

Universidade Federal do Rio Grande do Sul
Faculdade de Biblioteconomia e Comunicação
Departamento de Ciências da Informação

Luciana Monteiro Krebs

Sistema de recomendação para bibliotecas
universitárias

Porto Alegre

2013

Luciana Monteiro Krebs

Sistema de recomendação para bibliotecas universitárias

Trabalho de Conclusão de Curso de graduação, apresentado como requisito para obtenção do grau de bacharel em Biblioteconomia pelo Departamento de Ciências da Informação da Universidade Federal do Rio Grande do Sul.

Orientador: Prof. Dr. Rafael Port da Rocha

Coorientadora: Prof^a. Dra. Maria Cristina de Carvalho Alves Ribeiro

Porto Alegre

2013

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Reitor: Prof. Dr. Carlos Alexandre Netto

Vice-Reitor: Prof. Dr. Rui Vicente Oppermann

FACULDADE DE BIBLIOTECONOMIA E COMUNICAÇÃO

Diretora: Prof. Dra. Ana Maria Mielniczuk de Moura

Vice-Diretor: André Iribure Rodrigues

DEPARTAMENTO DE CIÊNCIAS DA INFORMAÇÃO

Chefe: Maria do Rocio Fontoura Teixeira

Chefe Substituto: Valdir José Morigi

COMISSÃO DE GRADUAÇÃO DO CURSO DE BIBLIOTECONOMIA

Coordenadora: Prof^ª Dra. Samile Andréa de Souza Vanz

Coordenadora Substituta: Prof^ª Me. Glória Isabel Sattamini Ferreira

K92s KREBS, Luciana Monteiro

Sistema de recomendação para bibliotecas universitárias /
Luciana Monteiro Krebs ; orientador: Rafael Port da Rocha ;
coorientadora: Maria Cristina de Carvalho Alves Ribeiro. – 2013. –
Monografia (Graduação em Biblioteconomia). Porto Alegre: Universidade
Federal do Rio Grande do Sul, Faculdade de Biblioteconomia e
Comunicação, 2013. 95 f. : il.

1. Ciência da Informação. 2. Sistemas de Recomendação. 3.
Catálogos on-line. I. Rocha, Rafael Port da. II. Ribeiro, Maria Cristina de
Carvalho Alves. III. Título.

CDU 025.002 855 3

Catálogo: Luciana Monteiro Krebs

Faculdade de Biblioteconomia e Comunicação – FABICO

Rua Ramiro Barcelos, 2705 – Bairro Santana – Porto Alegre – Rio Grande do sul

CEP 90035-007 – Tel.: (51) 3308-5146

e-mail: fabico@ufrgs.br

Luciana Monteiro Krebs

Sistema de recomendação para bibliotecas universitárias

Trabalho de Conclusão de Curso de graduação, apresentado como requisito para obtenção do grau de bacharel em Biblioteconomia pelo Departamento de Ciências da Informação da Universidade Federal do Rio Grande do Sul.

Orientador: Prof. Dr. Rafael Port da Rocha
Coorientadora: Prof^a. Dra. Maria Cristina de Carvalho Alves Ribeiro

Aprovada em 03 de julho de 2013.

BANCA EXAMINADORA

Prof. Dr. Rafael Port da Rocha – UFRGS (orientador)

Prof^a. Dra. Maria Cristina de Carvalho Alves Ribeiro – UP
(coorientadora)

Prof^a. Dr^a. Sônia Elisa Caregnato – UFRGS (examinadora)

Prof^a. Ms. Rita do Carmo Ferreira Laipelt – UFRGS (examinadora)

Aos meus pais,
seres de amor e retidão infindáveis,
a quem admiro profundamente.

*“In the nonstop tsunami of global information
librarians provide us with floaties and teach us to swim.”*

Linton Weeks

Agradecimentos

Aos meus pais, Martinho e Heloisa, pela candura, sabedoria e paciência com que me acolhem todas as vezes que lhes compartilho as dores e as delícias da vida acadêmica, mesmo quando já sabem exatamente como vai acabar. Silenciam me dando espaço para errar e trilhar meu próprio caminho. Obrigada. Aos meus irmãos Lisandra (*in memoriam*), Eduardo e Alexandre, por andarem ao meu lado, me incentivando a cada conquista. À minha vó, Loiva.

Ao Vicente, amor da minha vida. Obrigada pelas inúmeras vezes em que foste como uma ponte sobre águas turbulentas. Além da ajuda prática neste trabalho e em todos os anos de faculdade, por acreditar e incentivar quando eu mesma achava que não conseguiria. Aos meus sogros, Vicente e Mara.

À Universidade Federal do Rio Grande do Sul, por me oferecer o ambiente e os combustíveis necessários para realizar muito mais do que imaginava ser capaz. Por me colocar em contato com pessoas especiais. Aos mestres Rafael Port da Rocha e Maria Cristina Ribeiro pelo envolvimento maior do que o normal e pela generosidade de compartilhar comigo um pouco do muito que sabem, e às professoras Sônia Elisa Caregnato e Helen Rozados, pelas nobres trocas de ideias.

Aos colegas da UFRGS, que durante toda a jornada dividiram angústias e alegrias comigo, e àqueles que tiveram a maturidade de aceitar nossas diferenças e me escolherem como amiga.

Aos futuros colegas de profissão que apoiaram e forneceram informações que enriqueceram este trabalho: Matthew Phillips (que concebeu a ferramenta Related Books in Aleph OPAC e me colocou em contato com a *University of Dundee*), Karin Culley (que gentilmente me cedeu o *log* utilizado no estudo), todo o time de bibliotecários da PUC-RS, em especial à

Ginamara Lima pelo voto de confiança e orientação durante o estágio e Michelângelo Viana, por ter me apresentado à ferramenta e passado contatos que me ajudaram a encontrar meu objeto de estudo. À Caterina Pavão (UFRGS) pelos artigos e esclarecimentos sobre outros *softwares* de recomendação e Anderson Santana (USP), pelos contatos valiosos.

Ao Roberto Torres, cujo livro tangibilizou o que antes era só uma ideia abstrata.

Ao Fernando Marques Ferreira, pela irrestrita confiança no meu potencial e pela carinhosa e constante preocupação com meu futuro. Ao Fábio Giordani, pelo incentivo e contatos.

E a todos que de forma direta ou indireta contribuíram para a concretização deste trabalho. Meu sincero muito obrigada.

Resumo

Trabalho de conclusão de curso cujo tema é sistemas de recomendação em bibliotecas universitárias. Analisa o uso do *Related books in Aleph OPAC*, sistema de recomendação para catálogos *on-line* de bibliotecas universitárias. Apresenta conceitos de serviço de referência e marketing de serviços em unidades de informação, recuperação de informação, filtragem de informação, representação de documentos, personalização, recomendação e serendipidade. Utiliza abordagem quantitativa e metodologia descritiva e documental, em que se realizou a coleta, tratamento e análise dos dados provenientes do *log* fornecido pela Universidade de Dundee. Os 13.654 posts e 6.347 sessões analisados resultaram no seguinte: a recomendação foi utilizada em 11% das sessões, onde 43,9% dos registros de documentos visualizados foram oriundos de recomendação. As sessões com uso de recomendações foram, em média, 1 minuto e 18 segundos mais rápidas do que as sessões sem uso de recomendações. Nas sessões com recomendação foram visualizados em média 4,30 registros de 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 1 tipo de recomendação. Já o tipo de recomendação mais utilizado teve 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 posts (584). Em relação ao padrão de uso em dias da semana, identificou-se um comportamento semelhante no decorrer da semana, com pico na quinta-feira, tanto em sessões com recomendação quanto sem recomendação.

Palavras-chave: Ciência da Informação. Sistemas de recomendação. Catálogos on-line.

Abstract

Monograph which theme's recommender systems in academic libraries. Examines the use of "Related books in Aleph OPAC", recommendation system for academic libraries' online catalogs. Introduces concepts of service marketing and referral services in information units, information retrieval, information filtering, document representation, personalization, recommendation and serendipity. Uses a quantitative approach and methodology is descriptive and documentary, which was held in the collection, processing and analysis of data from the log provided by the University of Dundee. The 13,654 posts and 6,347 sessions analyzed resulted in the following: the recommendation was used in 11% of sessions, where 43.9% of records documents were viewed through recommendation. The use of recommendations sessions were on average, 1 minute 18 seconds faster than the sessions without the use of recommendations. In sessions with recommendation were viewed on average 4.30 records documents while in sessions without use of recommendation the average is 1.88. Using more than one type of recommendation had poor adherence, as 82% of sessions with recommendation register the use of only one type of recommendation. The most widely used type of recommendation had two results: "Related works include" appears in more sessions (348), but the type "People who borrowed this work also borrowed" registers the highest number of posts (584). Regarding the usage pattern on weekdays, identified a similar behavior during the week, peaking on Thursday, both in sessions with no recommendation as recommendation.

Keywords: *Information Science. Recommender systems. Online catalogs.*

Sumário

1	Introdução	13
1.1	Identificação e justificativa do problema de pesquisa	15
1.2	<u>Objetivos</u>	16
1.2.1	<i>Objetivo geral</i>	17
1.2.2	<i>Objetivos específicos</i>	17
1.3	<u>Estrutura do documento</u>	17
2	Revisão de literatura	18
2.1	<u>Serviço de referência em bibliotecas universitárias</u>	18
2.2	<u>Marketing de serviços em unidades de informação</u>	23
2.3	<u>Recuperação da informação (<i>Information Retrieval</i>)</u>	24
2.3.1	<i>Busca de informação versus Navegação</i>	27
2.3.2	<i>Recuperação de dados versus Recuperação de informação</i>	27
2.3.3	<i>O processo de recuperação da informação</i>	28
2.4	<u>Filtragem de informação</u>	34
2.5	<u>Personalização</u>	35
2.6	<u>Sistemas de recomendação</u>	36
2.6.1	<i>Identificação de usuário</i>	39
2.6.2	<i>Descobrendo hábitos de consumo</i>	40
2.6.3	<i>Técnicas de recomendação</i>	40
2.7	<u>Serendipidade</u>	46
2.8	<u>Tendências futuras em recomendação</u>	47
2.9	Soluções de recomendação existentes para bibliotecas	48
2.9.1	<i>Bx reccomender</i>	49
2.9.2	<i>Sciverse sciencedirect</i>	50
2.9.3	<i>Sistema de Recomendação Advanced IT</i>	51
2.9.4	<i>Related books links in Aleph OPAC</i>	53

3	Metodologia	57
3.1	<u>Classificação metodológica da pesquisa</u>	57
3.2	<u>Etapas de desenvolvimento do trabalho científico</u>	58
3.2.1	<i>Levantamento bibliográfico</i>	58
3.2.2	<i>Coleta de dados</i>	59
3.2.3	<i>Tratamento dos dados</i>	59
3.2.4	<i>Estrutura dos dados</i>	60
4	Análise dos dados	67
4.1	<u>Análise das sessões</u>	67
4.1.1	<i>Quanto ao tempo</i>	68
4.1.2	<i>Quanto ao uso nos dias da semana</i>	70
4.1.3	<i>Quanto ao número de registros de documentos visualizados</i>	72
4.2	<u>Recomendações</u>	75
4.2.1	<i>Uso de diferentes tipos de recomendação</i>	76
4.2.1	<i>Quanto ao número de registros de documentos visualizados por recomendação</i>	79
4.2.3	<i>Profundidade</i>	81
4.3	<u>Consultas</u>	82
4.4	<u>Recomendação versus Consultas</u>	85
5	Considerações finais	87
	Referências Bibliográficas	92

1 Introdução

Para desempenhar um serviço relevante, consultivo e agregador, especialmente no que se refere às atividades de referência, pressupõe-se que o bibliotecário deve possuir um perfil formado por diversas características: conhecer o usuário, ser genuinamente voltado para satisfazer as necessidades deste usuário, possuir conhecimento prévio do acervo e da produção intelectual disponível na sua unidade de informação ou em outras que possa vir a ter acesso, ter competência informacional para buscar e encontrar a informação desejada e, acima de tudo, ser pró-ativo para indicar leituras pertinentes que possam vir a atender às necessidades. É, sem dúvida, um desafio.

Em unidades de informação que o usuário não frequente presencialmente, ou em bibliotecas com cada vez mais usuários e cada vez menos bibliotecários, o problema começa já no passo inicial: conhecer o usuário. Infelizmente, nestes casos o serviço de referência tende a acontecer (ser entregue) apenas para aqueles usuários que efetivamente vão à biblioteca, ou consultam-na por e-mail, telefone ou outros canais de comunicação. Conseqüentemente, as decisões acerca de desenvolvimento de coleções, planejamento, políticas e *design* de serviços são pensados levando em conta estes – e somente estes – atendimentos. A realidade é que aqueles usuários que não procuram o bibliotecário de referência acabam ficando à margem de todo o planejamento da unidade de informação, que tende a afastá-lo cada vez mais.

Por isto, construir ou fazer uso de sistemas de recomendação pode ser um caminho para propor materiais informacionais aos usuários e sofisticar o serviço de referência de tal forma que se possam combinar indicações personalizadas com sugestões padronizadas (extensivas para todos os usuários), equalizando o serviço de referência. Ou seja, através de

sistemas que utilizam técnicas automáticas para analisar padrões de utilização e gerar indicadores de auxílio à gestão, podem-se criar serviços com base em seus relatórios, acompanhar mais de perto o uso da informação nos portais de conteúdo, e o mais importante: entregar informação de qualidade e de forma agregada, pensada exclusivamente para aquele usuário, vindo ele até a biblioteca ou não.

Prova de que a recomendação de livros e documentos é possível e traz resultados é a Goodreads¹ (AMAZON, 2013), recentemente adquirida pela Amazon, a maior companhia de comércio eletrônico do mundo. Goodreads permite aos usuários descobrir e discutir obras e, segundo nota publicada pela Amazon, contribuiu para **umentar a audiência** de novos autores. A gigante no *e-commerce* é historicamente conhecida pelo pioneirismo no uso de recomendação social (quem nunca ouviu falar no recurso “quem comprou este também comprou aquele?”). O anúncio da aquisição do site especializado em recomendação de livros surpreendeu positivamente o mercado de ações, e registrou alta de 0,45% nas ações da Amazon naquele dia. Mesmo que o valor da transação tenha sido mantido em sigilo, o mercado parece ter aprovado a aposta.

O bibliotecário no Brasil já sai da academia com boas noções de web, bases de dados, funcionamento da web 2.0 e em alguns casos até programação. Segundo SANTOS (1998, p. 8), atualmente a área de Tecnologia da Informação faz parte do núcleo principal de conhecimentos a serem abordados nos currículos das escolas de Biblioteconomia do país².

¹ Fundado em 2007 nos Estados Unidos, o site Goodreads possui pouco mais de 16 milhões de usuários cadastrados. O serviço reúne em torno de 30 mil clubes de livros. Nos últimos 90 dias, os usuários do site adicionaram em média quatro livros por segundo na lista de obras que querem ler.

² O contexto desta inclusão é um processo de harmonização curricular de vários Cursos dos Estados Parte e Associados do Mercosul, que chegou a estas diretrizes em novembro de 1997. O artigo citado indica que a Área 5 (Tecnologia da Informação) é composta pelos seguintes assuntos: Aplicações da tecnologia da informação e comunicação nas unidades de informação; análise, avaliação e desenvolvimento (hardware e software); gestão de bases de dados e bibliotecas virtuais; análise e avaliação de sistemas e redes de informação; informatização das unidades de informação.

Então, para os interessados, este estudo pretende aprofundar um pouco mais o estudo de possibilidades de aplicação dos sistemas de recomendação em bibliotecas digitais, fornecendo subsídios para que os bibliotecários possam compreender como os sistemas de recomendação operam, contribuir para a arquitetura desta solução, e fazer uso da ferramenta para promover uma experiência rica de navegação para os usuários.

1.1 Justificativa e identificação do problema de pesquisa

Com o aumento e aperfeiçoamento dos sistemas de informação digitais, especialmente os disponíveis na web, usuários de informação no mundo todo ficam cada vez mais exigentes em relação ao que esperam de um serviço. Empresas que desenvolvem motores de busca desenvolveram algoritmos sofisticados para trazer resultados relevantes aos usuários, baseando as buscas, principalmente, nos links de uma página web e utilizando esta informação para ordenação da página de resultados.

Enquanto isso, o discurso dos especialistas da Ciência da Informação versa sobre conhecer o usuário, a importância dos estudos de usuário e a efetiva aplicação deste conhecimento na prática diária do profissional bibliotecário. Felizmente, nunca se dispôs de tantos dados. Os bibliotecários, que estão à frente das unidades de informação, devem tomar para si a responsabilidade de planejar e entregar serviços com qualidade cada vez maior, surpreendendo o usuário cada vez mais exigente com o que há de melhor em termos de serviço personalizado. Os sistemas de recomendação parecem ser uma excelente solução.

Se não como gestores de unidades de informação, os bibliotecários podem envolver-se nesse processo enquanto especialistas na representação dos documentos e em todos os desafios que esta atividade apresenta. A

complexidade de escolha dos termos precisos para representação de conteúdos digitais é uma oportunidade para o bibliotecário, que pode enxergar este como um campo fértil de atuação e ávido por seu conhecimento específico, que vem sendo desenvolvido já há centenas de anos para a informação nos suportes tradicionais.

Até o momento, pouco foi publicado no Brasil sobre o uso de sistemas de recomendação em bibliotecas digitais e portais de conteúdo. Uma pesquisa prévia sobre o tema revelou que esta área é incipiente, e ainda mais com autoria ou colaboração de profissionais da Ciência da Informação. A oportunidade de munir a área de literatura sobre o tema foi um fator motivador.

Além disso, a escolha do tema surgiu de um interesse por sistemas de informação digitais com recomendação integrada. Por ser usuária da web e de algumas das ferramentas que lançam mão da personalização e recomendação, a autora deste trabalho desenvolveu certa curiosidade por conhecer as lógicas por trás destes sistemas. Os sistemas de recomendação são um valioso suporte para as atividades do bibliotecário, que se ocupa de estudar, entre outras questões, formas de satisfazer as necessidades informacionais dos usuários da unidade de informação em que atua, suprindo demandas, declaradas ou não. Além disto, encontrar formas de oferecer serviços diferenciados e que sejam valorizados por quem os utiliza também motivou este trabalho.

Desse modo, delinea-se a seguinte questão: Como os sistemas de recomendação operam e como são utilizados em unidades de informação?

1.2 Objetivos

Neste capítulo estão expressos os objetivos do presente trabalho, divididos em objetivo geral e específicos.

1.2.1 *Objetivo geral*

Analisar o uso do sistema de recomendação para catálogos on-line de bibliotecas universitárias, *Related books in Aleph OPAC*.

1.2.2 *Objetivos específicos*

- a) Analisar em que medida a recomendação é utilizada no catálogo da biblioteca;
- b) Investigar se o uso de recomendação resulta em sessões com duração maior ou menor em relação às sessões sem o uso de recomendação;
- c) Verificar se os usuários visualizam mais ou menos registros de documentos em sessões com recomendação do que em sessões sem recomendação;
- d) Caracterizar o uso de mais de um tipo de recomendação por sessão;
- e) Descobrir qual o tipo de recomendação mais utilizado;
- f) Identificar padrões de uso de recomendação por dias da semana.

1.3 Estrutura do documento

Este trabalho está estruturado da seguinte forma. O capítulo 1 apresenta introdução, os objetivos e a justificativa do trabalho. O capítulo 2 traz um apanhado da literatura que procura abarcar os assuntos contextuais dos sistemas de recomendação no ambiente das bibliotecas universitárias. O capítulo 3 fala da metodologia empregada no trabalho. O capítulo 4 traz a análise dos dados e o capítulo 5, as conclusões. Por fim, são apontadas as referências bibliográficas, apêndices e anexos.

2 Revisão de Literatura

Este trabalho baseia-se em conceitos da Ciência da Informação, Administração com ênfase em Marketing e Ciência da Computação. Neste capítulo apresentamos algumas ideias que apoiam o desenvolvimento do assunto Sistemas de Recomendação para Bibliotecas Universitárias.

2.1 Serviço de referência em bibliotecas universitárias

Diante da oferta de informação disponível atualmente, grande parte de forma gratuita, por vezes pensamos que há informação demais para lidar em escala humana. O usuário precisa, muitas vezes ao dia, tomar decisões para suprir suas necessidades informacionais. Por exemplo: Em que base irá procurar? Pode usar um motor de busca para encontrar o que precisa? Que expressão de busca deve usar? Qual, entre os artigos recuperados na busca, vai ler? Estas escolhas podem parecer banais, mas uma ou mais escolhas mal feitas pode acarretar em grande perda de tempo para o usuário, deixando-o frustrado.

O bibliotecário de referência é um profissional com competências específicas para capacitar e ajudar o usuário a buscar e encontrar aquilo que ele necessita no menor tempo e com o menor custo possível. Para Talavera Ibarra (1998, p.2) o serviço de referência pode ser definido do seguinte modo:

O serviço que facilita o acesso à informação, e que especialmente ajuda o utilizador a localizar a informação procurada é o serviço de referência. Este serviço, tradicionalmente definido como aquele que relaciona o utilizador que tem uma necessidade de informação com os recursos ou fontes que vão satisfazer as suas necessidades.

Interessa-nos também a afirmação de Bertrand Calenge (1996, p. 185, tradução nossa, grifo nosso), para quem o serviço de referência é “[...] uma função organizada de resposta personalizada a um pedido explícito de informação documental ou de documentação.” A personalização é um aspecto importante deste trabalho, como poderá ser visto a seguir.

Na visão europeia, os serviços dedicados ao público devem ser baseados em três aspectos: acolher, orientar e informar. (BERTRAND, 1996, tradução nossa) Já Richard Bopp (2000, tradução nossa), em “*Reference and information services: an introduction*” entende que os serviços de referência estão desdobrados em serviços de informação, de formação e de orientação.

Uma questão importante no serviço de referência é a especificidade do público ao qual se atende. Conhecendo bem o público é possível alcançar melhores níveis de satisfação, porque a interpretação da necessidade de informação do usuário é apurada com um conhecimento prévio do contexto onde este se insere e porque se pode antecipar o atendimento das necessidades informacionais trabalhando nos pedidos mais frequentes. Para Nice de Figueiredo (1991, p. 10),

[...] a longo prazo, a demanda no ambiente acadêmico é basicamente relacionada ao calibre intelectual e acadêmico da comunidade de usuários e à eficácia dos serviços de informação. A curto prazo, a demanda é em função da natureza das provas e trabalhos acadêmicos, do número e do tamanho da comunidade de usuários.

Os usuários das bibliotecas universitárias são frequentemente divididos conforme o vínculo que possuem com a instituição acadêmica, ou seja: podem ser alunos, docentes e pesquisadores. Para Mangas (2007, p. 3) “[...] os estudantes recorrerão ao serviço de referência à procura de bibliografia para estudar para as disciplinas e elaborar os trabalhos pedidos pelos professores.” Já os docentes e investigadores demandam um trabalho mais

demorado e complexo por parte do serviço de referência com suas perguntas e pesquisas. Segundo o autor, “Estes utilizadores exigirão um grau de pertinência, de especificidade e de exaustividade da informação solicitada muito superior à dos alunos de um curso de licenciatura.” (MANGAS, 2007, p. 3)

Nice de Figueiredo (1991, p. 14) afirma que “[...] idealmente, os sistemas de informação devem prover informações baseadas no conhecimento de como os seus usuários potenciais usam a informação”. O uso da informação deve servir de fonte para sofisticar os sistemas de informação e os bibliotecários devem estar atentos a isso. Mas que formas existem de medir o uso da informação nos dias de hoje?

A tecnologia trouxe para os serviços em geral nas bibliotecas, não somente para a referência, mudanças significativas. Baeza-Yates e Ribeiro-Neto (1999, p. 2, tradução nossa) dão uma ideia do que isto significa:

A web está se tornando um repositório universal de conhecimento e cultura humanos, o qual tem permitido o compartilhamento de ideias e informação em uma escala sem precedentes antes. Qualquer usuário pode criar seus próprios documentos na web (página web) e fazer hyperlinks para qualquer outra página sem qualquer restrição. E porque a web é formada por dados dispersos e desestruturados, encontrar informação útil na web é frequentemente uma tarefa tediosa e difícil.

Por outro lado, em bibliotecas digitais os usuários podem contar com a conveniência de encontrar materiais informacionais independentemente da hora e da sua localização. O usuário não tem que se preocupar com o horário de atendimento da biblioteca física nem que ir fisicamente até a biblioteca para acessar os recursos (SLOAN, 1998, tradução nossa). Com informação disponível através da web, o diferencial das unidades de informação tornou-se o serviço com valor agregado que elas podem

oferecer. O serviço de referência digital já foi alvo de diversos estudos, sendo definido por Janes, Carter e Memmott (1999, p. 146, tradução nossa) como:

um mecanismo pelo qual as pessoas podem enviar suas perguntas e tê-las atendidas por um membro da equipe da biblioteca através de alguns meios eletrônicos (e-mail, chat, formulário da web, etc.), não pessoalmente ou por telefone.

E a disponibilidade também causou impactos importantes no que diz respeito ao que se espera de uma unidade de informação. Para o leitor que desejar se aprofundar nestes impactos, existem estudos sobre a ansiedade informacional que a era digital provoca e sobre a expectativa que os usuários nutrem em relação ao tempo de resposta, instaurando certo “estado de emergência” nos setores de atendimento³.

Quanto à rapidez, nem sempre os usuários estão dispostos a aguardar pelo resultado de uma busca encomendada, feita especialmente para ele a partir de uma análise imediata. Às vezes, imediatamente não é suficiente. Devido aos motores de busca modernos, que utilizam largamente métodos automáticos para personalizar as buscas, já se espera que esta personalização esteja sendo feita enquanto ele – usuário – ainda nem sabe que realizará determinada consulta. A partir da constante coleta do histórico, o sistema estará pronto para atendê-lo quando for necessário, pois já conhece suas preferências.

Cabe ressaltar uma das facetas do serviço de referência que nos interessa especialmente neste trabalho, que é a disseminação seletiva da informação (DSI). Este serviço foi concebido por Hans Peter Luhn, da IBM

³ Um exemplo: OLIVEIRA, Natália Gastaud. Ansiedade Informacional : o caso dos estudantes de pós-graduação da Universidade Federal do Rio Grande do Sul. 2011. Trabalho de Conclusão de Curso. (Graduação em Biblioteconomia) - Universidade Federal do Rio Grande do Sul. Orientadora: Ida Regina Chitto Stumpf. Coorientadora: Sonia Elisa Caregnato.

Corporation, em 1958. Nasceu com a finalidade, segundo Bax (2004, p. 3), de “[...] aperfeiçoar serviços de alerta oferecidos por bibliotecas, centros de documentação e sistemas especializados de informações documentais.”

A grande expansão da literatura mundial, a proliferação das bases referenciais especializadas on-line e a necessidade de atualização dos usuários das inúmeras especialidades da comunidade científica e tecnológica foram fatores motivadores desse tipo de serviço automatizado de informação já nos anos 60. (BAX, 2004, p. 3)

Nice de Figueiredo (1992) afirma que o serviço de disseminação seletiva da informação tem como requisito que o bibliotecário saiba dar a cada usuário a informação desejada, atendendo a todos com imparcialidade e respeito à pessoa humana.

Baeza-Yates e Ribeiro-Neto (1999, p. 8, tradução nossa) resumem estes desafios e problematizaram a recuperação da informação, que será aprofundada ainda neste capítulo, com questões a serem respondidas há quase 13 anos atrás, e que hoje talvez estejam perto de uma solução:

No futuro, três questões principais devem ser abordadas. Primeiro, apesar da alta interatividade, as pessoas ainda acham difícil (se não impossível) recuperar informações relevantes para suas necessidades de informação. Assim, no mundo dinâmico da Web e de grandes bibliotecas digitais, que técnicas vão permitir a recuperação de maior qualidade? Segundo, com a crescente demanda por acesso, a resposta rápida está se tornando mais e mais um fator de pressão. Assim, que técnicas trarão índices mais rápidos e tempos menores de resposta da consulta? Em terceiro lugar, a qualidade da tarefa de recuperação é grandemente afetada pela interação do utilizador com o sistema. Assim, como é que um melhor entendimento do comportamento do usuário afetam o projeto e implantação das novas estratégias de recuperação de informação? (BAEZA-YATES; RIBEIRO-NETO, 1999, p. 8, tradução nossa)

Este trabalho propõe um paralelo entre a disseminação seletiva da informação (tradicionalmente realizada manualmente pelo bibliotecário) e sistemas de recomendação, cujo funcionamento conhecerá a seguir.

2.2 Marketing de serviços em unidades de informação

A ideia de que bibliotecas oferecem livros apenas em prateleiras já foi há muito ultrapassada. Grande parte desta informação (senão a maioria dela) já está em formato digital e não mais físico. Hoje se sabe que o serviço agregado a estes materiais é que faz a diferença. Nas unidades de informação, o bibliotecário é o ente por trás da preparação do material antes de sua disponibilização, para que seja encontrado com facilidade em uma busca. Ele está capacitado para orientar o usuário sobre que itens são os mais apropriados para a sua necessidade, além de apoiar a busca de informação nas fontes adequadas. Assim, são os serviços oferecidos nas unidades de informação que a diferenciam de uma base de dados qualquer e da própria web.

Uma característica importante dos serviços que os diferenciam de produtos é a **intangibilidade**. Serviços não podem ser vistos, tocados, nem medidos objetivamente, porque não têm existência física (FISK, 2008). A intangibilidade dos serviços enfraquece a percepção de satisfação dos usuários, mesmo quando a qualidade de serviço foi satisfatória. Por este motivo, frequentemente lança-se mão de recursos que visam tornar estes serviços mais tangíveis, palpáveis, para que o usuário possa ter uma percepção mais apurada do valor do serviço que acaba de consumir, e se lembrar disso. A oferta de serviços diferenciados e personalizados é capaz de aumentar a satisfação dos usuários e mantê-los em uma posição confortável enquanto consumidores de uma unidade de informação.

Outra característica dos serviços é a **simultaneidade de produção e consumo**, porque na maioria dos casos o usuário é coparticipante do serviço. (FISK, 2008) Esta simultaneidade entre produção e consumo do serviço está intrinsecamente ligada aos sistemas de recomendação e filtragem colaborativa, onde hábitos de consumo de um usuário o ligam a outros usuários pela similaridade, gerando recomendações. Mais detalhes desta técnica serão vistos mais adiante neste trabalho.

Além disto, segundo Fisk (2008), os serviços caracterizam-se também pela **heterogeneidade**, ou seja, por serem realizados por pessoas e não por máquinas, os serviços são bastante dependentes da ação das pessoas, e por isto é difícil padronizar a qualidade de performance em serviços, o que os torna heterogêneos. Isto pode prejudicar serviços à medida que podem perder-se informações por falha de comunicação ou esquecimento do profissional, algo perfeitamente compreensível tendo em vista o volume cada vez maior de informação disponível e a variedade de usuários a que deve atender. Incrementar serviços com ferramentas automáticas de recomendação podem trazer padronização na oferta de um serviço, dependendo do perfil do usuário em questão, ou seja, toda vez que uma oferta for conveniente para um usuário o sistema irá recomendar, sem prejuízos.

2.3 Recuperação da informação (*Information Retrieval*)

As primeiras instituições a adotar sistemas de recuperação da informação foram as bibliotecas, sendo estes, em sua maioria, desenvolvidos por instituições acadêmicas num primeiro momento e posteriormente por empresas comerciais. Sistemas de recuperação da informação são utilizados para atender as necessidades informacionais dos usuários. Segundo Baeza-Yates e Ribeiro-Neto (1999, p. 1),

Recuperação da informação lida com as representações, armazenamento, organizações e acesso aos itens de informação. A representação e organização dos itens de informação [documentos] devem fornecer ao usuário o acesso fácil à informação na qual ele está interessado.

A recuperação da informação, segundo Marquesuzaà et al. (2008, documento eletrônico, tradução nossa) lida com “[...] modelos, técnicas, procedimentos para extrair informações que já tenham sido processadas, organizadas e armazenadas (como bancos de dados, arquivos, arquivos XML, entre outros)”.

Os sistemas de recuperação, segundo Cazella (2006, p. 26), “[...] são uma ótima solução para os usuários que sabem realmente o que desejam buscar, ou seja, os que sabem definir com alguma exatidão os termos de sua consulta.” Segundo o autor, se o usuário não tem conhecimento de outros conteúdos que também seriam interessantes, muito pode ser perdido. (CAZELLA, 2006, p. 26)

Na primeira geração, estes sistemas consistiam apenas em uma automação de tecnologias anteriores (como catálogos de fichas) e basicamente permitiam buscas pelo nome do autor e título do documento. A segunda geração incrementou a funcionalidade de busca que passou a permitir buscas por cabeçalhos de assuntos, palavras-chave e algumas facilidades mais complexas. Na terceira geração que está atualmente [1999] sendo desenvolvida, o foco está na melhoria da interface gráfica, formulários eletrônicos, funcionalidades de hipertexto e arquiteturas abertas de sistemas. (BAEZA-YATES; RIBEIRO-NETO, 1999, p.7, tradução nossa)

Baeza-Yates e Ribeiro-Neto afirmam que os motores de busca na web atuais continuam a usar índices muito semelhantes aos utilizados por bibliotecários há mais de um século atrás. O que mudou, segundo eles, foram três pontos fundamentais, oriundos dos avanços da tecnologia moderna e do *boom* da web:

- a. Baixo custo do acesso à informação na web
- b. Maior acesso provocado pelos avanços tecnológicos dos meios de comunicação digital
- c. Liberdade de publicação

Primeiro, tornou-se muito mais barato ter acesso a várias fontes de informação. Isto permite chegar a um público mais amplo do que nunca possível antes. Segundo, os avanços em todos os tipos de comunicação digital proporcionaram maior acesso à rede. Isto implica que a fonte de informação está disponível, mesmo que remotamente localizada, e que o acesso pode ser feito rapidamente (frequentemente, em poucos segundos). Em terceiro lugar, a liberdade de publicar qualquer informação que alguém julgue útil tem contribuído grandemente para a popularidade da web. Pela primeira vez na história, muitas pessoas têm acesso gratuito a um meio de publicação em larga escala. (BAEZA-YATES; RIBEIRO-NETO, 1999, p.7-8, tradução nossa)

Portanto, baixo custo, maior acesso e liberdade de publicação permitiu que as pessoas usassem a web e as bibliotecas digitais como um meio altamente interativo. “A alta interatividade é a mudança fundamental e atual no paradigma da comunicação.” (BAEZA-YATES; RIBEIRO-NETO, 1999, p. 8, tradução nossa)

Atualmente, pode-se dizer que os sistemas de recuperação têm dado conta de ir além da resposta a uma consulta e englobam tecnologias mais aprimoradas para satisfazer as necessidades de informação dos usuários. Dados históricos, comportamento de navegação, comparação com usuários de perfis semelhantes e outros recursos são usados para incrementar e sofisticar os sistemas, trazendo resultados relevantes em meio a uma miscelânea cada vez mais densa e complexa de informação disponível.

Nos próximos parágrafos vamos esclarecer o funcionamento atual dos sistemas e discutir estas novas possibilidades.

2.3.1 *Busca de informação versus Navegação*

A diferença entre navegação e pesquisa (ou busca) é algo simples, mas importante para entendermos as particularidades das funções dos sistemas de recuperação da informação e sistemas de filtragem colaborativa.

Busca de informação (do inglês *searching*) significa busca cujo objetivo é claro, requer uma expressão de busca. Pressupõe que o usuário já consiga definir com alguma exatidão os termos da consulta para recuperar os documentos que deseja.

Navegação (em inglês *browsing*) significa busca cujos objetivos não estão bem claros no início da busca e cujos propósitos podem mudar durante o processo. Ações para tentar encontrar algo que talvez saibamos com algum rigor que existem. Esta forma de pesquisa permite serendipidade⁴, que será comentada no item 2.7 deste trabalho.

A diferença essencial entre pesquisa e navegação é que a pesquisa requer uma expressão de busca, e a navegação não.

2.3.2 *Recuperação de dados versus Recuperação de informação*

No contexto que estamos trabalhando, a recuperação de dados (*data retrieval*) consiste principalmente em determinar os documentos de uma coleção que contém as palavras-chave presentes na consulta (*query*) do usuário, e pressupõe dados estruturados. Porém, frequentemente esta função não é suficiente para satisfazer a necessidade de informação do usuário, que está mais preocupado em recuperar “informação” sobre um assunto do que recuperar dados que satisfaçam uma determinada consulta.

⁴ Serendipidade (do termo em inglês, *serendipity*), que significa encontrar algo útil ou agradável pelo qual não se esteja procurando.

Segundo Baeza-Yates e Ribeiro-Neto (1999), uma linguagem de recuperação de dados visa recuperar todos os documentos que satisfaçam condições claramente definidas como aquelas em uma expressão de busca comum ou em uma expressão de álgebra relacional. Um sistema de recuperação de dados (como um banco de dados relacional) lida com dados que têm uma estrutura e semântica bem definida. A recuperação de dados tem uma expressão lógica precisa, opera sobre informação estruturada e tem uma resposta correta única.

Por outro lado, a recuperação da informação opera sobre dados não estruturados, com necessidades de informação intrinsecamente imprecisas e as listas de resultados são sempre aproximações, sujeitas à apreciação do usuário quanto à sua relevância. Geralmente lida com linguagem natural, que pode ser semanticamente ambígua.

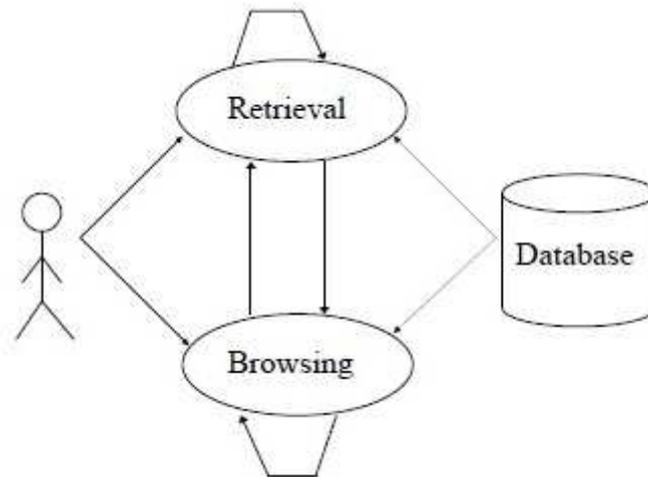
Assim, “A recuperação de dados, oferecendo uma solução para o uso de um sistema de banco de dados, não resolve o problema da recuperação de informações sobre um assunto ou tema.” (BAEZA-YATES; RIBEIRO-NETO, 1999, p. 2, tradução nossa).

2.3.3 O processo de recuperação da informação

O processo de recuperação da informação envolve a representação dos documentos, a formulação da consulta, a computação da resposta e a comunicação da resposta. Como se pode imaginar é uma tarefa desafiadora, pois envolve não só a interpretação da necessidade informacional do usuário (muitas vezes expressa apenas parcialmente), mas também a disponibilidade de recursos técnicos capazes de executar de forma satisfatória as etapas fundamentais do processo, citadas acima, e ordenar os resultados conforme relevância.

A figura 1 apresenta a interação do usuário com o sistema de recuperação.

Figura 1 - Interação do usuário com o sistema de recuperação através de tarefas distintas



Fonte: BAEZA-YATES e RIBEIRO-NETO (1999, p. 4)

O desafio começa na formulação de uma consulta (*query*), que deve expressar fielmente a necessidade de informação do usuário. A *query* ou consulta provê uma representação legível pelo sistema da necessidade de informação do usuário.

Para ser efetivo em sua intenção de satisfazer a necessidade de informação do usuário, o sistema de recuperação da informação precisa de alguma forma 'interpretar' os conteúdos dos documentos em uma coleção e ordená-los conforme um grau de relevância para a consulta do usuário. Esta 'interpretação' do conteúdo de um documento envolve extrair informação sintática e semântica do texto do documento e usar esta informação para combinar (*match*) com a necessidade de informação do usuário. (BAEZA-YATES; RIBEIRO-NETO, 1999, p.2, tradução nossa)

Uma questão chave aqui é a relevância, que varia de usuário para usuário. A relevância é uma medida de julgamento da importância de um documento a partir da análise do documento em relação à necessidade de informação do usuário. Por isso, um documento pode ser relevante para um usuário e não para outro.

A relevância, que precisa ser levada em consideração para ordenar os resultados da busca, é um verdadeiro desafio quando se trata de sistemas automatizados, pois pressupõe a interpretação da necessidade do usuário conforme mencionam os autores Baeza-Yates e Ribeiro-Neto (1999, p.2, tradução nossa):

A dificuldade não está apenas em saber como extrair esta informação mas também saber como usá-la para decidir a relevância. Então, a noção de relevância está no centro da recuperação da informação. Na verdade, o objetivo principal de um sistema de recuperação da informação é recuperar todos os documentos que são relevantes para uma consulta do usuário enquanto recupera o mínimo possível de documentos não relevantes.

Baeza-Yates e Ribeiro-Neto ressaltam a importância da validação da consulta ou fases de reformulação para melhorar a interpretação da necessidade informacional do usuário.

Em relação à **representação de documentos**, houve uma tendência de utilizar o texto integral ou completo do documento nos sistemas de recuperação da informação, mas em grandes coleções de documentos (como a web), aplicar uma busca no texto integral pode ser demasiado demorado ou custoso, pois implica em alta capacidade de processamento (e conseqüente alto custo computacional). Em decorrência da demanda por velocidade nos resultados, encontrou-se solução no pré-processamento para criação de índices e conseqüente redução na quantidade de termos que representem cada documento. O processamento rápido de uma consulta em

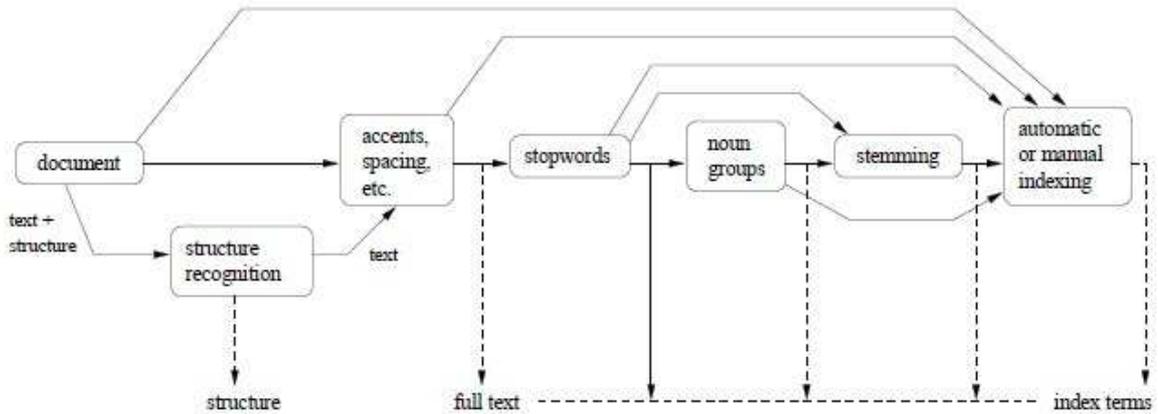
grandes coleções só é possível graças à estrutura de índice previamente construída.

A representação dos documentos fornece uma visão lógica do documento e permite melhorar o desempenho (provendo maior rapidez) nos sistemas de recuperação da informação. A representação de um documento através de termos indexadores, descritores ou palavras-chave pode ser feita por especialistas ou extraída automaticamente do texto integral do documento.

A representação de documentos feita pelos bibliotecários ou profissionais da informação permite que pequenos conjuntos de termos classifiquem o documento. Infelizmente, é impossível contar com bibliotecários suficientes para classificar toda a web, sem contar no custo que isto poderia ter e o tempo que levaria a indexação.

Na representação de documentos automatizada, usa-se o que chamamos de “*text operations*” (processamento do texto), que consistem na eliminação das *stopwords* (como artigos e conjunções), *stemming* ou radicalização (o que reduz palavras distintas para a sua raiz comum gramatical) e a identificação de grupos nominais (o que elimina adjetivos, advérbios e verbos). O objetivo das *text operations* é reduzir a complexidade da representação do documento permitindo passar a visão lógica do documento de seu texto na íntegra para termos indexadores. (BAEZA-YATES; RIBEIRO-NETO, 1999, p.5). A figura 2 apresenta representação de um documento sob uma perspectiva lógica.

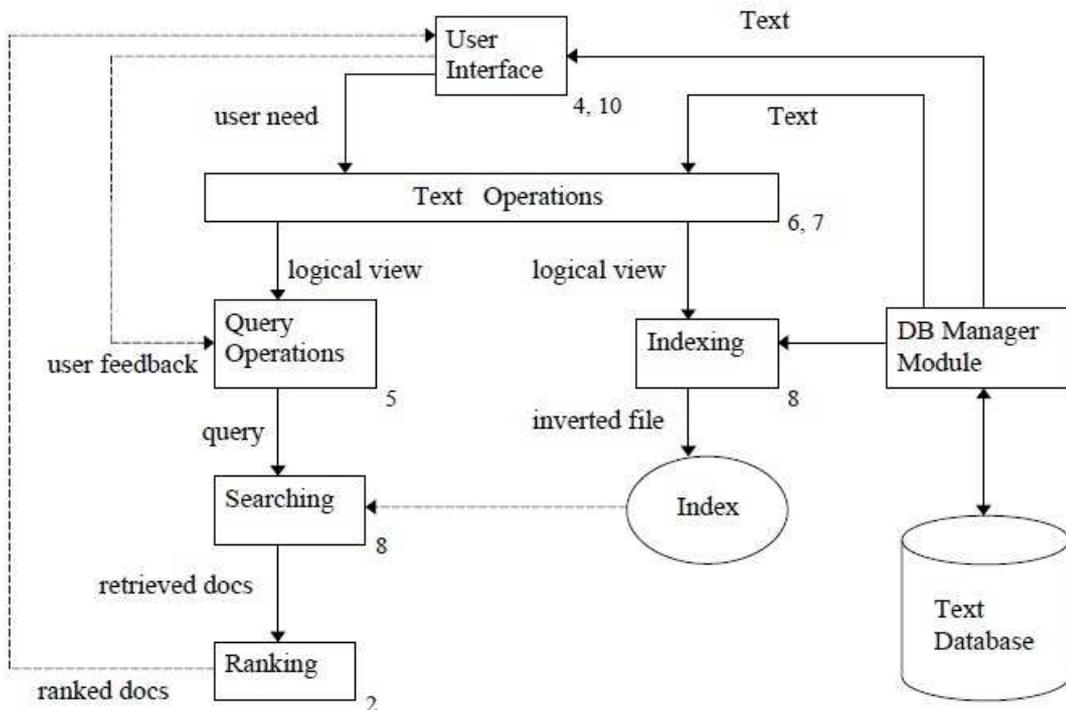
Figura 2 – Visão lógica de um documento: do texto integral aos termos indexadores



Fonte: Baeza-Yates e Ribeiro-Neto (1999, p.6)

Segundo Baeza-Yates e Ribeiro-Neto (1999, p.9, tradução nossa) o processo de recuperação da informação consiste nos passos descritos na figura 3.

Figura 3 – O processo de recuperação de informação



Fonte: Baeza-Yates e Ribeiro-Neto (1999, p.10)

Antes de iniciar o processo, deve-se estabelecer a base de dados que irá ser usada, definindo:

- a) Quais os documentos que compõem a coleção
- b) Quais as “*text operations*” que serão aplicadas aos textos
- c) O modelo de dados (estrutura de texto e quais elementos podem ser recuperados)
- d) O emprego das “*text operations*” transforma o texto completo dos documentos em uma visão lógica dos mesmos.

Depois disso, um índice do texto é construído. O índice é uma estrutura crítica de dados porque permite busca rápida sobre grandes volumes de dados. Quando a base de documentos está indexada, o processo de recuperação pode ter início. A figura 3 exibe o processo de recuperação de informação.

“Os recursos (tempo e espaço de armazenamento) investidos em definir a base de dados de texto e construir o índice são amortizados em muitas vezes pelo tempo de recuperação economizado.” (BAEZA-YATES; RIBEIRO-NETO, 1999, p.9, tradução nossa)

A **avaliação** é um importante tópico na área de Recuperação da Informação, e está amparada em duas unidades de medida: a Revocação (*Recall*) e a Precisão (*Precision*). Nesta seção trataremos destas duas formas de avaliação.

a) Revocação (*Recall*)

A recuperação é a proporção de documentos relevantes recuperados no conjunto de documentos da coleção. Mede-se a revocação (*recall*) para descobrir quanto a resposta apanha dos documentos relevantes da coleção

para aquela necessidade de informação, ou seja, quanto ela inclui no resultado recuperado os documentos relevantes para aquela pesquisa. Para calcular a revocação, usa-se a seguinte fórmula:

$$R = \frac{RA}{Rel}$$

Em que RA é o número de documentos relevantes na resposta e Rel o número de documentos relevantes na coleção.

b) Precisão (*Precision*)

A precisão, por outro lado, é a proporção de documentos relevantes na resposta, uma forma de medir o quanto de documentos não relevantes para a pesquisa foram excluídos da resposta. No âmbito dos sistemas de recomendação, “A precisão mede quanto um sistema é capaz de gerar recomendações de que os usuários gostem.” (TORRES, 2004, p.51)

A fórmula usada para calcular a precisão é:

$$P = \frac{RA}{A}$$

Onde RA é o número de documentos relevantes na resposta e A o número de documentos na resposta.

2.4 Filtragem de informação

A filtragem de informação leva em conta não apenas o interesse momentâneo do usuário, mas o seu perfil.

[...] em RI o usuário é responsável pela configuração da consulta, seja manualmente, seja utilizando uma interface, como ocorrem em uma biblioteca digital. Já em FI é o sistema que constrói o perfil do usuário e se encarrega de filtrar as informações de interesse. Em muitos sistemas, também é possível que o usuário crie ou altere seu perfil manualmente. (TORRES, 2004, p.77)

Assim, no contexto de um sistema de recomendação, não é necessário que o usuário formule uma consulta, pois uma vez o que seu perfil já está definido, torna-se possível oferecer ao usuário aquilo que se presume ser de seu interesse. O quadro 1 compara RI e FI:

Quadro 1 – Diferenças entre RI e FI

Recuperação de informações (RI)	Filtragem de informações (FI)
Uso esporádico com necessidade de informação momentânea	Uso constante por usuários com objetivos e interesses a longo prazo
Realização de consultas por meio de “queries”, criadas pelos usuários	O sistema constrói o perfil dos usuários e fornece as informações baseadas nele

Fonte: Tabela 5.1 – Recuperação de informações x filtragem de informações (Torres, 2004, p. 76)

Algumas técnicas funcionam bem com classificação explícita (como filmes e música), enquanto outras funcionam melhor com classificações implícitas (usando, por exemplo, visualizações de página, cliques, etc.).

2.5 Personalização

A personalização é uma evolução da segmentação de mercado, conceito tradicional do marketing que consiste em um grande grupo que é identificado a partir de suas preferências, poder de compra, localização geográfica, atitudes e hábitos de compra em comum (KOTLER, 2000, p.278). Para Torres (2004, p.25) “a personalização é uma técnica utilizada para

recomendar produtos aos consumidores, com base em seus perfis de consumo”.

À medida que métodos convencionais de propaganda (como mala-direta e catálogos) vêm encarecendo ao longo do tempo, a personalização ganhou força porque as pessoas dispõem de pouco tempo e querem mais conveniência para consumir (seja produtos ou informação), enquanto alimentam altas expectativas quanto a serviços de atendimento. Na web, a personalização pode ser largamente explorada porque oferece a possibilidade de rastreamento dos hábitos de navegação e consumo dos usuários, informações estratégicas para criação de ofertas personalizadas.

A personalização é uma técnica de marketing que objetiva aumentar a fidelização dos clientes e aumentar o retorno obtido em cada operação. Obtêm-se estes resultados porque oferece um serviço diferenciado a cada usuário, ajudando-os a encontrar bons produtos rapidamente e aumentar sua satisfação, o que fideliza. Com a personalização também é possível aumentar a venda casada, que eleva o número de vendas por transação, incrementando o lucro. (TORRES, 2004, p.27)

2.6 Sistemas de Recomendação

A capacidade de viver em sociedade permitiu que o homem evoluísse, porque poupamos esforços quando aprendemos com a experiência do outro, sem necessariamente ter de passar por ela. Nossas escolhas estão pautadas pelo que outros fizeram antes e, se tiveram êxito, este é o caminho seguro a seguir. Seguir recomendações é um traço marcante do indivíduo social. (ABEL, 2004, p.13)

Recomendação é uma forma particular de filtragem de informação, que explora os comportamentos do passado e semelhanças do usuário para

gerar uma lista de itens de informação que é pessoalmente adaptada às preferências de um usuário final.⁵

Resnick e Varian (1997) ensinam que os sistemas de recomendação auxiliam no aumento da capacidade e eficácia deste processo de indicação já bastante conhecida na relação social entre seres humanos.

O primeiro sistema de recomendação chamou-se Tapestry e foi criado para fins comerciais por Goldberg, em 1992 (TORRES, 2004, p.59). Os sistemas de recomendação permitem a personalização da oferta de produtos ou serviços para cada cliente (real ou potencial) através da análise de seu perfil e histórico e de técnicas – como a filtragem colaborativa, filtragem baseada em conteúdo ou frames de recomendação – que levam em consideração meta-informações da comunidade na qual o usuário está inserido e perfis similares.

O objetivo dos sistemas de recomendação é o de que os usuários não apenas recebam o retorno de itens a partir da formulação de consultas, mas que o possível interesse por um determinado item possa ser previsto, como uma forma de evitar a sobrecarga de informação (LICHTNOW *et al.*, 2006, p. 50). Agem de forma proativa sem que o usuário precise formular uma expressão de busca ou vasculhar dados estáticos a respeito dos dados de entrada para encontrar um produto ou serviço (SPEROTTO, 2010, p.34).

Segundo Chandra (2012, tradução nossa),

Sistemas de recomendação buscam prever a 'classificação' ou 'preferência' que um usuário poderia dar a um item (como música, livros ou filmes) ou elemento social (por exemplo, pessoas ou grupos) que ainda não tinha considerado.

São *softwares* responsáveis por receber pedidos de recomendação do servidor web, gerar as recomendações e devolvê-las. Têm como principais

5 Definição da ACM. Disponível em < <http://recsys.acm.org/> >. Acesso em 11 dez. 2012.

funções identificar o cliente, descobrir hábitos de consumo e gerar recomendações. (TORRES, 2004, p. 33-34) Barcellos *et al.* (2012) afirmam que “[...] 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.”

Para Barcellos *et al.* (2012), a ampla utilização dos sistemas de busca na web gera uma tarefa árdua para os usuários, que precisam filtrar dentro de muitos dados retornados, aqueles relevantes para si. E determinam que “[...] uma forma de auxiliar esse processo é o uso de sistemas de recomendação, que realizam indicações específicas para cada usuário, agilizando o processo de busca na Internet”. “Sistemas de recomendação procuram automatizar o processo social de indicação boca a boca”.⁶ (SHARDANAD; MAES, 1995 apud ROCHA, 2003, p. 2). Este processo é exemplificado por Terveen e Hill (2001, apud ROCHA, 2003, p. 2):

As pessoas geralmente recorrem a indicações de amigos, especialistas ou publicações especializadas quando precisam tomar uma decisão, escolher um produto que vão comprar ou até para decidir qual filme vão assistir no cinema. Assim como recorremos a esses recursos no dia a dia, saber a opinião de outras pessoas ajudaria a selecionar informações relevantes, em situações onde se tem um grande volume de informações. Os sistemas de recomendação automatizam esse processo, pois permitem que as pessoas compartilhem opiniões e possam se beneficiar com a experiência uma das outras⁷

6 SHARDANAD, U.; MAES, P. Social information filtering : algorithms for automating “word of mouth”. In: **Proceedings of the Conference on Human Factors in Computing Systems**. Denver: CHI 95, 1995. p. 210-127.

7 TERVEEN, L.; HILL, W. Beyond Recommender System : helping people to find each other. In: **Proceedings of HCI in the New Millennium**, [S.l.] : Addison-Wesley, 2001.

Os sistemas de recomendação utilizam modelos matemáticos e estatísticos para a predição, e alguns lançam mão inclusive de Inteligência Artificial⁸ no seu funcionamento.

2.6.1 Identificação de usuário⁹

Para que se possam recomendar leituras personalizadas relevantes, é necessário conhecer o perfil do usuário para o qual se está recomendando. Torres (2004, p. 36-37) apresenta duas formas de extração de perfil dos usuários:

- a) Explícita: nessa forma de extração, o usuário é responsável por informar ao sistema quais são suas preferências.
- b) Implícita: nessa forma de extração, o sistema extrai automaticamente o perfil do usuário, monitorando seus hábitos de consumo ou de navegação.

Na forma explícita, o usuário insere informações propositadamente para que o sistema as utilize para gerar recomendações. Uma forma comum de indicar as preferências são as avaliações (*ratings*), que veremos com mais detalhes a seguir. (TORRES, 2004, p.44)

⁸ A Inteligência Artificial é um campo de estudo da Ciência da Computação que estuda e cria sistemas com o objetivo de mimetizar a inteligência humana, programados a partir da ideia de que é possível reproduzir em máquinas a maneira de pensar, aprender e formar novos conhecimentos dos humanos. Produz softwares baseados em conhecimento de peritos para reproduzir o pensamento e comportamento humano, usando lógica, estatística e probabilidade para tomar decisões. São exemplos de sistemas que utilizam inteligência artificial: sistemas para jogar jogos (como em competições de xadrez), sistemas especialistas (tomam decisões para apoiar médicos no diagnóstico de pacientes), linguagem natural (compreendem linguagem de humanos), redes neurais (simulam inteligência, tentando reproduzir conexões físicas do cérebro de animais) e robótica (programação de computadores para ver, ouvir e reagir a outros estímulos sensoriais).

⁹ A identificação do usuário pode ser obtida por meio de *cookie* ou *login*. Segundo Torres (2004, p.34) "*cookie* é um arquivo gravado pelo site no computador do usuário, que vai guardando as interações de navegação." *Login*, por sua vez, é uma forma de reconhecer o usuário através da inserção manual de identificador e senha. Existem ressalvas em ambos os casos: a utilização de *cookies* pode ser desabilitada pelo usuário e o acesso via *login* frequentemente é indesejado pelos usuários, que preferem navegar anonimamente.

Na forma implícita, muitas vezes o usuário não sabe que essas informações estão sendo coletadas, e são obtidas através da sua navegação ou interações com aplicações. O acesso pode ser via arquivos de *log* de servidores ou monitoramento de páginas. Esta forma vem tomando espaço e tem sido alvo de estudo de muitos pesquisadores, “pois geralmente os usuários optam pela ‘lei do menor esforço’ no momento de receber recomendações”. (TORRES, 2004, p. 49)

A partir da criação do primeiro perfil do usuário (implícita ou explicitamente), este passa a ser modificado conforme mais interações ocorram. A construção e manutenção do perfil do usuário deve ser algo dinâmico ao longo do tempo, pois os interesses podem variar e o sistema deve ser capaz de “aprender” sobre o usuário à medida que ele utiliza o serviço.

2.6.2 *Descobrimdo hábitos de consumo*

Além de se identificar, o usuário precisa realizar alguma ação para que o sistema possa lhe recomendar conteúdos ou produtos. Adicionar itens a uma cesta, realizar a leitura de algum material ou avaliar um documento podem ser indicadores de consumo deste usuário. Em alguns sistemas até a navegação e o tempo de visualização de uma determinada página entram na conta do possível interesse do usuário no que está sendo exibido na página.

2.6.3 *Técnicas de recomendação*

Existem diferentes abordagens para a implementação de Sistemas de Recomendação e, eventualmente, ambas podem ser aplicadas simultaneamente. São elas:

- **Filtragem colaborativa**¹⁰ (*Collaborative Filtering*): técnica de recomendação que utiliza a similaridade entre os usuários para gerar recomendações (TORRES, 2004, p.58, grifo nosso).
- **Filtragem baseada em conteúdo** (*Content-based Filtering*): técnica de recomendação que utiliza a similaridade entre itens para gerar recomendações, mais usada para textos (artigos, notícias, etc.) (TORRES, 2004, p.75, grifo nosso).

O que determina a recomendação é a similaridade, que pode ser calculada entre itens ou entre usuários. Pode-se medir a similaridade entre itens ou usuários de varias formas, entre elas o coeficiente de Pearson¹¹ e o Co-seno¹².

a) Filtragem colaborativa

Segundo Reategui e Cazella (2005, p. 307), o termo foi cunhado pelos proponentes do primeiro sistema de recomendação, visando designar um tipo de sistema específico no qual a filtragem de informação era realizada com o auxílio humano, ou seja, pela colaboração entre os grupos de interessados.

Para Torres (2004, p. 58) a filtragem colaborativa (FC) é a “técnica de recomendação de produtos [ou documentos] que utiliza a similaridade entre os usuários para gerar recomendações.” Já Chandra (2012, tradução nossa) define os métodos de FC conforme segue:

¹⁰ Filtragem porque há um processo seletivo no momento de gerar as recomendações. Colaborativa porque as pessoas “colaboram” entre si para geração das recomendações utilizando a avaliação dada aos produtos que consomem. (TORRES, 2004, p.58)

¹¹ O coeficiente de Pearson mede a “força” ou o relacionamento entre duas variáveis, nesse caso os vetores que representam os perfis dos usuários, por meio de um valor entre [-1;1], onde -1 indica mínima correlação entre as variáveis e 1, máxima correlação. (TORRES, 2004, p.63)

¹² Há uma métrica que é baseada em co-seno e se aplica quando temos algum objeto (documento, usuário) representado por um vetor. Neste caso, mede-se o co-seno do ângulo existente entre os vetores, retornando um valor entre 0 e 1, onde 0 indica mínima correlação e 1, máxima correlação. (TORRES, 2004, p.65)

Os métodos de filtragem colaborativa (FC) são baseados na coleta e análise de uma grande quantidade de informações sobre os comportamentos dos usuários, atividades ou preferências - e predição do que os usuários vão gostar com base em sua similaridade com outros usuários.

Pode ser considerada uma forma automatizada de um processo que ocorre na vida real, conhecido como boca a boca. Para isso, é necessário que os usuários avaliem os documentos consumidos conforme sua satisfação. Alternativamente, o fato de ler, ou tomar emprestado um determinado documento, também pode ser considerado uma avaliação. A partir daí se pode determinar “vizinhos”, que são usuários similares ao usuário para o qual as recomendações estão sendo geradas (usuário ativo), e recomendar documentos que os vizinhos tenham lido e que não foram lidos pelo usuário ativo. (TORRES, 2004, p. 71)

Para Chandra (2012, tradução nossa),

uma vantagem chave da abordagem de filtragem colaborativa é que não se baseia no teor analisável por máquina e, portanto, é capaz de recomendar com precisão artigos complexos, tais como filmes, não requerendo uma "compreensão" [automática] do item em si.

Outras vantagens elencadas por Torres (2004, p. 68) são:

- independência de conteúdo¹³
- habilidade de recomendar produtos segundo características como qualidade e gostos¹⁴
- habilidade de produzir recomendações inesperadas e boas¹⁵.

¹³ “Como o sistema de FC é baseado nas avaliações que os usuários dão aos produtos, qualquer tipo de produto pode ser analisado, seja um filme, um livro, uma piada ou até um restaurante.” (TORRES, 2004, p. 68)

¹⁴ “[...] inerente ao processo de análise das avaliações, está a capacidade de recomendar um produto utilizando as análises subjetivas que os usuários tiveram, que foram expressas por meio de uma alta ou baixa avaliação. [...] Utilizando a análise do conteúdo do texto, extrair este tipo de informação não é uma tarefa bem-sucedida.” (TORRES, 2004, p. 68)

As desvantagens, segundo o autor, são:

- primeiro avaliador: apenas produtos já avaliados por outros usuários similares podem ser recomendados¹⁶,
- esparsidade do domínio: em grandes coleções, em que o número de documentos é muito maior do que o número de usuários, é difícil encontrar bons vizinhos pelo reduzido número de documentos em comum consumidos/avaliados pelos usuários (baixo *overlap*). Por exemplo, na Amazon, que possui 3 milhões de itens, um usuário para consumir apenas 0,1% dos produtos terá que comprar 3.000.¹⁷
- demanda por alta capacidade de processamento¹⁸.

Uma vantagem importante da filtragem colaborativa é que as vizinhanças não são pré-computadas. Cada vez que uma recomendação é requisitada para um usuário, uma nova vizinhança é determinada, ou seja, a vizinhança gerada para um usuário no período da manhã pode ser diferente da vizinhança gerada à tarde. (TORRES, 2004, p. 69) Isto permite que a geração de vizinhanças utilize dados atualizados das avaliações que os usuários dão aos produtos/documentos, mas exige uma alta capacidade de processamento computacional (embora a frequência de atualização do perfil possa ser configurável pela biblioteca). Torres (2004, p. 69) destaca que a sobrecarga no servidor pode ser reduzida através de outras técnicas:

¹⁵ “[...] FC pode produzir recomendações inesperadas, mas boas (em inglês isso é chamado *serendipity*). Essa é uma característica tipicamente da FC, que ocorre em virtude de representar os hábitos de consumo dos usuários.” (TORRES, 2004, p. 68)

¹⁶ “Tendo em vista que na FC as recomendações são produtos que foram consumidos pelos vizinhos, um produto jamais será recomendado sem que um usuário (vizinho) o consuma.” (TORRES, 2004, p. 69)

¹⁷ “Como geralmente existem milhares de produtos de uma loja virtual [ou documentos em uma biblioteca digital e na web], é pouco provável que um usuário consuma um grande percentual desses produtos. [...] Em virtude da esparsidade, torna-se difícil encontrar bons vizinhos, pelo reduzido número de produtos em comum consumidos pelos usuários (baixo *overlap*).” (TORRES, 2004, p. 69)

¹⁸ “Como a vizinhança é gerada para cada pedido de recomendação, FC requer bastante capacidade de processamento em virtude de seu custo computacional.” (TORRES, 2004, p. 69)

Outras técnicas, como frames de recomendação e geração de regras associativas utilizando mineração de dados, criam um modelo que demonstra como os usuários se comportam e atualizam esse modelo periodicamente, reduzindo a sobrecarga do servidor.

Há ainda a questão do falso bom vizinho. Enquanto um vizinho A avaliou 4 documentos de forma similar a B, e alcança um valor de similaridade alto (próximo a 1), o usuário C pode ter avaliado 30 documentos de forma mais ou menos similar a A e obter um coeficiente de similaridade igual a 0,6. Normalmente, o sistema iria indicar primeiro as opções de B, mas C teve uma relação mais estável de concordância e isso pode significar mais consistência na similaridade entre A e C, e as quatro avaliações semelhantes serem apenas coincidência. Para resolver este impasse, é necessário “atribuir um ‘peso’ aos coeficientes de similaridade, de tal forma que se o número de produtos [ou documentos] avaliados em comum for pequeno, esse ‘peso’ reduzirá o valor de similaridade.” (TORRES, 2004, p. 70)

b) Filtragem baseada em conteúdo

Filtragem baseada em conteúdo (FBC) é uma “técnica de recomendação de produtos que utiliza a similaridade entre os produtos para gerar recomendações. Geralmente, utilizada para recomendar texto (artigos, notícias, etc.)” (TORRES, 2004, p.75)

A FBC tem seu histórico originado nos sistemas de recuperação da informação (RI), que existem há décadas. (TORRES, 2004, p.75) A forte relação que esta abordagem guarda com área de recuperação da informação deve-se ao processo de comparação que pode ser feito entre os atributos dos documentos, pois quando os itens correspondem a artigos (ou

sites) os documentos podem ser considerados similares se compartilharem termos em comum. (LICHTNOW *et al.*, 2006, p.50)

Segundo Lichtnow *et al.* (2006, p.50), a filtragem baseada em conteúdo “[...] parte do princípio de que os usuários tendem a interessar-se por itens similares aos que demonstraram interesse no passado. Assim é definida a similaridade entre os itens.” Precisamos dela porque é inviável categorizar toda informação publicada na web. (TORRES, 2004, p.75)

Torres (2004, p.85) afirma que a principal técnica utilizada na FBC é TF-IDF em virtude de três componentes usados para medir similaridade entre textos: frequência de termos, discriminância das palavras e normalização.

Vantagens da filtragem baseada em conteúdo:

- não possui o problema do primeiro avaliador¹⁹,
- capacidade de recomendar todos os itens (redução da esparsidade)²⁰.

São desvantagens desta técnica:

- impossibilidade de análise de conteúdo em domínios não textuais: A análise de textos não é capaz de medir eficientemente áudio e vídeo²¹
- não consideração de aspectos como qualidade do texto e renome do autor: a qualidade e a clareza com que o texto é escrito são ignoradas, bem como o fato de um texto ter sido escrito por um autor renomado em uma determinada área²²
- superespecialização: por recomendar textos com base no casamento de palavras chave, essa técnica tende a recomendar sempre textos

¹⁹ “[...] um texto não precisa ser lido por um usuário para que seja recomendado a outro, pois basta que existam palavras em comum entre o texto e o perfil do usuário para que ocorra a recomendação. Dessa forma um texto novo no sistema pode ser recomendado a um dos usuários.” (TORRES, 2004, p. 84)

²⁰ “[...] para todo texto pode ser medida a similaridade com o perfil do usuário.” (TORRES, 2004, p. 84)

²¹ Alternativamente pode-se utilizar letra de música ou sinopse, mas a carência de texto nessas sinopses e indisponibilidade de todas as sinopses em meio digital prejudicam esta alternativa. (TORRES, 2004, p. 84-85)

²² (TORRES, 2004, p. 85)

semelhantes, de um determinado assunto, criando uma superespecialização de conteúdo²³

Outra desvantagem significativa é que, por ser completamente baseada em palavras, FBC pode não considerar similares palavras como “carro” e “automóvel”, mesmo que abordem o mesmo assunto. Para solucionar este problema, “podem-se utilizar dicionários de sinônimos para identificar quais palavras têm o mesmo significado”. (TORRES, 2004, p. 85)

2.7 Serendipidade

Os sistemas de recomendação têm sofrido algumas críticas, especialmente no que se refere à superespecialização. Os argumentos giram em torno da serendipidade (do termo em inglês, *serendipity*), que significa encontrar algo útil ou agradável pelo qual não se esteja procurando. No contexto dos sistemas de recomendação, serendipidade é a “Característica que mede se o sistema recomenda itens inesperados que os usuários gostam” (TORRES, 2004, p. 51). Sistemas que não trabalham com a correlação usuário para usuário (apenas com FBC, por exemplo) tendem a não recomendar itens que não estejam especificamente dentro do perfil previamente desenhado, e assim a “surpresa agradável” não aparece.

Por exemplo, dois usuários A e B podem gostar muito de música MPB. Ao verificar esta informação, FC pode identificar que A também gostou muito de um álbum de jazz ou de um livro policial, por algum motivo. Estatisticamente, existe a possibilidade de que o usuário B também goste desse álbum ou do livro, mesmo que Jazz e histórias policiais não façam parte de suas preferências. (TORRES, 2004, p. 68)

²³“Por exemplo, uma pessoa que gosta de receitas de tortas pode nunca receber uma receita de pudim como recomendação, apesar de ambas serem tipos de sobremesas.” (TORRES, 2004, p.85)

Sistemas híbridos de recomendação, que combinam filtragem baseada em conteúdo com filtragem colaborativa, tendem a amenizar as vantagens e desvantagens recíprocas, e são uma tendência para estes sistemas. O quadro 2 relaciona FC e FBC, apontando vantagens e desvantagens para cada item relacionado.

Quadro 2 – Vantagens e desvantagens de FC e FBC

FC		FBC	
<i>Vantagem</i>	Independência de conteúdo	Dependência de conteúdo	<i>Desvantagem</i>
<i>Vantagem</i>	Uso de qualidade e gosto	Não usa qualidade e gostos	<i>Desvantagem</i>
<i>Vantagem</i>	Serendipidade	Superespecialização	<i>Desvantagem</i>
<i>Desvantagem</i>	Primeiro avaliador	Não há problema do primeiro avaliador	<i>Vantagem</i>
<i>Desvantagem</i>	Esparsidade	Não há esparsidade	<i>Vantagem</i>

Fonte: adaptado pela autora da Tabela 7.1 – Vantagens e Desvantagens de FC e FBC (TORRES, 2004, p. 96)

Serendipidade é importante porque permite que itens diferentes do que o usuário costuma consumir podem ser recomendados, e mesmo assim ser de seu interesse.

2.8 Tendências futuras em recomendação

Verificou-se que algumas técnicas de recomendação (FC, FBC, etc.) são melhores do que outras para recomendar, por exemplo, artigos científicos mais novos, e se o usuário busca por artigos recentes deve preferir uma técnica em detrimento de outra. Porém, ele não tem poder sobre isso, porque o sistema age por trás da camada de interface.

Como foi comentado anteriormente, os estudos recentes na área de recomendação têm apontado para o uso de sistemas híbridos para melhorar a capacidade de recomendação. Torres (2004, p. 39) lista algumas tendências em pesquisa e desenvolvimento de software desta área:

- a) aprimoramento da qualidade das recomendações;
- b) aumento da velocidade das recomendações;
- c) recomendação de produtos a usuários com poucas informações no perfil;
- d) atualização das preferências dos usuários com o passar do tempo.

Para coletar informações implícitas, também podem ser usados movimentos de mouse e movimentação (*scroll*) da página. (TORRES, 2004, p. 43)

2.9 Soluções de recomendação existentes para bibliotecas

Elencamos a seguir alguns sistemas de recomendação para catálogos on-line que já estão em uso em bibliotecas universitárias. Nesta seção não há pretensão de se esgotar o assunto, apenas listar algumas das opções que podem interessar às instituições que vislumbram a utilização da recomendação em seus catálogos. Além das citadas, outras empresas também oferecem sistemas de recomendação, que podem ser utilizados para enriquecer a busca do usuário.²⁴

²⁴ No Brasil, Giseli Rabello Lopes, aluna do Instituto de Informática da Universidade Federal do Rio Grande do Sul, propôs em 2007 um Sistema de recomendação para bibliotecas digitais sob a perspectiva da web semântica. Segundo a autora, o sistema recomenda “[...] artigos científicos, na área da Ciência da Computação, que estejam de acordo com os interesses do usuário identificados a partir de informações presentes em seu currículo Lattes. Sob o ponto de vista tecnológico, o sistema proposto foi desenvolvido sob a perspectiva da Web Semântica, à medida que faz uso de suas tecnologias emergentes tais como: uso de meta-dados padrão para descrição de documentos – Dublin Core [...], uso de padrão XML para descrição do perfil do usuário – currículo Lattes [...], e utilização de provedor de serviços e dados para gerar a recomendação.” (LOPES, 2007, p. 14-15)

2.9.1 *bX Reccomender*

*bX Reccomender*²⁵²⁶ é um sistema de recomendação de literatura acadêmico-científica, criado por Johan Bollen e Herbert Van de Sompel em uma colaboração da empresa Ex Libris, desenvolvedora de sistemas de automação de bibliotecas, e do Los Alamos National Laboratory (LANL)²⁷.

Ele atua sobre o Primo-Central, que é um mega-índice centralizado de conteúdos acadêmico científicos. Tanto *bX* quanto Primo-Central centralizam os logs de vários catálogos on-line, e têm seus serviços hospedados na nuvem. Tem como objetivo criar novos conhecimentos, ajudando os usuários a descobrir a literatura relevante que eles não teriam encontrado de outra forma, e o faz expondo os padrões de uso, associações entre temas, autores e artigos, bem como popularidade e sucesso de sistemas de bibliotecas. Portanto, analisa o comportamento de leitura via log e recomenda artigos, ordenando-os por relevância.

O *bX Reccomender* utiliza a análise e mineração de dados de uso para enriquecer e ampliar a experiência de descoberta do pesquisador. Como seus dados ficam centralizados, é possível recomendar a partir da análise do comportamento de leitura (via *log*) de milhões de pesquisadores em todo o mundo, o que expande muito mais a pesquisa. A recomendação é feita a partir do artigo, e o usuário pode verificar que outros artigos foram utilizados em conjunto com ele através de uma lista de artigos relevantes.

Os usuários, ao se logarem na base Primo-Central, podem personalizar seus perfis, inserindo informações como nível de escolaridade e área do

²⁵ Disponível em < <http://www.exlibrisgroup.com/category/bXRecommender> >. Acesso em 12 dez. 2012.

²⁶ Vídeo sobre o *bX* está disponível no endereço < <http://www.youtube.com/watch?v=YvrPhATtGvY> >. Acesso em 12 dez. 2012.

²⁷ Um trabalho dos pesquisadores foi publicado e está disponível no link < http://old.cni.org/tfms/2005b.fall/abstracts/handouts/CNI_bx_bollen2.pdf >. Acesso em 12 dez. 2012.

conhecimento de sua linha pesquisa. Outra interação possível é a avaliação dos artigos pelos usuários.

As instituições que utilizam a base Primo-Central podem, através do envio dos logs de acesso realizados nas suas bibliotecas, contribuir para o aprimoramento do algoritmo e conseqüente melhoria da experiência de descoberta do usuário.

Focada exclusivamente no domínio acadêmico, é um serviço pioneiro na oferta de recomendações altamente granulares que apontam a determinados artigos acadêmicos. Os artigos podem ser de diferentes revistas, editoras e plataformas e podem ser representados por palavras-chave diferentes.

2.9.2 *Sciverse Sciencedirect*

Na base de dados SciVerse ScienceDirect²⁸ da Elsevier, o bX também é oferecido, chamado de artigos relacionados (*Related articles*). Para oferecer este serviço, o texto completo do artigo passa por um processo automático de varredura para extração de palavras-chave. Os termos recebem um peso em relação à frequência de ocorrência e local onde se encontram no artigo e, em seguida, são comparados com um banco de dados. Finalmente é realizada uma complexa busca, comparado o vetor resultante com o de outros artigos, e o resultado é uma lista de artigos classificados por similaridade.

Os artigos relacionados sugeridos são artigos aos quais o usuário pode ter acesso na íntegra.²⁹ (ELSEVIER, 2012, tradução nossa).

²⁸ Disponível em < <http://www.sciencedirect.com/> >. Acesso em 12 dez. 2012.

²⁹ **How related articles are determined.** *Related articles are determined as follows:*

- *An entire article, including content from tables, is processed and key terms are extracted.*
- *The terms are weighted for the frequency of occurrence and placement in the article and then matched against a database.*

Figura 4 – Tela de resultados de busca com link para artigos relacionados na base SciVerse Science Direct

Fonte: Catálogo online da SciVerse ScienceDirect, grifo da autora.

2.9.3 Sistema de Recomendação Advanced IT

O Sistema de Recomendação da Advanced IT faz a sintonia fina das classes de documentos e usuários dentro do ambiente onde os documentos são armazenados. Funciona como um “motor” de recomendação (*recommender engine*), e dispõe de relatórios e indicadores para potencializar o uso da ferramenta na biblioteca.

As recomendações que o sistema mobiliza são baseadas em critérios configuráveis (montagem de perfil do usuário, perfil dinâmico – através de comportamento) e têm como finalidade:

- The resulting "vector," which is essentially a complex search, is run against other articles, and the result is a list of articles ranked by similarity.

Note: The related articles suggested in the article suggestions pop-up are articles that you may have full access to. (Elsevier, 2012)

- a) **Diminuir a sobrecarga de informação** (causada por oferta muito grande de conteúdo) reduzindo conteúdos não relevantes;
- b) **Personalizar o consumo** através de ações específicas para indivíduos que possuem comportamentos similares
- c) **Aumentar satisfação dos usuários**, que pode ser acompanhada por meio da taxa de retorno

Através de pontuações (perfil = classe + pontuação pelo comportamento), o sistema colhe automaticamente as informações que compõem o perfil dinâmico do usuário. Suas interações com a biblioteca – como reserva, empréstimo, devolução, visualização, navegação no site – atualizam este perfil, tornando possível:

- a) Filtrar conteúdo por perfil de usuário (eliminar sobrecarga)
- b) Mapear comportamentos para apoio à decisão (identificar)
- c) Gerar ofertas de conteúdo personalizadas (indicar)

Por ser um motor e atuar na camada intermediária entre o banco de dados e a aplicação, a ferramenta é flexível e configurável, podendo ser utilizada tanto para documentos (portais e intranets, sites e catálogos eletrônicos) quanto para produtos ou serviços (bens duráveis, roupas, imóveis, etc.), no comércio eletrônico ou como apoio em pontos de venda.

O sistema envolve os módulos de gerenciamento das informações dos perfis (detalhamento de perfis funcionais e sociais), dos produtos (características, preços, promoções), das estatísticas (utilização do sistema de recomendação) e os mecanismos de recomendação (*engine* que propõe os produtos). Seu funcionamento está integrado a uma ontologia de domínio que serve para associar classes a usuários e a documentos – as classes podem estar estruturadas de forma hierárquica.

Pode ser utilizado para auxiliar em tarefas relacionadas a *e-commerce*, *CRM*, *e-mail marketing*, *call centers* e processos de Gestão do Conhecimento. Tem integração com Gerenciamento Eletrônico de Documentos (GED) e *Content Management*.

2.9.4 *Related books links in Aleph OPAC*

O *Related books links in Aleph OPAC*³⁰ é um recurso de recomendação em catálogos on-line criado por Matthew Phillips, então funcionário do *Library & Learning Centre (LLC)* da *University of Dundee*³¹ (Escócia), em 2010. Tem a função de recomendar documentos a partir do catálogo da biblioteca da universidade. Depois de implementado na universidade de Dundee, foi disponibilizado no ambiente de colaboração entre desenvolvedores do software Aleph, *El Commons CodeShare*³².

Esta ferramenta apresenta três possibilidades de recomendação de documentos:

- a. *People who borrowed this have also borrowed* (pessoas que retiraram esta obra também retiraram): similaridade usuário-usuário calculada com base na tabela de empréstimos
- b. *Related works include* (trabalhos relacionados incluem): similaridade item-item pela classificação de assunto atribuída pelos bibliotecários
- c. *Other editions include* (outras edições incluem): similaridade item-item onde o que varia é a edição de um mesmo documento

³⁰Disponível em <<http://www.exlibrisgroup.org/display/AlephCC/Related+books+links+in+Aleph+OPAC>>. Acesso em 12 dez. 2012.

³¹Disponível em <<http://www.dundee.ac.uk/library/>>. Acesso em 12 dez. 2012.

³² Neste ambiente os usuários e profissionais da ExLibris trocam experiências e compartilham códigos criados para melhoria no Aleph, pois partes do código estão disponíveis para este fim. Disponível em <<http://www.exlibrisgroup.org/display/CodeShare/Home>>. Acesso em 12 dez. 2012.

Ao acessar um registro completo do catálogo, a ferramenta mostra na mesma tela as recomendações: outras edições da obra, outros itens emprestados a pessoas que também tomaram emprestado aquele documento, e outras obras com os mesmos assuntos (relacionados pela indexação dos bibliotecários). As informações de recomendação são pré-calculadas em segundo plano, já que fazer este trabalho (medir similaridade) no momento da consulta pode tornar o sistema muito lento para o usuário.

A seguir explicamos um pouco melhor cada uma das funcionalidades:

a) *People who borrowed this have also borrowed* (pessoas que retiraram esta obra também retiraram)

Considera similares para recomendação aqueles documentos que foram emprestados para usuários em comum. O livro deve ter sido emprestado para pelo menos cinco usuários diferentes para aparecer nesta seção. O ordenamento dos livros na lista de trabalhos relacionados por empréstimo se dá pela quantidade de usuários mutuamente comuns àquele livro. Essas ligações são recalculadas semanalmente. Esta seção não é exibida 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. São “*linked by borrowings*” ou, ligadas pelos empréstimos, e por isso, no *log* do sistema, este tipo de recomendação é representado pela letra “B” (*borrowings*).

b) *Related works include* (trabalhos relacionados incluem)

Esta seção explora os cabeçalhos de assunto e os números de classificação nos registros MARC (por exemplo, a 080, 082, 093, 600, 650, 690 e áreas afins) e identifica os documentos que têm pelo menos três destas posições em comum. Tem de haver pelo menos três posições ou números de classificação em comum para que o sistema exiba livros

relacionados. Os *links* são recalculados a cada 14 dias. São “*linked by classification or subject headings*”, ou seja, os itens são ligados entre si pela classificação ou cabeçalhos de assunto atribuídos pelos bibliotecários. Por este motivo, no *log* do sistema, este tipo de recomendação é representado pela letra “H” (*headings*).

c) *Other editions include* (outras edições incluem)

Considera similares para recomendação aqueles documentos com o mesmo *International Standard Book Number* (ISBN), e que fazem parte do acervo da biblioteca. Este recurso utiliza um serviço na OCLC que identifica outras edições por ISBN. Portanto, só é possível encontrar *links* para outras edições se todos eles tiverem o mesmo ISBN. Entretanto, depende do banco de dados da OCLC, às vezes podem ocorrer problemas nesta recuperação. Livros desta seção estão listados em ordem decrescente de ano de publicação e até cinco serão mostrados. Uma vez calculados, esses dados só serão recalculados para qualquer livro em particular depois de passados 30 dias. No *log* do sistema, este tipo de recomendação é representado pela letra “E” (*editions*).

Um exemplo pode ser visualizado na figura 5.

Figura 5 – Tela de registro completo do catálogo da Universidade de Dundee, com destaque para as recomendações (*Related books in Aleph OPAC*)

The screenshot displays the University of Dundee Library Catalogue interface. The top navigation bar includes the 'CATALOGUE' logo, 'University of Dundee Library Catalogue', and 'LIBRARY & LEARNING CENTRE'. A search bar and various utility links like 'Sign-in', 'End Session', and 'My e-Shelf' are present. Below the navigation bar, the 'Full View of Record' for a book is shown. The main entry details include the title 'Notes on medical bacteriology / I. Douglas Sleight, Morag C. Timbury', 5th ed., published in 1998. The subject is 'Medical microbiology - outlines'. A red box highlights two sections: 'Other editions include:' and 'People who borrowed this have also borrowed:'. The 'Other editions include:' section lists 'Notes on medical bacteriology - 4th ed.' by Sleight, J. Douglas (1994). The 'People who borrowed this have also borrowed:' section lists several related titles, including 'Core paediatrics and child health' (2000), 'ABC of eyes - 4th ed.' (2004), 'Microbiology in clinical practice - 3rd ed.' (1999), 'Chest x-ray made easy - 2nd ed.' (2002), and 'Surgery 1 - 2nd ed.' (2002). Below this, the 'Related works include:' section lists 'Lecture notes on medical microbiology - 3rd ed.' (1997), 'Bergey's manual of determinative bacteriology - 9th ed.' (1994), 'Bacterial pathogenesis' (1994), 'Medical microbiology - 2nd ed.' (1998), and 'Jawetz, Melnick & Adelberg's medical microbiology - 19th ed.' (1991).

Fonte: Catálogo online da biblioteca da University of Dundee, grifo da autora.

Esta ferramenta foi objeto deste estudo e mais informações acerca dele serão fornecidas na seção 4 deste trabalho.

3 Metodologia

Este capítulo apresenta a classificação metodológica do trabalho que visa atender aos objetivos propostos no capítulo 1.

3.1 Classificação metodológica da pesquisa

- a) **com base na natureza:** quanto à natureza, o trabalho constitui-se de natureza aplicada, pois tem a intenção de encaminhar soluções para um problema prático, apresentado no item 1.1 deste trabalho. Segundo Boaventura (2009, p. 56), uma pesquisa aplicada gera conhecimentos úteis à solução de problemas sociais.
- b) **quanto à abordagem:** a abordagem da pesquisa é quantitativa, pois trabalha e se expressa através de números, apresentando porcentagem, média, [...]. (BOAVENTURA, 2009, p. 56). O método quantitativo caracteriza-se “pelo emprego da quantificação tanto nas modalidades de coleta de informações, quanto no tratamento dessas através de técnicas estatísticas”. (RICHARDSON, 1985, p. 29)
- c) **com base nos seus objetivos:** quanto aos objetivos, a pesquisa é descritiva, pois estas “têm como objetivo primordial a descrição das características de determinada população ou fenômeno ou, então, o estabelecimento de relações entre variáveis” (GIL, 2009, p. 42)
- d) **com base nos procedimentos técnicos utilizados:** a pesquisa é documental do ponto de vista dos procedimentos técnicos, pois estará baseada na análise de registros de ação (*logs* de acesso dos usuários a um sistema de recomendação). Segundo GIL (2009, p. 46) “[...] a

pesquisa documental vale-se de materiais que não recebem ainda um tratamento analítico, ou que ainda podem ser reelaborados de acordo com os objetos da pesquisa.”

- e) **com relação aos métodos para tratamento dos dados e apresentação dos resultados:** foram utilizados métodos matemáticos para tratamento dos dados. Boaventura (2009, p. 56) exemplifica uma pesquisa quantitativa como aquela que “trabalha e se expressa através de números, apresentando porcentagem, média [...]” Os resultados foram apresentados em gráficos e tabelas.

3.2 Etapas de desenvolvimento do trabalho científico

Nesta seção explanamos sobre as fases que tornaram a pesquisa possível. Quanto às etapas desenvolvidas, inicialmente foi concluído o levantamento bibliográfico.

3.2.1 *Levantamento bibliográfico*

As leituras do referencial teórico serviram para elencar os tipos de sistemas de recomendação existentes e as técnicas empregadas nos mesmos, além de sua relação com a literatura da Ciência da Informação.

O levantamento bibliográfico então teve foco nos temas

- a) serviço de referência em bibliotecas universitárias;
- b) recuperação da informação (*Information Retrieval*);
- c) personalização e sistemas de recomendação;
- d) soluções [sistemas de recomendação] existentes para bibliotecas;

3.2.2 Coleta de dados

A equipe escolheu trabalhar com a ferramenta “*Related Books in Aleph OPAC*”, e recebeu duas amostras de *log*³³ das bibliotecas da *University of Dundee* (que utilizam Aleph OPAC). O *log* era composto de URLs extraídas do servidor que registra os acessos ao catálogo da biblioteca. O período de levantamento compreendeu 30 dias: de 04 a 18/12/2012 e de 22/04/2013 a 06/05/2013.

Por questão de privacidade dos usuários, os arquivos texto recebidos tiveram omitidos o identificador IP do computador originário das visualizações de documentos e documentos recomendados. Também foram omitidos, os posts que correspondem às expressões de busca.

3.2.3 Tratamento dos dados

Para extração de dados do *log* (recebido em formato texto), foi criado um software extrator, que teve as seguintes funções: eliminar as URLs repetidas³⁴ do *log* (cerca de 30%), ordenar cronologicamente as URLs, contar o intervalo de tempo entre o primeiro e o último registro de uma sessão (em segundos) e exportar todos os dados para uma tabela (formato .CSV) organizando-os em colunas por tipo de informação. O cálculo de tempo foi feito a partir do campo data/hora em que o horário do último registro na sessão foi subtraído do horário do primeiro registro na mesma sessão registrado no *log*. Em algumas análises que envolvem o tempo das sessões, no entanto, as sessões com 0 segundos (aquelas com apenas um registro) e

³³Um *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)

³⁴Como URLs repetidas consideramos os *posts* que apresentam a mesma sessão, os mesmos parâmetros, todas as informações idênticas a outro, com exceção do campo “data”. Acreditamos que estes casos sejam um *refresh* da página, pois exibe poucos segundos ou minutos de diferença.

com mais de 24 horas (86.400 segundos) foram desconsideradas, para que os dados não fossem distorcidos. As sessões são contabilizadas normalmente nas demais análises.

Depois da extração obteve-se dois arquivos .CSV (planilha de dados com colunas separadas por ponto-e-vírgula) e dois arquivos .TXT (apenas com os tempos das sessões) que foram importados para tabelas de um banco de dados (Microsoft Access).

3.2.4 *Estrutura dos dados*

Primeiramente (a), os registros do *log* foram classificados em tipos de acordo com as características apresentadas. Esta classificação pode ser visualizada na alínea a (Caracterização do *log*) desta sessão. A seguir, na alínea b (Sessões) é apresentado o conceito de sessão e como se relaciona com os tipos de registros encontrados.

A partir deste trabalho foi possível dar início às consultas ao banco de dados, realizadas na ferramenta Microsoft Access e linguagem SQL. As consultas resultaram em tabelas (exportadas para Microsoft Excel) e transformadas em gráficos, que são exibidos na seção 4 deste trabalho.

a) Caracterização do *log*

O *log* utilizado para análise nesta pesquisa, conforme já foi mencionado, registrou acessos realizados pelos usuários da *Main Library* da *University of Dundee* ao catálogo desta, nos períodos entre 04/12/2012 a 18/12/2012 e 22/04/2013 a 06/05/2013, totalizando 30 dias. Para analisar os dados à luz dos conhecimentos estudados, dispomos de 16.806 registros (URLs) que abrangem registros de documentos visualizados, provenientes de consultas e recomendação.

Cada registro de *log* (que será chamado de *post* daqui para frente) é composto por informações que representam uma interação do usuário com o sistema, como, por exemplo, o acesso ao registro de um documento do catálogo a partir de uma consulta ao registro ou a partir de uma recomendação. Os elementos que compõem um *post* são chamados de parâmetros. Abaixo constam os parâmetros e seus significados.

- *doc_number*: o registro identificador do documento exibido. Significa que o *post* corresponde à exibição ao usuário de um registro de documento do catálogo. *Doc_number* representa o número do registro do documento
- *dunsid=Other*: representa que o usuário clicou em uma recomendação. Significa que o *post* corresponde à exibição ao usuário de um registro de documento do catálogo que teve como origem uma recomendação. O parâmetro possui a seguinte estrutura para seus valores:
 - *dunsid=OtherB1*
 - B: indica ligação por empréstimos³⁵. Indica que o registro clicado pelo usuário foi recomendado a partir de similaridade com usuários que emprestaram o mesmo documento. Este tipo de recomendação é chamado de **“Pessoas que retiraram esta obra também retiraram”**
 - 1: indica a profundidade³⁶ da recomendação, ou seja, quantas vezes o usuário clicou em um documento recomendado deste tipo.

³⁵ *Linked by borrowings (B)*.

³⁶ Chamamos profundidade o que o sistema *Related Books in Aleph Opac* registra no log após a tag *“dunsid”* quando o usuário clica em um documento recomendado. Este número recebe um incremento a cada clique que o usuário dá. Assim é possível acompanhar quantas vezes o usuário clicou em documentos recomendados. Quando um usuário clica em uma recomendação, o log registra algo como *“dunsid=OtherB1”*. Se na próxima página o usuário clicar novamente em uma recomendação, o sistema incrementa uma unidade a cada clique, registrando no log, por exemplo, *“dunsid=OtherB2”* e assim por diante. Nota-se que se o usuário usar outro tipo de recomendação, a profundidade persiste. No exemplo, se, após visualizar 2 documentos recomendados pelo tipo B, o usuário escolhesse recomendação do tipo E, o registro seria *“dunsid=OtherE3”*.

- *dunsid=OtherH1*: H indica ligação por classificação³⁷. Indica que o registro clicado pelo usuário foi recomendado a partir de cabeçalhos de assunto semelhantes. Este tipo de recomendação é chamado de **“Trabalhos relacionados incluem”**.
- *dunsid=OtherE1*: E indica outras edições³⁸. Indica que o registro clicado pelo usuário foi recomendado a partir de outra edição do mesmo documento. Este tipo de recomendação é chamado de **“Outras edições incluem”**.
- *session*: identificador da sessão do usuário.
- *set_number*: identificador da consulta³⁹. Quando o *post* corresponde à exibição ao usuário de um registro de documento do catálogo acessado a partir de uma consulta, identifica a consulta.
- *set_entry*: sequência daquele registro na lista de resultados. Quando o *post* corresponde à exibição ao usuário de um registro de documento do catálogo acessado a partir de uma consulta, indica a posição do registro na lista de resultado da consulta.

A partir destas informações, os *posts* foram classificados de acordo com o tipo de interação do usuário. Se o *post* corresponde a um registro de um documento do catálogo acessado a partir de uma consulta, este é chamado de REGDOCCON; se o *post* corresponde a um registro acessado a partir de uma recomendação, este é chamado de REGDOCREC; se o *post* corresponde a um registro de um documento do catálogo acessado em uma sessão, mas não tem relação com consulta ou recomendação, este é chamado de OUTROS DOC (COM SESSÃO). *Posts* correspondentes a interações não associadas às sessões foram desconsiderados neste estudo.

³⁷ *Linked by classification or subject headings (H).*

³⁸ *Variant editions (E).*

³⁹ Quando o usuário clica no item não abrindo em outra aba, ao retornar o *set_number* é outro, pois o sistema executa novamente a consulta.

Uma sessão corresponde a um ou mais registros de navegação (*posts*) feitos por um mesmo usuário, num determinado período de tempo.

Os *posts* que correspondem a um registro acessado a partir de uma recomendação (REGDOCREC) foram sub-classificados de acordo com o tipo de recomendação, isto é, “Pessoas que retiraram esta obra também retiraram”, “Trabalhos relacionados incluem” e “Outras edições incluem”.

Além desta classificação, a profundidade do uso das recomendações foi extraída dos *posts*. A profundidade é mensurada a partir do incremento de um número ao registro do *log* cada vez que o usuário clica em uma recomendação. Assim, identifica quantos documentos já foram visitados pelo usuário através de recomendação, em uma sessão.

O quadro 3 relaciona os parâmetros de *post* com os tipos estabelecidos. Por exemplo, mostra que o *post* do tipo REGDOCREC possui os parâmetros *session1*, *doc_number*, *dunsid=Other*.

Quadro 3 – Parâmetros presentes em cada tipo de *post*

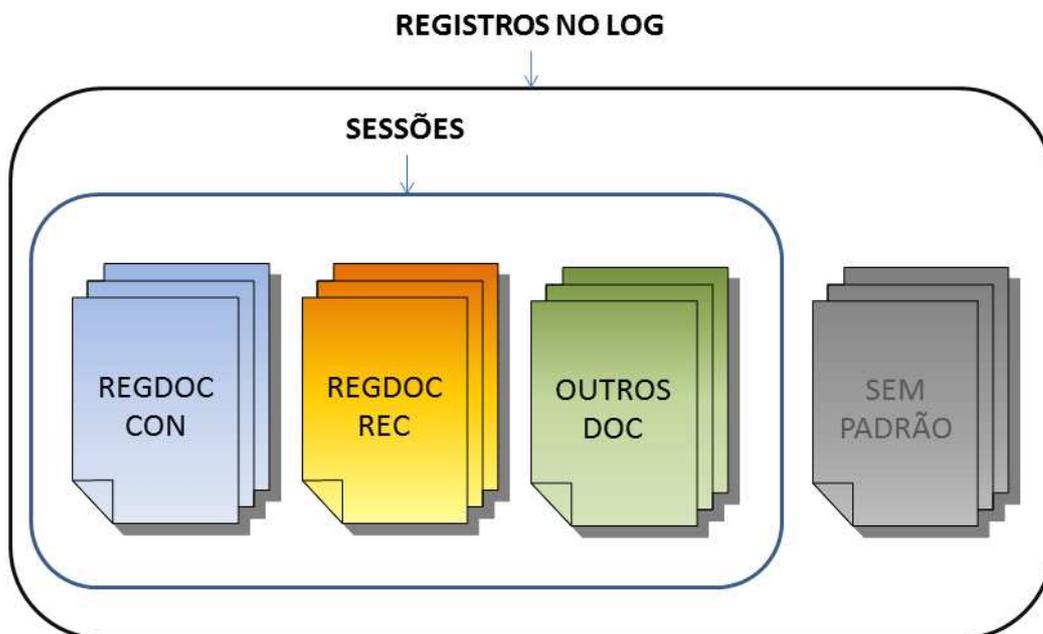
Parâmetro / Tipo	REGDOCCON	REGDOCREC	OUTROSDOC
Sessão (<i>session1</i>)	Possui	Possui	Possui
Registro de documento (<i>doc_number</i>)	Não possui	Possui	Possui
Ordenação do documento na consulta (<i>set_entry</i>)	Possui	Não possui	Não possui
Identificador da consulta (<i>set_number</i>)	Possui	Não possui	Não possui
Tipo de recomendação (<i>dunsid=Other</i>)	Não possui	Possui	Não possui

Fonte: a autora

Os tipos de registros presentes no *log* foram contabilizados e estas informações estão expressas na figura 6. O *log* fornecido é composto de 16.806 registros (*posts*). Dentro deste universo foram identificados *posts* que representam a visualização de registros de documentos. Destes, alguns são registros do tipo REGDOCCON, ou seja, originados a partir de uma consulta,

outros são do tipo REGDOCREC, isto é, originados a partir de uma recomendação. E há também *posts* do tipo OUTROS DOC, o que significa que representam a visualização de um registro de documento, porém sem registro de consulta ou recomendação, ou seja, pode ter sido clicado a partir de um link direto ou consulta a outra base que não o catálogo da biblioteca.

Figura 6 – Representação gráfica da estrutura de dados coletados a partir do *log* com diferentes tipos de registros



Fonte: a autora.

Este trabalho investiga a recomendação, que só ocorre em sessões. A sessão foi escolhida como unidade de medida, pois agrupa 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 à sessão.

b) Sessões

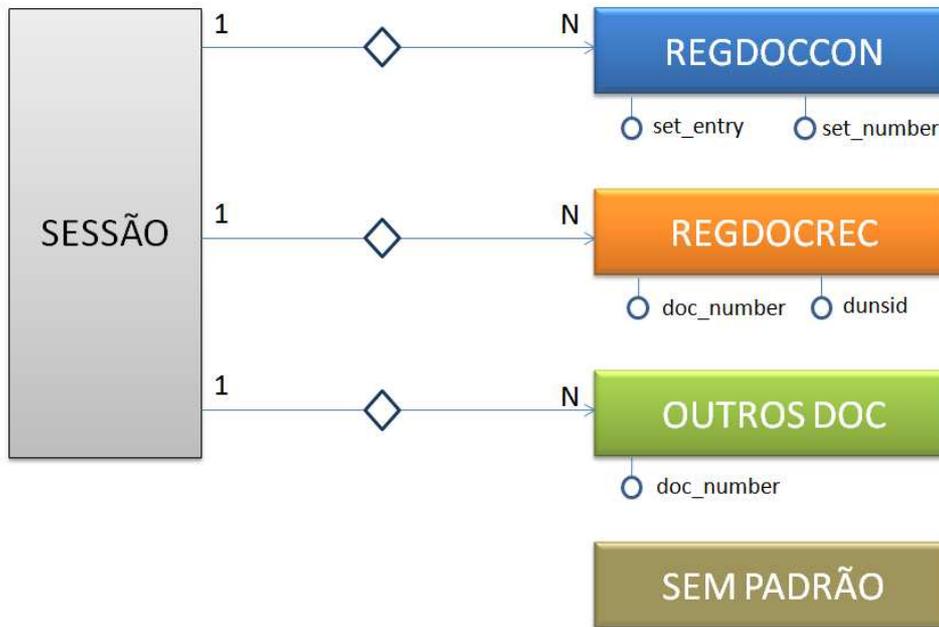
Uma sessão compreende várias interações de um usuário com sistema, como consulta e visualização de registro de catálogo a partir de consulta ou recomendação. Numa sessão podem ocorrer diversas e diferentes interações, e através dela é possível ligar consultas e a navegação decorrente destas. A sessão estabelece relação de tempo entre os posts (o que permite analisar e comparar períodos em que os usuários estão dedicados à busca de informação) e a ligação entre vários posts (consulta, recomendação, e outros), tornando possível agrupar várias interações do usuário com o sistema num determinado período de tempo. No *log*, a sessão é identificada no início de cada post⁴⁰. Cada vez que o navegador é fechado, o sistema encerra a sessão.

A figura 7 traz o Modelo Entidade-Relacionamento dos tipos de registros presentes no *log*.

⁴⁰ A seguir exibimos um exemplo de sessão onde cada linha representa um registro de interação do usuário com o catálogo (post). O número em negrito, que persiste em todas as linhas da sessão, é o que identifica que o acesso é feito por um mesmo usuário, de um mesmo local, num determinado intervalo de tempo:

```
[18/Dec/2012:02:22:06 +0000] "GET /F/5PAAB6D36DX1PIDIXEFENTFP2P6EBKAP21CE2MJS76HCAGQQ73-20903?func=full-set-set&set_number=796050&set_entry=000001&format=999 HTTP/1.1" 200 -
[18/Dec/2012:02:22:19 +0000] "GET /F/5PAAB6D36DX1PIDIXEFENTFP2P6EBKAP21CE2MJS76HCAGQQ73-19281?func=direct&doc_number=000670335&dunsid=OtherB1 HTTP/1.1" 200 -
[18/Dec/2012:02:25:25 +0000] "GET /F/5PAAB6D36DX1PIDIXEFENTFP2P6EBKAP21CE2MJS76HCAGQQ73-19408?func=direct&doc_number=000404536&dunsid=OtherH2 HTTP/1.1" 200 -
```

Figura 7 – Modelo ER dos dados coletados a partir do *log*



Fonte: a autora.

A figura 7 mostra que cada sessão pode conter zero ou mais registros do tipo REGDOCCON; e/ou zero ou mais registros do tipo REGDOCREC; e/ou zero ou mais registros do tipo OUTROS DOC. Isto significa que cada sessão pode conter, concomitantemente, registros dos três tipos (REGDOCCON, REGDOCREC e OUTROS DOC), mas cada *post* está associado a apenas uma sessão.

No *log* fornecido, não constam *posts* que contém as expressões de buscas efetuadas pelos usuários (consultas). Entretanto, informações referentes ao resultado das consultas estão presentes nos *posts* que representam registros de documentos visualizados a partir de consultas (REGDOCCON). Nestes *posts* são registrados o número identificador da consulta (*set_number*) e a posição na lista de resultados do registro de documento visualizado (*set_entry*). Nestes *posts* não são registrados o identificador do registro de documento visualizado (*doc_number*).

4 Análise dos dados

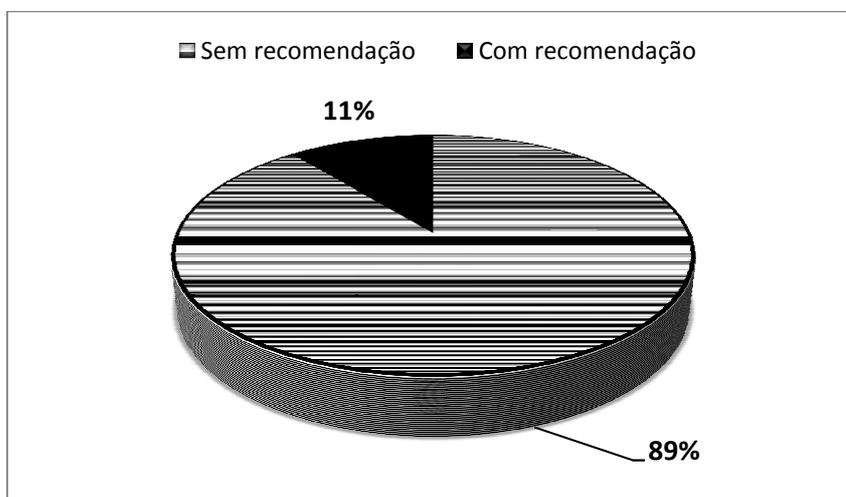
Este capítulo apresenta resultados de análises feitas sobre os dados coletados.

4.1 Análise das sessões

No *log* fornecido encontram-se 13.654 *posts* realizados em 6.347 sessões. Uma sessão pode conter diferentes tipos de registros, como registros de documentos visualizados a partir de consultas (REGDOCCON), registros de documentos visualizados a partir de recomendação (REGDOCREC) e registros de documentos visualizados a partir de outras fontes (OUTROS DOC). Isto está demonstrado no modelo ER (figura 7).

Para analisar o uso de recomendação em sessões, identificamos dois conjuntos de sessões: o conjunto “sessões com recomendação”, que contém sessões que possuem pelo menos um registro de uso de recomendação; e o conjunto “sessões sem recomendação”, que é composto por sessões que não contém nenhum registro de uso de recomendação.

Gráfico 1 – Sessões com recomendação e sem recomendação



Fonte: a autora.

Pode-se observar no gráfico 1 que as sessões com recomendação representam 11% do *log*, com 690 sessões. Já as 5.657 sessões sem recomendação representam 89% do *log*. Embora 11% pareça pouco, já justifica o uso de recomendação pela biblioteca. Especialmente quando se observa que as sessões com recomendação apresentam o uso intensivo do recurso (ver seção 4.1.3, referente à visualização de registros de documentos).

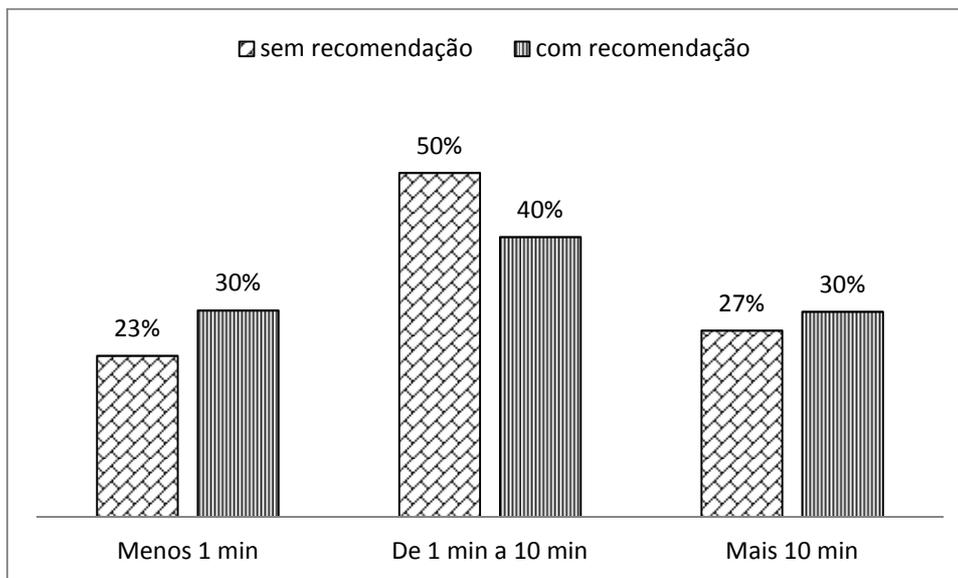
4.1.1 *Quanto ao tempo*

Para o cálculo de tempo, realizou-se uma limpeza que eliminou sessões com zero segundos ou mais de 24 horas, restando 2.522 sessões. Calculou-se então o intervalo de tempo entre o primeiro e o último acesso dentro da sessão e classificou-se a duração de cada sessão em:

- Sessões com menos de 1 minuto
- Sessões com duração de entre 1 e 10 minutos
- Sessões com duração de mais de 10 minutos

O gráfico 2 compara as sessões que tiveram alguma recomendação com aquelas que não tiveram nenhuma recomendação, com relação à sua duração.

Gráfico 2 – Sessões por intervalo de tempo

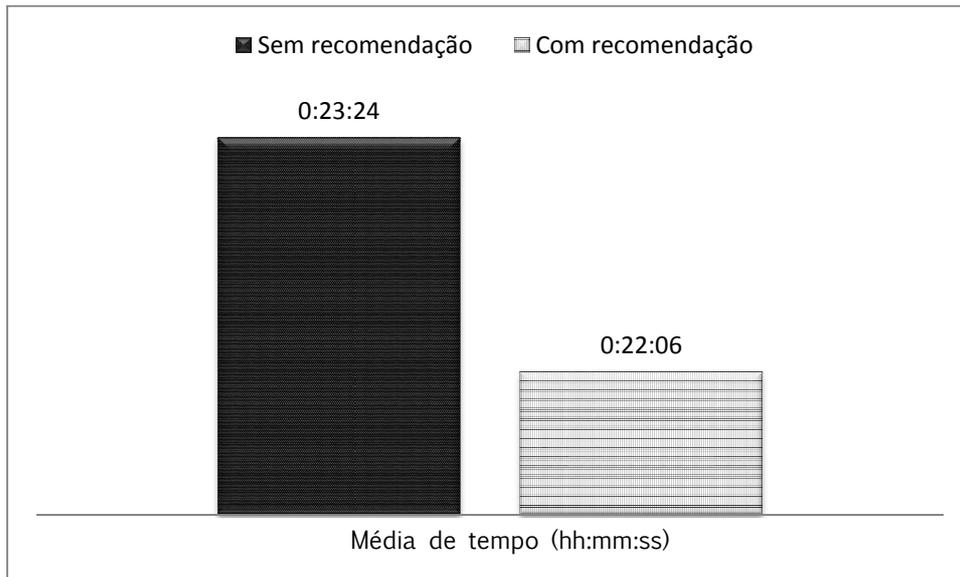


Fonte: a autora.

No gráfico 2 podemos observar que sessões apresentam uma duração predominante entre 1 e 10 minutos, independentemente se possui recomendação ou não (40,4% em seções sem recomendação e 49,7% em seções com recomendação). Em segundo lugar, predominam as sessões rápidas, com menos de 1 minuto de duração (29,8% em seções sem recomendação e 23,2% em seções com recomendação), e por fim, estão as sessões que duram mais de 10 minutos (29,6% em seções sem recomendação e 26,9% em seções com recomendação).

Já a média de tempo das sessões sem recomendações é 23 minutos e 24 segundos, enquanto as sessões com recomendação apresentam uma média de 22 minutos e 06 segundos. O gráfico 3 apresenta estes valores, a partir dos quais podemos afirmar que as sessões com registro de uso de recomendações da amostra são, em média, 1 minuto e 18 segundos mais rápidas do que as sessões sem registro de uso de recomendações. Essa diferença é pequena, isto é, o uso de recomendação não implica em uma redução substancial no tempo, quando comparado com sessões em que não houve recomendação.

Gráfico 3 – Média de tempo das sessões



Fonte: a autora.

Como já foi dito, para esta análise, as sessões com apenas um registro e as com mais de 24 horas de duração foram retiradas, para não distorcer os resultados. A título de curiosidade, antes da limpeza de dados, a maior sessão apresentou 345.071 segundos, que equivale a quase 95 horas e 51 minutos de duração. Esta ocorrência pode ter sido decorrente de uma seção aberta em terminal de consultas do catálogo on-line, que envolveu o uso do catálogo por vários usuários, sem que a aba do navegador fosse fechada.

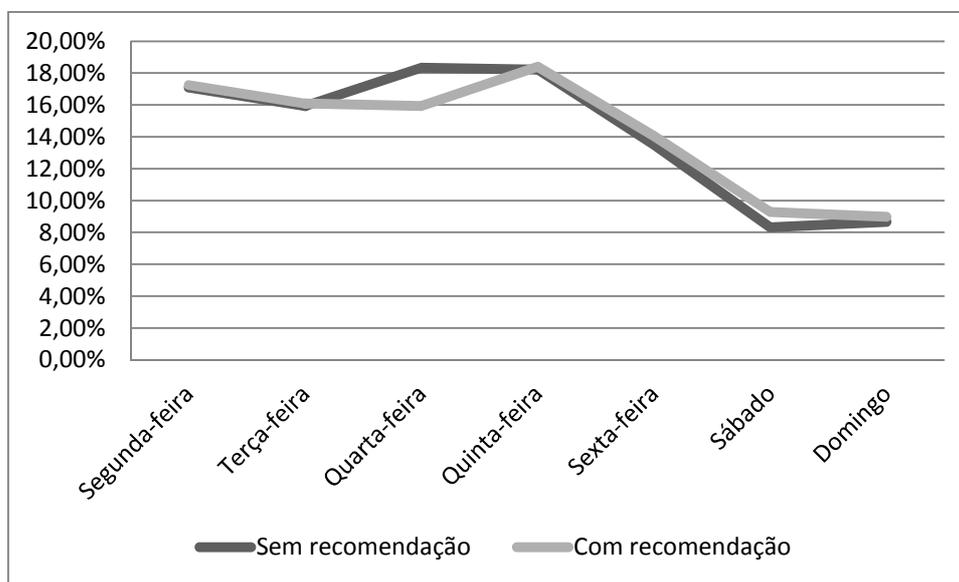
A partir das informações coletadas nesta seção do trabalho, pode-se entender que média de tempo das sessões é maior quando não se usa recomendação. Mas a diferença não é grande.

4.1.2 Quanto ao uso nos dias da semana

Com relação aos dias da semana em que ocorrem as sessões, este estudo identificou que quintas-feiras e quartas-feiras são os dias com maior número de acessos ao catálogo nas sessões sem recomendação, com 18%

das sessões cada um, seguidos da segunda-feira, com 17% e da terça-feira, com 16%. O sábado é o dia em que menos estão registradas sessões sem recomendações no log, com 8% das sessões realizadas neste dia, conforme gráfico 4. Nota-se que o período entre segunda-feira e quinta-feira possui um comportamento semelhante (entre 16% e 18%), caindo na sexta-feira, diminuindo significativamente no fim de semana.

Gráfico 4 – Sessões com e sem recomendação por dia da semana



Fonte: a autora.

Comportamento similar pode ser observado no gráfico 4, que se refere às sessões com recomendação. Nestas, a quinta-feira é também o dia em que ocorrem mais acessos ao catálogo, com 19% das sessões, seguido da segunda-feira, com 17%. No fim de semana ocorrem menos sessões com recomendações, com 9% tanto no sábado como no domingo.

Observa-se semelhança entre os acessos por dia da semana, envolvendo sessões com recomendação e sessões sem recomendação.

4.1.3 Quanto ao número de registros de documentos visualizados

Como já foi dito no início desta seção, foram contabilizados 13.654 *posts* em 6.347 sessões do nosso universo de estudo. Dividindo-se um valor pelo outro chegamos à média de 2,15 registros de documentos visualizados por sessão, conforme apresentado na tabela 1.

Nas 690 sessões que contém recomendação, foram encontrados 2.969 *posts* que representam a visualização de registros de documentos, totalizando uma média de 4,3 registros de documentos visualizados por sessão (tabela 1). Já considerando somente as sessões que não contém recomendação, temos um total de 5.657 sessões, e nestas sessões ocorrem 10.685 registros de visualização de documentos. Portanto, nas sessões que não contém recomendação, a média de documentos visualizados por sessão é 1,88.

Analisando a tabela 1, podemos observar que nas sessões com recomendação, embora representem 11% do total de sessões, ocorre a visita a uma quantidade maior de registros de documentos, 21,7% do total de visitas.

Tabela 1 - Registros de documentos visualizados em sessões com e sem recomendação

Sessões	Sessões	Sessões %	Documentos visitados	Documentos visitados %	Documentos 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: a autora.

Analisando somente as sessões com recomendação (tabela 2), dos 2.969 registros de documentos visualizados, 1.305 (43,9%) são provenientes

de recomendação. Esses registros correspondem a 9,5% do total de registros visualizados. Nas seções com recomendação, em média, são visitados 1,89 documentos via recomendação e 2,41 documentos não provenientes de recomendação, o que equivale a 12,8% do total de registros.

Tabela 2 – Registros de documentos visualizados nas sessões com recomendação

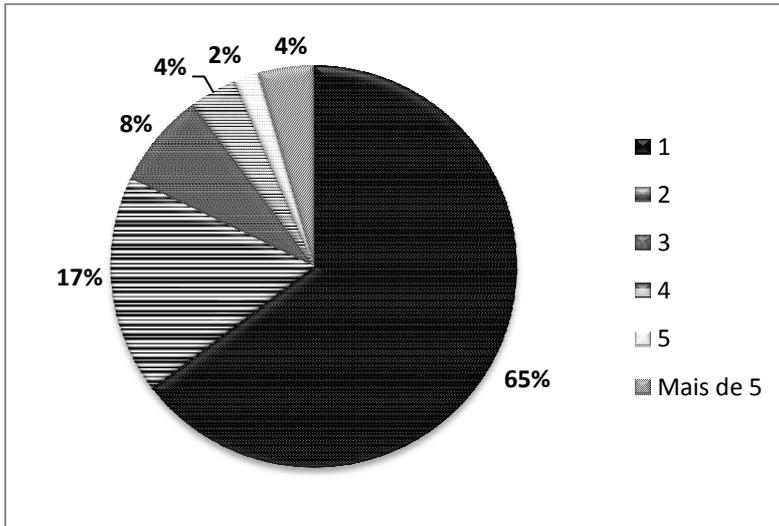
	Total	Percentual	Média de documentos por sessão	Percentual considerando o total de registros (13.654)
Registros de documentos visualizados via recomendação	1.305	43,9%	1,89	9,5%
Registros de documentos visualizados sem recomendação	1.664	56,0%	2,41	12,8%
Total	2.969	100%	4,3	21,74%

Fonte: a autora.

Na tabela 2, a coluna “Percentual” evidencia que nas sessões em que houve uso da recomendação, os usuários de fato usufruíram da recomendação em grande volume (43% dos registros de documentos visualizados são oriundos de recomendação - REGDOCREC). Embora do total de sessões uma parte pequena (11%) possui registro do uso de recomendação, nas sessões em que a recomendação é usada, ela é bastante utilizada (quase metade das visualizações). A seção 4.2 deste trabalho dedica-se a explorar mais as recomendações propriamente ditas.

Como podemos ver no gráfico 5, as sessões foram classificadas quanto ao número de registros visualizados na sessão.

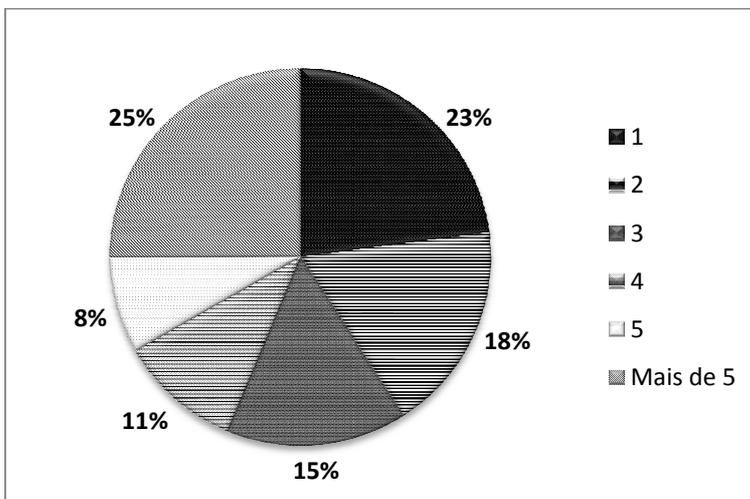
Gráfico 5 – Registros de documentos visualizados por sessão sem recomendação



Fonte: a autora.

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, conforme pode ser observado no gráfico 5. Já considerando somente as 690 sessões em que houve visita a documentos via recomendação, em 23% delas houve a visualização de somente um documento (gráfico 6).

Gráfico 6 – Registros de documentos visualizados por sessão com recomendação



Fonte: a autora.

Ao comparar sessões com recomendação com as sem recomendação, pode-se observar que 25% das sessões com recomendação concentra-se na visualização de mais 5 registros de documentos, ao contrário das sessões sem recomendação, onde a maioria das sessões (65%) apresenta visualização de apenas 1 registro de documento (gráfico 5). Disso, podemos observar um quadro diferente com relação aos dois tipos de sessão, onde as sessões com recomendação apresentam acessos menos pontuais e mais visitas a mais de um documento por sessão.

4.2 Recomendações

Como já mencionado, 690 sessões do *log* apresentam uso de recomendação, resultando na visualização de 1.305 registros de documentos que tiveram origem na recomendação. Esses registros são do tipo REGDOCREC.

As recomendações que levaram aos registros do tipo REGDOCREC, por sua vez, são classificadas em três tipos:

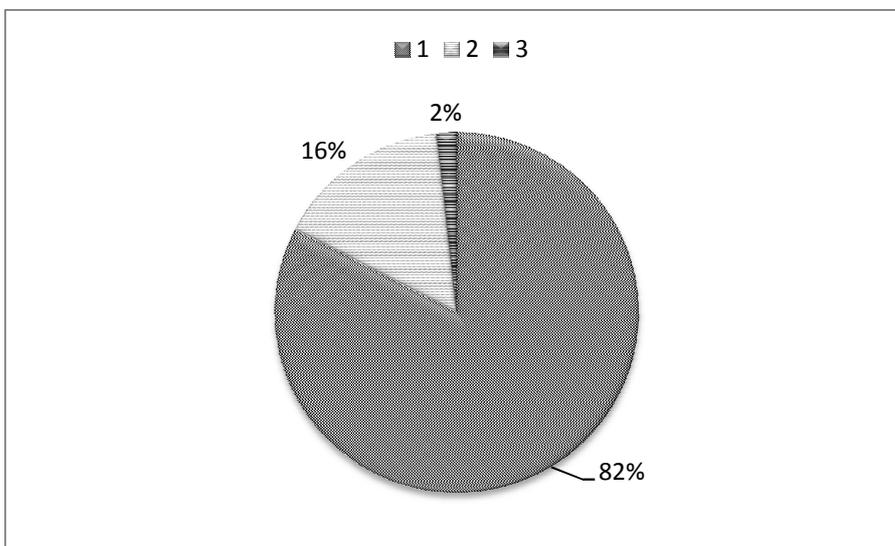
- “Pessoas que retiraram esta obra também retiraram” (similaridade usuário-usuário calculada com base na tabela de empréstimos);
- “Trabalhos semelhantes incluem” (similaridade item-item pela classificação de assunto atribuída pelos bibliotecários); e
- “Outras edições incluem” (similaridade item-item onde o que varia é a edição de um mesmo documento).

Nesta parte do trabalho serão explorados os diferentes tipos de recomendação.

4.2.1 *Uso de diferentes tipos de recomendação*

Dentre as 690 seções em que ocorre recomendação, em 570 sessões (82%) a visualização dos documentos por meio da recomendação ocorre através de um único tipo de recomendação, conforme demonstra o gráfico 7.

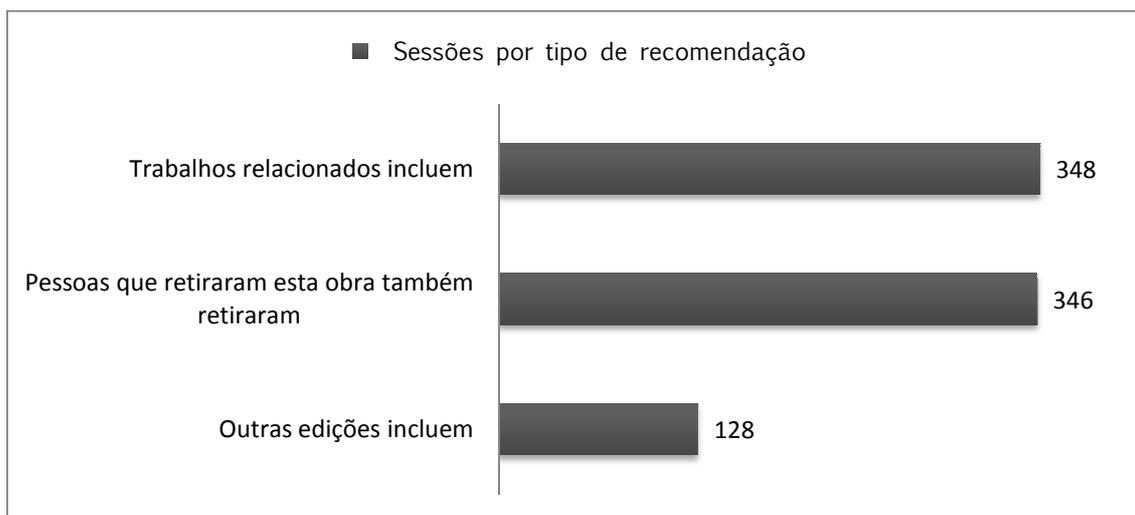
Gráfico 7 – Uso de mais de um tipo de recomendação na sessão



Fonte: a autora

O gráfico mostra que sessões onde o usuário utiliza mais de um tipo de recomendação representam 18% das sessões que contém recomendação. Em 2% (12 sessões) está registrado o uso dos 3 tipos de recomendação oferecidos pela ferramenta, e 108 sessões, ou seja, 16% das sessões com recomendação, registraram o uso de 2 tipos de recomendação diferentes.

O gráfico 8 apresenta os totais de sessões com recomendação 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).

Gráfico 8 – Sessões que contém recomendação por tipo de recomendação

Fonte: a autora.

No gráfico 8 verifica-se que o tipo de recomendação mais utilizado (348 sessões) é o tipo “Trabalhos relacionados incluem”, que recomenda documentos com assuntos semelhantes aos do registro visualizado.

A recomendação do tipo “Trabalhos relacionados incluem” sugere documentos com pelo menos três cabeçalhos de assunto ou números de classificação nos registros MARC em comum com o documento que está sendo visualizado. Este recurso permite que o usuário visualize documentos que podem não possuir o cabeçalho de assunto recuperado pela expressão de busca, mas que possui pelo menos três dos assuntos indexados no documento que deu origem à recomendação. Assim, a similaridade entre itens indica que é provável que este documento seja do interesse do usuário. Este recurso está ligado ao processo de recuperação da informação, do qual abordamos na seção 2.3.3 deste trabalho, que envolve, entre outros fatores, a representação da informação presente nos documentos. As palavras-chave ou cabeçalhos de assunto são de fundamental importância, e devem obedecer a um criterioso processo de representação dos assuntos do documento, a fim de evitar redundâncias na recomendação.

Este tipo também está relacionado com a filtragem baseada em conteúdo, técnica de recomendação mais usada para artigos e notícias que utiliza a similaridade entre itens para gerar recomendações (já vista na seção 2.6.3 alínea b deste trabalho).

Em segundo lugar, com quase o mesmo número de sessões (346), ocorre o tipo de recomendação “Pessoas que retiraram esta obra também retiraram”. Este tipo de recomendação utiliza a tabela de empréstimos para encontrar usuários similares (retiraram a mesma obra), e recomendar documentos que não necessariamente têm a mesma classificação, ou são variantes do mesmo documento. Neste caso, o que liga um documento ao outro são os usuários que os retiraram. Em 346 sessões este tipo de recomendação é usado. Com este dado, podemos afirmar que a recomendação ordenada através da co-ocorrência de cabeçalhos de assunto no registro de documentos (Trabalhos semelhantes incluem) é tão usada quanto a recomendação baseada em similaridade entre usuários (Pessoas que retiraram esta obra também retiraram).

O tipo de recomendação “Pessoas que retiraram esta obra também retiraram” está ligado ao conceito de serendipidade (da qual falamos na seção 2.7 deste trabalho), pois permite que o usuário encontre algo útil ou agradável pelo qual não estava procurando. A recomendação baseada nos empréstimos mostra que usuários que emprestaram o documento visualizado também emprestaram outros documentos, que não necessariamente têm o mesmo cabeçalho de assunto que retorna quando ele elabora uma expressão de busca. 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 também tem relação com a filtragem colaborativa, que é uma técnica de recomendação que utiliza a similaridade entre os usuários para gerar recomendações. Como leva em conta não apenas o interesse momentâneo do usuário (representado pela expressão de

busca), mas o seu perfil, é capaz de sugerir documentos que o usuário pode até não estar precisando no momento, mas que têm a ver com seu perfil.

Por fim, 128 sessões apresentam registros de uso de recomendação do tipo “Outras edições incluem”, que se refere à recomendação de outras edições do documento visualizado. Este tipo de recomendação, assim como o tipo “Trabalhos relacionados incluem”, utiliza a filtragem baseada em conteúdo para sugerir documentos aos usuários. Tem um aspecto administrativo no qual pode ajudar o usuário, que é a disponibilidade (ou indisponibilidade) de determinado item do acervo. Assim, quando o documento procurado não está disponível para empréstimo, o usuário rapidamente pode acessar o registro de outra edição do mesmo documento, o que, em alguns casos, pode atender a sua necessidade informacional. Livros muito populares e lançamentos tendem a criar uma longa fila de reservas quando a espera não é desejada. Em alguns casos, outra edição pode ser útil.

O fato de haver uma distribuição parelha entre os três tipos de recomendação (dois tipos com praticamente a mesma proporção de uso e outro chegando à metade dos dois maiores) indica que esses três tipos de recomendação são importantes e necessários para o sistema de recomendação de biblioteca.

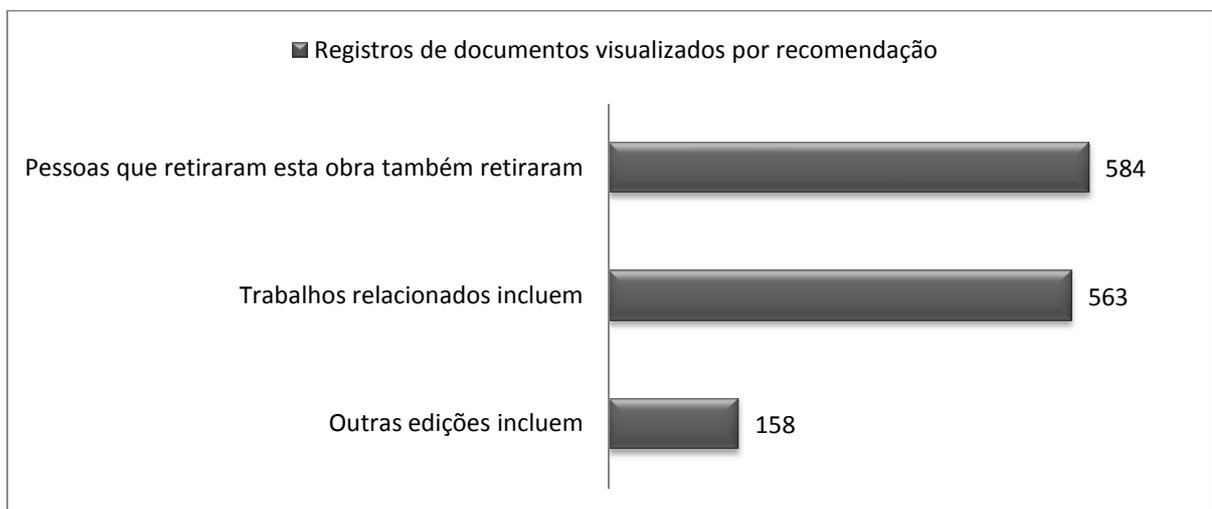
4.2.2 Quanto ao número de registros de documentos visualizados por recomendação

Nas sessões com recomendação, em média 1,89 registros de documentos visualizados são do tipo REGDOCREC. Se dividirmos este valor pelos tipos de recomendação encontramos esta média de registros de documentos visualizados por sessão:

- Pessoas que retiraram esta obra também retiraram = 0,85 registros de documentos por sessão
- Trabalhos relacionados incluem = 0,82 registros de documentos por sessão
- Outras edições incluem = 0,23 registros de documentos por sessão

Embora o número de sessões que contém recomendação do tipo “Trabalhos semelhantes incluem” supere o número de sessões que contém recomendação do tipo “Pessoas que retiraram esta obra também retiraram”, esta situação se inverte quando se trata de visualização de registros de documentos. No *log* fornecido, são 1.305 os registros de documentos visualizados a partir de recomendação. Destes, 584 (45%) foram oriundos da recomendação do tipo “Pessoas que retiraram esta obra também retiraram”, 563 (43%) foram oriundos da recomendação do tipo “Trabalhos relacionados incluem” e 158 (12%) oriundos da recomendação tipo “Outras edições incluem”.

Gráfico 9 – Número de registros de documentos visualizados por tipo de recomendação



Fonte: a autora.

Em números absolutos, o gráfico 9 indica que, no período de coleta, os usuários visualizaram mais registros de documentos do tipo “Pessoas que retiraram esta obra também retiraram” do que qualquer outro tipo de recomendação. Esta característica aponta uma preferência pela recomendação social, porque o tipo “Pessoas que retiraram esta obra também retiraram” é o tipo de recomendação inspirado sobre o conceito de similaridade entre usuários (aqueles que tenham retirado pelo menos 5 documentos em comum).

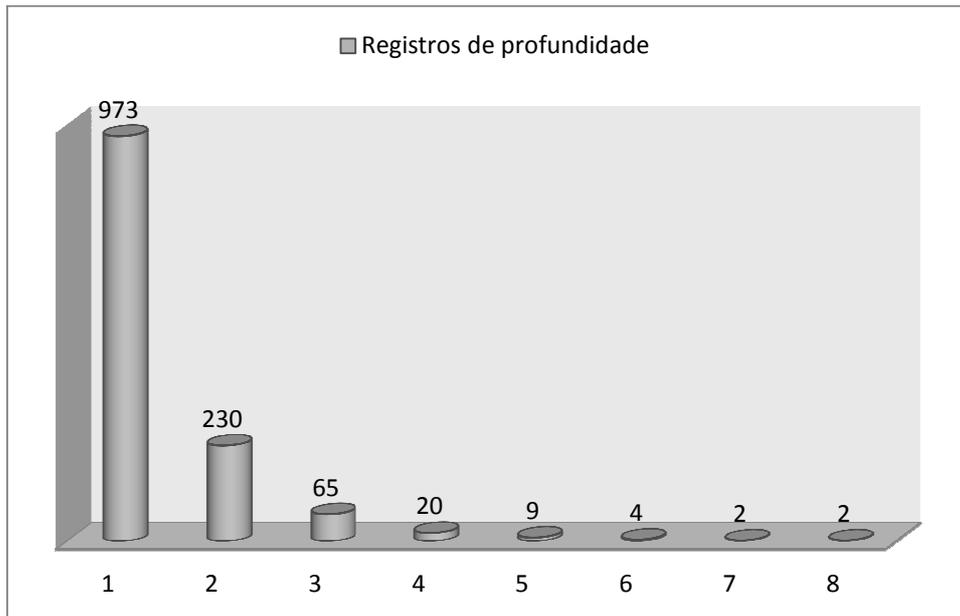
4.2.3 *Profundidade*

Outro aspecto que podemos observar do uso das recomendações é a profundidade. A ferramenta possui um recurso para “contar” o quanto o usuário avança no uso da recomendação. Assim, se uma sessão conta com 3 documentos visualizados por recomendação, sua profundidade é 3.

Porém, nesta ferramenta o incremento só ocorre se a visualização do registro recomendado se der, no navegador, na mesma aba do documento que o recomendou. Caso o usuário selecione a opção de abrir o registro de documento recomendado em outra aba, o incremento não ocorre e o registro volta a ser 1. Por esta razão, muitas sessões possuem vários registros com profundidade nível 1, apesar do usuário ter clicado em vários documentos e o uso da recomendação, de fato, ser mais “profunda” do que o registrado.

No gráfico 10, podemos visualizar o número de registros por nível de profundidade da recomendação.

Gráfico 10 – Profundidade por recomendação



Fonte: a autora.

Dos 1.305 registros de uso de recomendação, 973 (74,6%) têm profundidade nível 1, 230 (17,6%) têm profundidade 2 e 65 (5,0%) profundidade 3. Apenas 2 registros de profundidade 8 (o que equivale a 0,2% dos registros de uso de recomendação), mesma quantidade da profundidade 7.

A profundidade média de uso da recomendação registrada no *log* é de 1,38, e a profundidade máxima, 8.

4.3 Consultas

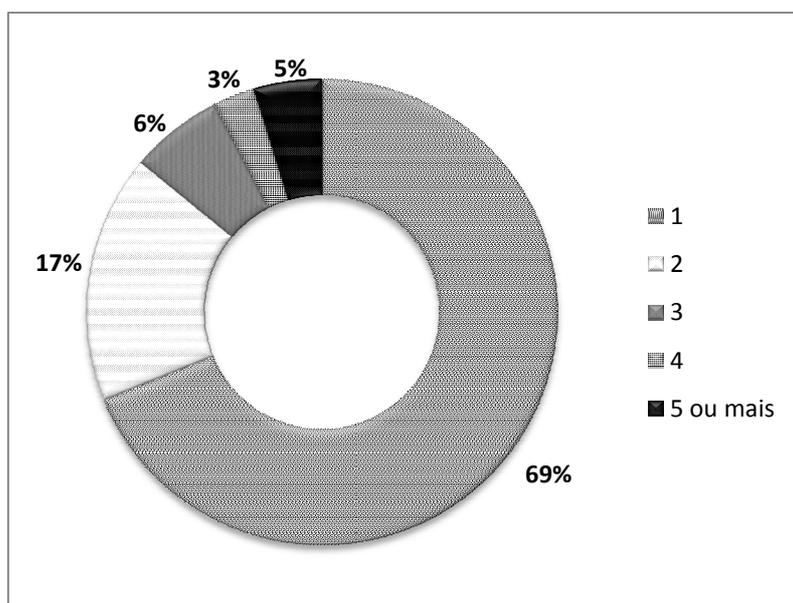
Algumas sessões contém *posts* que apresentam visualizações de registros de documentos que têm como origem resultados de consultas, ou seja, nestas sessões o usuário elaborou uma expressão de busca e a submeteu no campo de busca do catálogo, obtendo uma lista de resultados. É necessário lembrar que as sessões que não possuem recomendação não têm apenas registros oriundos de consultas (REGDOCCON), mas registros de

visualização de documentos oriundos de outras fontes, denominados OUTROS DOC. Também vale ressaltar que sessões que possuem consultas podem também possuir recomendação.

Foram contabilizadas 10.063 consultas em 5.979 sessões, numa média de 1,68 consultas por sessão. A sessão com o maior número de registros gerados a partir de consultas (tipo REGDOCCON) apresenta 44 consultas.

O gráfico 11 mostra quantas consultas foram realizadas por sessão, considerando somente aquelas sessões em que ocorrem consultas.

Gráfico 11 – Consultas por sessão



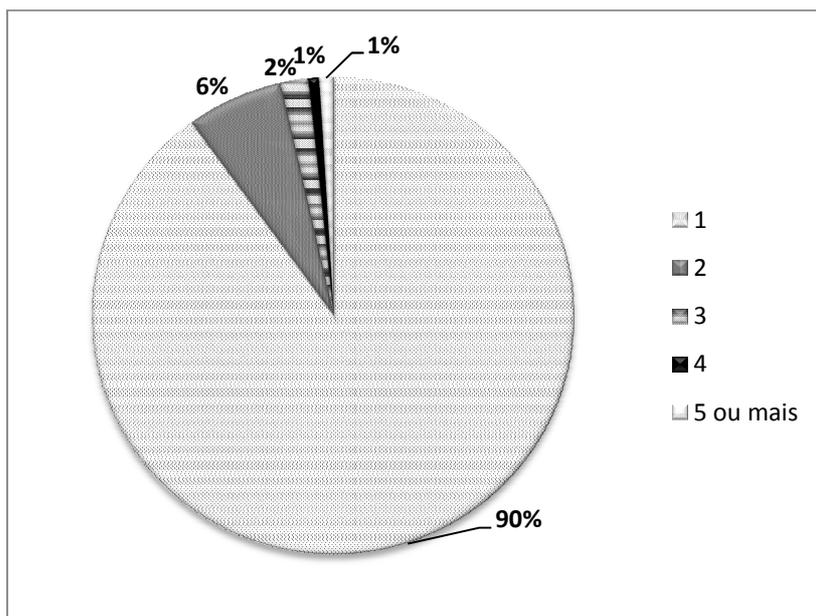
Fonte: a autora.

Como se pode observar no gráfico 11, sessões com uma única consulta são maioria, com 69% das sessões do *log*. Sessões com 2 consultas representam 17% das sessões que contém consultas, enquanto a classe de 5 ou mais consultas agrupa 5% das sessões com consultas.

A partir da consulta, é exibida uma lista de resultados ao usuário, que clica em um ou mais dos resultados e chega a um registro de documento, que é a ficha do catálogo que exhibe todas as informações relativas ao

documento. O gráfico 12 apresenta o número de registros de documentos clicados pelo usuário na lista de resultados de uma consulta, separados por classes.

Gráfico 12 – Registros de documentos visualizados por consulta



Fonte: a autora

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 do *log*). Apenas 1% das consultas resulta em visualização de 5 ou mais documentos.

Foram visualizados 11.959 registros a partir de consulta (REGDOCCON), e 10.063 consultas, o que resulta em uma média de 1,19 visualizações de registros de documentos por consulta.

Já a média de registros de documentos visualizados por sessão é de 2,00 nas sessões com consulta e 4,60 nas sessões sem consulta (tabela 3). A visita a 2 registros de documentos, em média, nas sessões com consulta, pode indicar que o usuário encontra rapidamente o que ele necessita. O aproveitamento dos documentos encontrados, no entanto, só poderia ser

medido com um estudo de usuários, pois como foi exposto na seção 2.3.3 deste trabalho, a relevância varia de usuário para usuário.

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

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: a autora.

Assim tem-se que, apesar das sessões sem consulta representarem 5,8% do total de sessões, nelas 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 12, pois apresenta, nas sessões com consulta, 90% de sessões com visita a apenas um documento. Assim pode-se concluir que as sessões onde ocorre mais navegação (*browsing*) são as sessões sem consulta.

4.4 Recomendação versus Consultas

Apesar de a pesquisa lidar com dados relativos a apenas 30 dias e não possuir informações mais detalhadas da navegação que seriam desejáveis, uma análise de recomendação *versus* consultas pode ser ensaiada, tentando-se relacionar as duas formas de pesquisa.

a) Sessões com recomendação

Entre todas as sessões registradas, 11% possuem recomendação, e contém 21% dos registros de documentos visualizados, sendo que destes documentos, 43,9% tiveram origem em uma recomendação. Estas sessões

apresentam, em média, 4,30 registros de documentos visualizados (incluindo registros oriundos de recomendados e não oriundos de recomendados).

b) Sessões com consulta

Entre todas as sessões registradas, 94% apresentam consulta, e nestas sessões estão 87,6% dos registros de documentos visitados. A média de registros de documentos visitados é de 2,00 por sessão nestes casos.

Um alto grau de registros de documentos visitados por sessões com recomendação (4,30) demonstra a ocorrência de serendipidade. Os usuários navegam em mais registros de documentos similares àqueles inicialmente acessados, e, “permanecem na recomendação” em 43,9% dos registros visitados.

A média baixa de registros de documentos visitados por sessões com consulta (2,00) demonstra que os resultados são, em média, assertivos para os que buscam informação.

5 Considerações finais

Com a disponibilidade cada vez maior de informação, em diversos formatos e de forma não estruturada, os sistemas de recomendação têm sido usados com crescente frequência e evoluído de forma gradual à medida que os usuários exigem maior precisão e personalização em seus resultados de busca. A ferramenta *Related Books in Aleph OPAC* se apresenta como uma opção viável para implantação dos conceitos de recomendação nas bibliotecas universitárias que já utilizam o sistema Aleph.

As características dos registros presentes no *log* fornecido e o método adotado (análise de *log*) permitiram analisar o uso de recomendação e consultas nas bibliotecas da *University of Dundee*.

Verificou-se que a recomendação foi utilizada em 11% das sessões registradas. Nestas sessões, em média, foram visitados 1,89 registros de documentos via recomendação (REGDOCREC), o que representa 43,9% dos posts nestas sessões. Já os registros de documentos não provenientes de recomendação (REGDOCCON e OUTROS DOC) representam 56,0% dos posts nas sessões com recomendação, com uma média de 2,41 por sessão.

Embora apenas 11% das sessões apresentem o uso de recomendação, 43,9% dos registros de documentos visualizados nestas sessões são oriundas de recomendação, o que demonstra um alto uso de recomendação nestas sessões. Este percentual justifica o uso de recomendação e traz à tona um questionamento que poderia ser respondido em um estudo futuro: o baixo índice de sessões com recomendação está relacionado com baixa satisfação ou utilidade percebida pelos usuários ou é apenas uma questão de divulgação do serviço e capacitação para seu uso? A alta aderência dos usuários ao recurso nas sessões sugere que aqueles que utilizam o serviço veem valor, pois o exploram bastante.

O estudo comprovou que as sessões com uso de recomendações foram (no período de amostra) em média 1 minuto e 18 segundos mais rápidas do que as sessões sem uso de recomendações.

A média geral de registros de documentos visualizados por sessão é 2,15. Nas sessões sem recomendação, 1,88 registros de documentos são visualizados por sessão. A média sobe 4,30 nas sessões com recomendação (incluindo registros oriundos de recomendados e não oriundos de recomendados). Então, embora apenas 11% das sessões contenham recomendação, nelas ocorre a visita a uma quantidade maior de registros de documentos (21,7% do total de visualizações de registro de documentos).

Este dado, aliado a informação de que as sessões com recomendação são em média mais curtas, evidencia maior eficácia nas sessões que apresentam uso da recomendação, pois foi possível visualizar, em média, mais registros de documentos em menos tempo do que em sessões sem o uso da recomendação.

Também se mostrou que a maioria (25%) das sessões com recomendação apresentaram a visualização de mais de 5 documentos por sessão, enquanto a maioria (65%) das sessões sem recomendação apresenta visualização de apenas 1 registro de documento. Daí conclui-se que as sessões sem recomendação têm acessos mais pontuais (visitas a menos documentos por sessão) do que as sessões com recomendação.

Registrou-se que 18% dos usuários de recomendação exploram mais de um tipo de recomendação, e 82% preferem utilizar apenas um tipo de recomendação.

A preferência por recomendações do tipo “Trabalhos relacionados incluem” se mostra na quantidade de sessões que registram seu uso (348).

No entanto, a recomendação do tipo “Pessoas que retiraram esta obra também retiraram” quase empata na quantidade de sessões (346) e ainda supera em número de registros de documentos visualizados (584) o tipo “Trabalhos relacionados incluem” (563). A recomendação do tipo “Pessoas que retiraram esta obra também retiraram” é um tipo de filtragem colaborativa e inspirado na serendipidade, enquanto recomendação do tipo “Trabalhos relacionados incluem” apresenta-se como uma filtragem baseada em conteúdo, ancorada na classificação realizada pelos bibliotecários. Analisando o número de acessos podemos afirmar que ambos os tipos de recomendação são bem aceitos pelos usuários.

O baixo número de sessões que registram o uso da recomendação do tipo “Outras edições incluem” (128) pode indicar duas coisas: que a biblioteca possui poucos títulos repetidos com edições diferentes ou que os usuários preferem outros tipos de recomendação.

Em relação ao dia da semana, identificou-se a quinta-feira como dia de maior uso das recomendações.

A profundidade média de uso da recomendação registrada no log é de 1,38, e a profundidade máxima, 8.

Em relação às consultas, registrou-se uma média de 1,68 consultas por sessão. A grande maioria das consultas (90%) teve como resultado a visualização de apenas um documento, e a média de registros de documentos visualizados por consulta chegou em 1,19.

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.

Com base na análise realizada, esboçam-se algumas sugestões para a ferramenta *Related Books in Aleph OPAC*. Com relação ao cálculo de profundidade, seria interessante que o sistema incrementasse +1 à profundidade da recomendação independentemente se o usuário abriu o registro em outra aba ou não. Atualmente, se um usuário clica em um link e abre em outra aba a contagem não recebe incremento, e o grande volume de registros com profundidade 1 pode estar distorcendo os dados por esta característica.

Documentos on-line (como *e-books*) não entram nos cálculos que gerenciam a similaridade do sistema: a tabela de empréstimos só considera materiais físicos retirados, mas não considera documentos baixados pelo computador, e que de certa forma, também foi “emprestado” ao leitor. Esta informação pode fazer falta, levando-se em conta que os acessos a documentos digitais vêm crescendo mais e mais a cada dia.

Nas bibliotecas universitárias poderia ser útil uma funcionalidade que recomendasse com base no curso ou nível de escolaridade dos usuários, (facilmente obtido pelo vínculo no cadastro da universidade). Neste caso, um empecilho seria o anonimato (preferido pela grande maioria dos usuários), mas no caso de computadores que já exigem o login do usuário para acesso a outros sistemas, esta informação poderia ser aproveitada.

Outra sugestão é que, com a informação de qual funcionalidade faz maior sucesso entre os usuários, esta receba maior destaque nos sites das bibliotecas. A capacitação dos usuários para aumentar a adesão também pode ser útil e ampliar sua satisfação, já que a recomendação é tão amplamente utilizada em outros recursos com os quais muitos dos usuários já estão acostumados (como e-mail, máquinas de busca, *e-commerce*, entre outros).

Como este trabalho consiste em uma análise preliminar, exige evolução em trabalhos futuros. Sugere-se que haja um refinamento do estudo de recomendações (com uma amostra maior, por exemplo) para uma experiência mais esclarecedora para o público das bibliotecas universitárias.

Apesar de a obtenção de um *log* para análise não ser uma tarefa das mais simples, por envolver uma série de questões técnicas e até políticas, acredita-se que uma comparação de uso desta ferramenta entre as universidades que também a disponibilizam no catálogo pode ser interessante no sentido de identificar tendências e disparidades entre os usuários. Em etapas iniciais deste trabalho identificou-se o uso do recurso em pelo menos três países.

Referências bibliográficas

ABEL, Mara. Apresentação. In: TORRES, Roberto. **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. p. 13-14.

AMAZON compra site de recomendação de livros Goodreads. **Globo.com**, S.l., 28 mar. 2013. Disponível em < <http://m.g1.globo.com/economia/noticia/2013/03/amazon-compra-site-de-recomendacao-de-livros-goodreads.html> >. Acesso em 06 abr. 2013.

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. **Modern Information Retrieval**. New York: ACM Press, 1999. 501 p.

MARQUESUZAÀ, Christophe, et al. Accessing heritage documents according to space criteria within digital libraries. **Journal of Digital Information Management**, v. 6, n. 1, p. 102-117, feb. 2008. Disponível em: < http://hal.inria.fr/docs/00/35/30/89/PDF/Marquesuzaa_alJDIM.pdf >. Acesso em: 17 nov. 2012.

BARCELLOS, Carla Duarte; BRANDÃO, André Luiz; MUSA, Daniela Leal. Sistema de Recomendação Acadêmico para Apoio a Aprendizagem. **CINTED-UFRGS Novas Tecnologias na Educação**. Porto Alegre, v.5, n.2, 10p, dez. 2007. Disponível em < <http://www.cinted.ufrgs.br/ciclo10/artigos/3fDaniela.pdf> >. Acesso em 9 mai. 2012.

BAX, M. P.; ALVARENGA, L.; PARREIRAS, F. S.; BRANDÃO, W. C. Sistema automático de disseminação seletiva. In: IFLA M&M, 2004, São Paulo, **Anais...**, São Paulo: USP. 2004. Disponível em: < http://www.fernando.parreiras.nom.br/publicacoes/dsi_ifla.pdf >. Acesso em: 09 dez. 2012.

BOAVENTURA, Edivaldo M. **Metodologia da Pesquisa** : monografia, dissertação, tese. São Paulo: Atlas, 2009. 160 p.

CALENGE, Bertrand. **Accueillir, orienter, informer** : l'organisation des services aux publiques dans les bibliothèques. Paris: Éditions du Cercle de la Librairie, 1996. 129 p.

CHANDRA, Nishant. The Evolution of Recommendation Systems. In: **DZone** : big data / BI zone. 19 nov. 2012. Disponível em <

<http://architects.dzone.com/articles/evolution-recommendation> >. Acesso em: 05 dez. 2012.

ELSEVIER PROPERTIES S.A. **SciVerse Science Direct** : how related articles are determined. [S.l], 2012. Disponível em <
http://help.sciencedirect.com/flare/sdhelp_Left.htm#CSHID=related_art_lrn_more.htm|StartTopic=Content%2Frelated_art_lrn_more.htm|SkinName=svs_SD >. Acesso em 04 dez. 2012.

FIGUEIREDO, Nice Menezes de. A modernidade das cinco leis de Ranganathan. **Ciência da Informação**, Brasília, v. 21 n. 3, p. 186-191, set./dez. 1992. Disponível em: <
<http://revista.ibict.br/cienciadainformacao/index.php/ciinf/article/view/1277/911> >. Acesso em: 10 dez. 2012.

FISK, Raymond P. GROVE, Stephen J. JOHN, Joby. **Interactive Services Marketing**. Boston: Houghton Mifflin Company, 2008. 276 p.

GIL, Antonio Carlos. **Como Elaborar Projetos de Pesquisa**. São Paulo: Atlas, 2009. 184 p.

JANES, Joseph; CARTER, David; MEMMOTT, Patricia. Digital Reference Services in Academic Libraries. **Reference & User Services Quarterly**, v. 39, n. 2, p. 145-150, winter 1999. Disponível em: <
http://go.galegroup.com/ps/retrieve.do?retrieveFormat=PDF_FROM_CALLISTO&inPS=true&prodId=AONE&userGroupName=capes58&workId=PI-0HDS-1999-WIN00-IDS1-24.JPG%7CPI-0HDS-1999-WIN00-IDS1-25.JPG%7CPI-0HDS-1999-WIN00-IDS1-26.JPG%7CPI-0HDS-1999-WIN00-IDS1-27.JPG%7CPI-0HDS-1999-WIN00-IDS1-28.JPG%7CPI-0HDS-1999-WIN00-IDS1-29.JPG&docId=GALE%7CA61755295&callistoContentSet=PER >. Acesso em 03 nov. 2012.

JANSEN, Bernard 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 03 maio 2013.

LICHTNOW, Daniel; LOH, Stanley; KAMPFF, Adriana Justin Cerveira; PRIMO, Tiago; GARIN, Rodrigo Saldaña; OLIVEIRA, José Palazzo Moreira de; LIMA, José Valdeni de; PALAZZO, Luiz A. Moro. 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 09 mai. 2012.

LOPES, Giseli Rabello. **Sistema de Recomendação para Bibliotecas Digitais sob a Perspectiva da Web Semântica**. 2007. 69p. Dissertação (Mestrado em Ciência da Computação)–Instituto de Informática, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2007.

MANGAS, Sergio Filipe Agostinho. Como planificar e gerir um serviço de referencia. **Biblios**, n. 28, 31 p., abr./jun. 2007. Disponível em: < <http://sisbib.unmsm.edu.pe/bibvirtualdata/publicaciones/biblios/n28/a02n28.pdf> >. Acesso em: 10 dez. 2012.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Metodologia do Trabalho Científico** : procedimentos básicos, pesquisa bibliográfica, projeto e relatório, publicações e trabalhos científicos. São Paulo: Atlas, 2001. 219 p.
REATEGUI, Eliseo Berni; CAZELLA, Sílvio César. Sistemas de Recomendação. In: Congresso Da Sociedade Brasileira De Computação, XXV, 2005, São Leopoldo. **Anais do Encontro Nacional de Inteligência Artificial**. São Leopoldo: Unisinos, 2005, p. 306-348. Disponível em: < citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.92.2811&rep=rep1&type=pdf >. Acesso em: 9 mai. 2012.

RESNICK, P. Varian, H. R. **Recommender Systems**. Communications of the ACM, New York, v.40, n.3, pp. 55-58, Mar. 1997.

RICHARDSON, Roberto Jarry. **Pesquisa social** : métodos e técnicas. São Paulo: Atlas, 1985. 287 p.

ROCHA, Catarina Carneiro. **RECDOC: um sistema de recomendação para uma biblioteca digital na web**. 2003. 110 f. Dissertação (Mestrado em Engenharias de Sistemas e Computação)-Programa de Pós-Graduação de Engenharia, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2003. Disponível em < wiki.xexeo.org/tiki-download_file.php?fileId=5 >. Acesso em 9 mai. 2012.

SANTOS, Jussara Pereira. **Reflexões sobre currículo e legislação na área da biblioteconomia**. Porto Alegre: ABEED, 1998.

SLOAN, Bernie. Service Perspectives for the Digital Library Remote Reference Services. **Library Trends**, [Baltimore], v. 47, n. 1, p. 117-[144], summer 1998. Disponível em: < <http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=1401584&lang=pt-br&site=ehost-live> >. Acesso em 04 nov. 2012.

SPEROTTO, Fabio Aiub. **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/000061/00061D6.pdf> > Acesso em 9 mai. 2012.

TALAVERA IBARRA, Ana María. El bibliotecólogo electrónico : consideraciones en torno al servicio de referencia en la era digital. **Biblios: revista electrónica de bibliotecología, archivología y museología**, Peru, n. 3, 17 p., jan./mar. 2000. Disponível em: < <http://dialnet.unirioja.es/servlet/articulo?codigo=283510> >. Acesso em 10 dez. 2012.

TORRES, J.M.; Parkes, A. User modelling and adaptivity in visual information retrieval systems. In: Workshop on Computational Semiotics for New Media, 2000, Guilford, (UK). **Final program**, University of Surrey, jun. 2000, Disponível em: < <http://www.scm.tees.ac.uk/users/p.c.fencott/newMedia/torresParkes.html> > Acesso em: 05 dez. 2012.

TORRES, Roberto. **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.