

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

User journey: log analysis as a methodology for the study of information searching behavior of users of the Portal LexML

User journey: el análisis de los logs como metodología para estudiar el comportamiento de búsqueda en los sistemas de información de los usuarios del Portal LexML

Samuel Santos da Rosa

Universidade Federal do Rio Grande do Sul, Brasil

Rita do Carmo Ferreira Laipelt

Universidade Federal do Rio Grande do Sul, Brasil

Luciana Monteiro-Krebs

Universidade Federal do Rio Grande do Sul, Brasil

Submetido em: 23/04/2021

Aceito em: 14/06/2021

Publicado em: 28/10/2021

Licença:



Autor para correspondência: Samuel Santos da Rosa

Email: samuel.sdrosa@gmail.com

ORCID: <https://orcid.org/0000-0001-7209-6644>

Como citar este artigo:

ROSA, Samuel Santos da; LAIPELT, Rita do Carmo Ferreira; MONTEIRO-KREBS, Luciana. 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. **REBECIN**, São Paulo, v. 8, edição especial, p. 01-12, 2021. DOI: 10.24208/rebecin.v8i.255

RESUMO

O crescimento exponencial na produção e armazenamento da informação tem gerado uma constante modificação no comportamento informacional dos usuários, sendo necessária a utilização de métodos que auxiliem na coleta de informações sobre esses usuários. A análise de *logs* é um desses métodos ao permitir identificar as ações dos usuários de um sistema de informação. O presente trabalho tem como objetivo analisar, por meio da metodologia de análise de *logs*, o comportamento de busca dos usuários em sistemas de informação do Portal LexML. Compõem o corpus de análise 882 *logs* divididos em 236 jornadas de usuário. Percurso metodológico ocorreu em quatro etapas: Estruturação e Organização; Classificação e Seleção; Validação; Análise Estatística; e Análise das Jornadas dos Usuários. Como resultados foram observados: (a) baixa interação com o sistema e com alto índice de expressões modificadas na construção das expressões e estratégias de busca dos usuários; (b) a construção de estratégias que em sua maioria partem do geral para o específico; (c) uso de numeração jurídica como o recurso terminológico mais utilizado; (d) um baixo uso de booleanos e modificadores. Identificamos um perfil de usuário caracterizado por pesquisas simples com baixo índice de solicitação ao sistema. Conclui-se que a análise de *logs* é uma fonte de dados valiosa para o estudo do comportamento de busca em sistemas de informação, ao trazer uma gama de informações disponíveis passíveis de análise, fazendo desta uma metodologia importante no aperfeiçoamento do SRI e processo de indexação.

Palavras-Chave: Análise de Logs; Comportamento de Busca em Sistemas de Informação; Recuperação da Informação; Sistemas de Recuperação da Informação.

ABSTRACT

The exponential growth in the production and storage of information has generated a constant change in the informational behavior of users, requiring the use of methods that assist in the collection of information about these users. The log analysis is one of those methods, as it allows identifying the actions of users of an information system. The present work aims to analyze, using the log analysis methodology, the information searching behavior of users of the Portal LexML. They compose the

analysis corpus 882 logs divided into 236 user journeys Methodological path occurred in four stages: Structuring and Organization; Classification and Selection; Validation; Statistical analysis; and User Journey Analysis. As results were observed: (a) low interaction with the system and with high index of modified expressions in the construction of users expressions and search strategies; (b) the construction of strategies that mostly start from the general to the specific; (c) use of legal numbering as the most used terminological resource; (d) a low use of booleans and modifiers. We identified a user profile characterized by simple searches with a low rate of request to the system. It is concluded that the log analysis is a valuable data source for the study of information searching behavior, by bringing a range of available information that can be analyzed, making this an important methodology in the improvement of the SRI and indexing process.

Keywords: Log Analysis; Information Searching Behaviour; Information Retrieval; Information Retrieval Systems.

RESUMEN

El crecimiento exponencial en la producción y almacenamiento de información ha generado una constante modificación en el comportamiento informativo de los usuarios, siendo necesario el uso de métodos que ayuden a la recolección de información sobre estos usuarios. El análisis de registros es uno de estos métodos, que permite identificar las acciones de los usuarios de un sistema de información. El presente trabajo pretende analizar, mediante la metodología de análisis de logs, el comportamiento de búsqueda de los usuarios en los sistemas de información del Portal LexML. El corpus de análisis consta de 882 registros divididos en 236 jornadas de usuario. El curso metodológico se desarrolló en cuatro etapas: estructuración y organización, clasificación y selección, valoración, análisis estadístico y análisis de las jornadas de los usuarios. Como se observaron los resultados: (a) baja interacción con el sistema y con alto índice de expresiones modificadas en la construcción de las expresiones y estrategias de búsqueda de los usuarios; (b) la construcción de estrategias que parten mayoritariamente de lo general a lo específico; (c) uso de la numeración legal como recurso terminológico más utilizado; (d) un bajo uso de booleanos y modificadores. Identificamos

un perfil de usuario caracterizado por búsquedas sencillas con un bajo índice de solicitudes al sistema. Concluimos que el análisis de registros es una valiosa fuente de datos para el estudio del comportamiento de búsqueda en los sistemas de información, al aportar una serie de información disponible susceptible de ser analizada, lo que la convierte en una metodología importante en la mejora de los SRI y del proceso de indexación.

Palabras clave: Análisis de registros; comportamiento de búsqueda en sistemas de información; recuperación de información; sistemas de recuperación de información.

1 INTRODUÇÃO

As Tecnologias de Informação e Comunicação tem proporcionado um crescimento exponencial na produção e armazenamento da informação disponível na *internet*. Isso se deve principalmente pela utilização dos catálogos *on-line* das bibliotecas, repositórios digitais e bases de dados em geral, que permitem aos usuários fazerem suas pesquisas remotamente. Por outro lado, a falta de contato gera um desafio para a compreensão do processo de busca do usuário.

Torna-se importante, com isso, que os Sistemas de Recuperação da Informação (SRI), indexem considerando um público heterogêneo, de diferentes níveis de conhecimento, como destacado nos trabalhos de Laipelt (2015a) e Monteiro-Krebs (2016), que revelam a importância de envolver as linguagens especializada e leiga para o aperfeiçoamento das linguagens documentárias.

A análise de *logs* se estabelece como um método que permite identificar as ações dos usuários através das expressões de busca, ao registrar a interação direta do usuário com o sistema. Estas expressões

podem ser vistas como pegadas de informação digital do usuário, meio que este emprega para se comunicar com o sistema de informação.

A pesquisa parte de uma abordagem que reúne aspectos quantitativos, qualitativos e apresenta caráter empírico de análise e observação de dados. Seu *corpus* de análise é constituído pelas expressões de busca utilizadas pelos usuários do Portal LexML¹ (882 *logs* divididos em 236 jornadas de usuário) extraídas dos arquivos *log* no período de 2 a 24 de setembro de 2017.

No Portal LexML, cada pesquisa realizada pelos usuários gera um registro, chamado de *log*, que informa o número de IP (*Internet Protocol*), expressão de busca, entre outras informações. Ao recebermos os *logs*, o IP foi anonimizado através de embaralhamento automático, sem possibilidade de identificação do usuário. O percurso metodológico ocorreu através da análise dos *logs* em quatro etapas: Estruturação e Organização; Classificação e Seleção; Validação; Análise Estatística; e Análise das Jornadas dos Usuários.

A presente pesquisa busca, portanto, analisar, por meio da metodologia de análise de *logs*, o comportamento de busca dos usuários em sistemas de informação do Portal LexML, através das estratégias utilizadas e identificar as características de busca destes usuários.

2 COMPORTAMENTO DE BUSCA EM SISTEMAS DE INFORMAÇÃO

No estudo do comportamento de busca em sistemas de informação (*information searching behaviour*) são analisados todos os tipos de interações com o sistema, no nível da interação homem-computador (por

¹ Disponível em: <http://www.lexml.gov.br/>.

exemplo, clique em *links*) ou em nível intelectual (por exemplo, adotando uma estratégia de busca booleana) (WILSON, 2000). Em nossa análise observamos por ambos os níveis a relação do usuário com um SRI, que se estabelece pelo conjunto de expressões que representam a estratégia de busca de um usuário, constituindo assim o que chamamos de “Jornada do Usuário”.

O termo “estratégia de busca” é utilizado para se referir à forma mais global da tomada de decisão no processo de busca de informação, onde o usuário faz uso de diversos artifícios, como operadores booleanos ou outros conectivos sintáticos combinados com os termos da pesquisa e utilização direta de termos ou conjunto de termos (XIE; JOO, 2010).

O usuário ao interagir com o SRI delinea sua busca e a concretiza junto ao SRI, tendo como resultado a formulação de expressões. Desta forma, a observação e análise destas estratégias de busca tornam-se fator essencial para o estudo do comportamento de busca do usuário ao permitir verificar as ações realizadas.

O SRI deve se relacionar às etapas de armazenamento e indexação da informação (representação do conteúdo) para possibilitar ao usuário o desenvolvimento da melhor estratégia de busca junto aos sistemas de informação. O processo de RI de forma simplificada se dá por meio da especificação de três elementos: a representação dos documentos; a expressão de busca utilizada na pesquisa; e a função de busca. O segundo elemento do processo, a expressão de busca, é o meio que o usuário utiliza para se comunicar com o sistema utilizando para tal os mecanismos existentes nos sistemas. Essa expressão pode ser especificada em linguagem natural ou linguagem artificial, dependendo dos recursos oferecidos pelo sistema (CHOWDHURY, 2015).

Sendo assim observar e analisar o comportamento de busca em um SRI, utilizando a análise de *logs*, se torna relevante para verificar a potencialidade desta metodologia para conhecer e compreender as características dos usuários.

A análise de *logs* se coloca como uma metodologia que possibilita verificar os caminhos percorridos pelos usuários junto ao SRI (monitorar a interação entre o usuário e a base de dados) (JAMALI; NICHOLAS; HUNTINGTON, 2005; LAIPELT, 2015a, 2015b).

Vemos assim, que as informações geradas por esta interação, possuem potencialidades de nos fornecer um panorama de uso de determinada base de dados proporcionando a geração de estatísticas, e observação de aspectos do comportamento de busca dos usuários.

3 RESULTADOS E DISCUSSÃO

Nossas análises foram realizadas desconsiderando o “porquê de buscar”, mas sim “como buscou”, visamos com isso verificar a realidade dos usuários através dos rastros de informação registrados nos *logs*.

Inicialmente verificamos o número de consultas por usuário (quantidade de vezes que este interagiu com o sistema). Identificamos uma média de 3,74 solicitações por usuário. Tal resultado se assemelha ao encontrado por Jansen, Spink e Saracevic (2000), que ao analisarem 51.474 expressões de busca de 18.113 usuários do buscador Excite identificaram uma média de 2,84 solicitações por jornada (conjunto de estratégias de determinado usuário), tendo a maior parte dos usuários realizado de 1 a 4 consultas junto ao sistema.

Seguindo nossa análise, verificamos que 48,38% dos usuários modificaram sua expressão de busca, ou seja, os usuários realizam

buscas pelo mesmo assunto alterando a expressão utilizada quanto às lexias. O alto índice de modificação pode representar a presença de usuários com experiência na utilização dos mecanismos de busca ou insatisfação com relação ao que foi recuperado, conforme destaca Jansen, Spink e Saracevic (2000).

Posteriormente verificamos de que forma as modificações ao longo da sessão de busca ocorriam, comparando as expressões entre si para verificar se ficavam mais específicas ou gerais. Os resultados encontrados mostram que 31,78% dos usuários iniciam sua busca junto ao sistema com uma expressão abrangendo determinado assunto de forma geral e conforme realiza novas buscas vai modificando sua estratégia tornando-a mais específica. Isso ocorre possivelmente na tentativa de diminuir o número de documentos recuperados.

Verificamos também a utilização de recursos terminológicos, operadores lógicos e modificadores. Diante dos dados identificamos que o recurso terminológico com maior índice de utilização pelos usuários é a numeração jurídica, presente em 24,58% das jornadas. Também observamos uma baixa porcentagem de uso de operadores booleanos e modificadores, o que nos possibilita inferir que o usuário desconhece sua utilização ou forma de utilizá-los. Tal constatação corrobora com os resultados encontrados por Jansen, Spink e Saracevic (2000), que obtiveram baixos índices de utilização destes recursos.

Com os resultados encontrados nas análises anteriores conseguimos traçar um perfil dos usuários do Portal LexML, na perspectiva da sua interação com o SRI, sendo caracterizado por: realização de pesquisas simples com baixo índice de solicitação ao sistema; estratégia direcionada do geral para o específico; alto índice de modificação no decorrer da jornada; baixo uso de operadores booleanos

e modificadores; e uso da numeração jurídica como principal recurso para constituição das expressões de busca.

As características observadas nos possibilita inferir que o perfil dos usuários se assemelham aos encontrados nos trabalhos de Jansen, Spink e Saracevic (2000); e Spink e Xu (2000) que se caracterizam por realizarem pesquisas simples com baixo índice de solicitações ao sistema, formulações de consultas com poucos termos e pouco uso de operadores booleanos e modificadores. As principais diferenças observadas em relação a estes estudos dizem respeito ao índice de modificações nas expressões de busca, que em nosso estudo foi alto e nos estudos citados baixo. Este fato pode estar relacionado ao nosso contexto de análise, caracterizado pela linguagem jurídica que possui alto índice de especialidades e grande volume de informação e complexidade.

4 CONCLUSÃO

Nesta nova realidade de busca pela informação de forma remota, temos como principal consequência a dificuldade de conhecer e compreender como o usuário realiza seu processo de busca, como ele constrói sua estratégia, sendo necessário o uso de metodologias para analisar esta nova realidade. No presente trabalho, utilizamos uma destas metodologias, a análise de *logs*.

Com os resultados foi possível a elaboração de um perfil dos usuários do Portal LexML abrangendo todas as características identificadas, sendo caracterizado por: realização de pesquisas simples com baixo índice de solicitação ao sistema; estratégia direcionada em sua maioria do geral para o específico; alto índice de modificação no decorrer da jornada; baixo uso de operadores booleanos e modificadores; e uso

da numeração jurídica como principal recurso para constituição das expressões de busca.

Vemos assim que a análise de *logs* nos permitiu verificar de que forma os usuários realizam suas pesquisas junto a um SRI. Acreditamos que tal metodologia é uma fonte valiosa de coleta de termos e observação do comportamento de busca em sistemas de informação (análise e identificação de características dos usuários), ao possibilitar traçar os caminhos do usuário para obtenção da informação sem a interferência do bibliotecário e/ou do pesquisador.

Portanto, as informações extraídas dos *logs* nos proporcionaram uma visão rica do comportamento de busca em sistemas de informação ao expor o retrato da realidade da interação entre os usuários e os SRI, além de fornecer estatísticas detalhadas sobre as características de busca destes usuários possibilitando a comparação com trabalhos futuros.

REFERÊNCIAS

CHOWDHURY, G. Usuários da recuperação da informação. **InCID: Revista de Ciência da Informação e Documentação**, Ribeirão Preto, v. 5, n. 2, p. 4-33, set. 2014. Disponível em: <http://www.revistas.usp.br/incid/article/view/83858/87523>. Acesso em: 15 abr. 2021.

JAMALI, H. R.; NICHOLAS, D.; HUNTINGTON, P. The use and users of scholarly e-journals: a review of log analysis studies. **Aslib Proceedings: New Information Perspectives**, v. 57, n. 6, p. 554-571, 2005. Disponível em: <https://www.emeraldinsight.com/doi/abs/10.1108/00012530510634271>. Acesso em: 12 abr. 2021.

JANSEN, B. J.; SPINK, A.; SARACEVIC, T. Real life, real users, and real needs: a study and analysis of user queries on the web. **Information**

Processing and Management, Elmsford, NY, v. 36, p. 207-227, 2000.

Disponível em:

<https://www.sciencedirect.com/science/article/abs/pii/S0306457399000564>. Acesso em: 5 mar. 2021.

LAIPELT, R. C. F. **Metodologia para seleção de termos equivalentes e descritores de tesouros**: um estudo no âmbito do Direito do Trabalho e do Direito Previdenciário. 2015. Tese (Doutorado em Linguística Aplicada) – Programa de Pós-graduação em Linguística Aplicada, Universidade do Vale do Rio dos Sinos, São Leopoldo, 2015a.

Disponível em:

http://www.repositorio.jesuita.org.br/bitstream/handle/UNISINOS/4853/RI%20DO%20CARMO%20FERREIRA%20LAIPELT_.pdf?sequence=1&isAllowed. Acesso em: 12 abr. 2021.

LAIPELT, R. C. F. A análise de logs como estratégia para a realização da garantia do usuário. **Em Questão**, Porto Alegre, v. 21, n. 3, p. 150-170, 2015b. Disponível em:

<http://seer.ufrgs.br/index.php/EmQuestao/article/view/59806/36047>. Acesso em: 12 abr. 2021.

MONTEIRO-KREBS, L. **Terminologia e variação conceitual**: um estudo de interface com ontologias. 2016. Dissertação (Mestrado em Linguística Aplicada) – Programa de Pós-graduação em Linguística Aplicada, Universidade do Vale do Rio dos Sinos, São Leopoldo, 2016.

Disponível em:

http://www.repositorio.jesuita.org.br/bitstream/handle/UNISINOS/5053/Luciana+Monteiro+Krebs_.pdf?sequence=1. Acesso em: 13 abr. 2021.

SPINK, A; XU, J. L. Selected results from a large study of web searching: the excite study. **Information Research**: an international electronic journal, v. 6, n. 1, out. 2000. Disponível em:

<http://www.informationr.net/ir/6-1/paper90.html>. Acesso em: 15 abr. 2021.

WILSON, T. D. Human information Behavior. **Informing science research**, v. 3, n. 2, p. 49-53, 2000. Disponível em:

https://www.researchgate.net/publication/270960171_Human_Information_Behavior. Acesso em: 13 abr. 2021.



XIE; JOO, S. Transitions in search tactics during the web-based. **Journal of the American Society for Information Science and Technology**, v. 61, n. 11, p. 2188-2205, 2010. Disponível em: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.21391>. Acesso em: 13 abr. 2021.