



Eixo 6 – O mundo digital: apropriações e desafios

Representação, Interação e Garantia do Usuário: Padrões de Busca em Catálogo Bibliográfico Online

Representation, Interaction, and User Warrant: Search Patterns in an Online Bibliographic Catalog

Maristella Soares dos Santos – Universidade Estadual de Campinas (Unicamp) – mariss@unicamp.br

Resumo: Este estudo analisa 214.057 buscas no catálogo de um sistema de bibliotecas universitárias, para entender como os usuários formulam suas consultas. A pesquisa combina análise quantitativa das frequências e análise qualitativa de 800 termos. Os resultados mostram muitas buscas únicas, erros de digitação, variações ortográficas e expressões incompletas, indicando influência de motores de busca comuns. Há evidência de distanciamento entre a linguagem do usuário e a do catálogo. Conclui-se que interfaces mais flexíveis podem melhorar a recuperação e fortalecer a garantia do usuário.

Palavras-chave: Catálogo bibliográfico. Comportamento de busca. Análise de logs. Garantia do usuário. Recuperação da informação.

Abstract: This study analyzes 214,057 searches in the catalog of a university library system, to understand how users formulate their queries. The research combines quantitative frequency analysis and qualitative examination of 800 terms. Results show many unique searches, typing errors, spelling variations, and incomplete expressions, suggesting influence from general search engines. There is evidence of a gap between users' natural language and the catalog's structure. The study concludes that more flexible interfaces can improve retrieval and strengthen user warrant.

Keywords: Bibliographic catalog. Search behavior. Log analysis. User warrant. Information retrieval.



1 INTRODUÇÃO

Nas bibliotecas universitárias, a investigação baseada no uso real do comportamento de busca ainda é limitada frente à expansão e complexidade crescente dos acervos digitais. Observa-se um claro descompasso prático: embora os registros bibliográficos sigam rigorosos padrões de representação, os usuários raramente formulam suas consultas utilizando esses mesmos elementos descritivos. Essa lacuna entre a linguagem documentária e a linguagem natural suscita uma questão central: as expressões utilizadas em buscas correspondem às estruturas de metadados do catálogo? Nesse contexto, os termos digitados constituem registros diretos das necessidades informacionais, revelando padrões essenciais para compreender o comportamento de busca e aprimorar os mecanismos de mediação e recuperação da informação.

Compreende-se que a catalogação consiste numa ação de mediação indireta da informação, e, logo, exige que o profissional conheça o seu público e contexto para viabilizar e facilitar a recuperação (Nascimento; Sousa, 2021). Quando os descritores e demais atributos diferem dos termos de busca, a recuperação da informação torna-se mais dispendiosa (Laipelt, 2015). Investigar como os usuários formulam suas buscas contribui, portanto, para aprimorar práticas de representação, reduzir barreiras informacionais e efetivar o princípio de “poupar o tempo do leitor”, presente na teoria das Cinco Leis da Biblioteconomia, publicadas em 1931 por Ranganathan (2009). Nesse sentido, compreender a necessidade informacional, entendida como o ponto de partida da interação entre indivíduo e sistema, conforme Le Coadic (2004), é mandatório para que o catálogo cumpra sua função mediadora.

Nesse sentido, esta pesquisa tem como objetivo analisar os padrões de formulação de buscas em catálogos bibliográficos com base na intenção informacional inferida dos termos digitados pelos usuários. Para isso, propõe-se a classificar semanticamente as expressões, identificar suas estruturas recorrentes e interpretar como essas escolhas traduzem as estratégias cognitivas e as expectativas dos usuários frente ao sistema.

Como forma de aproximar a representação da informação e a linguagem do público, é fundamental adotar métodos empíricos que revelem o comportamento real



de busca (Bastos; Vidotti, 2014). Nesse contexto, a análise de *logs*, estes descritos por Monteiro-Krebs, Laipelt e Rocha (2022), como dados gravados virtualmente a partir das interações usuários-sistemas, que registram a requisição realizada, data, horário, local/IP, duração, entre outros, se destaca como uma fonte privilegiada por capturar as “pegadas de informação digital” (Rosa; Laipelt; Monteiro-Krebs, 2021), de forma espontânea. É a partir desses “rastros” que a Arquitetura da Informação (AI) e a Experiência do Usuário (UX, do inglês *User Experience*) podem integrar-se na construção de interfaces mais acessíveis e alinhadas às expectativas do público, minimizando as barreiras de encontrabilidade, ou seja, de recuperação da informação (Torino *et al.*, 2022).

Essa adequação das interfaces é o que viabiliza a garantia do usuário (Lancaster, 2004; Laipelt, 2015), no cenário digital, minimizando a ausência direta da mediação humana presencial. Essa abordagem vai ao encontro do modelo conceitual dos Requisitos Funcionais para Registros Bibliográficos (FRBR), distanciando o foco histórico no bibliotecário como “usuário-meio” para priorizar o usuário final em suas tarefas de encontrar, identificar, selecionar e obter informações, as chamadas “tarefas do usuário” (Moreno; Arellano, 2005). Logo, investigar a formulação dessas consultas é o caminho prático para que o catálogo cumpra verdadeiramente sua função mediadora.

2 METODOLOGIA

Trata-se de uma pesquisa bibliográfica, documental e descritiva, baseada na análise dos registros de busca (*logs*) do Catálogo Acervus, que reúne a totalidade dos acervos das 30 bibliotecas que compõem o Sistema de Bibliotecas da Universidade Estadual de Campinas (SBU/UNICAMP), obtido por meio do relatório “Termos pesquisados no Terminal Web”, composto pelas colunas “Termo” e “Quantidade”. A investigação combinou abordagens quantitativas e qualitativas. A etapa quantitativa mapeou frequências e distribuições no conjunto de 214.057 registros, enquanto a etapa qualitativa analisou uma amostra aleatória de 800 termos, examinando sua estrutura, a intenção informacional presumida e características linguísticas, como padrões de escrita, erros de digitação, fragmentos de referências, siglas ambíguas, aproximações fonéticas e tentativas de busca por similaridade. O embasamento teórico foi construído



a partir de um levantamento na Base de Dados Referenciais de Artigos de Periódicos em Ciência da Informação (Brapci).

A coleta abrangeu logs de março, junho, agosto e novembro de 2025, permitindo observar o comportamento dos usuários em diferentes momentos do calendário acadêmico. Após a limpeza dos relatórios, consolidaram-se 214.057 registros válidos. Embora não reconstruam sessões completas (MONTEIRO-KREBS; LAIPELT; ROCHA, 2022) nem isolem acessos automatizados ou de rotinas administrativas e treinamentos, esse volume expressivo garante a identificação de padrões gerais de busca.

Como apoio técnico, recorreu-se pontualmente ao uso de ferramentas de Inteligência Artificial (IA)¹. O Microsoft Copilot auxiliou na construção fórmulas e rotinas de planilhas no Excel e Power Query e, posteriormente, na condensação do texto final, garantindo clareza à exposição das ideias. Para testar as categorias qualitativas estabelecidas e confirmar sua consistência, a IA Institucional “CACia” (DETI/UNICAMP) foi empregada. Reitera-se, contudo, que a condução intelectual, bem como análises e interpretações permaneceram sob responsabilidade e condução da autora.

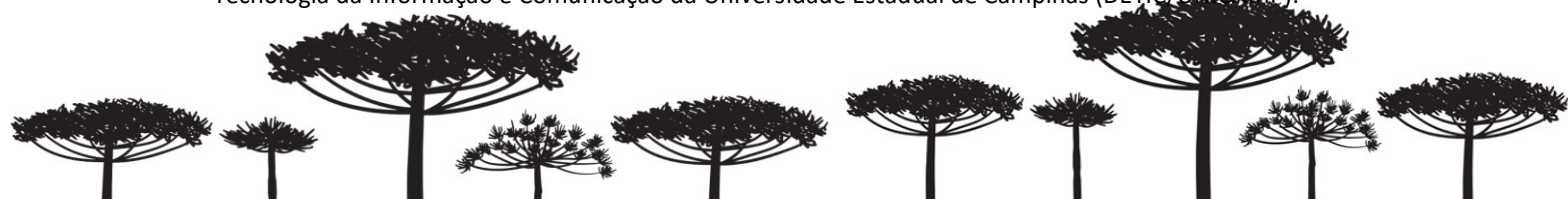
3 RESULTADOS E DISCUSSÕES

A verificação dos 214.057 permitiu identificar padrões de uso do catálogo e estratégias de formulação de necessidades informacionais empregadas pelos usuários nesses usos. Embora os dados tenham sido coletados em quatro meses distintos, não se observaram variações temporais significativas. Razão pela qual apresenta-se os dados de maneira consolidada

Tabela 1 — Distribuição dos campos de busca utilizados

Campo de busca	Quantidade	Frequência (%)
Todos os campos	120.020	56,07%
Título	35.997	16,82%
Autoria*	29.893	13,96%
Autoridades**	17.120	8,00%
Assunto	8.797	4,11%

¹ O Microsoft Copilot é um assistente de inteligência artificial generativa desenvolvido pela Microsoft. O sistema CACia é uma ferramenta baseada em IA desenvolvida internamente pela Diretoria Executiva de Tecnologia da Informação e Comunicação da Universidade Estadual de Campinas (DETI/UNICAMP).



Outros campos*** ²	2.230	1,04%
Total	214.057	100,00% ³

Fonte: Elaborada pela autora

A tabela 1 sintetiza a distribuição das categorias de busca. Observa-se que “todos os campos” concentra mais da metade dos registros (56,07%), seguida por “título” (16,82%) e “autoria” (13,96%). Juntas, essas três categorias representam mais de 86% das interações, evidenciando a preferência por estratégias amplas e diretas de recuperação da informação. Esse comportamento está alinhado aos achados de Bastos e Vidotti (2014), que associam o uso massivo de “todos os campos” à busca por interações rápidas e à adoção do campo padrão do catálogo (default). As buscas por título e autoria mostram tendência de procurar “itens conhecidos”, estratégia que, segundo Lancaster (2004), demanda menor esforço cognitivo do que a pesquisa formal por assunto. De modo geral, os dados confirmam padrões já observados em estudos de logs, com formulações simples, diretas e com baixo uso de recursos avançados (Aires; Aluísio, 2003; Rosa; Laipelt; Monteiro-Krebs, 20212021).

Tabela 2 - Frequência de ocorrência dos termos

Faixa de ocorrência	Quantidade	Frequência (%)
1 ocorrência	175.543	82,01%
2 a 3	31.114	14,54%
4 a 5	3.725	1,74%
6 a 10	2.054	0,96%
11 a 50	1.245	0,58%
51+	376	0,18%
Total	214.057	100,00%

Fonte: Elaborada pela autora

A tabela 2 demonstra a alta dispersão dos termos, verificando-se que 82,01% ocorreram uma única vez e 16,28% entre 2 e 5 vezes, enquanto apenas 0,18% superaram 50 ocorrências. Essa enorme variedade de termos configura um autêntico fenômeno de cauda longa (*long tail*), refletindo a pluralidade de perfis da UNICAMP. A predominância de buscas únicas reforça o comportamento exploratório, marcado por estratégias

² **Nota:** (*) Refere-se aos índices gerais de responsabilidade intelectual (autor, organizador, etc.). (**) Buscas realizadas especificamente na interface do índice controlado de autoridades. (***) Contempla opções de baixa frequência, como ISBN, ISSN, editora e ano de publicação.



informais, fragmentos e erros, expondo o desafio das unidades de informação em lidarem com a linguagem natural. Esse padrão reflete o modelo de *berrypicking* (coleta seletiva) proposto por Bates (1989), no qual a busca não se dá por uma correspondência exata e estática, mas como um processo onde o usuário recolhe informações aos poucos e adapta continuamente suas estratégias.

Essa singularidade nas buscas, de certa forma, atualiza a quase centenária teoria de Ranganathan (2009) para o ambiente digital, reafirmando na prática que há “para cada leitor o seu livro” e “para cada livro o seu leitor”, não apenas no item, mas na interação que viabiliza esse encontro na atualidade, independente do suporte ou tipo de informação. Como bem elaborado pelo teórico, as bibliotecas permanecem em crescimento, todo o seu vasto conteúdo deseja ser encontrado, e o entendimento de toda essa dinâmica é o que permite, fundamentalmente, otimizar o tempo do leitor.

Tabela 3 – Tipo de entrada de dados

Tipo de entrada	Quantidade	Frequência (%)
Digitável	187.477	87,58%
Digitável + Seleção	25.917	12,11%
Seleção	390	0,18%
Data	222	0,10%
Navegação	51	0,02%
Total	214.057	100,00%

Fonte: Elaborada pela autora

A Tabela 3 apresenta os tipos de entrada utilizados. A esmagadora maioria (87,58%) ocorre exclusivamente por campos digitáveis. Somando-se às entradas combinadas (12,11%), nota-se que 99,69% das buscas exigem a inserção textual livre pelos usuários. Em contrapartida, recursos baseados em seleção (0,18%), data (0,10%) e navegação (0,02%), frequentemente atrelados à busca avançada ou a menus de página inicial, são subutilizados. Essa baixíssima adesão a cliques e filtros estruturados confirma a preferência pela caixa de busca simples e direta.

Quanto às estruturas de formulação, as buscas simples são amplamente majoritárias, seguidas pelas buscas combinadas e, em proporção muito reduzida, pelas buscas com operadores booleanos digitados. Esse padrão confirma a tendência de uso de estratégias naturais de busca, sem aplicação de sintaxe avançada, mesmo quando o sistema oferece suporte a operadores. Aires e Aluísio (2003) e Rosa; Laipelt; Monteiro-



Krebs, (2021) já haviam observado que usuários raramente utilizam operadores lógicos, preferindo expressões curtas e diretas.

A análise qualitativa dos 800 termos permitiu identificar padrões que aprofundam os achados quantitativos. Observou-se grande presença de erros de digitação, variações ortográficas, sinais gráficos irregulares e junções incomuns entre nomes, títulos e assuntos, indicando buscas rápidas, pouco revisadas e frequentemente guiadas por "vagas lembranças". Também foram recorrentes abreviações, siglas ambíguas e fragmentos de referências bibliográficas, muitos deles provavelmente extraídos de planos de ensino ou ementas, sugerindo que o catálogo é utilizado como extensão de rotinas acadêmicas. Esses elementos revelam tanto estratégias informais de busca quanto dificuldades de traduzir a necessidade informacional para a lógica do catálogo. As buscas analisadas trazem desde títulos que parecem frases em linguagem natural, como "o cara que estou a fim não é um cara", até consultas ruidosas, como "HISTORIADOR,SEUS,FATOS,CARR", que exigem maior refinamento manual. Também surgem erros simples, como "churchul", revelando a ausência de correção automática. Vale destacar que, eventualmente, tais padrões não impedem a recuperação, mas podem demandar maior refinamento por parte do usuário, dada a interpretação literal do catálogo (isto é, busca exatamente o que foi digitado, sem correções automáticas ou aproximações semânticas). Esses exemplos ajudam a entender como expectativas formadas em outros ambientes de busca influenciam o uso do catálogo.

Os termos válidos foram agrupados em seis macrocategorias, sendo elas Autor, Obra, Assunto, Híbrida, Metadado e Ruído, que sintetizam os principais modos de formulação observados. Essa classificação evidencia a diversidade de caminhos adotados pelos usuários e ajuda a compreender quais tipos de entrada predominam no catálogo. A tabela a seguir apresenta a distribuição dessas macrocategorias:

Tabela 4 – Padrões de Formulação de Busca no Catálogo – Macrocategorias

Macrocategoria	Quantidade	Frequência (%)
Autor	556	69,50%
Obra	133	16,63%
Assunto	69	8,63%
Híbrida	21	2,63%
Metadado	14	1,75%
Ruído	7	0,88%



Total	800	100,00%
-------	-----	---------

Fonte: Elaborada pela autora

Durante a análise, chamou atenção o fato de muitas ocorrências envolverem tentativas de buscas combinadas, nas quais o usuário insere duas ou mais palavras de categorias diferentes, sem inserção de operadores, talvez confiando que o sistema possa interpretar a relação entre elas. Também surgiram, em menor número, buscas com operadores booleanos digitados, frequentemente misturados a expressões em português ou aplicados de forma inconsistente. Esses padrões sugerem que, embora o catálogo ofereça suporte a operadores, a maior parte dos usuários adota estratégias semelhantes às utilizadas em motores de busca generalistas e plataformas comerciais, que realizam correções automáticas, ampliam resultados e interpretam aproximações semânticas. Em tais ambientes, mecanismos de busca mais flexíveis acabam compensando imprecisões e “poupando o tempo do leitor”.

Outro conjunto de entradas revela incerteza na formulação da busca: termos vagos, aproximações sucessivas, variações semânticas de um mesmo conceito e tentativas de refinamento lexical. Esses indícios apontam para diferentes níveis de familiaridade com o catálogo e com o próprio processo de consulta, mostrando que o sistema é usado tanto para buscas diretas quanto para exploração temática. Integrados, tais achados ajudam a explicar a predominância de buscas simples e a elevada dispersão dos termos observada na análise quantitativa, reforçando a necessidade de interfaces que tolerem variações linguísticas, erros e estratégias informais.

A ligação dos resultados quantitativos e qualitativos mostra que as buscas simples predominam porque os usuários recorrem a termos curtos e fragmentados. Além disso, que o uso mínimo de operadores reflete maior tendência a estratégias naturais de busca. Já a variedade de padrões observados indica diferentes níveis de familiaridade tanto com o catálogo quanto com a própria necessidade informacional.

4 CONSIDERAÇÕES FINAIS

Este estudo analisou os padrões de formulação de buscas no catálogo da Unicamp, combinando evidências quantitativas e qualitativas para compreender como os usuários expressam suas necessidades informacionais. Os resultados mostram que a predominância de buscas simples, a ampla dispersão dos termos e o uso quase exclusivo



de campos digitáveis refletem um comportamento de busca orientado por consultas livres, rápidas e pouco estruturadas.

A análise qualitativa revelou erros, variações ortográficas, fragmentos de referências, siglas e combinações híbridas, além de indícios de termos copiados de planos de ensino. Esses elementos aproximam o comportamento dos usuários de modelos mentais formados em motores de busca generalistas e plataformas comerciais, que oferecem correções automáticas e interpretações flexíveis, no entanto, essas expectativas nem sempre correspondem ao funcionamento literal do catálogo, o que não impede necessariamente a recuperação da informação, mas gera atritos na interação e demanda múltiplas tentativas de ajuste por parte do público.

Em conjunto, os achados indicam que os usuários recorrem a estratégias variadas, desde lembranças parciais até aproximações sucessivas, e que diferentes níveis de familiaridade com o sistema influenciam diretamente a formulação das consultas. Esses padrões reforçam a necessidade de interfaces mais tolerantes a variações linguísticas, capazes de apoiar a construção da busca e dialogar com práticas já consolidadas em outros ambientes digitais.

Os achados também apontam para a importância de explorar formas de ampliar a representação dos registros, de modo a apoiar buscas menos estruturadas e reduzir o impacto de variações linguísticas e imprecisões de escrita. Qualquer abordagem desse tipo, porém, exige critérios rigorosos para evitar ruídos, reduzir a necessidade de refinamento excessivo e manter a consistência da indexação. Reforça-se ainda o papel central da gestão de autoridades, especialmente as de Pessoa, já exercido na qualificação contínua do catálogo. Também se destaca o potencial de práticas de escuta ativa nos momentos de interação com o catálogo, como breves enquetes sobre termos e dificuldades, para retroalimentar a representação e enriquecer registros. Essa participação contínua dos usuários contribui para tornar a encontrabilidade mais inteligente e alinhada às tarefas dos usuários.

Diante de um cenário em que as pessoas realizam buscas por múltiplos dispositivos e modalidades, como voz, imagem e inteligência artificial, os catálogos ainda oferecem experiências limitadas, baseadas em interpretações literais e pouca mediação semântica. As unidades de informação têm, portanto, a oportunidade de aprimorar interfaces e mecanismos de recuperação, fortalecendo a garantia do usuário,



ampliando a autonomia de busca e reduzindo barreiras de acesso, especialmente diante da diversidade de competências, necessidades especiais, neurodivergências e modos de interação presentes nos públicos que atendem.

REFERÊNCIAS

- AIRES, R. V. X.; ALUISIO, S. M. Como incrementar a qualidade dos resultados das máquinas de busca: da análise de logs à interação em português. **Ciência da Informação**, v. 32, n. 1, 2003. Acesso em: 1 maio 2026. Disponível em: <https://doi.org/10.1590/S0100-19652003000100001>
- BASTOS, F. M.; VIDOTTI, S. A. B. G. A interação do usuário com catálogos bibliográficos on-line de bibliotecas universitárias. **Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação**, v. 15, 2014. Acesso em: 1 maio 2026. Disponível em: <https://brapci.inf.br/v/189575>
- BATES, M. J. The design of browsing and berrypicking techniques for the online search interface. **Online Review**, [S.L.], v. 13, n. 5, p. 407-424, 1989. Emerald. <https://doi.org/10.1108/eb024320>
- DIRETORIA EXECUTIVA DE TECNOLOGIA DA INFORMAÇÃO E COMUNICAÇÃO. CACia 2.0: plataforma de Inteligência Artificial da UNICAMP. 2026. Acesso em: 1 maio 2026. Disponível em: <https://plataformaia.unicamp.br/>. LAIPELT, R. C. F. A análise de logs como estratégia para a realização da garantia do usuário. **Em Questão**, v. 21, n. 3, 2015. Acesso em: 1 maio 2026. Disponível em: <https://doi.org/10.19132/1808-5245213.150-170>.
- LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2. ed. Brasília, DF: Briquet de Lemos/Livros, 2004. 452 p.
- LE COADIC, Y. F. **A ciência da informação**. 2. ed. Brasília, DF: Briquet de Lemos, 2004. 124p.
- MICROSOFT. **Microsoft Copilot**. [S. l.]: Microsoft, c2026. Acesso em 1 maio 2026. Disponível em: <https://copilot.microsoft.com/>.
- MORENO, F. P.; ARELLANO, M. A. M. Requisitos funcionais para registros bibliográficos - FRBR: uma apresentação. RDBCI: Revista Digital de Biblioteconomia e Ciência da Informação, v. 3, n. 2, 2005. Acesso em: 1 maio 2026. Disponível em: <https://periodicos.sbu.unicamp.br/ojs/index.php/rdbci/article/view/2052>.
- BARROS, T. H. B.; LAIPELT, R. C. F. Organização e representação do conhecimento em múltiplas abordagens. In: *MONTEIRO-KREBS, L.; LAIPELT, R. C. F.; ROCHA, R. P. Metodologia de análise de logs na ciência da informação: revisão de literatura e melhores práticas*. São Paulo, SP: Pimenta Cultural, 2022, cap. 6, p. 181. Acesso em: 1 maio. 2026. Disponível em: <http://dx.doi.org/10.31560/pimentacultural/2022.95613.6>.



NASCIMENTO, A. P. J.; SOUSA, A. C. M. A catalogação enquanto uma ação indireta de mediação da informação: um estudo a partir da percepção das bibliotecárias do SIBI da UFBA. **Biblionline**, v. 17, n. 3, 2021. Acesso em: 1 maio. 2026. Disponível em: <https://doi.org/10.22478/ufpb.1809-4775.2021v17n3.60323>.

RANGANATHAN, S. R. **As cinco leis da Biblioteconomia**. Brasília, DF: Briquet de Lemos Livros, 2009. 336 p.

ROSA, S. S.; LAIPELT; MONTEIRO-KREBS. Jornada do usuário: a análise de logs como metodologia para o estudo do comportamento de busca em sistemas de informação dos usuários do Portal LexML. **Revista Brasileira de Educação em Ciência da Informação**, [S.L.], v. 8, p. 1-12, 28 out. 2021. Acesso em: 1 maio. 2026. Disponível em: <http://dx.doi.org/10.24208/rebecin.v8i.255>.

SISTEMA DE BIBLIOTECAS DA UNICAMP. SBU. 2026. Acesso em: 1 maio. 2026. Disponível em: <https://www.sbu.unicamp.br/>.

TORINO, E.; *et al.* A relação entre arquitetura da informação e experiência do usuário sob a ótica dos pesquisadores da ciência da informação brasileira. **BIBLOS - Revista do Instituto de Ciências Humanas e da Informação**, v. 36, n. 1, 2022. Acesso em: 1 maio. 2026. Disponível em: <https://doi.org/10.14295/biblos.v36i1.13769>.

