdian; LABEGALINI, Andréia Cristina Fregate Baraldi. **Pesquisa em educação: passo a passo**. Marília: Edições M3T, 2007. p.83-94.

MOREIRO GONZÁLEZ, José Antonio. **El contenido de los documentos textuales: su análisis y representación mediante el language natural**. Gijón: Trea, 2004.

PANUTO, Jéssica Cristina **A abordagem do controle de vocabulário nos repositórios institucionais e sua importância para a Arquivologia**. Marília: Faculdade de Filosofia e Ciências, 2021. 63p. (Relatório final de pesquisa IC-CNPq)

SAYÃO, Luís Fernando. Repositórios digitais confiáveis: conceitos, tecnologias e padrões. *In*: FÓRUM DE CIÊNCIA E TECNOLOGIA: REPOSITÓRIOS CONFIÁVEIS DE DOCUMENTOS ARQUIVÍSTICOS DIGITAIS, 2011. **Tópico temático** [...]. Campinas: Unicamp, 2011.

TARTAROTTI, Roberta Cristina. Dal'Evedove. **Avaliação do processo de indexação de assuntos em repositórios institucionais pela abordagem da recuperação da informação**. 2019. 370p. Tese (Doutorado em Ciência da Informação) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista “Júlio de Mesquita Filho” (Unesp), Marília, 2019. Disponível em: <https://bit.ly/3wYVrU>. Acesso em: 06 de março de 2022.

UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO". **Portaria Unesp nº 88, de 28 de fevereiro de 2013**. Dispõe sobre a criação do Grupo Gestor da Política do Repositório Institucional UNESP (GRI-UNESP). São Paulo: Unesp, 2006. Disponível em: <https://bit.ly/38EAfCQ>. Acesso em: 22 de março de 2022.

WEBOMETRICS. **Ranking web of repositories**. Disponível em: <https://docs.python.org/pt-br/3/library/json.html>. Acesso em: 06 de março de 2022.

Objetivo: obter a visão sobre os procedimentos à respeito das atividades do Repositório, bem como sobre a necessidade e importância do controle de vocabulário e uso de vocabulários controlados

Método de elaboração:

- Análise da literatura e da documentação sobre o Repositório Institucional Unesp (RI- Unesp);
- Análise das respostas de questionário sobre política de indexação preenchido pela Equipe Executiva do RI- Unesp;

A- Questões da análise de literatura e da documentação sobre o Repositório Institucional Unesp:

1- “No tocante à política de indexação já estabelecida para o catálogo online Athena, a gestora do repositório da Unesp acredita que pode ser utilizada futuramente também para o repositório institucional.” (TARTAROTTI, 2019, p.162)

Comentar sobre o aproveitamento da validação de assuntos realizada em dissertações e teses por bibliotecários para o catálogo Athena;

2- “Para sua implementação, foram definidas quatro metas que refletiam o objetivo do Repositório (ASSUMPÇÃO; SILVA; FERREIRA, BASTOS 2014, p. 4):

1. inclusão da produção científica institucional publicada no período de 2008 a 2012 e indexada na Web of Science;
2. inclusão da produção científica institucional publicada em periódicos da SciELO;
3. inclusão da produção científica institucional publicada no período de 1976 a 2007 e indexada na Web of Science;
4. inclusão da produção científica institucional indexada na Scopus.”

2.1 Após 2014 foram fixadas outras metas?

2.2 Quantas metas foram alcançadas até a atualidade?

3 - “Em agosto de 2021, o repositório dispõe de 171337 registros dos quais 56% são artigos (95392), 17% são dissertações de mestrado (29389), 9% são teses de doutorado (15738), 5% são trabalhos de conclusão de curso (9350) e outros materiais como trabalhos apresentados em evento (8902), resumo (6579), resenha (1984), editoriais (754), cartas (643), capítulos de livro (549), livros (483), patentes (393), podcasts (274), erratas (254), notas (209), teses de livre-docência (167) boletim (78), revista (74), jornal (34), relatório (26), data paper (13), biografia (7), objeto educacional (6), plano de gestão de dados (5), partitura (2), regulamento (1) e vídeo (1).” (PANUTO, 2021, p.53)

3.1 Esses quantitativos correspondem à produção científica dos docentes ou parte dela?

3.2 Teses e dissertações estão todas no repositório?

3.3 Porque o tesauro Unesp não é citado nos tutoriais para atribuição de palavras-chave?

B – Questões elaboradas a partir das respostas ao questionário de política de indexação

O “Questionário do Subgrupo de Trabalho da Rede Sudeste: política de indexação em repositórios” foi respondido por Flávia Maria Bastos, gestora do Repositório Institucional Unesp.

4- A resposta de que 36 profissionais da Unesp são dedicados a tarefas de indexação e catalogação significa que o RI- Unesp realiza indexação de quais tipos de documentos?

5- Explique a função de cada um dos softwares utilizados no tratamento da informação:

Duplicate Checker:

Oxygen:

Adobe:

MarcEdit:

Libre office:

- 6- Qual o tipo de padronização de metadados é utilizada para definição de campos obrigatórios, repetitivos e de descrição?
- 7- Considerando-se que os procedimentos práticos do processo de indexação de assuntos estão cobertos por algum manual da instituição, significa que realizam indexação? Para quais tipos documentais? Descreva o processo de indexação realizado e cite o manual que o contém.
- 8- o Tesouro Unesp foi citado como auxílio automático ou semiautomático para facilitar o processo de indexação, porém, ele não tem nenhum auxílio semiautomático ou automático incorporado ao RI-Unesp. Isso é possível?
- 9- Várias respostas indicam que seguem um manual de política de indexação ou um manual de serviços, mas não é citado, por quê?
- 10- É permitido o uso de mais de uma linguagem de indexação, indicadas na resposta sobre a descrição de linguagens de indexação: utilizam 4 linguagens? Para todos os tipos documentais?
- 11- Comente a última resposta: “A universidade possui um grupo de bibliotecários que estuda a linguagem Unesp e a indexação que, entre suas atividades, realiza estudos sobre estes procedimentos, porém este grupo e estas atividades não estão vinculados a equipe do Repositório Institucional”
- 12- Existe algum planejamento para que isso aconteça no futuro?