

Quem leu este também leu...: sistema de recomendação na biblioteca universitária

Luciana Monteiro Krebs

Mestre em Linguística Aplicada (Universidade do Vale do Rio dos Sinos) e Bacharel em Biblioteconomia (Universidade Federal do Rio Grande do Sul).

Rafael Port da Rocha

Doutor em Computação pela Universidade Federal do Rio Grande do Sul e Professor Associado do Departamento de Ciência da Informação da Universidade Federal do Rio Grande do Sul (Brasil).

Cristina Ribeiro

Doutora em Informática pela Universidade Nova de Lisboa e Professora Auxiliar do Departamento de Engenharia Informática da Universidade do Porto (Portugal).

<http://dx.doi.org/10.1590/1981-5344/2496>

O Related books in Aleph OPAC é um sistema de recomendação para catálogos online de bibliotecas universitárias. O seu uso foi analisado utilizando abordagem quantitativa e metodologia descritiva. Foram realizados coleta, tratamento e análise dos dados provenientes de um log fornecido pela Universidade de Dundee. Os 13.654 registros e 6.347 sessões analisados resultaram no seguinte: a recomendação foi utilizada em 11% das sessões, nas quais 43,9% dos documentos visualizados foram oriundos de recomendação. De todos os registros de documentos visualizados, 9,6% foram oriundos de recomendação. As sessões com uso de recomendação foram, em média, dez segundos mais curtas do que as sessões sem uso de recomendação. Nas sessões com recomendação foram visualizados em média 4,30 documentos, enquanto nas sessões sem uso de recomendação a média é de 1,88. O uso de mais de um tipo de recomendação teve baixa adesão, pois 82% das sessões com recomendação apresentam o uso de apenas

um tipo de recomendação. Para o tipo de recomendação mais utilizado há dois resultados: "Trabalhos relacionados incluem" aparece em mais sessões (348), porém o tipo "Pessoas que retiraram esta obra também retiraram" registra o maior número de registros no log (584).

Palavras-chave: *Sistemas de recomendação; OPACs; Avaliação de sistemas de recuperação da informação.*

People who borrowed this have also borrowed: recommender system in academic library

The paper analyses the use of recommender systems in academic libraries, examining the use of the "Related books in Aleph OPAC" recommendation system for academic libraries' online catalogues. A quantitative approach and descriptive methodology is used to collect, process and analyse the data from a usage log provided by the University of Dundee. The analysis of 13,654 posts and 6,347 sessions provided the following observations: the recommendation was used in 11% of the sessions, and 43.9% of the recorded document views on those sessions were generated by recommendation. 9.6% of the records of document views, were derived from recommendation. Sessions using recommendations were on average 1 minute 18 seconds shorter than the sessions without recommendations. In sessions with recommendation 4.30 records were viewed on average while in sessions without recommendation the average is 1.88. Using more than one type of recommendation is not common, as 82% of the sessions with recommendation have recorded the use of only one kind of recommendation. The analysis of recommendations by kind provided two results: "Related works include" appears in more sessions (348), while "People who borrowed this work also borrowed" has the highest number of posts (584).

Keywords: *Recommender systems; OPACs; Evaluation of information retrieval systems.*

Recebido em 31.08.2015 Aceito em 02.02.2017

1 Introdução

Catálogos *online* OPACs (*Online Public Access Catalogues*) são recursos que auxiliam os usuários na busca e recuperação da informação em bibliotecas. A utilização de sistemas de recomendação em conjunto com os OPACs amplia as possibilidades de recuperação para além da busca, através de sugestões de leitura para os usuários.

A recomendação é uma forma particular de filtragem de informação que explora os comportamentos do passado e o perfil do usuário para gerar uma lista de itens de informação que é particularmente adaptada às preferências de um usuário final. (ASSOCIATION FOR COMPUTING MACHINERY, 2012). Os sistemas de recomendação permitem a personalização da oferta de produtos ou serviços para cada usuário (real ou potencial), agindo de forma proativa sem que o usuário precise formular uma expressão de busca para encontrar um produto ou serviço (SPEROTTO, 2010). São utilizados “[...] para identificar usuários, armazenar suas preferências e recomendar itens que podem ser produtos, serviços ou conteúdos, de acordo com suas necessidades e interesses.” (BARCELLOS; BRANDÃO; MUSA, 2012, documento *online*).

Entre as técnicas utilizadas nos sistemas de recomendação, destacam-se a **filtragem baseada em conteúdo (FBC)** e a **filtragem colaborativa (FC)**. A primeira utiliza a similaridade entre itens para gerar recomendações e é mais usada para textos (livros, notícias, etc.), partindo do princípio de que os usuários tendem a interessar-se por itens similares aos que despertaram o seu interesse no passado. Já a segunda utiliza a similaridade entre os usuários para gerar recomendações, gerando predições através da análise de informações sobre seu comportamento (TORRES, 2004; LICHTNOW *et al.*, 2006).

A presente pesquisa investigou durante o ano de 2013 o uso do sistema de recomendação *Related Books in Aleph OPAC* (PHILLIPS, [2009?])¹ com base em dados das bibliotecas da Universidade de Dundee (Escócia), sendo o objetivo geral analisar o uso desse sistema de recomendação para catálogos *online* de bibliotecas universitárias. A pesquisa apresenta os seguintes objetivos específicos:

a) Verificar em que medida a recomendação é utilizada;

Identificar eventuais diferenças do uso ou não de recomendação nas interações do usuário com o sistema, no que diz respeito ao tempo da interação e à quantidade de registros visualizados;

¹ Criado por Matthew Phillips, foi implementado primeiramente na Universidade de Dundee e posteriormente adotado em Portugal e Brasil. Disponível em: <<http://www.exlibrisgroup.org/display/AlephCC/Related+books+links+in+Aleph+OPAC>>. Acesso em: 10 jun. 2014. (PHILLIPS, 2010). Outro estudo feito a partir dos dados coletados na University of Dundee pode ser visualizado no link: <<http://igelu.org/wp-content/uploads/2010/09/recommendations-in-aleph.pdf>>. Acesso em: 18 ago. 2015.

b) Verificar o uso dos tipos de recomendação, levando em consideração as interações dos usuários com o sistema e o número de registros visualizados;

c) Relacionar o uso de recomendação e de consultas nas interações dos usuários com o sistema, considerando a visualização de registros.

As interações dos usuários com o sistema, conforme será detalhado na Seção 2 deste artigo (Aspectos metodológicos) são registradas e agrupadas pelo que chamamos de sessão, compreendendo um determinado conjunto de ações (representadas nos *posts*) realizadas pelo mesmo usuário em um determinado período de tempo. O número de sessão fica registrado no *post*, o que permite agrupar todos os *posts* da mesma sessão para analisar as interações do usuário.

O sistema de recomendação *Related Books in Aleph OPAC* possui três tipos de recomendação para o item que está sendo visualizado:

Outras edições incluem (EDIÇÕES): FBC que considera similares ao item visualizado aqueles documentos com o mesmo *International Standard Book Number* (ISBN) e que estão presentes no acervo da biblioteca. As recomendações são listadas em ordem decrescente de ano de publicação, com um limite de cinco itens. Os cálculos são atualizados mensalmente.

Pessoas que retiraram esta obra também retiraram (EMPRÉSTIMO): FC que considera recomendáveis aos usuários do item visualizado aqueles documentos que foram emprestados aos usuários que também retiraram este mesmo item. O item recomendado deve ter sido emprestado a pelo menos cinco usuários diferentes. O ordenamento dos itens na lista se dá pela quantidade de usuários mutuamente comuns ao item visualizado. As ligações são recalculadas semanalmente².

Trabalhos relacionados (ASSUNTO): FBC que considera similares ao item visualizado aqueles documentos com pelo menos três cabeçalhos de assunto ou três números de classificação (registro MARC) em comum com o item visualizado. Os *links* são recalculados quinzenalmente.

Salienta-se que estas recomendações não são geradas em tempo real, mas calculadas periodicamente conforme indicado em cada uma, com base no histórico de empréstimos. Os três tipos de recomendação serão identificados a seguir por EDIÇÕES, EMPRÉSTIMO e ASSUNTO, respectivamente.

A Seção 2 deste artigo apresenta aspectos metodológicos envolvendo a coleta, a preparação e a análise dos dados. Nas seções 3 a 6 os dados são apresentados, contextualizados e discutidos, para análise dos resultados obtidos na pesquisa no que diz respeito, respectivamente, à adesão geral dos usuários à recomendação, ao uso de recomendações em sessões pelos usuários, ao uso dos tipos de recomendações e ao uso de

² Esta recomendação não é realizada para itens que não podem ser emprestados, tais como e-books e obras de referência, pois o cálculo é feito sobre a tabela de empréstimos da biblioteca.

recomendações em consultas. Na Seção 7 são apresentadas considerações finais.

2 Aspectos metodológicos

Este estudo tem como fonte de análise os *logs*³ cujas entradas descrevem acessos a registros do catálogo das bibliotecas da Universidade de Dundee⁴. A escolha dos autores em trabalhar com dados desta instituição deve-se ao pioneirismo da Universidade de Dundee (Escócia), uma vez que foi a primeira biblioteca a implementar a ferramenta de recomendação *Related Books in Aleph OPAC*, especialmente criada para funcionar no sistema Aleph. A ferramenta foi disponibilizada no ambiente de colaboração entre desenvolvedores e clientes do Aleph, denominado *EL Commons CodeShare Beta* e adotada em outras universidades ao redor do mundo, incluindo Reino Unido e Brasil. (PHILLIPS, 2010).

O método de análise adotado, de natureza quantitativa, segue uma abordagem de *transaction log analysis* (TLA) (JANSEN, 2006), que envolve as etapas de coleta, preparação e análise. A relevância do uso de *logs*, segundo Nicholas *et al.* (2004, p. 119, tradução nossa), é que eles “[...] disponibilizam um registro direto e imediato do que as pessoas têm feito no sistema: não o que dizem que poderiam ou deveriam ter feito, não o que lhes foi solicitado dizer, não o que eles acham que fizeram.”

Para a coleta foram reunidas duas amostras de *log* da *Main Library* da *University of Dundee*, compostas por URL que registram os acessos ao catálogo da biblioteca. O período de levantamento compreendeu 30 dias: de 04 a 18 de dezembro de 2012 e de 22 de abril de 2013 a 6 de maio de 2013. No total foram coletados 16.806 registros (URL) que abrangem registros de documentos visualizados, provenientes de consultas e de recomendação. Respeitando a privacidade dos usuários, os arquivos texto recebidos não continham os identificadores IP dos computadores utilizados.

Na etapa de preparação foi utilizado o software extrator⁵ que limpou os registros de *log* e os organizou em tabelas de banco de dados relacional. Foram realizadas consultas ao banco de dados, através da linguagem SQL, cujos resultados foram exportados para planilhas e transformados em gráficos, que serviram de base para a análise de dados.

No nosso trabalho os registros de documentos analisados correspondem a *posts* armazenados no *log*. Cada *post* corresponde a uma tela que exibe um registro do catálogo. Cada *post* do *log* contém diversas informações, entre elas: a sessão, a data da interação e a operação do usuário que determinou o acesso à tela, que pode ter sido uma consulta

³ Um *log* de transações é um arquivo [...] que registra as comunicações (transações) entre um sistema e os usuários desse sistema. (JANSEN, 2006, p. 408, tradução nossa).

⁴ Mais informações: < <http://www.dundee.ac.uk/> >. Acesso em: 9 mar. 2015.

⁵ O software elimina as URLs repetidas do *log*, ordena cronologicamente as URLs, conta o intervalo de tempo entre o primeiro e o último registro de uma sessão (em segundos) e exporta os dados para uma tabela (formato .CSV) organizando-os em colunas por tipo de informação. Agradecemos a Vicente Grassi Filho pela construção do software, exclusivamente para esta pesquisa.

(situação 1), uma recomendação (situação 2), ou outro acesso, conforme detalhado a seguir.

A situação 1 (ver Figura 1) ocorre quando o usuário está navegando sobre a lista de itens do resultado de uma consulta e solicita a visualização de um determinado item dessa lista. Nesse caso, o *post* correspondente ao item da lista visualizado contém um parâmetro que identifica a consulta. Um *post* desse tipo será chamado neste trabalho de RegCON, para registro de consulta⁶.

A situação 2 (ver Figura 1) ocorre quando o usuário está observando um determinado item e solicita a visualização de uma recomendação para este item. Nesse caso, o *post* correspondente ao item recomendado visualizado contém um parâmetro que indica qual o tipo de recomendação que motivou o acesso ao registro, isto é, EDIÇÕES, EMPRÉSTIMO ou ASSUNTO. Um *post* que corresponde a uma visualização de um item que teve como origem uma recomendação é chamado de RegREC.

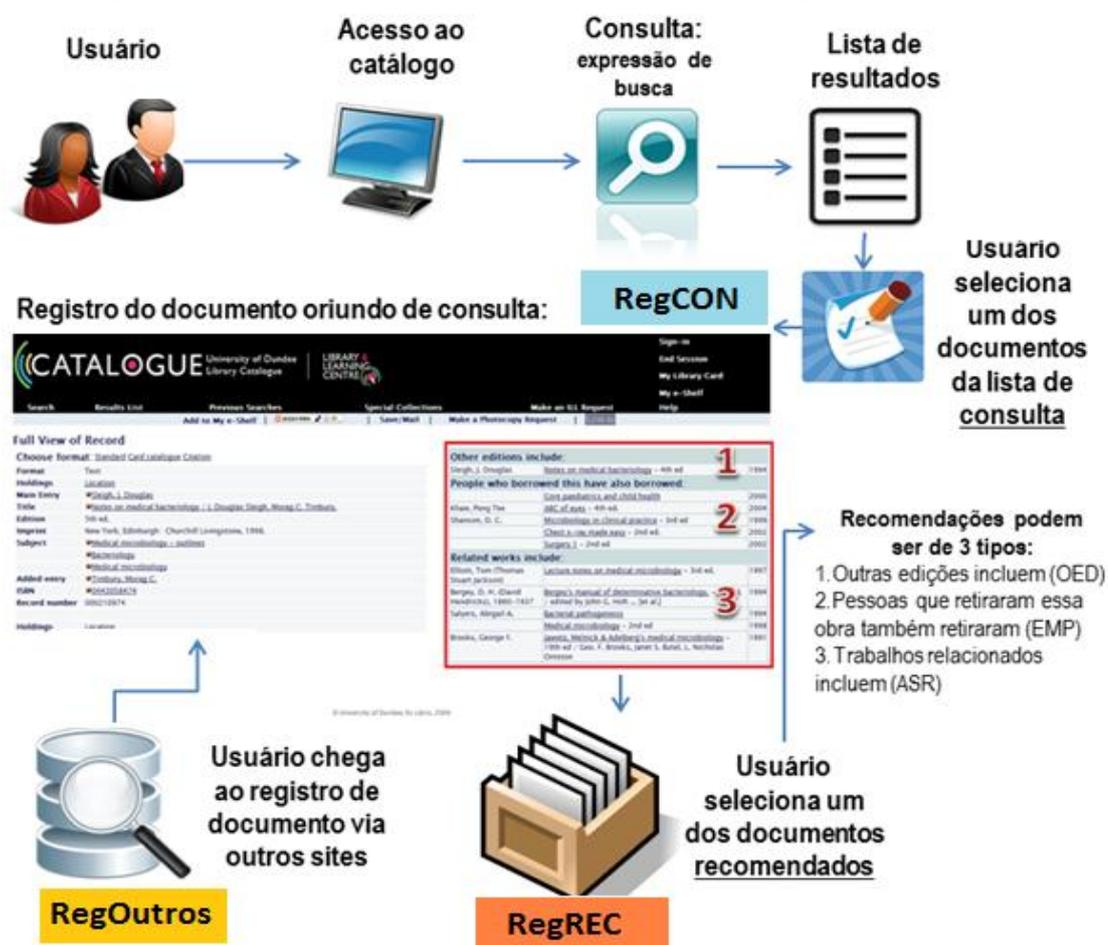


Figura 1 – Esquema de navegação do usuário no catálogo, consulta e recomendação

Fonte: Dados da pesquisa.

⁶ No log fornecido não constam posts que correspondem a telas que registram as expressões de busca, nem a listas que contêm resultados de consultas. Entretanto, as consultas podem ser identificadas, pois seus códigos identificadores estão presentes nos posts do tipo RegCON.

No *log* há também *posts* que correspondem à visualização de registros de documentos que não tiveram como origem uma consulta ou uma recomendação. Esses *posts* foram classificados como RegOUTROS. Um *post* desse tipo representa a visualização, por exemplo, de um item cujo acesso teve origem em um buscador (como Google) ou ocorreu através de *link* direto de uma página qualquer da web.

Os *posts* podem conter informações de sessão. O conjunto das interações que ocorreram entre o usuário e o catálogo *online* em um determinado período de tempo é chamado de sessão. Uma sessão pode conter várias interações de que são exemplo as consultas, visualizações de registros de documentos do catálogo e reserva de itens do acervo, entre outras. Uma sessão é identificada pelo sistema através de um número, é iniciada após a ativação do navegador no momento do primeiro acesso ao sistema, e é encerrada quando a aba de visualização do catálogo no navegador é fechada. Este trabalho investiga somente interações que ocorrem em sessões. A sessão foi escolhida como unidade de medida porque engloba várias interações de um usuário com o sistema, permitindo uma análise baseada em período de tempo e na ligação entre as várias interações realizadas neste intervalo. Por isso, para este trabalho não foram considerados *posts* que não estão ligados a uma sessão.

Os *posts* considerados (ou seja, gerados em sessão) contêm parâmetros que indicam a data da interação (em dia, hora, minuto e segundos). Com isso, o tempo de uma sessão pode ser calculado pelo intervalo de tempo decorrente entre o primeiro e o último *post* da sessão.

Em resumo, cada *post* contém: (a) a identificação da sessão; (b) a data e hora do acesso; e (c) a origem do acesso, isto é, se o *post* é do tipo RegCON, RegREC ou RecOUTROS. *Posts* do tipo RegREC contêm também o tipo de recomendação, isto é, EDIÇÕES, EMPRÉSTIMO ou ASSUNTO; e *posts* do tipo RegCON contêm ainda o identificador da consulta.

Na etapa de análise, para atender aos objetivos desta pesquisa, o construto metodológico apresentado no Quadro 1 foi elaborado. Para maior clareza e facilidade de localização dos resultados da pesquisa ao longo do texto, cada análise feita para atingir os objetivos foi identificada com um indicador alfanumérico (A1, A2, A3, B1, B2, etc.).

Quadro 1 - Objetivos específicos e métodos de análise

Objetivo específico	Análise
(a) Verificar em que medida a recomendação é utilizada (Seção 3);	<ul style="list-style-type: none"> • Calcular a proporção de registros visualizados a partir de recomendação considerando o universo de registros visualizados (A1) • Verificar se sessões com recomendação ocorrem em maior ou menor quantidade que sessões sem recomendação (A2) • Considerando somente sessões em que houve recomendação, verificar dentre todos os registros visualizados a proporção daqueles que tiveram como origem uma recomendação (A3)

<p>(b) Identificar eventuais diferenças do uso ou não de recomendação nas interações do usuário com o sistema, no que diz respeito ao tempo da interação e à quantidade de registros visualizados (Seção 4);</p>	<ul style="list-style-type: none"> • Verificar se sessões com recomendação possuem duração menor ou maior que sessões sem recomendação (B1) • Verificar se sessões com recomendações possuem maior ou menor quantidade de registros visitados que sessões sem recomendação (B2)
<p>(c) Verificar o uso dos tipos de recomendação, levando em consideração as interações dos usuários com o sistema e o número de registros visualizados (Seção 5);</p>	<ul style="list-style-type: none"> • Verificar a quantidade de sessões por tipo de recomendação (C1) • Verificar a quantidade de registros visitados por tipo de recomendação (EDIÇÕES, EMPRÉSTIMO e ASSUNTO) (C2)
<p>(d) Relacionar o uso de recomendação com o uso de consultas nas interações dos usuários com o sistema, considerando a visualização de registros (Seção 6).</p>	<ul style="list-style-type: none"> • Verificar número de registros visualizados por consulta (D1) • Verificar se em sessões com consulta houve mais ou menos registros visualizados que em sessões sem consulta (D2) • Verificar a relação de busca por consulta e uso de recomendação, através da comparação dos registros visualizados em sessões em que houve consulta com os registros visitados em sessões em que houve recomendação (D3)

Fonte: Dados da pesquisa.

Neste instrumento, considera-se sessão com recomendação uma sessão em que pelo menos um item foi visualizado através de uma recomendação (que contém pelo menos um *post* do tipo RegREC) e considera-se uma sessão com consulta aquela em que pelo menos um item visualizado tem como origem uma consulta (*post* do tipo RegCON).

3 Medindo o uso geral da recomendação

No total dispôs-se de 16.806 *posts*. Dentro deste universo, 13.654 *posts* foram realizados em 6.347 sessões. O *log* apresenta 10.063 consultas, e em relação aos registros de documentos visualizados, 10.685 são oriundos de consultas (RegCON), 1.305 são oriundos de recomendação (RegREC) e 1.664 de outras fontes, como sites de busca e *links* externos (RegOUTROS). Salienta-se que do total de 13.654 registros de documentos visualizados, 1.305 são oriundos de recomendação, o que resulta em 9,6% dos acessos **(A1)**.

A análise das sessões permite identificar se o usuário, num processo motivado de busca, utilizou ou não serviço de recomendação. Tendo como foco a sessão, a análise envolve a proporção de sessões em que houve uso de recomendação, tendo como base o total de sessões. No *log* fornecido encontram-se 13.654 *posts* em 6.347 sessões⁷. Para analisar o uso de recomendação em sessões, foram identificados dois conjuntos de sessões: o conjunto "sessões com recomendação", que é composto por sessões que possuem pelo menos um acesso a registro de documento que teve como origem uma recomendação (RegREC); e o conjunto "sessões sem recomendação", composto por sessões que não contêm nenhum

⁷ Os demais foram acessos sem registro de sessão, conforme já explicado na Seção 2 deste trabalho.

acesso de registro a partir de recomendação. As sessões com recomendação representam 11% do *log*, totalizando 690 sessões. Já as 5.657 sessões sem recomendação representam 89% do *log* (**A2**).

Considerando somente as sessões com recomendação (690), analisou-se também a proporção de registros de documentos visualizados nessas sessões que tiveram como origem a recomendação (RegREC), tendo como base o total de registros de documentos visualizados.

Tabela 1 – Registros de documentos visualizados nas sessões com recomendação (A3)

	Total	Percentual	Média de documentos por sessão	% considerando o total de registros (13.654)
Registros de documentos visualizados via recomendação (RegREC)	1.305	43,9%	1,89	9,5%
Registros de documentos visualizados sem recomendação (RegCON e RegOutros)	1.664	56,0%	2,41	12,8%
Total	2.969	100%	4,3	21,74%

Fonte: Dados da pesquisa.

Analisando as sessões em que houve recomendação (Tabela 1) observa-se que, dos 2.969 registros de documentos visualizados, 1.305 (43,9%) são provenientes de recomendação, enquanto 1.664 (56%) não são. Ou seja, quem utiliza a recomendação, usa bastante, visitando documentos do tipo RegREC em quase metade de toda a sessão (**A3**).

Em média, nas sessões com recomendação, são visitados 1,89 documentos provenientes de recomendação (o que equivale a 9,5% do total de registros visualizados, considerando também a sessões sem recomendação), e 2,41 documentos não provenientes de recomendação (o que equivale a 12,8% do total de registros).

4 Recomendações em sessões de usuários

Esta seção do trabalho está voltada exclusivamente para análise das sessões dos usuários, levando em conta a sua duração (sessões rápidas, curtas e longas) e o tempo médio de duração das sessões. Essa proporção também é investigada com relação à quantidade de registros visitados, isto é, se houve um menor/maior acesso a registros de documentos em sessões em que houve o uso da recomendação.

Para o cálculo da duração das sessões, foram excluídas sessões com 0 (zero) segundos ou mais de 24 horas⁸, restando 2.522 sessões (44,5% do total). Quanto à duração, as sessões foram classificadas em: (i) sessões com menos de 10 segundos; (ii) sessões com duração entre 10 segundos e 10 minutos; e (iii) sessões com duração de mais de 10 minutos. O Gráfico 1 compara as sessões que tiveram alguma

⁸ Sessões de mais de 24 horas são aquelas em que o usuário deixou o navegador sempre ativo por mais de um dia. As sessões de zero segundos correspondem a interações de um único acesso.

recomendação com aquelas que não tiveram recomendação, com relação à sua duração. Pode-se observar que tanto as sessões com recomendação quanto as sessões sem recomendação apresentam comportamento similar.

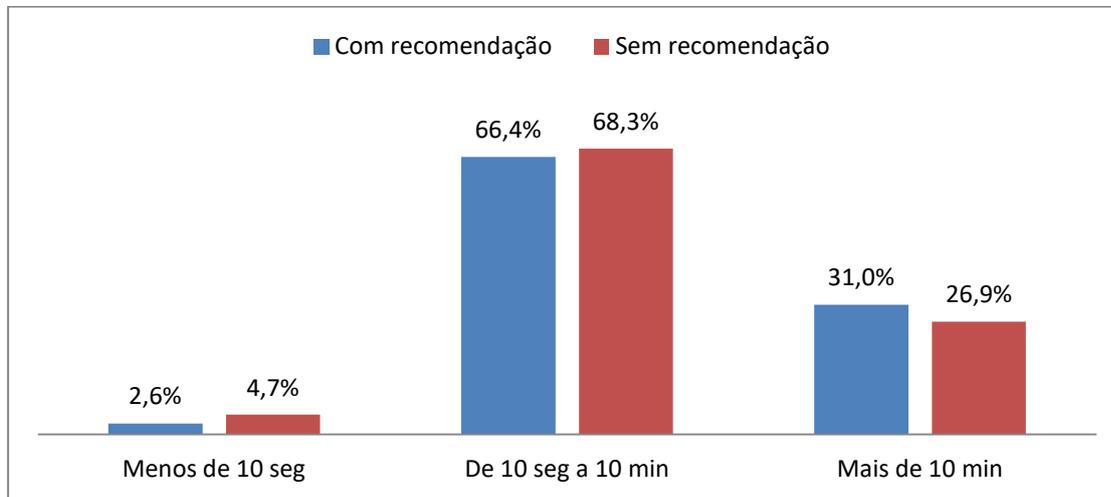


Gráfico 1 – Sessões por intervalo de tempo

Fonte: Dados da pesquisa.

Em primeiro lugar, estão as sessões de média duração, com no mínimo 10 segundos e no máximo 10 minutos. No *corpus* analisado, as sessões de média duração dominam tanto as sessões com recomendação quanto as sessões sem recomendação, representando 66,4% e 68,3% das sessões, respectivamente. As sessões de longa duração vêm em segundo lugar, representando 31% das sessões com recomendação e 26,9% das sessões sem recomendação. As sessões longas têm mais de dez minutos de duração. Por fim, as sessões curtas representam a menor parcela de sessões e são aquelas que têm até 10 segundos de duração. Das sessões com recomendação apenas 2,6% são sessões curtas, e das sessões sem recomendação as sessões curtas representam 4,7% do *corpus*.

Assim, vê-se interessante aprofundar a análise temporal das sessões de média duração, aquelas entre 10 segundos e 10 minutos de duração, por representarem a maior parcela de sessões do conjunto de registros coletados. Nesta contagem, portanto, foram desconsideradas as sessões muito curtas (menores de 10 segundos) e muito longas (maiores de 10 minutos), para evitar distorção da média.

Foram 352 sessões com recomendação com no mínimo 10 segundos e no máximo 10 minutos de duração. Nesse grupo, a média de duração foi de 176 segundos, ou seja, dois minutos e 57 segundos. Já as sem recomendação com média duração somaram 1.369 sessões. A média de duração nesse grupo foi de 186 segundos, ou três minutos e seis segundos. Pode-se afirmar que as sessões com registro de uso de recomendação da amostra são, em média, dez segundos mais rápidas do que as sessões sem recomendação. **(B1)**. Essa diferença é pequena, isto é, o uso de recomendação não implica em uma redução substancial no

tempo da sessão, quando comparado com sessões em que não houve recomendação. A média de tempo das sessões é maior quando não se usa recomendação, mas sem apresentar diferença significativa.

Tabela 2 – Registros de documentos visualizados em sessões com e sem recomendação (B2)

Sessões	Sessões	% de Sessões	Documentos visitados	% Doc. Visitados	Doc. Visitados por sessão
Com recomendação	690	10,8%	2.969	21,7%	4,30
Sem recomendação	5.657	89,1%	10.685	78,2%	1,88
Total	6.347	100%	13.654	100%	2,15

Fonte: Dados da pesquisa.

Quanto ao número médio de registros de documentos visitados por sessão, considerando todos os 13.654 documentos visitados nas 6.347 sessões, o valor obtido foi de 2,15 registros (Tabela 2). Nas 690 sessões que contêm recomendação, a média de registros visitados por sessão é bastante superior, atingindo o valor de 4,3 registros. Já considerando somente as 5.657 sessões em que não houve o uso da recomendação, a média de documentos visualizados por sessão é 1,88 (**B2**). Analisando ainda a Tabela 2, pode-se observar que nas sessões com recomendação, embora as mesmas representem 11% do total de sessões, ocorre visita a uma quantidade maior de registros de documentos, 21,7% do total de visitas.

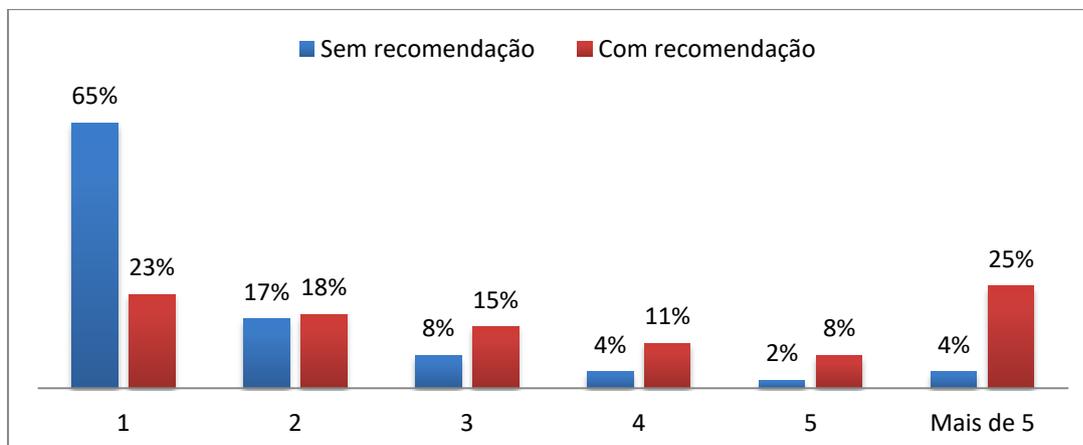


Gráfico 2 – Percentual de sessões com e sem recomendação com os respectivos números de registros de documentos visualizados

Fonte: Dados da pesquisa.

As sessões foram analisadas quanto ao número de registros nelas visualizados (Gráfico 2). Considerando somente as 5.657 sessões em que não há visita a documentos via recomendação, em 65% das sessões houve a visualização de somente um documento. Já considerando somente as 690 sessões em que houve visita a documentos via recomendação, em 23% delas houve a visualização de somente um documento.

Ao comparar sessões com e sem recomendação, pode-se constatar que 25% das sessões com recomendação se concentram na visualização de mais de cinco registros de documentos, ao contrário das sessões sem recomendação, onde a maioria das sessões (65%) apresenta visualização de apenas um registro de documento (Gráfico 2). Disso, pode-se perceber um quadro diferente com relação aos dois tipos de sessão, uma vez que as sessões com recomendação tendem a apresentar visitas a mais de um documento por sessão, enquanto apenas 4% das sessões sem recomendação possui visualização de mais de cinco registros de documentos.

5 Uso dos tipos de recomendação

Como já mencionado, 690 sessões do *log* apresentam uso de recomendação, resultando na visualização de 1.305 registros que tiveram origem na recomendação. Esses registros são do tipo RegREC. As recomendações que levaram aos registros do tipo RegREC, por sua vez, são classificadas em três tipos: EDIÇÕES, EMPRÉSTIMO e ASSUNTO. Primeiramente investigou-se o comportamento desses três tipos de recomendação em sessões, envolvendo o número de sessões que usam pelo menos uma recomendação de cada tipo. A seguir, investigou-se a proporção do uso de cada tipo tendo como base o total de registros que foram visualizados via recomendação.

A parte A do Gráfico 3 apresenta os totais de sessões de acordo com o tipo de recomendação. Como algumas sessões apresentam mais de um tipo de recomendação, a soma de todos os números do gráfico supera o número total de sessões com recomendação (690). No caso das sessões, verifica-se que o tipo de recomendação mais utilizado (348 sessões ou 42,3%) é ASSUNTO, que recomenda documentos com assuntos semelhantes. Essa filtragem baseada no conteúdo permite que o usuário visualize documentos que podem não possuir um assunto usado na expressão de busca formulada **(C1)**.

Com quase o mesmo número de sessões (346 ou 42,1%) ocorre a recomendação do tipo EMPRÉSTIMO. Esta recomendação utiliza a tabela de empréstimos. Neste caso, o que liga um documento ao outro são os usuários que os retiraram (filtragem colaborativa). Podemos afirmar que recomendação por ASSUNTO é tão usada quanto a recomendação baseada em similaridade entre usuários (EMPRÉSTIMO).

Dos 1.305 registros visualizados a partir de recomendação, 584 (44,8%) são tipo EMPRÉSTIMO, 563 (43,1%) são do tipo ASSUNTO e 158 (12,1%) são do tipo EDIÇÕES (conforme representado na parte B do Gráfico 3). A preferência por registros do tipo EMPRÉSTIMO (44,8%) aponta uma escolha pela recomendação social, porque esse tipo de recomendação é inspirado no conceito de similaridade entre usuários (aqueles que tenham retirado, por empréstimo, pelo menos cinco documentos em comum) **(C2)**.

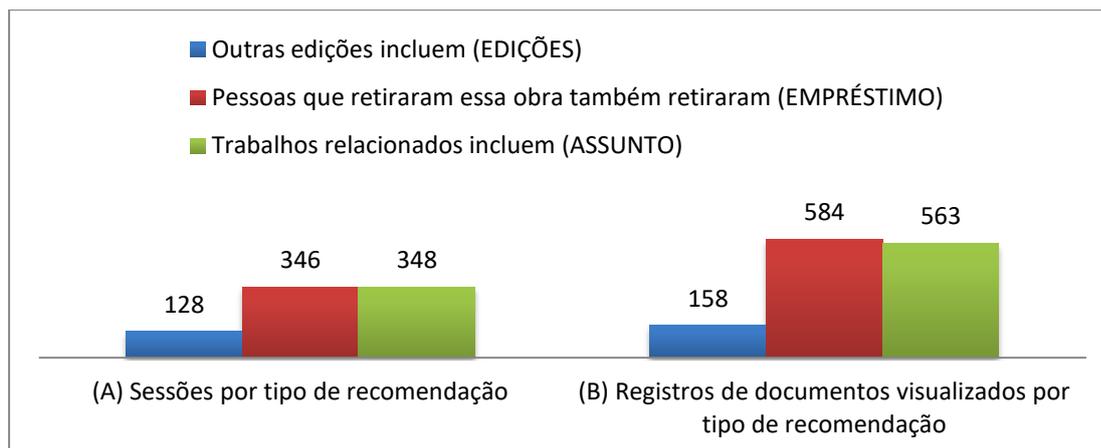


Gráfico 3 – Uso dos tipos de recomendação: registros de documentos visualizados e sessões (C1, C2)

Fonte: Dados da pesquisa.

Nas 690 sessões com recomendação, em média são visitados 1,89 registros de documentos originários de recomendação (RegREC) (ver Tabela 1). Desses 1,89 registros, em média, 0,85 são registros de documentos oriundos da recomendação do tipo EMPRÉSTIMO, 0,82 são registros provenientes da recomendação do tipo ASSUNTO e 0,23 são registros originários de recomendações do tipo EDIÇÕES.

Em suma, a análise do tipo de recomendação mais utilizado foi realizada com base nas sessões e nos documentos visualizados. Na análise por sessão, o tipo de recomendação mais utilizado é ASSUNTO (presente em 348 sessões), seguido de perto do tipo EMPRÉSTIMO (presente em 346 sessões) **(C1)**. Na análise por registros de documentos visualizados, constatou-se que os usuários visualizaram mais registros de documentos do tipo EMPRÉSTIMO do que qualquer outro tipo de recomendação (584). O segundo, por ASSUNTO, apresentou 563 registros, e o terceiro, EDIÇÕES, 158 registros de documentos visualizados **(C2)**.

Os números indicam que as recomendações do tipo ASSUNTO e EMPRÉSTIMO são as mais usadas, em proporção similar. Embora o número de sessões que contêm recomendação do tipo ASSUNTO supere o número de sessões que contêm recomendação por EMPRÉSTIMO, esta situação se inverte discretamente quando se trata de visualização de registros de documentos. Ou seja, os usuários preferiram utilizar mais frequentemente o tipo de recomendação ASSUNTO, mas os que optaram por utilizar o tipo de recomendação EMPRÉSTIMO acabam visualizando mais documentos na mesma sessão. Pode-se afirmar que os três tipos de recomendação são importantes, embora EDIÇÕES tenha proporção menor (pouco menos da metade dos outros), o que pode ser causado por um número menor de recomendações disponíveis desse tipo (pois depende de fatores como variedade do acervo e coesão na indexação).

A recomendação por EMPRÉSTIMO está ligada ao conceito de serendipidade (do termo em inglês, *serendipity*), pois permite que o

usuário encontre algo útil ou agradável pelo qual não estava procurando. Esse tipo de recomendação mostra que usuários que fizeram o empréstimo do documento visualizado também fizeram o empréstimo de outros documentos, que não necessariamente têm o mesmo cabeçalho de assunto. Isto significa que o usuário pode visualizar sugestões de documentos de usuários que têm algo em comum com quem faz a pesquisa. Esta similaridade entre usuários enquadra-se na filtragem colaborativa e leva em conta não apenas o interesse momentâneo do usuário (representado pela expressão de busca), mas o seu perfil. Nos dados disponíveis, a recomendação por ASSUNTO é tão usada quanto a recomendação por EMPRÉSTIMO. Esse tipo de recomendação permite também ao usuário a descoberta de documentos correlacionados por assuntos que não foram especificados em suas expressões de busca.

6 Recomendações em consultas

Nesta seção, investiga-se a visualização de documentos envolvendo sessões em que houve consulta *versus* aquelas em que a consulta não foi utilizada⁹. Foram visualizados 11.959 registros (RegCON) a partir de 10.063 consultas, o que resulta em uma média de 1,19 visualizações de registros de documentos por consulta (**D1**). Foram contabilizadas 10.063 consultas realizadas em 5.979 sessões, numa média de 1,68 consultas por sessão.

O Gráfico 4 apresenta o número de registros de documentos clicados pelo usuário na lista de resultados de uma consulta, separados por classes. Neste gráfico fica evidente que a grande maioria das consultas tem como resultado a visualização de apenas um documento (representando 90% ou 9.058 consultas). Apenas 1% das consultas resulta em visualização de cinco ou mais documentos (**D1**).

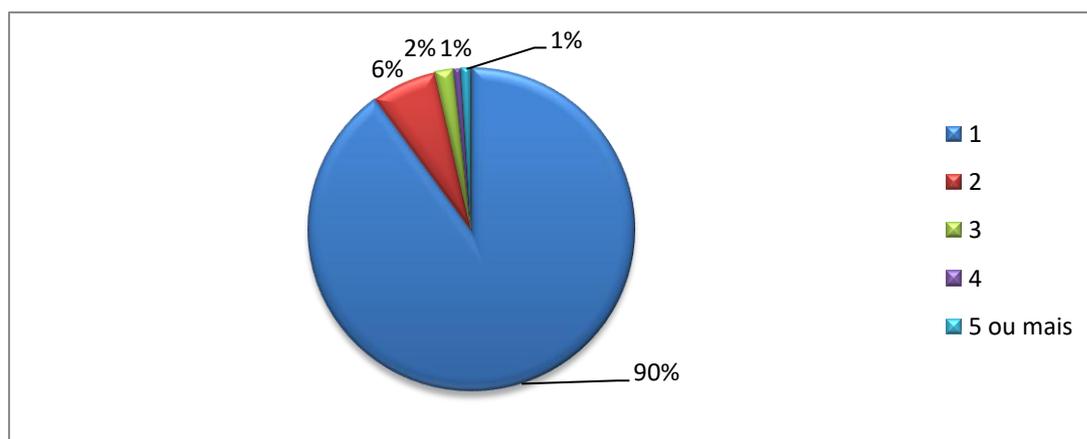


Gráfico 4 – Registros de documentos visualizados por consulta (D1)

Fonte: Dados da pesquisa.

⁹ As sessões em que não houve consulta são aquelas em que o utilizador não faz uma busca por palavras-chave no catálogo. Nesses casos ele pode chegar até o registro de documento através de navegação na web (link externo).

São consideradas sessões em que houve consulta aquelas em que o usuário elaborou pelo menos uma expressão de busca e, a partir da lista de resultados, selecionou pelo menos um registro dessa lista. Conforme a Tabela 3, nas sessões em que houve consulta a média de registros visualizados por sessão é de 2,00, enquanto nas sessões em que não houve consulta, a média é de 4,60 registros de documentos visualizados (**D2**). A visita a dois registros de documentos, em média, nas sessões com consulta, pode indicar que o usuário encontra rapidamente o que ele necessita. Assim tem-se que nas sessões sem consulta, apesar de estas representarem 5,8% do total de sessões, os usuários visitam, em média, mais que o dobro de registros de documentos do que nas sessões com consulta, o que condiz com o gráfico 5, pois apresenta, nas sessões com consulta, 90% de sessões com visita a apenas um documento.

Tabela 3 – Registros de documentos visualizados por sessão com ou sem consulta (D2)

Sessões	Sessões	Sessões %	Reg. Doc. Visualizados	Reg. Doc. Visualizados %	Média de reg. Doc. Visualizados por sessão
Com consulta	5.979	94,2%	11.959	87,6%	2,00
Sem consulta	368	5,8%	1.695	12,4%	4,60
Total	6.347	100%	13.654	100%	2,15

Fonte: Dados da pesquisa.

Em relação às consultas, a grande maioria (90%) teve como resultado a visualização de apenas um documento, e a média de registros de documentos visualizados por consulta chegou a 1,19 (**D1**). Quando a unidade de medida é a sessão, tem-se a média de 2,00 visualizações de registros de documentos nas sessões com consulta. Nas sessões sem consulta os usuários visitaram, em média, mais que o dobro (4,60) de registros de documentos do que nas sessões com consulta (**D2**).

A partir dos dados apresentados, pode-se então analisar recomendação *versus* consultas, relacionando recomendações em sessões, com recomendações em sessões em que houve consulta. Analisando as recomendações em sessões, entre todas as sessões registradas, 11% possuem recomendação e contêm 21,7% dos registros de documentos visualizados (Tabela 2), sendo que, destes documentos, 43,9% tiveram origem em uma recomendação (Tabela 1). Estas sessões apresentam, em média, 4,30 registros de documentos visualizados (incluindo registros oriundos de recomendação e não oriundos de recomendação) (Tabela 1). Analisando sessões com consultas (Tabela 3), entre todas as sessões registradas, 94% apresentam consulta, e nessas sessões estão 87,6% dos registros de documentos visitados. A média de registros de documentos visitados é de 2,00 por sessão nesses casos.

Um alto grau de registros de documentos visitados em sessões com recomendação (4,30)¹⁰ sugere a existência de serendipidade (do termo em inglês *serendipity*, que significa encontrar algo útil ou agradável pelo qual não se esteja procurando). Os usuários navegaram em mais registros

¹⁰ Tabela 2.

de documentos com algum grau de similaridade com aqueles inicialmente acessados e “permaneceram na recomendação” em 43,9% dos casos. A média baixa de registros de documentos visitados por sessões com consulta (2,00)¹¹ sugere que os resultados são, em média, assertivos para os que buscam informação, ou seja, é possível que eles encontrem rapidamente o que estavam buscando (embora não se possa afirmar em definitivo, pois há a possibilidade de abandono de sessão) **(D3)**.

7 Considerações finais

A principal aspiração dos autores neste trabalho foi analisar o uso do sistema de recomendação *Related Books in Aleph OPAC*, desenvolvido para catálogos *online*, mais especificamente o Aleph OPAC. Os dados obtidos e a metodologia utilizada mostraram-se satisfatórios para atingir os objetivos da investigação, cujos resultados se analisam sinteticamente a seguir.

Quanto ao primeiro objetivo, verificar em que medida a recomendação é utilizada, a análise dividiu-se em três vetores: número total de registros visualizados a partir da recomendação, número de sessões com e sem recomendação e proporção, nas sessões com recomendação, de registros de documentos oriundos de recomendação. Descobriu-se que 9,6% do *log* (1.305 *posts*) correspondem a registros de documentos visualizados a partir de recomendação **(A1)**. Também se identificou que a recomendação foi utilizada em 11% das sessões (690), enquanto em 89% das sessões (5.657) a recomendação não foi utilizada **(A2)**. Considerando somente sessões em que houve recomendação, a proporção de registros visualizados a partir de recomendação (RegREC) foi de 43,9% (1.305) **(A3)**.

Considerando o segundo objetivo específico, eventuais diferenças do uso ou não de recomendação nas interações do usuário com o sistema, no que diz respeito ao tempo da interação e à quantidade de registros visualizados, pôde-se identificar que as sessões com uso de recomendações foram em média 10 segundos mais rápidas do que as sessões sem uso de recomendações (89%) **(B1)**. Quanto à visualização de registros de documentos, nas sessões com recomendação a média é de 4,30 registros visitados por sessão. Já naquelas sem uso de recomendação apresenta-se média de 1,88 registros **(B2)**. Embora representem 11% do total, nas sessões com recomendação são visitados 21,7% do total de registros.

Quanto ao terceiro objetivo específico, verificar o uso dos tipos de recomendação levando em consideração as interações dos usuários com o sistema e o número de registros visualizados, o estudo aponta para mais do que um tipo preferido: do ponto de vista das sessões (interações feitas pelo mesmo usuário, no mesmo computador, em um determinado período de tempo) o tipo de recomendação mais utilizado é o ASSUNTO, filtragem baseada em conteúdo que considera similares ao item visualizado aqueles

¹¹ Tabela 3.

documentos com pelo menos três cabeçalhos de assunto ou três números de classificação – no registro MARC - em comum com o item visualizado. Este tipo de recomendação apareceu em 348 sessões **(C1)**. Por outro lado, analisando pelo ponto de vista de quantidade de registros visualizados, o tipo mais utilizado é o EMPRÉSTIMO, filtragem colaborativa que considera similares ao item visualizado aqueles documentos que foram emprestados aos usuários que também retiraram o item visualizado. Foram 584 registros de documentos visualizados a partir deste tipo de recomendação **(C2)**. Ambos são proporcionalmente relevantes no uso da recomendação. Com isto pode-se afirmar que nem a filtragem colaborativa nem a filtragem baseada em conteúdo predominam massivamente no uso da recomendação.

Por fim, a pesquisa viabilizou uma análise da relação entre o uso da recomendação e de consultas nas interações dos usuários com o sistema, atingindo o quarto objetivo específico. Identificou-se que a média de registros de documentos visualizados por consulta é de 1,19 **(D1)**. Usando as sessões como medida, identificou-se que menos registros de documentos foram visualizados nas sessões com consulta (2,00 em média) do que nas sessões sem consulta (4,60 em média)¹² **(D2)**.

Visualizou-se, em média, 2,00 registros de documentos nas sessões com consulta. Nas sessões com recomendação a média sobe para 4,30¹³ registros de documentos por sessão. Considerando, ainda, a visualização de registros, foi possível entender que o uso da recomendação incentiva a serendipidade, pois nas sessões com uso de recomendação em 43,9% dos casos os usuários de recomendação navegaram em mais registros de documentos com alguma similaridade ao documento acessado inicialmente. Já as sessões com consulta apresentaram uma média baixa de documentos visitados por sessão, o que podemos interpretar como alta assertividade por parte dos usuários (que encontram rapidamente o que estavam procurando)¹⁴ **(D3)**.

Nota-se que nas sessões com recomendação os usuários tendem a visitar mais de um documento por sessão. Este dado, aliado à informação de que as sessões com recomendação são em média mais curtas, põe em evidência as sessões que apresentam uso da recomendação, pois foi possível visualizar, em média, mais registros em menos tempo, do que em sessões sem o uso da recomendação. A proporção justifica o uso da recomendação nas bibliotecas, especialmente quando se observa que as sessões com recomendação apresentam o uso intensivo do recurso, pois mostrou-se que nelas há um maior acesso a registros de documentos, e que destes, quase a metade tem como origem uma recomendação (Tabela 1).

A partir da análise dos dados, de modo geral, este trabalho evidencia a adesão dos usuários à ferramenta de recomendação aplicada a catálogos *online* e suas preferências na utilização, demonstrando que esta temática é um relevante campo de pesquisa. Hoje a recomendação é

¹² Tabela 3.

¹³ Tabela 2.

¹⁴ Sem excluir a possibilidade de abandono de sessão.

ainda pouco usada em catálogos *online* de bibliotecas, frente à larga utilização em outros mercados, como o *comércio eletrônico*, por exemplo. No entanto, conforme demonstrado neste trabalho, os usuários têm interesse em utilizar a recomendação. Ela fornece aos catálogos *online* possibilidades que vão além das consultas, enriquecendo a navegação do usuário através de recursos como a descoberta prazerosa de documentos inesperados (serendipidade).

Este estudo mostra que este campo pode ainda ser muito mais explorado, tanto intensificando o uso de sistemas de recomendação em bibliotecas do Brasil, como ampliando pesquisas que busquem compreender o comportamento dos usuários, suas preferências, assim como a aplicação de novas técnicas de recomendação (filtragem colaborativa e filtragem baseada em conteúdo). O estudo também pode ser útil a gestores interessados em implementar sistemas de recomendação, uma vez que: (a) indica os diferentes tipos de recomendação que podem ser utilizados (sem a pretensão de esgotá-los, visto que novas formas de cálculo de similaridade surgem a cada dia); e (b) apresenta o uso real da ferramenta, através da análise de *logs*, considerado um dos modos de maior fidelidade para análise do comportamento do usuário em interação com o sistema.

À medida que os usuários exigem maior precisão e personalização em seus resultados de busca por informação, a ferramenta *Related Books in Aleph OPAC* pode tornar a recomendação uma realidade nas bibliotecas.

Por fim, destaca-se que o estudo foi baseado em registros de interação ligados à recomendação e os pesquisadores não obtiveram acesso às expressões de busca usadas nas consultas, não sendo possível analisar diferenças de comportamento dos usuários em relação à natureza de sua pesquisa, o que pode ser determinante para a escolha do modo de acesso ao sistema e também no uso da recomendação. Estas ideias podem ser exploradas em trabalhos futuros.

Referências

ASSOCIATION FOR COMPUTING MACHINERY (ACM). *Recomendação*. Disponível em: <<http://recsys.acm.org/>>. Acesso em: 11 dez. 2012.

BARCELLOS, C. D.; BRANDÃO, A. L.; MUSA, D. L. Sistema de Recomendação Acadêmico para Apoio a Aprendizagem. *CINTED-UFRGS Novas Tecnologias na Educação*, Porto Alegre, v. 5, n. 2, 10 p., dez. 2007. Disponível em <<http://www.cinted.ufrgs.br/ciclo10/artigos/3fDaniela.pdf>>. Acesso em: 9 nov. 2014.

JANSEN, B. J. Search log analysis: what it is, what's been done, how to do it. *Library & Information Science Research*, Pennsylvania, v. 28, p. 407-432, 2006. Disponível em: <<http://lincs.hum.iit.edu/sites/default/files/JansenSearchLog.pdf>>. Acesso em: 3 nov. 2014.

LICHTNOW, D. *et al.* O uso de técnicas de recomendação em um sistema para apoio à aprendizagem colaborativa. *Revista Brasileira de informática na educação (RBIE)*, v. 14, n. 3, p. 49-59, 2006. Disponível em: <<http://www.br-ie.org/pub/index.php/rbie/article/view/46/40>>. Acesso em: 9 out. 2014.

NICHOLAS, D. *et al.* Opening the digital box: what deep log analysis can tell us about our digital journal users. In: STRAUCH K.; BAZIRJIAN R.; SPECK V. *Charleston 2003 Conference Proceedings*. Charleston: Libraries Unlimited, 2004. Disponível em: < <http://discovery.ucl.ac.uk/185267/> >. Acesso em: 24 fev. 2015.

PHILLIPS, M. *Every book its reader: recommendations in Aleph*. University of Dundee. [2009?]. Disponível em: <<http://igelu.org/wp-content/uploads/2010/09/recommendations-in-aleph.pdf>>. Acesso em: 21 set. 2015.

PHILLIPS, M. *Related books links in Aleph OPAC*. Dundee: University of Dundee, 2010. Disponível em: <<http://www.exlibrisgroup.org/display/AlephCC/Related+books+links+in+Aleph+OPAC>>. Acesso em: 12 set. 2014.

SPEROTTO, F. A. *Protótipo de um sistema de recomendação no auxílio da gestão de competências em disciplinas na área acadêmica*. 2010. 83f. Monografia (Graduação em Ciência da Computação) - Universidade Comunitária Regional de Chapecó, Chapecó, 2010. Disponível em: <<http://www5.unochapeco.edu.br/pergamum/biblioteca/php/imagens/000061/000061D6.pdf>> Acesso em: 9 nov. 2014.

TORRES, R. *Personalização na Internet: como descobrir os hábitos de consumo dos seus clientes, fidelizá-los e aumentar o lucro de seu negócio*. São Paulo: Novatec, 2004.