





Aprimoramento do Annif para indexação automática de publicações em português

Enhancement of Annif for automatic indexing of publications in portuguese



Bruno Carlos da Cunha Costa¹

 <http://lattes.cnpq.br/8162384236816488>
 <https://orcid.org/0000-0001-9655-8698>

Ana Carolina Simionato Arakaki²

 <http://lattes.cnpq.br/9896600626524397>
 <https://orcid.org/0000-0002-0140-9110>

Renan Luiz da Silva Nascimento³

 <http://lattes.cnpq.br/9394350252301981>
 <https://orcid.org/0009-0001-9785-8833>

Resumo

A indexação automática desempenha um papel essencial na organização e recuperação da informação, especialmente em bibliotecas e repositórios institucionais que lidam com grandes volumes de documentos. O presente estudo investiga a aplicação de uma ferramenta baseada em inteligência artificial para a indexação automática de publicações em língua portuguesa, com o objetivo de aprimorar a precisão e a coerência dos descritores atribuídos aos documentos. Para isso, foram conduzidos testes com uma ferramenta desenvolvida originalmente para a indexação de textos em outro idioma, avaliando sua capacidade de adaptação a um novo contexto linguístico. A metodologia adotada incluiu a realização de testes exploratórios iniciais, seguidos da criação de um conjunto de treinamento específico composto por trinta mil documentos extraídos de bases institucionais. O treinamento foi conduzido de maneira incremental, permitindo a avaliação da evolução do desempenho da ferramenta. Os resultados indicaram que a qualidade das sugestões geradas pela ferramenta está diretamente relacionada à diversidade e estrutura do conjunto de treinamento. Inicialmente, a indexação apresentou baixa precisão para documentos em língua portuguesa, mas, após o refinamento dos dados e a ampliação do conjunto de treinamento, observou-se uma melhora significativa na coerência dos descritores atribuídos. O estudo demonstra que a adaptação da ferramenta para a indexação de documentos em língua portuguesa é viável, desde que sejam realizados ajustes metodológicos no treinamento do modelo. Conclui-se que o aprimoramento contínuo do conjunto de treinamento pode fortalecer a padronização da indexação,

¹ Doutor, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, RJ, Brasil Professor, Instituto Federal de Educação, Ciência e Tecnologia (IFRJ), Rio de Janeiro, RJ, Brasil. brunocosta@ibict.br.

² Doutora, Universidade Estadual Paulista Júlio de Mesquita Filho, UNESP, Brasil. Coordenadora de Serviços Bibliográficos, Instituto Brasileiro de Informação em Ciência e Tecnologia (COBIB/IBICT), Brasília, DF, Brasil. anaarakaki@ibict.br.

³ Graduação em andamento em Estatística, Universidade de Brasília, UnB, Brasília, DF, Brasil. Bolsista, Instituto Brasileiro de Informação em Ciência e Tecnologia (COBIB/IBICT), Brasília, DF, Brasil. renannascimento@ibict.br.

melhorar a recuperação da informação e reduzir a necessidade de intervenção manual em bibliotecas e repositórios institucionais.

Palavras-chave: indexação automática; Inteligência Artificial; processamento de linguagem natural.

Abstract

Automatic indexing plays a fundamental role in information organization and retrieval, especially in libraries and institutional repositories that manage large volumes of documents. This study investigates the application of an artificial intelligence-based tool for the automatic indexing of publications in Portuguese, aiming to improve the accuracy and coherence of descriptors assigned to documents. To achieve this, tests were conducted with a tool originally developed for indexing texts in another language, assessing its ability to adapt to a new linguistic context. The methodology included initial exploratory tests, followed by the creation of a specific training set consisting of thirty thousand documents extracted from institutional databases. Training was conducted incrementally, allowing for the evaluation of the tool's performance over time. The results indicated that the quality of the generated suggestions is directly related to the diversity and structure of the training set. Initially, the indexing showed low accuracy for documents in Portuguese. However, after refining the data and expanding the training set, a significant improvement was observed in the coherence of the assigned descriptors. The study demonstrates that adapting the tool for indexing documents in Portuguese is feasible, provided that methodological adjustments are made in model training. It is concluded that the continuous improvement of the training set can enhance indexing standardization, improve information retrieval, and reduce the need for manual intervention in libraries and institutional repositories.

Keywords: automatic indexing; Artificial Intelligence; natural language processing.

1 INTRODUÇÃO

O *Annif*⁴ foi desenvolvido em 2017 como um protótipo experimental para automatizar a indexação documental, utilizando *scripts* em *Python* e um índice do *Elasticsearch* para associar termos extraídos dos títulos de documentos a conceitos de um vocabulário controlado. Diferentemente dos sistemas tradicionais de recuperação da informação, em que o usuário busca documentos a partir de um tema, o *Annif* inverte essa lógica: recebe um documento como entrada e retorna os assuntos mais relevantes atribuídos ao seu conteúdo.

Seu nome, um anagrama de "*Finna*", serviço de descoberta usado em bibliotecas finlandesas, reflete essa inovação. Em 2018, a ferramenta foi ampliada com a adoção dos *frameworks Flask* e *Connexion*, permitindo a criação de uma interface *Web* interativa e uma API REST, o que facilitou sua integração a outros sistemas. A classificação automática do *Annif* baseia-se na execução de diferentes *backends*, ou seja, algoritmos que operam

⁴ Disponível em: <https://annif.org/> e seu código <https://github.com/NatLibFi/Annif>.

isoladamente ou combinados para melhorar a precisão das sugestões. Essa arquitetura modular permite sua adaptação a diversos vocabulários controlados e conjuntos documentais (Suominen, 2019).

Sua principal vantagem é a flexibilidade, possibilitando treinamento com diferentes bases de conhecimento e tornando-se útil para bibliotecas, repositórios institucionais e outros sistemas que necessitam de uma indexação escalável e padronizada. No Brasil, o Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict) tem explorado o *Annif* para aprimorar bases como a Rede Bibliodata e o Catálogo Coletivo Nacional de Publicações Periódicas (CCN), buscando maior interoperabilidade e normalização de registros bibliográficos por meio da Inteligência Artificial e do aprendizado de máquina.

A eficiência do *Annif* depende de conjuntos de treinamento adequados à área do conhecimento em questão. Embora permita testes com documentos de exemplo, seu desempenho é otimizado com dados específicos. O formato recomendado de entrada é .txt, e a escolha do *backend* influencia diretamente a qualidade da indexação. Para este estudo, foi adotado o *Term Frequency-Inverse Document Frequency* (TF-IDF), um método estatístico amplamente utilizado para atribuir pesos a termos, identificando os mais representativos em um conjunto documental.

A introdução do *Annif* na indexação e catalogação documental representa um avanço para a Biblioteconomia, ao automatizar a atribuição de descritores e reduzir a subjetividade do processo manual, frequentemente influenciado por variáveis cognitivas e interpretativas dos indexadores (Fujita, 1988). Como a indexação é essencial para a organização da informação e afeta diretamente a recuperação de documentos, a consistência e precisão dos termos atribuídos tornam-se aspectos críticos (Rubi e Fujita, 2003). Ferramentas automatizadas, como o *Annif*, mitigam essas limitações, proporcionando maior uniformidade terminológica e escalabilidade na indexação documental.

A qualidade da recuperação da informação depende da coerência dos metadados, sendo que sua estruturação adequada melhora a busca e a interoperabilidade entre bases de dados (Haynes, 2018). Conforme Pomerantz (2015), a

integração de modelos estatísticos e vocabulários controlados permite que sistemas como o Annif gerem sugestões mais precisas e alinhadas ao conteúdo dos documentos. Além disso, a normalização terminológica, conforme Svenonius (2000), é um fator decisivo para evitar ambiguidades e variações linguísticas, garantindo um controle mais eficaz dos termos indexados.

A automação da indexação não apenas reduz o esforço manual dos profissionais da informação, mas também possibilita a análise de grandes volumes documentais com maior rapidez e precisão. A adoção de ferramentas baseadas em inteligência artificial fortalece a infraestrutura da gestão da informação, tornando os processos de organização e recuperação da informação mais eficientes e beneficiando pesquisadores, bibliotecários e usuários finais.

Dessa forma, este estudo tem como objetivo avaliar o desempenho do *Annif* na indexação automática de publicações em português e propor estratégias para aprimorar sua eficácia, por meio do treinamento com dados especializados extraídos de bases de informação do Instituto Brasileiro de Informação em Ciência e Tecnologia (Ibict).

2 PROCEDIMENTOS METODOLÓGICOS

Este estudo adota uma abordagem experimental e aplicada, com o objetivo de avaliar a eficácia do *Annif* na indexação automática de publicações em português e propor estratégias para aprimorar seu desempenho. A pesquisa segue um delineamento quantitativo e qualitativo, combinando testes controlados da ferramenta com análise crítica dos resultados. A metodologia foi estruturada em três etapas principais: preparação do corpus de treinamento, configuração da indexação e análise dos resultados gerados.

A investigação caracteriza-se como um estudo experimental, no qual diferentes configurações do *Annif* foram testadas para determinar sua capacidade de indexação em português. Além disso, trata-se de uma pesquisa aplicada, pois visa desenvolver uma solução prática para aprimorar a indexação automatizada em bibliotecas e repositórios nacionais.

Inicialmente, foram conduzidos testes exploratórios com publicações de exemplo fornecidas pelo próprio *Annif*, todas em inglês. O objetivo era validar o funcionamento da ferramenta em um idioma previamente treinado e verificar a coerência das sugestões geradas. Os resultados foram satisfatórios, com alta correspondência entre os termos sugeridos e os conteúdos dos documentos.

Na segunda fase, o *Annif* foi testado com publicações científicas em português, selecionadas a partir das revistas do Ibict. No entanto, os resultados foram significativamente inferiores, com baixa precisão na atribuição dos descritores. Esse desempenho insatisfatório sugeriu dificuldades na adaptação do modelo para a língua portuguesa, possivelmente devido à ausência de um vocabulário treinado para esse idioma.

Diante disso, foi elaborado um corpus de treinamento especializado, contendo aproximadamente 30 mil publicações em português extraídas das bases do Ibict. Os documentos foram estruturados no formato *tab-separated values* (.tsv,) exigido pelo *Annif*, no qual cada linha contém o título e, quando disponível, o resumo da publicação, associados a *Uniform Resource Identifiers* (URIs) correspondentes aos descritores indexados na ontologia da ferramenta.

Para otimizar a indexação automática, foi realizada uma análise dos vocabulários utilizados pelo Ibict, mapeando-os para as ontologias do *Annif* a fim de garantir maior coerência na atribuição dos termos. Esse processo permitiu que o treinamento fosse ajustado às particularidades semânticas do português e às necessidades específicas das bibliotecas e repositórios nacionais.

O *backend* selecionado para a etapa de treinamento foi o *Term Frequency-Inverse Document Frequency* (TF-IDF), amplamente utilizado para análise de relevância de termos em documentos. Esse método estatístico atribui pesos às palavras com base na frequência com que ocorrem dentro do corpus, permitindo que o *Annif* identifique os termos mais representativos para cada publicação.

O treinamento do modelo foi realizado de forma incremental, no período de julho de 2024 a março de 2025, com avaliações semanais à medida que novos dados eram incorporados ao corpus, cabe destacar que o modelo enfrentou desafios específicos

relacionados à adaptação ao idioma português, o que exigiu estratégias adicionais de refinamento. Durante o processo de treinamento incremental, foram aplicadas técnicas de limpeza e normalização textual, padronização de grafias e tratamento de siglas recorrentes nas publicações acadêmicas. Além disso, buscou-se equilibrar o corpus entre diferentes áreas temáticas para evitar o viés de indexação em domínios mais representados. Também foram realizadas análises de erros a partir das sugestões inadequadas geradas, o que possibilitou ajustes interativos nos parâmetros do *backend* TF-IDF e na estrutura dos arquivos de entrada. Essas ações contribuíram significativamente para aprimorar a sensibilidade do modelo às especificidades linguísticas do português e aumentar a relevância dos descritores sugeridos.

A avaliação do desempenho do *Annif* foi baseada em testes comparativos, conduzidos semanalmente com um conjunto fixo de cinco publicações em português. Para cada teste, as sugestões geradas foram analisadas em relação à coerência semântica, precisão e recall dos termos atribuídos. Os principais indicadores utilizados na análise foram: Precisão, a partir do percentual de descritores sugeridos que realmente correspondem aos conteúdos dos documentos, revocação: percentual de termos relevantes efetivamente recuperados pelo modelo, e F1-score: métrica que combina precisão e revocação, avaliando o equilíbrio entre ambas.

Além das métricas quantitativas, foi realizada uma avaliação qualitativa, na qual especialistas em Biblioteconomia e Ciência da Informação examinaram a pertinência dos termos atribuídos pelo *Annif*, considerando sua adequação às práticas de indexação bibliográfica.

3 RESULTADOS

A saída do *Annif* segue um formato simples e estruturado, retornando, por padrão, dez sugestões de assuntos organizadas em ordem decrescente de relevância. A pontuação atribuída a cada termo varia de 0 a 1, indicando o grau de confiança do sistema na associação entre o assunto sugerido e o conteúdo analisado. Valores mais próximos de 1 sinalizam maior precisão na indexação, permitindo que os usuários avaliem a qualidade das recomendações.

Para o treinamento do modelo, foram utilizadas três fontes de dados: Bibliodata, publicações do Encontro Nacional de Pesquisa em Ciência da Informação (ENANCIB) – Grupo de Trabalho 8 (Dado, Informação e Tecnologia) e uma seleção de textos da Sociedade Brasileira de Computação (SBC). Essa diversidade garantiu um conjunto representativo de documentos, cobrindo distintas áreas do conhecimento e fortalecendo a base de aprendizado da ferramenta. No entanto, nas primeiras semanas de testes, a escassez de dados corretamente formatados comprometeu os resultados, evidenciando a necessidade de um corpus mais robusto.

As limitações iniciais tornaram-se evidentes na análise das sugestões geradas. Em diversos casos, os termos atribuídos não apresentavam correspondência coerente com o conteúdo dos textos, especialmente em documentos que abordavam conceitos técnicos e específicos. Essa inconsistência revelou dificuldades na adaptação do modelo ao contexto semântico dos documentos, reforçando a importância de um treinamento mais refinado. A seguir, na Figura 1, são apresentadas imagens ilustrando essa limitação, com títulos das publicações, palavras-chave e sugestões geradas pelo *Annif*.

Figura 1: Sugestões do Annif para três publicações utilizando o “arquivo de treinamento inicial”

Título: Publicação de dados de pesquisa no contexto da comunicação científica: cenário terminológico-conceitual	Título: Arquitetura da informação para publicação de dados abertos conectados
Palavras-chave: dados de pesquisa; ciência aberta; comunicação científica	Palavras-chave: metadados de negócio, dados abertos conectados, arquitetura de informação para processos de negócio, web semântica
Annif:	Annif:
<http://www.yso.fi/onto/ysop3056> databases 0.3221	<http://www.yso.fi/onto/ysop3056> databases 0.3193
<http://www.yso.fi/onto/ysop3286> social work 0.1598	<http://www.yso.fi/onto/ysop16380> music therapy 0.1959
<http://www.yso.fi/onto/ysop4> fiction and poetry 0.1292	<http://www.yso.fi/onto/ysop5643> neuroses 0.1818
<http://www.yso.fi/onto/ysop2945> learning 0.1156	<http://www.yso.fi/onto/ysop4> fiction and poetry 0.1740
<http://www.yso.fi/onto/ysop13084> nature 0.1153	<http://www.yso.fi/onto/ysop14783> national literature 0.1682
<http://www.yso.fi/onto/ysop27522> fanzines 0.1087	<http://www.yso.fi/onto/ysop21217> travel 0.1682
<http://www.yso.fi/onto/ysop3333> alternative movements 0.1087	<http://www.yso.fi/onto/ysop27522> fanzines 0.1594
<http://www.yso.fi/onto/ysop608> racism 0.1087	<http://www.yso.fi/onto/ysop3333> alternative movements 0.1594
<http://www.yso.fi/onto/ysop14524> algorithms 0.1057	<http://www.yso.fi/onto/ysop608> racism 0.1594
<http://www.yso.fi/onto/ysop183> research 0.1041	<http://www.yso.fi/onto/ysop6702> credits 0.1534
Título: Aplicando o ChatGPT na condução de revisões sistemáticas da literatura	
Palavras-chave: revisão sistemática da literatura; ChatGPT; inteligência artificial	
Annif:	
<http://www.yso.fi/onto/ysop27522> fanzines 0.2967	
<http://www.yso.fi/onto/ysop3333> alternative movements 0.2967	
<http://www.yso.fi/onto/ysop608> racism 0.2967	
<http://www.yso.fi/onto/ysop4> fiction and poetry 0.2391	
<http://www.yso.fi/onto/ysop1178> entrepreneurs 0.2295	
<http://www.yso.fi/onto/ysop10315> scientific journals 0.2234	
<http://www.yso.fi/onto/ysop3286> social work 0.2138	
<http://www.yso.fi/onto/ysop2921> religions 0.1981	
<http://www.yso.fi/onto/ysop8113> literature 0.1964	
<http://www.yso.fi/onto/ysop21530> software development 0.1886	

Fonte: Elaboração própria.

Inicialmente, apenas a sugestão *databases* apresentou alguma relação com a publicação original, enquanto os demais termos atribuídos não possuíam conexão

semântica relevante. Esse resultado evidenciou a necessidade de ajustes no corpus de treinamento.

Após aproximadamente dois meses de refinamento, incluindo a ampliação da base de dados e a padronização da formatação dos registros, observou-se uma melhoria significativa no desempenho do *Annif*. No novo processamento da mesma publicação, as sugestões geradas demonstraram maior coerência temática e precisão, refletindo o impacto positivo das otimizações realizadas. Na Figura 2 são apresentadas imagens que ilustram essa evolução.

Figura 2: Sugestões do *Annif* para três publicações utilizando o “arquivo de treinamento final”

Título: Publicação de dados de pesquisa no contexto da comunicação científica: cenário terminológico-conceitual	Título: Arquitetura da informação para publicação de dados abertos conectados
Palavras-chave: dados de pesquisa; ciência aberta; comunicação científica	Palavras-chave: metadados de negócio, dados abertos conectados, arquitetura da informação para processos de negócio, web semântica
Annif:	Annif:
<http://www.yso.fi/onto/ysop27006> research data 0.6508	<http://www.yso.fi/onto/ysop21716> semantic web 0.7205
<http://www.yso.fi/onto/ysop26655> open data 0.5913	<http://www.yso.fi/onto/ysop9319> metadata 0.6158
<http://www.yso.fi/onto/ysop21716> semantic web 0.5726	<http://www.yso.fi/onto/ysop1140> data storage 0.5832
<http://www.yso.fi/onto/ysop17649> open access 0.5256	<http://www.yso.fi/onto/ysop5521> information management 0.5830
<http://www.yso.fi/onto/ysop12336> data archives 0.5226	<http://www.yso.fi/onto/ysop26655> open data 0.5767
<http://www.yso.fi/onto/ysop1140> data storage 0.5080	<http://www.yso.fi/onto/ysop27006> research data 0.5615
<http://www.yso.fi/onto/ysop9319> metadata 0.4996	<http://www.yso.fi/onto/ysop26001> linked open data 0.5004
<http://www.yso.fi/onto/ysop27163> visual literacy 0.4941	<http://www.yso.fi/onto/ysop3056> databases 0.4887
<http://www.yso.fi/onto/ysop3056> databases 0.4915	<http://www.yso.fi/onto/ysop27202> big data 0.4793
<http://www.yso.fi/onto/ysop27250> data 0.4865	<http://www.yso.fi/onto/ysop27163> visual literacy 0.4744
Título: Aplicando o ChatGPT na condução de revisões sistemáticas da literatura	
Palavras-chave: revisão sistemática da literatura; ChatGPT; inteligência artificial	
Annif:	
<http://www.yso.fi/onto/ysop2616> artificial intelligence 0.5559	
<http://www.yso.fi/onto/ysop9319> metadata 0.5252	
<http://www.yso.fi/onto/ysop2964> information retrieval 0.4941	
<http://www.yso.fi/onto/ysop2339> technology (information) 0.4936	
<http://www.yso.fi/onto/ysop25167> data models 0.4839	
<http://www.yso.fi/onto/ysop21716> semantic web 0.4827	
<http://www.yso.fi/onto/ysop5521> information management 0.4804	
<http://www.yso.fi/onto/ysop17836> information science 0.4735	
<http://www.yso.fi/onto/ysop27250> data 0.4702	
<http://www.yso.fi/onto/ysop21978> computational science 0.4678	

Fonte: Elaboração própria.

A evolução do desempenho do *Annif* não se refletiu apenas na qualidade das sugestões geradas, mas também na consistência das pontuações atribuídas a cada termo recomendado. Com o aprimoramento do “arquivo de treinamento final”, as sugestões passaram a ser mais precisas e alinhadas ao conteúdo das publicações, enquanto as pontuações demonstraram maior estabilidade, indicando um refinamento na confiabilidade do modelo. Essa melhoria não se limitou aos exemplos previamente analisados, mas foi constatada em todas as publicações testadas, evidenciando o impacto positivo da expansão e estruturação do corpus na precisão e coerência dos resultados.

4 CONSIDERAÇÕES FINAIS

Os testes demonstraram que a qualidade e diversidade do arquivo de treinamento inicial e final, são fatores determinantes para a precisão das sugestões geradas pelo *Annif*. Inicialmente, os resultados iniciais para publicações em português foram insatisfatórios, devido à escassez de dados adequados e à necessidade de ajustes no formato dos registros. No entanto, com a expansão e refinamento do corpus, observou-se uma melhora significativa na coerência dos termos sugeridos, tornando a indexação mais alinhada ao conteúdo dos textos analisados.

Além disso, a evolução das pontuações atribuídas aos termos evidenciou um ganho na confiabilidade do modelo, garantindo maior consistência na atribuição dos descritores. Essa melhoria não se limitou a casos isolados, mas foi consistente em todas as publicações testadas, reforçando a importância de um treinamento contínuo e bem estruturado.

A adaptação do *Annif* para publicações em português mostrou-se uma estratégia promissora para automatizar a indexação, reduzindo o esforço manual e garantindo maior padronização na recuperação da informação. O aprimoramento do arquivo de treinamento deve seguir um processo contínuo, incorporando novas fontes de publicações e ampliando o conjunto de dados, o que permitirá um refinamento progressivo dos resultados. Além disso, a experimentação com diferentes *backends* pode aumentar a eficiência da recuperação da informação e contribuir para a normalização dos metadados.

Apesar dos avanços, alguns desafios persistem, como a necessidade de expandir o vocabulário controlado para ampliar a cobertura terminológica e minimizar ambiguidades semânticas. Futuras pesquisas poderão explorar a integração do *Annif* com modelos baseados em redes neurais, aprimorando sua capacidade de generalização e adaptação ao contexto linguístico do português.

REFERÊNCIAS

FUJITA, M. L. Sistema de indexação PRECIS: perspectiva histórica e técnica de seu desenvolvimento e aplicação. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 21, n. 1/2, p. 46-62, jan./jun. 1988.

HAYNES, D. **Metadata for information management and retrieval: understanding metadata and its use**. London: Facet Publishing, 2018.

POMERANTZ, J. **Metadata**. Cambridge: MIT Press, 2015.

RUBI, M. P.; FUJITA, M. S. L. Elementos de política de indexação em manuais de indexação de Sistemas de Informação especializados. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 8, n. 1, p. 66-77, jan./jun. 2003. Disponível em: <https://periodicos.ufmg.br/index.php/pci/article/view/23460/18920>. Acesso em: 19 maio 2025.

SUOMINEN, Osma. Annif: DIY automated subject indexing using multiple algorithms. **Liber Quarterly: the Journal of the Association of European Research Libraries**, [S. l.], v. 29, n. 1, p. 1-26, 2019. Disponível em: <https://doi.org/10.18352/lq.10285>. Acesso em: 19 maio 2025.

SVENONIUS, E. **The intellectual foundation of information organization**. Cambridge, MA: MIT Press, 2000.



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

Como citar este trabalho:

COSTA, Bruno Carlos da Cunha; ARAKAKI, Ana Carolina Simionato; NASCIMENTO, Renan Luiz da Silva. Aprimoramento do Annif para indexação automática de publicações em português. In: WORKSHOP DE INFORMAÇÃO DADOS E TECNOLOGIA, 8., 2025, Marília, SP. **Anais [...]**. Marília, SP: Universidade de Marília, 2025. DOI: <https://doi.org/10.22477/viii.widat.273>.