

GT 2 – Organização e Representação do Conhecimento

ISSN 2177-3688

AVALIAÇÃO DA INDEXAÇÃO SEMI-AUTOMÁTICA PARA O ACESSO À INFORMAÇÃO CIENTÍFICA EM ITALIANO

EVALUATION OF SEMI-AUTOMATIC INDEXING FOR ACCESS TO SCIENTIFIC INFORMATION IN ITALIAN

Cecília Abrahão Nascimento de Santi - Universidade Federal do Pará (UFPA)

Franciele Marques Redigolo - Universidade Estadual Paulista (Unesp)

Stela Andrade Vasconcelos - Universidade Federal do Pará (UFPA)

Luciana Di Paula Andrade da Fonseca - Universidade Estadual Paulista (Unesp)

Modalidade: Trabalho Completo

Resumo: A avaliação da indexação, torna-se fundamental para que a produção contida nos sistemas de informação seja aproveitada pela comunidade usuária, uma vez que permite verificar a qualidade da representação e a recuperação da informação em relação às solicitações de pesquisa dos usuários. Nesse sentido, as indexações semi-automáticas e automatizadas podem auxiliar o gerenciamento das produções de forma otimizada, facilitando o trabalho para a instituição provedora. Esta pesquisa tem como objetivo geral avaliar a indexação semi-automática mediante análise comparativa entre os descritores obtidos pelo Sistema de Indexação Semi-Automático (SISA) e as palavras-chave dos autores dos artigos. Trata-se de uma pesquisa de natureza básica, caracteriza-se como do tipo descritiva, com abordagem quali-quantitativa, por meio do método de estudo de caso. Os resultados obtidos se referem à análise da indexação com o SISA, assim, foi possível observar nos quatro artigos analisados que a quantidade de descritores sugeridos por artigo é alta, porém com o uso das etiquetas, as representações pelo SISA coincidiram com as realizadas pelos autores e as médias dos índices de consistência apresentaram de normal para baixa. Conclui-se que a linguagem natural usada pelos autores e a linguagem controlada do SISA obtiveram um bom desempenho na indexação dos artigos, assim, o vocabulário do SISA necessita estar em constante atualização para que a representação possa ser fiel ao conteúdo do texto.

Palavras-chave: acesso à informação; indexação semi-automatizada; avaliação da indexação; Sistema de Indexação Semiautomático (SISA).

Abstract: The evaluation of indexing is essential if the production contained in information systems is to be used by the user community, since it allows the quality of the representation and retrieval of information to be verified in relation to users' search requests. In this sense, semi-automatic and automated indexing can help manage production in an optimized way, making work easier for the provider institution. The general aim of this research is to evaluate semi-automatic indexing by comparing the descriptors obtained by the Semi-Automatic Indexing System (SISA) with the keywords used by the authors of the articles. This is basic research, characterized as descriptive, with a qualitative and quantitative approach, using the case study method. The results obtained refer to the analysis of indexing with SISA, so it was possible to observe in the four articles analyzed that the number of descriptors suggested per article is high, but with the use of labels, the representations by SISA coincided with those made by the authors and the averages of the consistency indices ranged from normal to low. It can be concluded that the natural language used by the authors and SISA's

controlled language performed well in indexing the articles, so SISA's vocabulary needs to be constantly updated so that the representation can be faithful to the content of the text.

Keywords: access to information; semiautomatic indexing; indexing evaluation; Semiautomatic Indexing System (SISA).

1 INTRODUÇÃO

A avaliação da indexação desempenha um papel fundamental na área da Ciência da Informação, de modo que permite estabelecer a qualidade e a eficácia dos sistemas de indexação utilizados na organização e recuperação da informação. Nesse sentido, por meio da avaliação da indexação, é possível medir e analisar o desempenho dos sistemas de indexação em relação a diferentes critérios.

Dessa maneira, Gil Leiva (1997) identifica a ampla gama de terminologias utilizadas para descrever conceitos semelhantes acerca da indexação semi-automática e/ou automática, em que ambas são tratadas na literatura de forma relacionada, porém é importante compreender suas definições. Dessa forma, destaca-se que a indexação semi-automática é “realizada por programas computacionais e finalizada pelo indexador humano” (NASCIMENTO; MARTINS; ALBUQUERQUE, 2023, p. 8), em que sua vantagem consiste na otimização das análises linguísticas e estatísticas nos textos, identificando os termos potencialmente representativos dos documentos e ser assistida por um indexador humano.

Por outro lado, Gil Leiva (1997) define os sistemas de indexação automática como aplicações que dispensam a necessidade de validação, ou seja, os termos sugeridos são registrados imediatamente como descritores do documento correspondente. Ademais, ainda segundo o autor, a indexação semi-automática realiza análises automáticas de documentos, contudo, com os termos de indicação avaliados e autorizados por humanos, bem como pode ser visto no Sistema para Indexação Semi-Automático (SISA) que foi introduzido em meados de 1990 (GIL LEIVA, 1999) e pode ser utilizado nesse processo. Dito isso, o SISA é um sistema utilizado para indexação semi-automática e/ou automática de artigos científicos, tal como para avaliação da indexação.

Assim, a indexação automática e/ou semi-automática pode contribuir para a organização, representação e recuperação, objetivando a preservação digital em repositórios institucionais. A partir desse contexto, o problema de pesquisa visa discutir sobre a avaliação da indexação para a representação da informação de textos científicos realizada por meio do Sistema de Indexação Semi-Automático (SISA). Desse modo, o trabalho tem como objetivo

avaliar a indexação semi-automática mediante análise comparativa entre os descritores obtidos pelo software SISA e as palavras-chave dos autores dos artigos indexados em Repositórios Institucionais Universitários no idioma italiano da área de Ciências Sociais Aplicadas, do curso de Turismo.

Justifica-se o uso do SISA por ser o único que permite indexação automática por atribuição, além de ser utilizado de forma unânime em trabalhos relacionados à indexação automática e/ou semi-automática.

2 REPRESENTAÇÃO DA INFORMAÇÃO PARA PRESERVAÇÃO E O ACESSO À INFORMAÇÃO CIENTÍFICA: INDEXAÇÃO SEMI-AUTOMÁTICA E A INDEXAÇÃO AUTOMÁTICA

Esta seção permite discutir como a avaliação da indexação automática e da indexação semi-automática pode contribuir para a representação e organização da informação, visando o acesso e a preservação e da informação científica, no contexto dos repositórios institucionais, como um fator favorável ao acesso livre e irrestrito à informação científica, uma vez que essa ferramenta possibilita maior visibilidade das produções com menor custo, aumenta relevância das instituições e os autores das produções que os adotam, assim, acarretando benefícios em âmbitos científicos e socioeconômicos.

Observa-se no âmbito internacional um aporte teórico consolidado acerca do tema de indexação semi-automática e/ou automática, assim, destacam-se os pesquisadores defensores dessa prática, como Rodríguez e González (1999), Gil Leiva (2017) e Golub (2019). Por outro lado, a busca em âmbito nacional, revelou que apesar de existirem trabalhos importantes sobre o assunto estudado como as produções de Robredo (1982), Rubi (2009), Narukawa (2011), a literatura brasileira sobre indexação semi-automática e/ou automática continua significativamente menor em comparação com a estrangeira, “o que, provavelmente, reflete o pouco desenvolvimento desta técnica no país” como afirma Vieira (1988, p. 43), evidenciando a necessidade de novos estudos a serem realizados em nível nacional acerca da automatização dos processos de tratamento dos conteúdos temáticos dos documentos.

A partir da literatura estudada, torna-se possível compreender a avaliação da indexação semi-automática e/ou automática para sistema de informações, como os repositórios institucionais, em que esse processo passa a ser crucial para os profissionais envolvidos na gestão e desenvolvimento desse sistema, uma vez que contribui com retorno

de informações valiosas de como encontra-se o acesso a informação ao seu conjunto de usuários. Sendo assim, a mesma pode contribuir para eficiência, precisão, abrangência, velocidade, escalabilidade, consistência, padronização, redução de recursos humanos necessários para executar essa função e maior flexibilidade para adaptação às necessidades específicas de um contexto.

Dessa maneira, considerando que os repositórios institucionais estão “constituídos por um conjunto de arquivos digitais representando produtos científicos e acadêmicos que podem ser acessados por usuários” (TEXIER, 2013, p. 4, tradução nossa), nota-se a grande capacidade do RIs de armazenamento e promoção a comunicação científica. Assim, o bom funcionamento do mesmo, é uma tarefa complexa, pois há um acúmulo das informações que o alimentam, evidenciando-se a importância de que procedimentos estratégicos para representação temática sejam delimitados.

Segundo Tartarotti, Dal'Evedove e Fujita (2020, p. 16) “[...] as primeiras produções científicas brasileiras a considerarem a indexação como uma importante atividade de representação e recuperação de informações no âmbito dos repositórios digitais datam de 2011 [...]”, sendo assim, configura-se como uma temática recente, o que justifica a produção sobre indexação semi-automática e/ou automática no repositório institucional ser menos expressiva na literatura brasileira.

Portanto, a aplicação desse tipo de tratamento temático da informação em repositórios institucionais, auxilia o gerenciamento das produções de forma otimizada, facilitando o trabalho para a instituição provedora, uma vez que a quantidade de artigos, teses, dissertações, livros ou outros tipos de documento não é uma problemática para realização de uma indexação consistente, resultando em uma preservação digital e ao combate do volume documental ligado a paradigmas digitais. Dito isso, no próximo parágrafo apresenta-se o percurso metodológico utilizado na pesquisa.

3 METODOLOGIA

A fim de analisar os descritores obtidos pela indexação semi-automática do Sistema de Indexação Semi-Automática a partir das palavras-chave dos autores, verifica-se que esta pesquisa é de natureza básica e caracteriza-se como do tipo descritiva, com abordagem qualitativa, por meio do método de estudo de caso, com uma análise comparativa. Fez-se inicialmente uma pesquisa bibliográfica sobre as temáticas de repositórios institucionais e

indexação semi-automática e/ou automática, em seguida, aplicou-se a indexação semi-automática em artigos científicos utilizando o Sistema de Indexação Semi-Automática (SISA). Para isso definiu-se alguns critérios de escolha dos Repositórios Institucionais selecionados para o estudo.

3.1 Repositórios Institucionais: critérios de seleção

Os Repositórios Institucionais foram determinados a partir dos critérios de seleção a seguir:

- Repositório Institucional Universitário;
- Que contenha artigos no idioma italiano da área de Ciências Sociais Aplicadas, curso de Turismo indexado em seu acervo.

Mediante isso, a busca deu-se a partir da lista de repositórios institucionais disponibilizada no Portal do Instituto Brasileiro de Informação em Ciência e Tecnologia (IBICT), no qual foi coletada a por meio do Portal Brasileiro de Acesso Aberto à Informação Científica (OASISBR). Nesta lista foi realizada uma busca de forma simples e direta, apenas verificando os nomes em ordem alfabética e, por conseguinte, as que possuíam artigos científicos sobre turismo no idioma italiano.

A partir disso, em âmbito nacional foram encontrados somente artigos científicos no idioma italiano e não acerca da área do Turismo, os quais foram o Repositório de Produção USP com o periódico Revista de Italianística, o Repositório Institucional da UFRJ - PANTHEON com a Revista Italiano UERJ.

Entretanto, na intenção de realizar uma busca mais específica a partir dos critérios estabelecidos na pesquisa, foi realizada uma pesquisa avançada no Oasisbr por meio do uso do descritor “turismo” com todos os campos selecionados, com limitação do idioma para o italiano e a seleção do tipo de documento “artigo”. Assim, o resultado da busca ficou reduzido, trazendo exclusivamente um repositório que atendesse aos critérios de seleção: Repositórios Científicos de Acesso Aberto de Portugal (RCAAP).

No RCAAP, devido a densidade de informações geradas pelo SISA, foram escolhidos cinco artigos científicos para a análise dos descritores gerados pela indexação semi-automática, a partir dos seguintes critérios: a) ser artigo científico; b) estar dentro da área de assunto

compatível com vocabulário controlado a ser incluído no SISA, explicado no parágrafo seguinte.

3.2 SISA

O SISA é um sistema semi-automático de indexação, desenvolvido por Gil Leiva (1997, 1999) que permite “adaptar sua configuração para aplicar a qualquer área, desde que possua uma linguagem documentária” (NARUKAWA; GIL LEIVA; FUJITA, 2009, p. 106). Desta forma, o sistema processa documentos em idiomas variados dependendo do vocabulário controlado utilizado e a lista de palavras vazias.

Deste modo, é possível observar em Silva e Corrêa (2020, p. 12) a descrição dos requisitos de entrada de dados do SISA e aplicados nesta pesquisa da seguinte forma:

- a) Lista alfabética de termos e descritores: arquivo texto contendo os termos autorizados e termos alternativos no idioma italiano, sendo para cada termo alternativo indicado o respectivo termo autorizado;
- b) Lista alfabética de termos e respectivos termos gerais: arquivo opcional contendo o termo autorizado e a relação de termo geral;
- c) Lista de palavras vazias em italiano, para fins de eliminação das palavras consideradas vazias (stopwords), como conectivos e artigos;
- d) Marcação das partes estruturais constituintes do documento a ser indexado;
- e) Todos os arquivos de entrada, incluindo os textos a serem indexados, devem estar no formato .txt.

Com base nos requisitos acima, foi possível determinar a escolha da área de conhecimento, mediante o vocabulário controlado, e, os artigos científicos para utilização do SISA, discutidos na seção seguinte.

3.2.1 Área do conhecimento dos artigos analisados: critérios de escolha

A área do conhecimento foi determinada a partir do vocabulário controlado disponível para indexação no SISA, específico do domínio do turismo no idioma italiano, o qual foi adotado para esta pesquisa.

3.2.2 Títulos dos artigos indexados pelo SISA

Devido ao resultado da busca seguindo os critérios de seleção da pesquisa apresentados na seção 3.1, foram selecionados quatro artigos científicos para a análise dos descritores gerados pela indexação semi-automática, a saber:

- Artigo 1 - Turismo religioso e sviluppo delle aree rurali.considerazioni sul caso della sardegna;
- Artigo 2 - Turismo lungo le aree costiere euromediterranee: dalla scoperta, al consumo, al progetto del paesaggio;
- Artigo 3 - Il turismo enogastronomico in italia: domanda, offerta e trend futuri;
- Artigo 4 - Il turismo culturale in italia tra tradizione e innovazione martinotti universita di sassari Fondazione mons. giovannino pinna (italia).

3.2.3 Critérios de aplicação no SISA e avaliação da indexação

Os critérios de aplicação seguiram as regras que o SISA impõe para indexação. Abaixo, o quadro que explica as definições e as aplicações do que foi reproduzido pela indexação semi-automática e pelas autoras, no momento da análise.

Quadro 1 — Definição das variáveis da indexação semi-automática no SISA

	Definições	Aplicação
Linguagem documentária	Uso da linguagem controlada de acordo com a área do conhecimento	Uso do vocabulário no idioma italiano disponível no SISA para definição dos descritores representativos do conteúdo dos documentos.
Lista de palavras vazias	Lista de palavras vazias	Uso da lista de palavras vazias para exclusão de termos que não são significativos para a representação.
Regras para a indexação no SISA	Etiquetas de marcação do texto utilizadas para a indexação no SISA	TI; RE; PC; EP; PP; OP; TT; TF; CO; RF. Nesta etapa, todos os textos foram marcados com estes filtros em formato txt.
Critérios de filtragem para análise	Descritores de indexação obtidos pelo SISA e palavras-chave dos autores	Comparação dos descritores indexados pelo SISA e as palavras-chave informadas em cada texto selecionado

Fonte: Elaborado pelas autoras (2023).

No quadro 1, acerca da indexação no SISA foi realizada com as seguintes regras de filtragem, a saber: TI (Título), RE (Resumo), PC (Palavras-chave), EP (Epígrafe), PP (Primeiro parágrafo), OP (Outros parágrafos), TT (Título-tabela), TF (Título-figura), CO (Conclusões) e RF (Referência). Após a análise dos textos, considerando os descritores indexados pelo SISA e as palavras-chave escolhidas pelas autoras, de forma a comparar os termos.

Além da comparação entre a indexação automática e a indexação feita pelos autores, buscou-se relatar os possíveis fatores intervenientes durante o processo de indexação, segundo Bandim e Corrêa (2019). Os autores também utilizaram o SISA para comparar os dois tipos de indexação supracitados, na área de Ciência da Informação. Tais fatores percebidos por eles foram: 1. uso de termos no singular e no plural; 2. dificuldade em atribuir termos compostos; 3. diferenças nas estruturas dos termos de indexação; 4. dificuldade em atribuir conceitos implícitos; 5. atribuição de termo específico a termo geral. Logo, esses cinco fatores também foram relacionados ao processo desta pesquisa, uma vez que o SISA pode demonstrar divergências ou convergências durante a indexação dos artigos em italiano e os artigos em português usados por Bandim e Corrêa (2019).

Para este trabalho foram coletadas todas as palavras-chave introduzidas pelos autores e comparadas com os termos de indexação atribuídos pelo SISA. Justifica-se o uso das palavras-chave na comparação entre os termos adotado pelo SISA, uma vez que são consideradas instrumentos de representação da informação contida em documentos, o que pode ser examinado como um processo de indexação intelectual realizado pelos autores.

Em razão disso, torna-se possível calcular o grau de semelhança na consistência dos termos adotados tanto pelo sistema como pelos autores. Desse modo, entende-se a consistência como “grau de semelhança na representação da informação documental de um documento por meio de termos de indexação selecionados por um ou vários indexadores, resultando em um índice de consistência” (GIL LEIVA; RUBI; FUJITA, 2008, p. 234).

Nesse sentido, para uma recuperação da informação adequada, faz-se necessário refletir quais termos são relevantes e quais não são no momento da indexação documental. Para analisar essa questão, utilizou-se a fórmula de consistência com a adaptação da fórmula de Hooper (1965) por Gil Leiva, Rubi e Fujita (2008) para calcular os índices de consistência rígido e o relaxado dos termos adotado nos artigos indexados pelo SISA e as palavras-chave dos autores.

Segundo Sousa *et al.* (2020, p. 10) o índice rígido ocorre “quando o assunto determinado coincide completamente atribuindo-se valor 1 e atribuiu-se 0 quando não há coincidência”, de modo que o índice relaxado “atribui o peso 1 quando um cabeçalho ou subcabeçalho de assunto de um documento coincide com o assunto de outro documento. Quando ocorre coincidência somente no cabeçalho ou subcabeçalho, considera-se a metade (0,5) e quando não há nenhuma coincidência o valor é 0” (SOUSA *et al.*, 2020, p. 10).

Ademais, destacam-se também para elaboração deste trabalho, os artigos de Gil Leiva, Rubi, Fujita (2008), Tartarotti, Dal’Evedove e Fujita (2017). Entretanto, adotou-se de acordo com Bandim e Correa (2019), a consistência entre as palavras-chave selecionadas pelos autores dos artigos e os termos escolhidas pelo SISA, de forma que pode ser mensurada pela seguinte fórmula:

$$C = \frac{Tco}{(A + B) - Tco}$$

Da qual se lê: C (consistência), Tco (termos comuns entre a seleção dos autores e do software), A (indexação feita pela variável A), B (indexação feita pela variável B). Foi considerado como indexação A o número de palavras-chave escolhidas e os termos inseridos nos títulos pelos autores, e a indexação B o número de palavras-chave indexadas pelo SISA.

4 ANÁLISE E DISCUSSÃO SOBRE A INDEXAÇÃO SEMI-AUTOMÁTICA NOS ARTIGOS CIENTÍFICOS UTILIZANDO O SISA EM COMPARAÇÃO COM AS PALAVRAS-CHAVE E OS TERMOS INSERIDOS NOS TÍTULOS PELOS AUTORES

Nesta seção, apresenta-se a análise dos quatro artigos científicos mediante os critérios definidos para utilização do SISA compilados na metodologia. Sendo que os artigos científicos foram analisados na ordem descrita no item 3.2.2. Além disso, alguns fatores foram identificados durante o processo de indexação pelo SISA, de forma que não se impediu de a indexação ser feita, porém alguns pontos valem serem discutidos, em comparação ao processo que Bandim e Corrêa (2019) também identificaram, como mostra a Tabela 1:

Tabela 1 — Fatores intervenientes durante o processo de indexação

Fatores intervenientes	Termo das Palavras-chave	Termos do SISA
Termos no singular e no plural	x	x
Dificuldade em atribuir termos compostos	Percorsi religiosi	Percorsi
Diferença na estrutura de termos de indexação	x	x
Dificuldade em atribuir termos implícitos	-Valorizzazione aree rurali; -Aree costiere euromediterranee; -Paesaggi Protetti;	-Destinazioni -Turismo alternativo -Sviluppo turistico -Turismo sostenibile -Paesaggi costieri
Atribuição de termo específico a termo geral	-Italia; -Promozione di beni culturali; -Politiche di pianificazione e gestione	-Beni culturali; -Politiche del turismo

Fonte: Dados da pesquisa (2023) baseado em Bandim e Correa (2019).

Os fatores intervenientes foram: Termos no singular e no plural; Dificuldade em atribuir termos compostos; Diferença na estrutura de termos de indexação; Dificuldade em atribuir termos implícitos e Atribuição de termo específico a termo geral. Nesse caso, não se teve dados em relação aos fatores de termos no singular e no plural, pois os termos atribuídos pelos autores e os termos atribuídos pelo SISA foram diferentes. Bandim e Correa (2019) haviam identificados termos que não coincidiam nesses dois fatores, o que pode demonstrar uma melhora na indexação realizada pelo SISA, com menos variações identificadas entre os termos. O mesmo ocorre em relação à estrutura de termos de indexação, com nenhum termo variando com o uso de artigos ou preposição que liga os substantivos nos termos.

Por outro lado, foi percebido a dificuldade de atribuir termos compostos em relação a um artigo. Nas palavras-chave constava “Percorsi religiosi”, e o SISA atribuiu “Percorsi”. Porém, em outros artigos é possível identificar que o SISA consegue atribuir termos compostos, por exemplo, “Turismo culturale” e “Turismo balneare”, o que pode levar a inferir que o SISA indexa com termos compostos quando são mencionados com frequência no corpo do texto.

Ainda, houve o fator sobre a atribuição de termos específicos a termos gerais pelo SISA que Bandim e Corrêa (2019) identificaram. Nesse experimento foi diferente, o SISA atribuiu termos gerais como “Beni culturali”, quando nas palavras-chave constava “Promozione di beni culturali”, um termo mais específico. No lugar do termo “Politiche di pianificazione e gestione” escolhido pelo autor, o SISA indexou “Politiche del turismo”, trocando completamente os termos, mas ainda atribuindo um termo que melhor se relaciona ao assunto do artigo.

No que tange ao fator atribuir termos implícitos, o SISA atribuiu os termos “Turismo alternativo”, “Sviluppo turistico” e “Turismo sostenibile” para os conceitos referentes às áreas costeiras, “Aree costiere euromediterranee”, e a valorização de áreas rurais, “Valorizzazione aree rurali”. Percebe-se que o SISA simplificou a linguagem dos termos, tornando-os mais parte da linguagem controlada. Neste sentido, o primeiro artigo para comparação entre os descritores do SISA e as palavras-chaves dos autores, consta na tabela abaixo, a qual foi gerada a partir dos termos extraídos do título e das palavras-chave encontradas no texto.

Tabela 2 — Análise comparativa do Artigo 1- “Turismo religioso e sviluppo delle aree rurali. Considerazioni sul caso della Sardegna”

Título	Palavras-chave dos autores	Descritores do SISA	Termos comuns
1-Turismo religioso e sviluppo delle aree rurali. Considerazioni sul caso della Sardegna	1-Percorsi religiosi 2-Valorizzazione aree rurali 3-Territorio 4-Rappresentazioni	1-Percorsi 2-Turismo religioso 3-Destinazioni	1-Turismo religioso

Fonte: Dados da pesquisa (2023).

Na indexação do Artigo 1 acima, foram recuperadas um total de 9978 termos pelo SISA, todavia, por meio das etiquetas de marcação do texto e do índice de frequência do termo no texto, o SISA destacou três descritores importantes para a representação do artigo “Turismo religioso e sviluppo delle aree rurali. Considerazioni sul caso della Sardegna”. Em relação a comparação com as palavras-chave do autor, o SISA foi menos específico do que os descritores estabelecidos pelo autor do texto.

O SISA utiliza dois descritores semelhantes “percorsi” e “destinazioni” e genéricos, que podem prejudicar na representação e na recuperação deste artigo. Dado que, o autor usa da exaustividade nas palavras-chave, sendo quatro termos escolhidos e dois deles com mais especificidade do que os termos do SISA: “percorsi religiosi” e “valorizzazione aree rurali”. Porém, o autor também usa do termo “rappresentazioni”, o qual não contribui para a representação de assunto, pois é muito amplo. Deste modo, o segundo artigo, na Tabela 2, contém a comparação dos termos extraídos do SISA e os termos extraídos do título e das palavras-chave encontradas no texto.

Tabela 3 — Análise comparativa do Artigo 2 - “Turismo lungo le aree costiere euromediterranee: dalla scoperta, al consumo, al progetto del paesaggio”

Título	Palavras-chave dos autores	Descritores do SISA	Termos comuns
1-Turismo lungo le aree costiere euromediterranee: dalla scoperta, al consumo, al progetto del paesaggio	1-Turismo balneare 2-Aree costiere euromediterranee 3-Politiche di pianificazione e gestione 4-Paesaggi Protetti	1-Turismo balneare 2-Turismo alternativo 3-Sviluppo turistico 4-Turismo sostenibile 5-Politiche del turismo 6-Paesaggi costieri	1-Turismo balneare 2-Aree costiere euromediterranee

Fonte: Dados da pesquisa (2023).

Na indexação do Artigo 2 acima, foram recuperados um total de 7401 termos pelo SISA, porém, com base nas etiquetas de marcação do texto e do índice de frequência do termo no texto, o SISA destacou seis descritores relevantes para a representação do artigo “Turismo lungo le aree costiere euromediterranee: dalla scoperta, al consumo, al progetto del paesaggio”. Desta vez, o SISA aplicou os seis descritores com maior especificidade.

Já o autor escolheu quatro termos também condizentes com o seu trabalho. Nota-se, então, que os termos aplicados pelo SISA e pelo autor foram similares, uma vez que “politiche del turismo” e “politiche di pianificazione e gestione” podem remeter a um mesmo assunto. E o termo “turismo balneare” foi indexado tanto pelo sistema quanto pelo autor, apontando que a indexação semi-automática e a manual, neste caso, tiveram o mesmo desempenho. Na Tabela 4, constam os termos comparados do SISA com os termos selecionados pelos autores do artigo.

Tabela 4 — Análise comparativa do Artigo 3 - “Il turismo enogastronomico in Italia: domanda, offerta e trend futuri”

Título	Palavras-chave dos autores	Descritores do SISA	Termos comuns
1-Il turismo enogastronomico in Italia: domanda, offerta e trend futuri	Sem palavras-chave do autor.	1-Turisti 2-Turismo enogastronomico	1-Turismo enogastronomico

Fonte: Dados da pesquisa (2023).

Na indexação do Artigo 3 acima, foram recuperadas um total de 1453 termos pelo SISA, mas a partir das etiquetas de marcação do texto e do índice de frequência do termo no texto, o SISA destacou dois descritores importantes para a representação do artigo “Il turismo enogastronomico in Italia: domanda, offerta e trend futuri”.

O artigo analisado não possui palavras-chave o que dificultou a análise, no entanto, em comparação com o título, o SISA apresentou dois descritores representativos do documento “turisti” e “turismo enogastronomico”, ambos os termos estão presentes no título do artigo de forma a compreendermos que a representação do SISA está sendo fiel ao assunto

principal do artigo. E no último artigo analisado também foi feita a comparação entre a indexação semi-automática e a intelectual, constam os seus dados na Tabela 5:

Tabela 5 — Análise comparativa do Artigo 4 - “Il turismo culturale in Italia tra tradizione e innovazione”

Título	Palavras-chave dos autores	Descritores do SISA	Termos comuns
1-II turismo culturale in Italia tra tradizione e innovazione	1-Turismo culturale 2-Italia 3-Tradizione 4-Innovazione 5-Gestione 6-Valorizzazione 7-Promozione di beni culturali	1-Beni culturali 2-Turismo culturale	1-Turismo culturale

Fonte: Dados da pesquisa (2023).

Na indexação do Artigo 4 acima, o SISA recuperou um total de 2758 termos, no entanto com a aplicação das etiquetas de marcação do texto e do índice de frequência do termo no texto, o SISA destacou dois descritores importantes para a representação do artigo “Il turismo culturale in Italia tra tradizione e innovazione”.

Embora os dois descritores que o SISA marcou referem-se ao assunto do artigo, não especificam que é em âmbito italiano. Uma vez que duas das palavras-chave do autor condizem com os dois descritores obtidos no SISA “Beni culturali” e “turismo culturale”, as outras cinco palavras-chave do autor são muito genéricas, o que não é indicado para a representação, dado que pode levar a um ruído grande na busca do documento: “Italia” “tradizione” “innovazione” “gestione” “valorizzazione”.

Diante disso, mediante a análise geral dos artigos, foi possível visualizar a discrepância, o índice de consistência rígido, que variou entre 11% a 50%, com uma média de 24,30% de consistência entre as indexações. Já o índice de consistência relaxado variou entre 10% e 50%, com uma média de 21, 25% (Tabela 6) entre os termos atribuídos pelo SISA e as palavras-chave adotadas pelos autores, como demonstra a Tabela 6:

Tabela 6 — Índices de consistência rígido e relaxado dos quatro artigos.

Artigos	Índice de consistência rígido (%)	Índice de consistência relaxado (%)
Artigo 1	14	14
Artigo 2	22,22	10
Artigo 3	50	50
Artigo 4	11	11
Média	24,30	21,25

Fonte: Dados da pesquisa (2023).

Dessa maneira, nota-se que houve um mínimo de 0 e um máximo de 7 palavras-chave, o que equivale a uma média de 3,75 termos atribuídos em cada artigo pelos autores dos trabalhos. Por outro lado, o SISA obteve um mínimo de 2 e um máximo de 6 termos atribuídos, com uma média de 3,25 termos designados. Portanto, observa-se como o número de palavras-chave e o número de termos selecionados pelo SISA influenciam diretamente o valor resultante de ambas as consistências. Com base nos resultados encontrados neste estudo, na seção seguinte, trata-se das considerações finais.

5 CONSIDERAÇÕES FINAIS

Com base nos resultados encontrados, é possível inferir que a quantidade de descritores sugeridos por artigo é alta, porém, com o uso das etiquetas, as representações pelo SISA coincidiram com as realizadas pelos autores, de modo que confirma capacidade desse software de cruzar as palavras dos artigos com as existentes no vocabulário controlado para o idioma estudado, além de evidenciar como a indexação semi-automática e sua avaliação podem contribuir com otimização do acesso à informação científica em italiano, tornando-o mais eficiente, preciso e abrangente para a comunidade usuária dos sistemas de informação que utilizam em suas pesquisas e estudos o língua italiana, assim como em qualquer outra idioma.

Nesse sentido, verifica-se que as médias dos índices ficaram bem próximas, de modo que se obteve 21,25% no índice de consistência relaxado e 24,30% no índice de consistência rígido, o que demonstra, respectivamente, uma consistência de normal para baixa.

Portanto, constatou-se que a linguagem natural utilizada pelos autores obteve um bom desempenho, assim como a controlada empregada pelo SISA. Dito isso, considera-se que o vocabulário utilizado pelo SISA precisa estar em constante atualização e ser abrangente, para que a representação possa ser fiel ao conteúdo do texto, assim, espera-se que este estudo sirva como ponto de partida para novas investigações e contribua na discussão sobre a avaliação da indexação para a representação de artigos científicos.

REFERÊNCIAS

BANDIM, M. A. S.; CORREA, R. F. Indexação automática por atribuição de artigos científicos em português da área de Ciência da Informação. **Transinformação**, Campinas, v. 31, 2019.

GIL LEIVA, I. **La automatización de la indización, Propuesta teórico-metodológica**: aplicación al área de Biblioteconomía y Documentación. Tese (Doutorado) Universidade de Múrcia, Múrcia, 1997.

GIL LEIVA, I. **La automatización de la indización de documentos**. Gijón: Trea, 1999

GIL-LEIVA, I. La indización de artículos científicos con el sistema de indización automática SISA comparada con la indización en las Bases de datos Agrícola, WoS y SCOPUS. *In*: CONGRESSO ISKO ESPANHA, 13., 2017. **Atas [...]**, Universidad de Coimbra, Portugal, 2017.

GIL LEIVA, I.; RUBI, M. P.; FUJITA, M. S. L. Consistência na indexação em bibliotecas universitárias brasileiras. **Transinformação**, Campinas, v. 20, n. 3, p. 233–253, set. 2008.

GOLUB, K. Automatic Subject Indexing of Text. **Knowledge Organization**, [s. l.], v. 46, n. 2, p. 104-121, 2019.

HOOVER, R. S. **Indexer consistency tests: origin, measurement, results, and utilization**. Bethesda: IBM Coporation, 1965.

NARUKAWA, C. M. **Estudo de vocabulário controlado na indexação automática**: aplicação no processo de indexação do Sistema de Indización Semiautomática (SISA). 2011. 222f. Dissertação (Mestrado) – Faculdade de Filosofia e Ciências, Universidade Estadual Paulista, Marília, 2011.

NARUKAWA, C. M.; LEIVA, I. G.; FUJITA, M. S. L. Indexação automatizada de artigos de periódicos científicos: análise da aplicação do software SISA com uso da terminologia DeCS na área de Odontologia. **Informação & Sociedade: Estudos**, [s. l.], v. 19, n. 2, 23 ago. 2009.

NASCIMENTO, G. D.; MARTINS, G. K.; ALBUQUERQUE, M. E. B. C. Automação da indexação: evidências e tendências da produção científica indexada na brapci. **Encontros Bibli**, Florianópolis, SC, v. 28, 2023.

ROBREDO, J. A indexação automática de textos: o presente já entrou no futuro. **Estudos Avançados em Biblioteconomia e Ciência da Informação**, [s. l.], v. 1, n. 1, 1982.

RODRÍGUEZ, E. M. M. GONZÁLEZ, J. A. M. Lenguaje natural e indización automatizada. **Ciencias de la Información**, Havana, vol. 30, n. 3, 1999.

RUBI, M. P. Os princípios da política da indexação na análise de assunto para catalogação: especificidade, exaustividade, revocação e precisão na perspectiva dos catalogadores e usuários. *In*: FUJITA, Mariângela Spotti Lopes (Org.). **A indexação de livros**: a percepção de catalogadores e usuários de bibliotecas universitárias. Um estudo de observação do contexto sociocognitivo com protocolos verbais. São Paulo: Cultura Acadêmica, 2009.

SILVA, S. R. B.; CORREA, R. F. Sistemas de Indexação automática por atribuição: uma análise comparativa. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, SC, v. 25, p. 01-25, 2020.

XXIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB
Aracaju-SE – 06 a 10 de novembro de 2023

SOUSA, L. de; BARROS, H. B.; REDIGOLO, M.; FUJITA, S. L. Avaliação extrínseca mediante a interconsistência nos catálogos das bibliotecas universitárias da amazônia brasileira: UFPA, UFOPA, IFPA E UFRA. **Encontros Bibli**: revista eletrônica de biblioteconomia e ciência da informação, Florianópolis, SC, v. 25, p. 01–26, 2020.

TARTAROTTI, R. C. D. Representação temática em repositórios institucionais de bibliotecas universitárias. **RDBCI**: Revista Digital de Biblioteconomia e Ciência da Informação, [s. l.], v. 18, 2020.

TARTAROTTI, R. C. D. E.; DAL' EVEDOVE, P. R.; FUJITA, M. S. L. Avaliação da consistência da indexação em bibliotecas universitárias federais da região nordeste do Brasil. **Anales de Documentación**, v. 20, n. 1, 2017.

TARTAROTTI, R. C. D.; DAL'EVEDOVE, P. R.; FUJITA, M. S. L. Indexação em repositórios digitais: uma incursão na produção científica brasileira de Ciência da Informação. **Scire**: representación y organización del conocimiento, [S. l.], v. 26, n. 2, p. 13-18, 2020.

TEXIER, J. Los repositorios institucionales y las bibliotecas digitales: una somera revisión bibliográfica y su relación en la educación superior. *In*: LATIN AMERICAN AND CARIBBEAN CONFERENCE FOR ENGINEERING AND TECHNOLOGY, 1., 2013. **Anais [...]**, Cancún, 2013.

VIEIRA, S. B. Indexação automática e manual: revisão de literatura. **Ciência da Informação**, Brasília, v. 17, n. 1, p. 43-57, 1988.