



24° ENANCIB
Encontro Nacional de Pesquisa em Ciência da Informação
Perspectivas Contemporâneas na Ciência da Informação
• Vitória - ES • Ancib • PPGCI/UFES



XXIV ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – XXIV ENANCIB

ISSN 2177-3688

GT 2 – Organização e Representação do Conhecimento

ANÁLISE DO METADADO ASSUNTO NO REPOSITÓRIO INSTITUCIONAL DA UNIVERSIDADE FEDERAL DE ALAGOAS: UM ESTUDO DE CASO DOS CURSOS DE ENFERMAGEM E SERVIÇO SOCIAL

ANALYSIS OF THE SUBJECT METADATA IN THE INSTITUTIONAL REPOSITORY OF THE FEDERAL UNIVERSITY OF ALAGOAS: A CASE STUDY OF NURSING AND SOCIAL WORK COURSES

Valter dos Santos Andrade – Universidade Federal de Alagoas (UFAL)

Robéria de Lourdes Vasconcelos Andrade – Universidade Federal de Alagoas (UFAL)

Modalidade: Trabalho Completo

Resumo: A representação da informação destaca-se como parte fundamental na tarefa de garantir a recuperação de um item informacional, principalmente quando se trata do meio digital como os repositórios institucionais. A pesquisa tem como objetivo analisar a representação da informação no Repositório Institucional da Universidade Federal de Alagoas, no que concerne ao metadado assunto. A análise foi sobre o metadado assunto presente nas monografias depositadas através do autoarquivamento pelos alunos dos cursos de Enfermagem e Serviço Social. Trata-se de uma pesquisa exploratória com característica descritiva e com abordagem qualitativa. Foram mapeadas 160 monografias, sendo 84 do curso de Enfermagem e 76 do curso de Serviço Social. As monografias de Enfermagem totalizaram 252 metadados de assunto, nos quais 201 apresentaram conformidade com o catálogo de assunto do Sistema de Biblioteca da universidade e 29 em linguagem natural, 15 sem padronização, 3 com erro de grafia e 4 com erro de digitação; as monografias de Serviço Social totalizaram 210, sendo 84 metadados autorizados em conformidade com o catálogo, e 86 em linguagem natural, 23 sem padronização, 2 com erros de grafia e 15 com erros de digitação. Percebe-se que o uso autorizado de ferramentas como tesouros, bem como a orientação do profissional bibliotecário, contribuem para a padronização e a qualidade dos metadados de assunto, proporcionando uma maior eficácia e eficiência na busca. Conclui enfatizando a importância para o repositório adotar uma política de indexação, disponibilizando meios para auxiliar os usuários na utilização de ferramentas como tesouros e índices que possam guiá-los nas escolhas dos termos, assim melhorando a qualidade de descrição dos conteúdos documentais, visando à precisão no processo de busca e recuperação da informação.

Palavras-chave: repositório; autoarquivamento; representação da informação; indexação.

Abstract: The representation of information stands out as a fundamental part of the task of ensuring the recovery of an information item, especially when it comes to digital media such as institutional repositories. The research aims to analyze the representation of information in the Institutional Repository of the Federal University of Alagoas, with regard to subject metadata. The analysis was on the subject metadata present in monographs deposited through self-archiving by students of the Nursing and Social Work courses. This is an exploratory research with descriptive characteristics and a qualitative approach. A total of 160 monographs were mapped, 84 from the Nursing course and 76 from the Social Work course. The Nursing monographs had a total of 252 subject metadata, of which 201 were in accordance with the subject catalog of the university's Library System and 29 in natural language, 15 without standardization, 3 with spelling errors and 4 with typing errors; the Social Work monographs had a total of 210, of which 84 metadata were authorized in accordance with the catalog and 86 in natural language, 23 without standardization, 2 with errors spelling errors and 15 with typing errors. It is clear that the authorized use of tools such as thesaurus, as well as guidance from a professional librarian, contribute to the standardization and quality of subject metadata, providing greater effectiveness and efficiency in searches. The article concludes by emphasizing the importance of the repository adopting an indexing policy, providing means to assist users in using tools such as thesaurus and indexes that can guide them in choosing terms, thus improving the quality of description of documentary content, aiming at precision in the process of searching and retrieving information.

Keywords: repository; self-archiving; representation of information; indexing.

1 INTRODUÇÃO

Os Repositórios Institucionais (RIs), oriundos do movimento da Ciência Aberta, foram concebidos com o intuito de disponibilizarem as produções científicas em acesso aberto, sem barreira ou entraves. Compartilhando deste mesmo entendimento, uma das preocupações principais na Ciência da Informação (CI) consiste em tornar a informação acessível e recuperável para aqueles que dela necessitam. Leite (2009, p. 21) define RI como “um serviço de informação científica – em ambiente digital e interoperável – dedicado ao gerenciamento da produção intelectual de uma instituição”. Diante disso, surgem novos desafios para as universidades e seus RIs: produzir, organizar, preservar e principalmente disseminar os seus conteúdos informacionais.

Os RIs são ferramentas adequadas para a gestão do conhecimento científico, pois potencializam a condução de processos que maximizam a criação, o comportamento, a disseminação e o uso do conhecimento científico, bem como agilizam os processos de comunicação científica.

Para sua completude, o RI precisa prover a interoperabilidade e o autoarquivamento. Conforme Villalobos e Gomes (2015, p. 4), a interoperabilidade consiste na “comunicação entre diferentes sistemas de informação”, disponibilizando os metadados, os quais identificam, descrevem e localizam os recursos informacionais armazenados, possibilitam a interação do sistema com o usuário, permitem que tenha acesso aos dados, podendo

compartilhá-los com outros sistemas de diferentes *hardwares* e *softwares*. Já o autoarquivamento vem se tornando uma das etapas operacionais fundamentais, porque possibilita ao autor realizar o depósito do seu documento diretamente na base de dados. Esse processo agiliza a visibilidade e a disponibilidade da publicação.

No entanto, esse procedimento pode interferir na qualidade e, principalmente, na confiabilidade do sistema. Freitas (2019) ressalta que o autoarquivamento pode comprometer a qualidade dos descritores de assunto, dificultando o processo de busca da informação.

De fato, essa preocupação faz sentido, visto que o autor do trabalho depositado pode não ter utilizado padrões metodológicos para escolha do metadado assunto, além de uso inadequado de alguns termos. É preciso que os RIs garantam a descrição de documentos com metadados padronizados.

Nesse aspecto, faz-se necessária uma análise criteriosa da forma da representação temática da informação nos ambientes digitais, principalmente nos RIs. A representação temática consiste na análise do conteúdo de um documento, ou seja, o tema ou assunto abordado por ele, objetivando a sua disseminação. Para Albuquerque, Gaudêncio e Santos (2019, p. 19), a representação da informação “consiste em uma tarefa indispensável para a organização, a recuperação, o acesso e o uso dos registros do conhecimento. Para tanto, são criados processos, produtos e instrumentos que corroborem para essa finalidade”. Dentre os instrumentos estão os tesouros, listas de cabeçalhos de assuntos e glossários, os quais podem ser aperfeiçoados, atualizados e implantados nas estruturas dos sistemas de recuperação da informação na *web*. Esses instrumentos auxiliam na padronização do metadado assunto.

O Repositório Institucional da Universidade Federal de Alagoas (Riufal) ainda não implantou a sua política de indexação. Até o ano de 2020, orientava ao uso do metadado assunto extraído do resumo e da ficha catalográfica. No caso das dissertações, teses e as monografias dos cursos da área de Saúde, já orientavam, obrigatoriamente, o uso de ficha. Para os demais cursos, o uso da ficha catalográfica não era obrigatório, portanto a recomendação era o uso do metadado assunto extraído do resumo. Só a partir de 2023, o Riufal passa a adotar a obrigatoriedade da ficha catalográfica para todos os Trabalhos de Conclusão de Curso, seja da pós-graduação ou da graduação. Os cursos de Enfermagem e Serviço Social foram escolhidos por serem os primeiros a fazerem essa modalidade de depósito.

A hipótese deste estudo é de que o cadastramento do metadado “assunto” por parte dos autores pode não seguir uma padronização e apresentar ruídos, vindo a comprometer a recuperação dos dados relevantes no momento da pesquisa na Base de Dados do Riufal.

Nesse sentido, a presente pesquisa levanta o seguinte questionamento: os metadados de assunto cadastrados pelos usuários no Riufal, no período anterior ao ano de 2022, no momento do autoarquivamento de seus trabalhos, foram qualificados e validados tendo como referencial os cabeçalhos autorizados utilizados pelo Catálogo de Assunto do Sistema Integrado de Biblioteca da Universidade Federal de Alagoas (Sibi/Ufal)? Diante do exposto, o objetivo geral da presente pesquisa é analisar a representação da informação no Riufal, no que concerne ao metadado assunto indexado na referida base de dados, tendo como objetivos específicos: a) Coletar os TCC dos cursos de Enfermagem e Serviço Social cadastrados através do autoarquivamento; b) Avaliar os metadados de assunto cadastrados nos TCC coletados; c) Comparar os metadados utilizados com cabeçalhos autorizados do catálogo de assunto do Sibi/Ufal; d) Identificar se, dentre os metadados cadastrados, existe algum que apresente ruídos que possam interferir no processo da busca e recuperação da informação no sistema.

2 O AUTOARQUIVAMENTO E A REPRESENTAÇÃO DA INFORMAÇÃO

Com a “via verde”, o Movimento de Acesso Aberto estabelece a estratégia do autoarquivamento, o qual tem como objetivo incentivar os autores a disponibilizarem diretamente as suas produções científicas em repositórios de acesso aberto. Conforme argumenta Assis (2013, p. 217), “apesar de não ser ainda uma prática comum nos RIs de todo o mundo, o arquivamento não demanda esforço de trabalho excessivo para o autor”. O processo é bastante prático, cabendo ao autor preencher os metadados solicitados na base de dados do repositório, haja vista que os RIs incorporam, em suas políticas informacionais, a obrigatoriedade do depósito em suas bases de dados, as publicações científicas acadêmicas produzidas nos âmbitos dessas instituições, aparecendo as teses e dissertações e os trabalhos de conclusão de cursos como os tipos mais comuns de documentos depositados. O autoarquivamento é relevante, visto que aproxima a interação entre o autor, a instituição e um universo de usuários que utiliza o ambiente *web* para produção e compartilhamento de pesquisas científicas, bem como agiliza o processo de visibilidade, devido à inclusão imediata de todos os conteúdos informacionais no repositório.

Entretanto, apesar das diversas possibilidades atualmente para o acesso à informação, o fator agilidade no processo nem sempre é sinônimo de qualidade e relevância das informações recuperadas. A possibilidade do próprio autor autoarquivar a sua produção dá a liberdade de escolher os descritores que nem sempre estão abalizados por uma fonte de referência autorizada, como catálogos de assuntos ou tesouros, além das inconsistências de uso de plural ou singular, termos muito genéricos e até mesmo erro de grafia. Fatos esses que interferem na busca, possibilitando uma não exatidão no momento da pesquisa e interferindo diretamente na qualidade e na confiabilidade da base. Santos (2017, p. 8) ressalta que “as informações disponíveis nos RI são descritas com palavras provenientes da linguagem natural sem o auxílio de um instrumento de representação da informação”, ou seja, não seguem uma padronização determinada.

Os usuários representam conteúdos com termos que usam e/ou acreditam que entendem, sendo assim alguns termos podem fazer sentido apenas a quem os elencou e nem sempre para a comunidade ou grupo de pessoas que utilizam o sistema. Com a liberdade da atribuição dos descritores um mesmo termo pode ter diferentes significados para aqueles que atribuíram as etiquetas/descriptores para fins de representação do conteúdo (Santos, 2017, p. 8).

Fica evidente que o uso da linguagem natural, geralmente, apresenta a visão “subjetiva” dos usuários produtores dos conteúdos informacionais, muitas vezes não se atentando para a linguagem do sistema e, com isso, descredibilizando-o, visto que a utilização de termos não padronizados, ou não controlados em uma determinada área do conhecimento, poderá dificultar a recuperação entre os pares.

Freitas (2019, p. 31) ressalta que o “uso de linguagem natural pode trazer problemas no que se refere à recuperação do material armazenado”, pois isso pode ocorrer devido à falta de padronização e à liberdade de escolha de termos de indexação, o que torna a pesquisa muito abrangente, cuja busca no sistema de informação necessitará de um esforço maior do usuário na escolha dos termos, prejudicando a busca e o acesso ao item pesquisado.

Nesse aspecto, o uso excessivo da linguagem natural nos metadados assunto nos RIs passa a ser um grande desafio para a representação, o acesso e a recuperação da informação.

A representação da informação destaca-se como parte fundamental na tarefa de garantir a recuperação de um item informacional, principalmente quando se trata do meio digital. Arakaki e Castro (2020) argumentam que a intensa produção e a disseminação de dados através do meio digital proporcionaram inúmeros benefícios para a sociedade no que

diz respeito ao acesso e à democratização da informação, à medida que se ampliam as formas de disponibilização e de disseminação dos recursos informacionais.

Os meios digitais proporcionam a facilidade para as pessoas publicarem seus conteúdos informacionais em uma multiplicidade de formatos, ocasionando ambiguidades e heterogeneidade. Esses fatores dificultam o acesso e a recuperação das informações em ambientes digitais. A representação da informação possibilita identificar os aspectos fundamentais de um documento, os quais direcionam o usuário para uma tomada de decisão assertiva sobre a pertinência ou não do seu conteúdo. Marcondes (2001, p. 61) enfatiza: “de nada adianta a informação existir, se quem dela necessita não sabe da sua existência, ou se ela não puder ser encontrada”.

A informação precisa estar bem representada para ser recuperada pelos usuários que buscam para satisfazerem as suas necessidades informacionais. A representação da informação apresenta-se de duas formas: representação descritiva e representação temática. A primeira refere-se às características específicas do documento, ou seja, a descrição bibliográfica, e permite a individualização dele; já a segunda destaca-se pela representação dos assuntos dos documentos, com o objetivo de tornar mais fácil a sua recuperação. Para Ferreira e Albuquerque (2013), a representação da informação busca atender prontamente a todos os usuários, disponibilizando-os instrumentos que possam descrever as informações da forma mais clara possível, garantindo eficiência e eficácia na realização das pesquisas.

Nessa visão, a indexação torna-se ponto essencial para a representação da informação, parte determinante para o sucesso ou insucesso em uma busca. Daí a importância na escolha dos metadados de assuntos que identificam um determinado item, pois eles possibilitam a sua mensuração quanto à pertinência e à relevância para uma determinada pesquisa, principalmente nos ambientes digitais, onde se encontra um grande volume de informação.

3 A INDEXAÇÃO E A IMPORTÂNCIA DE UMA POLÍTICA DE INDEXAÇÃO NOS REPOSITÓRIOS

O acesso aos conteúdos informacionais é uma das preocupações constantes da CI. Dispor de mecanismos eficientes para a localização e a recuperação dos registros documentais é fator preponderante para o sucesso de uma pesquisa. Dentre esses mecanismos, encontra-se a indexação, ponto determinante para o sucesso ou insucesso em uma busca.

Freitas (2019) define a indexação como um processo fundamental para a identificação dos conteúdos de um documento. Para ela, a indexação caracteriza-se em:

Um processo essencial para as atividades informativas, ao passo que visa identificar o conteúdo do documento e descrevê-lo mediante termos provenientes de uma linguagem de indexação que o represente, isto é, que consiga favorecer a busca e o entendimento do usuário sem a necessidade de leitura do seu conteúdo. Seu objetivo é, portanto, atender as necessidades dos usuários na busca por informação (Freitas, 2019, p. 26).

Para Santos (2017, p. 2), “a indexação consiste na descrição do conteúdo de um documento, de forma concisa e condensada, por meios de emprego de termos – também denominados como palavras-chave ou descritores”. A indexação tem como finalidade, caracterizar de forma concisa os conteúdos informacionais de um item, sendo ainda uma ponte de acesso entre o usuário e o item.

Lancaster (2004, p. 9) considera a indexação em duas etapas: análise conceitual e tradução. A análise conceitual refere-se à extração de conceitos que representem o conteúdo de um documento, “implica decidir do que trata um documento – isto é, qual o seu assunto”; enquanto a tradução consiste na representação dos documentos em termos oriundos de instrumentos de indexação, “envolve a conversão da análise conceitual de um documento num determinado conjunto de termos de indexação”.

Segundo Dias e Naves (2007, p. 6), esses instrumentos de indexação são chamados de linguagens de indexação ou linguagens documentárias e “depois de traduzidos, para uma linguagem de indexação, passam a ser chamados de descritores de assunto, rótulos, cabeçalhos de assunto, termos de indexação ou enunciados”.

No que concerne à tradução, Lancaster (2004) a tipifica em indexação por extração, ou indexação derivada, e indexação por atribuição. No primeiro caso, as palavras ou expressões empregadas pelo autor no item documental são selecionadas pelo indexador para representar seu conteúdo. Ao utilizar-se desse recurso, o indexador faz uso do que é chamado de uma linguagem natural, pois ele seleciona palavras encontradas no título, nas palavras-chave ou no próprio conteúdo informacional do documento. No segundo caso, indexação por atribuição, como o próprio nome diz, são atribuídos descritores ao documento, utilizando-se uma linguagem oriunda de fonte externa, não necessariamente o próprio documento. Sendo assim, o indexador utilizar-se de vocabulários controlados, tesouros, entre outros, fazendo uso de uma linguagem controlada e/ou artificial, ou seja, buscando uma padronização.

Apesar de no ambiente *web* a linguagem natural ocorrer com maior intensidade, devido à facilidade e à liberdade das escolhas dos termos, faz-se necessária a busca de padronização para melhor aproveitamento entre a linguagem dos usuários e a linguagem dos sistemas, e isso é possível através do uso dos vocabulários controlados.

Embora os RIs tenham avançado no que concerne à política de funcionamento, definindo as estruturas de organização de suas comunidades e coleções, a maioria dos RIs aqui no Brasil ainda não implantaram a sua política de indexação. Galvino e Pereira (2021), em um estudo realizado com 11 RIs na região Nordeste, constataram que todos os repositórios analisados não tinham ainda uma política de indexação definida e implantada. Outro ponto relevante, ainda segundo os autores, apesar da existência de uma vasta literatura a respeito de indexação e política de indexação em unidades de informação, a temática política de indexação em repositórios ainda é escassa, “ao fazer uma busca rápida nos sistemas de busca, pouco se encontra sobre a temática ainda” (Galvino; Pereira, 2021, p. 243). Os autores ressaltam que “algo de grande relevância para um bom funcionamento do RI no sentido da recuperação da informação vinha sendo esquecido ou negligenciado, a Política de Indexação (PI) de metadados para os RI (p. 243).

Nesse aspecto, definir uma política de indexação é de fundamental importância para os RIs, pois além de definir as formas das entradas do metadado assunto, contribui também para a padronização desses metadados. Conforme Oliveira e Santos (2022), uma boa indexação deve estar atrelada a uma política pré-estabelecida pela equipe do RI, contemplando diretrizes e normas padronizadas, definindo os critérios de exaustividade, especificidade, precisão, revocação e linguagem de indexação.

4 PROCESSO METODOLÓGICO

Elemento essencial no processo investigativo, o processo metodológico é uma tarefa preponderante para que se possa atingir os objetivos da pesquisa, pois é através dele que serão apresentados, de forma descritiva, os métodos utilizados pelo pesquisador durante o estudo.

Quanto ao alcance dos objetivos propostos, deu-se por meio de um estudo de caráter exploratório e descritivo na Base de Dados do Riufal, com a análise do metadado “assunto”, especificamente os assuntos cadastrados através do autoarquivamento pelos alunos concluintes dos cursos de Enfermagem e Serviço Social da Ufal.

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

No que se refere à modalidade, a pesquisa fez uma abordagem qualitativa e quantitativa a partir de informações coletadas na Base de Dados do Riufal, no tocante ao metadado “assunto”.

Para coleta e para sistematização dos dados, foram divididos nas seguintes fases: Primeira fase: busca direta na base de dados do Riufal, na qual foram mapeadas as monografias depositadas através do autoarquivamento, no período de 2020 a 2022. Através da pesquisa nas comunidades “curso de Enfermagem e Serviço Social”, dentro da subcomunidades “TCC”, obteve-se uma amostra de 160 monografias. Foram recuperados 475 metadados de assunto, sendo 260 de Enfermagem e 215 de Serviço Social. Segunda fase: os dados foram transcritos em planilhas do *Excel* e *Word*, ambos programas da *Microsoft*, possibilitando assim a organização, tabulação e análise dos dados. Terceira fase: os dados foram tratados e depurados, eliminando o metadado “assunto” repetido, perfazendo um total de 462 cabeçalhos de assunto, 252 nas monografias de Enfermagem e 210 nas monografias de Serviço Social. Na quarta fase, foram feitas as análises dos metadados “assunto” (cabeçalhos de assunto), com os cabeçalhos de assunto autorizados pelo catálogo de assunto do Sibi/Ufal.

Foram definidas as seguintes variáveis: metadado de assunto autorizado, metadado de assunto em linguagem natural, metadado de assunto sem padronização, metadado de assunto com erro de grafia e metadado de assunto com erro de digitação.

Quadro 1 – Variáveis da pesquisa

VARIÁVEL	DESCRIÇÃO
Metadado de assunto autorizado	Utilizado pelo SIBI/UFAL
Metadado de assunto linguagem natural	Usado pelo discente (extração do resumo)
Metadado de assunto sem padronização	Registrado com várias variantes
Metadado de assunto com erro de grafia	Ortografia
Metadado de assunto com erro de digitação	Digitação

Fonte: Dados da pesquisa (2023).

Essas variáveis foram escolhidas pelo fato de serem os ruídos mais comuns apresentados nos metadados “assunto” presentes nos RIs. Também foi estipulado o índice de compatibilidade ou concordância dos metadados cadastrados, o qual possibilitará a mensuração de pertinência dos metadados com os utilizados pelo Riufal, ficando definidos os

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

percentuais de: a) 80 % a 100% - EXCELENTE, b) 70% a 79% - ALTO; c) 50% a 69% - MÉDIO; Abaixo de 50% - BAIXO.

É a partir desses valores que podemos definir se a linguagem empregada pelo usuário está compatível com a linguagem do sistema do Riufal, além de determinar o grau de compatibilidade.

Para a coleta dos dados, foi realizada uma busca diretamente na base de dados do Riufal, onde foi feita a seleção de 160 (cento e sessenta) monografias autoarquivadas pelos discentes concluintes dos cursos, sendo 84 do curso de Enfermagem e 76 do curso de Serviço Social, as quais foram coletadas e analisados 462 metadados de assunto. Conforme a Tabela 1.

Tabela 1 – Metadado de assunto cadastrado no RIUFAL pelos cursos de Enfermagem e Serviço Social

METADADOS “ASSUNTO” DOS TCCs DE ENFERMAGEM E SERVIÇO SOCIAL	TOTAL DE TCCs	SIBI/UFAL %
Metadado de assunto autorizado	285	61,7%
Metadado de assunto em linguagem natural	115	25%
Metadado de assunto sem padronização	38	8,3%
Metadado de assunto com erro de grafia	5	1%
Metadado de assunto com erro de digitação	19	4%
Total	462	100%

Fonte: Dados da pesquisa, 2023.

5 APRESENTAÇÃO E DISCUSSÃO DOS RESULTADOS

Das monografias do curso de Enfermagem, foram coletados 252 metadado assunto, enquanto das de Serviço Social foram 210. Os metadados assunto coletados foram comparados com os cabeçalhos de assuntos autorizados pelos catálogos de assuntos do Sibi/Ufal.

O curso de Enfermagem apresentou 201 (duzentos e um) metadados assuntos autorizados, 29 (vinte e nove) de linguagem natural incluídos pelos autores, 15 (quinze) metadados assunto sem padronização, metadados assunto com erro de grafia 3 (três) e metadado assunto com erro de digitação 4 (quatro), conforme Tabela 2.

XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024

Tabela 2 – Metadados de assunto cadastrados no Riufal (Enfermagem)

METADADO “ASSUNTO” DOS TCCs DE ENFERMAGEM	TOTAL DE TCCs	SIBI/UFAL %
Metadado de assunto autorizado	201	79,8%
Metadado de assunto em linguagem natural	29	11,5%
Metadado de assunto sem padronização	15	6%
Metadado de assunto com erro de grafia	3	1,20%
Metadado de assunto com erro de digitação	4	1,5%
Total	252	100%

Fonte: Dados da pesquisa, 2023.

Percebe-se que há muitos cabeçalhos autorizados, correspondendo a aproximadamente 79,8% de concordância, enquanto 11,5% cabeçalhos provenientes da linguagem natural, 6% sem padronização, 1,20% com erro de grafia e 1,5% com erro de digitação. Tais valores demonstram, assim, o alto grau de compatibilidade dos metadados assunto extraídos dos TCCs de Enfermagem com os utilizados no catálogo de assunto do Sibi/Ufal.

Enquanto Serviço Social apresentou 210, sendo 84 (oitenta e quatro) metadado de assunto autorizado, 86 (oitenta e seis) metadado assunto em linguagem natural, 23 (vinte e três) metadado assunto sem padronização, 2 (dois) metadado com erro de grafia, e 15 (quinze) metadado assunto com erro de digitação. Conforme Tabela 3, seguem dados descritos com valores e percentuais.

Tabela 3 – Metadados de assunto cadastrados no RIUFAL (Serviço Social)

METADADO “ASSUNTO” DOS TCCs DE SERVIÇO SOCIAL	TOTAL DE TCCs	SIBI/UFAL %
Metadado de assunto autorizado	84	40%
Metadado de assunto em linguagem natural	86	41%
Metadado de assunto sem padronização	23	11%
Metadado de assunto com erro de grafia	2	1%
Metadado de assunto com erro de digitação	15	7%
Total	210	100%

Fonte: Dados da pesquisa, 2023.

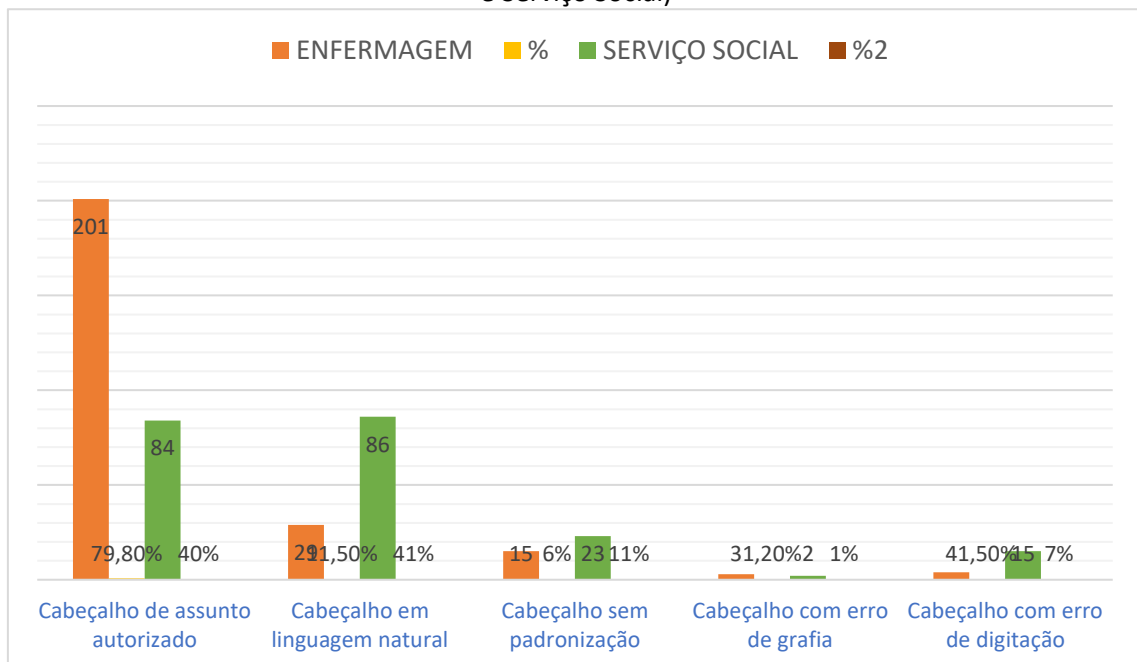
Quanto aos metadados de assunto extraídos dos TCCs de Serviço Social, percebe-se que há poucos cabeçalhos autorizados, correspondendo a, aproximadamente, 40% de

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

concordância, enquanto 41% cabeçalhos provenientes da linguagem natural, 11% sem padronização, com erro de grafia 1% e com erro de digitação 7%. Portanto, esses trabalhos apresentaram um baixo índice de concordância com os metadados assunto do catálogo de assunto do Sibi/Ufal.

Comparando as tabelas 2 e 3, percebe-se a discrepância entre os dois cursos. Enquanto os TCCs de Enfermagem apresentam um alto índice de compatibilidade entre os metadados assunto extraídos deles com metadados assunto pertencentes ao catálogo de assunto do Sibi/Ufal, o mesmo não ocorre com os TCCs de Serviço Social.

Gráfico 3 – Comparativo dos metadados de assunto cadastrados (discentes de Enfermagem e Serviço Social)



Fonte: Dados da pesquisa, 2023.

Algo relevante para destacar é o bom grau de concordância nas monografias de Enfermagem, porque se acredita que isso se dá pelo fato de os alunos de Enfermagem utilizarem com frequência a Base de Dados de Descritores de Saúde (DeCS), para coletar termos da área da Saúde, bem como o uso dos cabeçalhos de assuntos constantes na ficha catalográfica, as quais são elaboradas pelos bibliotecários do Sibi/Ufal, sendo item obrigatório em suas monografias.

Os alunos do curso de Enfermagem, ao fazerem o autoarquivamento, foram orientados a incluírem, no momento do preenchimento dos metadados de assuntos, além das palavras-chave do seu resumo, os cabeçalhos de assuntos da ficha, o mesmo não aconteceu com o curso de Serviço Social, que só adotou o uso da ficha catalográfica a partir do ano de

2023. Nos metadados de assuntos autorizados apresentados nos TCCs de Serviço Social dos 210 metadados de assuntos analisados, 84 foram compatíveis com os cabeçalhos do catálogo de assunto do SIBI/UFAL, ou seja, apresentou um número muito baixo de concordância.

Nesta análise, percebe-se que o índice de discordância em relação aos termos autorizados fica muito baixo quando o usuário do sistema tem algum auxílio que os oriente nas escolhas dos termos. Apesar de o autoarquivamento possibilitar a livre escolha do usuário para inserção dos cabeçalhos, a opção por termos ou cabeçalhos não padronizados deste grupo é muito pequena, entretanto o grupo que não utilizou a ficha apresenta o maior uso de termos em linguagem natural. Outro ponto a ser analisado é o fato de apesar de serem termos autorizados, às vezes o autor peca na quantidade de termos utilizada, não explorando no trabalho outros cabeçalhos ou termos que o trabalho apresenta, limitando-se apenas das palavras-chave do resumo, que geralmente tem um número limitado.

6 CONSIDERAÇÕES FINAIS

O Riufal se constitui como um instrumento fundamental para preservação, disseminação e divulgação das produções acadêmicas e científicas, pois cumpre o seu papel enquadrado na filosofia da Ciência Aberta. Entretanto, para que seus conteúdos sejam disseminados, é preciso que sejam descritos e representados de forma que os usuários do sistema possam localizá-los em suas buscas. Nesse sentido, a representação da informação é fator fundamental para o processo de organização, acesso e recuperação. Outro ponto preponderante é a utilização do autoarquivamento como um processo contínuo, possibilitando a interação do autor, seu conteúdo informacional e o sistema de informação.

Diante disso, a pesquisa buscou analisar a representação da informação nos metadados “assunto” cadastrados no Riufal através do autoarquivamento dos TCCs, realizados pelos alunos dos cursos de Enfermagem e Serviço Social. Percebeu-se que o uso da linguagem natural foi mais predominante no curso de Serviço Social, o qual apresentou 41%, enquanto o curso de Enfermagem 11,5%, demonstrando que, quem tem como base o uso de ferramentas auxiliares como catálogos autorizados, tesouros especializados e orientação do profissional bibliotecário para a elaboração dos metadados “assunto”, não utiliza tanto a linguagem natural.

No que se refere à falta de padronização, 6% foram detectados no curso de Enfermagem, aparecendo novamente o curso de Serviço Social com o maior percentual, 11%;

no tocante ao erro de grafia e digitação, foi apresentado um baixo índice, apenas 1% para erro de grafia e 4% para erro de digitação. Ao comparar os metadados utilizados com cabeçalhos autorizados do catálogo de assunto Sibi/Ufal, dos 462 metadados de assunto analisados, 285 apresentaram concordância, perfazendo o percentual de 61,7%, o que equivale a um índice mediano.

Quando analisados os dados isolados por cada curso, o índice de concordância apresenta uma grande variação, enquanto o curso de Enfermagem, dos 252 metadados de assunto, 201 apresentaram concordância com o catálogo Sibi/Ufal, ou seja, 79,8%, um alto índice; enquanto o curso de Serviço Social, dos 210 metadados “assunto”, apenas 84 apresentaram concordância, 40%, ou seja, um índice baixo.

Com base na visão geral dos metadados assunto analisados, foi possível identificar alguns ruídos que podem interferir no processo da busca e recuperação da informação no sistema. Foram constatados vários ruídos, desde variações de uso de palavras no plural e singular, erros de digitação e grafia, bem como variações de forma de apresentação do mesmo assunto, caso dos metadados de assunto “própolis vermelha e própolis vermelho”, no qual o primeiro apresenta a recuperação de 57 documentos, enquanto o segundo, a recuperação de sete (7) documentos, ocasionando, assim, uma grande falha que pode prejudicar a busca de conteúdos informacionais, visto que, dependendo da escolha do metadado “assunto”, poder-se-á deixar de fora determinados documentos na pesquisa recuperada.

Ficou muito evidente, nesta pesquisa, que a utilização da linguagem natural em grande proporção apresenta uma maior ocorrência, com variações de cabeçalhos com falhas de padronização, erro de digitação e, principalmente, não concordância com o catálogo do Sibi/Ufal. Na outra vertente, monografias que utilizaram ferramentas como tesouros e os cabeçalhos de assunto, presentes na ficha catalográfica, apresentaram um excelente grau de concordância e confiabilidade, contribuindo para a qualidade do sistema.

Ainda assim, não há dúvidas de que o uso da linguagem natural é importante também para a aproximação do autor com o conteúdo produzido, contudo a padronização se torna algo essencial para a qualidade e para credibilidade de qualquer sistema de informação. Ressalta-se que a falta de uma política de indexação por parte do Riufal também contribui para a falta de padronização dos metadados de assunto, pois foi constatado que o catálogo de assunto do Sibi/Ufal ora usava termos retirados do Catálogo da BN, ora da base DeCS, apresentando o mesmo metadado “assunto” registrado com diferentes formas de entrada.

Recomenda-se a implantação de uma política de indexação no Riufal, criação de mecanismos que contribuam para a padronização, qualidade e pertinência dos metadados “assunto”, tais como índice ou catálogos de assunto, elaboração de manuais e guias para inserção de metadados, tutoriais para treinamentos dos usuários e a implantação definitiva do autoarquivamento para a toda comunidade.

Por fim, que este estudo estimule pesquisas futuras que contribuam para o desenvolvimento do Riufal e, principalmente, dos temas “representação da informação e indexação nos repositórios”, de suma importância para a recuperação da informação em ambientes digitais.

REFERÊNCIAS

ALBUQUERQUE, M. E. C.; GAUDÊNCIO, S. M.; SANTOS, R. F. Reflexões teóricas em representação da informação. *In*: ALBUQUERQUE, M. E. C.; MARTINS, G. K.; MOTA, D. A. R. (org.). **Organização e representação da informação e do conhecimento**: intersecções teórico-sociais. João Pessoa: Editora UFPB, 2019.

ARAKAKI, A. C. S.; CASTRO, F. F. Aspectos da representação da informação na curadoria digital. **Encontros Bibli**: Revista Eletrônica de Biblioteconomia e Ciência da Informação, Florianópolis, v. 25, p. 01-21, 2020.

ASSIS, T. B. Análise das políticas de autoarquivamento nos repositórios institucionais brasileiros e portugueses. **InCID**: Revista de Ciência da Informação e Documentação, Ribeirão Preto, v. 4, n. 2, Ed. Esp., p. 212-227, jul./dez. 2013.

DIAS, E. W.; NAVES, M. M. **Análise de assunto**: teoria e prática. Brasília: Thesaurus, 2007.

FERREIRA, L. C.; ALBUQUERQUE, M. E. C. A representação da informação para a organização do acervo em bibliotecas universitárias. *In*: ALBUQUERQUE, M. E. C. *et al.* (org.). **Representação da informação**: um universo multifacetado. João Pessoa: Editora UFPB, 2013.

FREITAS, M. P. **Autoarquivamento e representação de assunto**: estudo analítico de teses e dissertações do Repositório Institucional da UFSCar. 2019. 89 f. Dissertação (Mestrado em Ciência da informação) – Centro de Educação e Ciências Humanas, Programa de Pós-Graduação em Ciência da informação, Universidade Federal de São Carlos, São Carlos, 2019.

GALVINO, C. C. T.; PEREIRA, G. M. Práticas de indexação nos repositórios institucionais do Nordeste do Brasil. **Páginas A&B**: Arquivos e Bibliotecas, Braga, S.3 n. esp. CONFOA, p. 243-246, 2021.

**XXIV Encontro Nacional de Pesquisa em Ciência da Informação – XXIV ENANCIB
Vitória-ES – 04 a 08 de novembro de 2024**

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

LEITE, F. C. L. **Como gerenciar e ampliar a visibilidade da informação científica brasileira: repositórios institucionais de acesso aberto**. Brasília: IBICT, 2009.

MARCONDES, C. H. Representação e economia da informação. **Ciência da Informação**, Brasília, v. 30, n. 1, p. 61-70, jan./abr. 2001.

SANTOS, R. F. Indexação em repositórios digitais: uma abordagem sobre o metadado assunto da Biblioteca Digital de Monografias da UFRN. **Revista Informação na Sociedade Contemporânea**, Natal, n. Especial, p. 11-22, 2017.

OLIVEIRA, T. N.; SANTOS, R. F. Política e diretrizes de indexação em repositórios institucionais das universidades federais brasileiras. **Revista Informação na Sociedade Contemporânea**, Natal, v. 6, p. 1-24, 2022. Disponível em: <https://periodicos.ufrn.br/informacao/article/view/29444>. Acesso em: 20 abr. 2024.

VILLALOBOS, A. P. O.; GOMES, F. A. Padronização de metadados na representação da informação em repositórios institucionais de universidades federais brasileiras. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 26., 2015, João Pessoa. **Anais [...]**. João Pessoa: ANCIB, 2015. Disponível em: <http://hdl.handle.net/20.500.11959/brapci/187093>. Acesso em: 17 abr. 2023.