

UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO (UNIRIO)
CENTRO DE CIÊNCIAS HUMANAS E SOCIAIS (CCHS)
ESCOLA DE BIBLIOTECONOMIA (EB)

DÉBORA MILENA NIEDZEILSKI FERREIRA

**PERSPECTIVAS DA FILTRAGEM COLABORATIVA PARA A INDEXAÇÃO E SUA
RELAÇÃO COM O ESTUDO DE USUÁRIOS EM BIBLIOTECAS
ESPECIALIZADAS**

Rio de Janeiro
2018

DÉBORA MILENA NIEDZEILSKI FERREIRA

**PERSPECTIVAS DA FILTRAGEM COLABORATIVA PARA A INDEXAÇÃO E SUA
RELAÇÃO COM O ESTUDO DE USUÁRIOS EM BIBLIOTECAS
ESPECIALIZADAS**

Trabalho de Conclusão de Curso apresentado como requisito parcial para a obtenção do título de Bacharel em Biblioteconomia pela Escola de Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro.

Orientadora: Prof^ª. Dr^ª. Ludmila dos Santos Guimarães.

Rio de Janeiro

2018

Catálogo informatizada pelo(a) autor(a)

F383 Ferreira, Débora Milena Niedzeilski
Perspectivas da Filtragem Colaborativa para a
Indexação e sua Relação com o Estudo de Usuários em
Bibliotecas Especializadas / Débora Milena
Niedzeilski Ferreira. -- Rio de Janeiro, 2018.
71 f.

Orientadora: Ludmila dos Santos Guimarães.
Trabalho de Conclusão de Curso (Graduação) -
Universidade Federal do Estado do Rio de Janeiro,
Graduação em Biblioteconomia, 2018.

1. Sistemas de Recomendação. 2. Filtragem
Colaborativa. 3. Indexação. 4. Estudo de Usuários.
5. Bibliotecas Especializadas. I. Guimarães,
Ludmila dos Santos, orient. II. Título.

DÉBORA MILENA NIEDZEILSKI FERREIRA

**PERSPECTIVAS DA FILTRAGEM COLABORATIVA PARA A INDEXAÇÃO E SUA
RELAÇÃO COM O ESTUDO DE USUÁRIOS EM BIBLIOTECAS
ESPECIALIZADAS**

Trabalho de Conclusão de Curso apresentado como requisito parcial para a obtenção do título de Bacharel em Biblioteconomia pela Escola de Biblioteconomia da Universidade Federal do Estado do Rio de Janeiro.

Orientadora: Prof^ª. Dr^ª. Ludmila dos Santos Guimarães.

Rio de Janeiro, _____ de _____ 2018.

BANCA EXAMINADORA

Prof^ª. Dr^ª. Ludmila dos Santos Guimarães (Orientadora)
Universidade Federal do Estado do Rio de Janeiro – UNIRIO

Prof.^a Dr^a. Dayanne da Silva Prudêncio (Avaliadora)
Universidade Federal do Estado do Rio de Janeiro – UNIRIO

Prof. Dr. Carlos Alberto Ferreira (Avaliador)
Universidade Federal do Estado do Rio de Janeiro – UNIRIO

AGRADECIMENTOS

Agradeço aos meus pais e ao meu irmão, meus maiores incentivadores, pelos conselhos e por me manterem firme sempre. Minha vida pela de vocês! Agradeço também a todos os meus familiares.

Agradeço ao meu gafanhoto, melhor [e único] sobrinho, Guilherme, por ter sido a pessoa que mais me acalmou em momentos difíceis da faculdade. Às vezes é necessário abstrair; e nós fazíamos isso assistindo anime juntos. Tia te ama, criancinha!

Agradeço à minha orientadora, Prof.^a Dr.^a Ludmila dos Santos Guimarães, por todo incentivo ao longo da iniciação científica, monitoria e desenvolvimento deste trabalho; por todos os ensinamentos, puxões de orelha e palavras amigas. Obrigada por tudo!

Agradeço ao corpo docente da Unirio, em especial à professora Naira Christofolletti Silveira pela orientação na iniciação científica em Representação Descritiva, e por ser sempre muito gentil e solícita, à professora Dayanne da Silva Prudêncio pelos conselhos e incentivos durante a disciplina de Biblioteconomia Digital, e ao professor Carlos Alberto (Café) pelos ensinamentos preciosos durante as aulas de A.I., onde pude finalmente encontrar meu caminho dentro da Biblioteconomia. São professores como vocês que nos mantêm motivados durante a graduação.

Agradeço à comunidade R-Ladies Rio, em especial à Celina Rebello pelas conversas, conselhos e pela troca de experiência. Encontrá-las e fazer parte desta comunidade incrível é muito gratificante! Viva o R!!!!

Agradeço aos bibliotecários supervisores, Ten. Ana e Ten. Carlos, ao gestor da biblioteca, Cel. Velôzo, à Sgt. Naiara e aos parceiros de trabalho, Magnólia Félix, Ivanildo Conceição e Giovana Saboia, durante o estágio na Escola de Comando e Estado-Maior do Exército (ECEME). Serei eternamente grata pelas experiências e pessoas que conheci nesse local.

Agradeço aos amigos da faculdade: Tessali Nabekura, Fernanda Mattos, Priscila Faria, Fabrício Magno, Natália Araújo, Flávio Pereira, Aldair Diniz,

especialmente Maria Rosangela, Eliane Maceió, Heloísa Proença e Ana Karolina.
Vocês foram essenciais no decorrer desses anos!

Agradeço aos meus amigos da vida, em especial à Cinara Lima, Emily Mutti,
Lorrayne Luz e Leonardo Andrade pelo apoio de sempre.

“Carpe diem. Seize the day, boys. Make your lives extraordinary.”
(John Keating – Dead Poets Society)

RESUMO

Os Sistemas de Recomendação surgem em meados da década de 90 como uma área de pesquisa independente da Ciência da Computação, atuando como uma alternativa diante da sobrecarga de informação na *web*. Dentre suas abordagens, tem-se a Filtragem Colaborativa, que atua como automatização do processo natural de recomendação *boca a boca*, utilizando das preferências dos usuários para a geração de recomendações a outros usuários de perfil semelhante. Têm seu desenvolvimento sobretudo no âmbito do *e-commerce* e *e-business*, no entanto, atualmente são também aplicados em catálogos de acesso público online (OPACs) de bibliotecas, sejam elas físicas ou digitais, a nível internacional. Embora direcionados a questão comercial e estudados de forma significativa na área da computação, os sistemas de recomendação, de modo geral, têm forte relação com a Biblioteconomia. Diante disso, esse trabalho, por meio de uma pesquisa exploratória, busca analisar de que forma a Filtragem Colaborativa, aplicada em bibliotecas especializadas, pode auxiliar na indexação e no estudo de usuários.

Palavras-chave: Sistemas de Recomendação. Filtragem Colaborativa. Estudo de Usuários. Indexação. Bibliotecas Especializadas.

ABSTRACT

The Recommendation Systems appeared in the mid-1990s as an independent research area of Computer Science, acting as an alternative to the overload of information on the *web*. Among its approaches, we have the Collaborative Filtering, which acts as automation of the natural process of *mouth-to-mouth* recommendation, using the preferences of users to generate recommendations to other users of similar profile. They have their development mainly in the field of *e-commerce* and *e-business*, however, they are currently also applied in online public access catalogs (OPACs) of libraries, traditional or digital, internationally. Although targeted to the commercial issue and studied significantly in the computing, the recommendation systems, in general, have a strong relationship with Library Science. Therefore, this work through an exploratory research, seeks to analyze how Collaborative Filtering applied in specialized libraries can help in the indexing and the user studies.

Keywords: Recommender Systems. Collaborative Filtering. User Studies. Indexing. Specialized Libraries.

LISTA DE ILUSTRAÇÕES

Figura 1 -	Recomendações do Spotify (Daily Mix).....	22
Figura 2 -	Recomendação por gênero (Netflix).....	25
Figura 3 -	Recomendação por similaridade (Netflix).....	25
Figura 4 -	Recomendação de itens relacionados aos interesses do usuário (Saraiva).....	26
Figura 5 -	Recomendação de itens geralmente comprados por quem procura o item em questão (Saraiva)	26
Figura 6 -	Recomendação de itens também acessados por quem procura o item em questão (Saraiva).....	27
Figura 7 -	Exemplo de filtragem colaborativa baseada em usuário.....	31
Figura 8 -	Exemplo de filtragem colaborativa baseada em item.....	31
Figura 9 -	Representação do processo de recuperação da informação.....	43
Figura 10 -	Representação da ciclicidade do processo de busca e recuperação da informação.....	43
Figura 11 -	Busca sob o termo “Romance” (Netflix).....	53
Figura 12 -	Busca sob o termo “Anos 80” (Netflix).....	54
Figura 13 -	Busca sob o termo “Romântico” (Spotify).....	54
Figura 14 -	Estatísticas personalizadas de uso (Spotify).....	55
Figura 15 -	Esquema de navegação do usuário no catálogo.....	59

LISTA DE QUADROS

Quadro 1 -	Quadro teórico.....	17
Quadro 2 -	Características utilizadas e analisadas pelo Spotify.....	23
Quadro 3 -	Técnicas de recomendação.....	29
Quadro 4 -	Vantagens e desvantagens (feedback implícito/explicito).....	32
Quadro 5 -	Classificação das técnicas utilizadas em Sistemas de Recomendação, adaptado de Adomavicius e Tuzhilin (2005).....	35
Quadro 6 -	Vantagens e desvantagens da filtragem colaborativa.....	36
Quadro 7 -	Adaptação do quadro de relação entre as vantagens/desvantagens e técnicas adicionais.....	37
Quadro 8 -	Diferenças entre as abordagens tradicional e alternativa.....	40
Quadro 9 -	Modelos quantitativos de recuperação da informação.....	48
Quadro 10 -	Modelos dinâmicos de recuperação da informação.....	49
Quadro 11 -	Fatores relacionados a busca e recuperação da informação na <i>web</i>	52

LISTA DE ABREVIATURAS E SIGLAS

CI -	Ciência da Informação
EU -	Estudo de Usuários
FC -	Filtragem Colaborativa
HTML -	<i>Hypertext Markup Language</i>
HTTP -	<i>Hypertext Transfer Protocol</i>
SR -	Sistemas de Recomendação
SRI -	Sistemas de Recuperação da Informação
OPAC -	<i>Online Public Access Catalog</i>
URL -	<i>Universal Resource Locator</i>

SUMÁRIO

1	INTRODUÇÃO	13
2	JUSTIFICATIVA	14
3	OBJETIVOS	15
3.1	Objetivo geral	15
3.2	Objetivos específicos	15
4	METODOLOGIA	16
5	REFERENCIAL TEÓRICO	17
5.1	Sistemas de Recomendação	18
5.1.1	Cenário dos Sistemas de Recomendação	22
5.1.2	Técnicas de Filtragem Aplicadas a Sistemas de Recomendação	27
5.2	Filtragem Colaborativa	30
5.2.1	Vantagens e Desvantagens da Filtragem Colaborativa	35
5.3	Estudo de Usuários	38
5.3.1	Abordagens tradicional e alternativa	39
5.4	Busca e Recuperação da Informação	41
5.5	Indexação	44
5.5.1	Indexação automática	47
5.5.2	Indexação na <i>web</i>	51
5.6	Cenário dos sistemas de recomendação em bibliotecas	56
6	ANÁLISE E DISCUSSÃO DOS RESULTADOS	63
7	CONSIDERAÇÕES FINAIS	67
	REFERÊNCIAS	69

1 INTRODUÇÃO

Os Sistemas de Recomendação (SR) surgem em meados da década de 90 como uma alternativa diante da sobrecarga de informações na *web*. Seu desenvolvimento se dá, sobretudo, no contexto do *e-commerce*, e hoje se expande para outros segmentos como *e-business*, bibliotecas digitais, inclusive em catálogos de acesso público online (OPACs) de bibliotecas.

A Filtragem Colaborativa (FC), uma das abordagens dos sistemas de recomendação, surge como uma forma de automatizar o processo de recomendação “boca a boca”, permitindo que os usuários recebam recomendações de acordo com seu perfil de interesse. A filtragem colaborativa pode ser aplicada sob duas perspectivas: a do usuário (*user-based*) e a do item (*item-based*). Na primeira, a recomendação é gerada a partir da similaridade entre os usuários, e na segunda, a recomendação é gerada por meio dos itens previamente classificados pelo usuário.

Embora muito presentes em ambientes de *e-commerce*, os sistemas de recomendação, mais especificamente os baseados em filtragem colaborativa, têm forte relação com temáticas da Biblioteconomia, sobretudo estudo de usuários, uma vez que se utiliza de informações do usuário para criação e manutenção de seu perfil para gerar recomendações. E, também, a indexação uma vez que em OPACs, a recomendação ocorre após a busca, resultados e escolha do item mais relevante de acordo com a expressão de busca utilizada; sendo de extrema importância a qualidade da indexação nesse processo, dentre outras atividades, tais como o serviço de referência - em virtude da possibilidade de um usuário descobrir itens novos, pouco circulados, pouco conhecidos.

Neste estudo o foco é a indexação centrada em uma abordagem que o processo de indexação precede e é a causa da busca e recuperação da informação.

Sendo assim, esse trabalho visa analisar a Filtragem Colaborativa e sua relação com a indexação e com o estudo de usuários em bibliotecas especializadas, uma vez que os sistemas de recuperação da informação utilizados em bibliotecas muitas vezes não identificam os interesses de seus usuários, nem os distinguem individualmente.

2 JUSTIFICATIVA

A importância deste estudo se dá em decorrência das mudanças presenciadas na sociedade atual no âmbito tecnológico, como o estudo e desenvolvimento do campo da inteligência artificial, o uso da lógica booleana para a otimização do processo de busca e recuperação da informação, desenvolvimento de técnicas de mineração de dados e textos, a qual aproximou-se da Ciência da Informação (CI) em virtude da automação de processos documentários, tais como a indexação e elaboração de resumos. O uso e estudo dos sistemas de recomendação surgiu e permanece no contexto nacional, majoritariamente, nos segmentos de e-commerce, muito embora sejam de grande utilidade para bibliotecas, sobretudo as especializadas, visto que, de acordo com sua definição organizam-se sobre disciplinas ou áreas específicas do conhecimento, atendendo a certas categorias de usuários (CUNHA, 2008). E neste sentido podem fornecer subsídios às atividades biblioteconômicas, tais como o estudo de usuários e o processo de indexação para a recuperação da informação na web. Diante disso, parte-se da seguinte questão: *de que forma a filtragem colaborativa pode corroborar para o processo de indexação e estudo de usuários em bibliotecas especializadas?*

3 OBJETIVOS

3.1 Objetivo geral

Identificar as possibilidades de aplicação de Sistemas de Recomendação baseados em filtragem colaborativa na indexação e sua relação com o estudo de usuários em bibliotecas especializadas.

3.2 Objetivos específicos

- a) Identificar e analisar as técnicas levantadas sob a perspectiva da indexação e recuperação da informação, e sua relação com o estudo de usuários;
- b) Apresentar as possibilidades da utilização de Sistemas de Recomendação baseados em Filtragem Colaborativa para a indexação e sua relação com o estudo de usuários em bibliotecas especializadas.

4 METODOLOGIA

Trata-se de uma pesquisa exploratória, no que tange aos objetivos, uma vez que visa “[...] desenvolver, esclarecer e modificar conceitos e ideias, tendo em vista a formulação de problemas mais precisos ou hipóteses pesquisáveis para estudos posteriores” (GIL, 2008, p. 27), além de propiciar uma visão mais ampla sobre o assunto a ser tratado. Quanto aos procedimentos técnicos trata-se de uma pesquisa bibliográfica desenvolvida a partir de material elaborado anteriormente (GIL, 2008) - em fontes¹ nas áreas das Ciências da Informação e Computação, englobando diferentes tipologias documentais (artigos, teses, dissertações). A técnica em questão foi pesquisada por meio de fundamentação teórica cujas temáticas abordadas foram a dos Sistemas de Recomendação baseados em Filtragem Colaborativa, a indexação, a busca e a recuperação da informação e o estudo de usuários. A busca ocorreu sob os termos Filtragem Colaborativa, Sistemas de Recomendação, estudo de usuários, indexação, busca e recuperação da informação, também em língua estrangeira [inglês] e utilizando-se de mecanismos de busca [busca por expressões, lógica booleana, truncamento]. Foram realizadas também buscas a fim de identificar sistemas que utilizam de recomendação, sendo escolhidas as plataformas *Spotify* e *Netflix*, para estudo comparativo e posterior análise de sua relação com a indexação e estudo de usuários.

¹ AIRCC Digital Library (Biblioteca digital de editora acadêmica, *open access*, nas áreas da Ciência da Computação e Engenharia), BDTD (Biblioteca Digital Brasileira de Teses e Dissertações), BRAPCI (Base de Dados em Periódicos da Ciência da Informação), LISA (Library and Information Science Abstracts), Portal de Periódicos da CAPES.

5 REFERENCIAL TEÓRICO

O referencial teórico adotado e a revisão de literatura efetuada permitiram apresentar e discutir as temáticas dos Sistemas de Recomendação baseados em Filtragem Colaborativa (I), e sua relação com a indexação, busca e recuperação da informação (II), e o estudo de usuários (III), conforme apresentado no quadro abaixo.

Quadro 1 - Quadro teórico

	Conceitos	Autores
I	Sistemas de Recomendação; Filtragem Colaborativa; Filtragem baseada em conteúdo; Filtragem híbrida; Técnicas aplicadas a Filtragem Colaborativa; <i>Feedback</i> implícito e explícito; Geração e manutenção do perfil do usuário; Formação da vizinhança	Aldomavicius e Tuzhilin (2005); Bernartt (2008); Zanette (2008); Resnick e Varian (1997); Morvile e Sindhvano (2010); Ricc, Rokach e Shapira (2011); Alvarez et al. (2016); Sampaio e Ramalho (2006); Madadipouya (2015); Banati e Metha (2010); Sampaio e Ramalho (2006); Chan e Jung (2015); Montaner, López e De La Rosa (2003)
II	Indexação; Busca e recuperação da informação; Indexação automática; Indexação na web; Modelos quantitativos de recuperação da informação; Modelos dinâmicos da recuperação da informação	Calvin Mooers (1951); Saracevic (1996); Ferneda (2003); Lancaster (2004); Chaumier (1988); Araújo Júnior (2007); Silva, Santos e Ferneda (2013)
III	Estudo de usuários; Abordagem tradicional do estudo de usuários; Abordagem alternativa do estudo de usuários	Figueiredo (1994); Dervin e Nilan (1983); Araújo (2008)

Fonte: a autora (2018).

A busca das fontes deu-se a partir de revisão de literatura, sendo escolhidos os materiais considerados relevantes e pertinentes para embasamento do trabalho. Com relação a temática dos sistemas de recomendação, a cobertura temporal teve um peso maior; foram selecionados principalmente trabalhos mais recentes [2012-2017] a fim de identificar o que há de novo com relação às técnicas utilizadas. Em se tratando das demais temáticas (ligadas a CI), foram utilizados principalmente

trabalhos de autores considerados referência em suas respectivas áreas de pesquisa, não excluindo trabalhos mais recentes e interdisciplinares, a fim de apresentar avanços e novas relações entre as temáticas.

5.1 Sistemas de Recomendação

Os Sistemas de Recomendação (SR), segundo Aldomavicius e Tuzhilin (2005) e Bernartt (2008), surgem em meados da década de 90 como uma área de pesquisa independente, “quando a partir do sistema de recomendação *Tapestry*, os pesquisadores passaram a focar em problemas de recomendação que explicitamente invocavam estruturas de avaliação (ratings)” (*op. cit*, 2008, p.16).

O sistema *Tapestry* foi projetado sob a justificativa de alavancar a participação social a fim de impedir que os usuários do sistema fossem inundados com a quantidade de informação disponível.

O *Tapestry* rastreava a reação das pessoas aos emails recebidos em massa – quais mensagens eram abertas, quais eram respondidas e quais eram apagadas – e então usava essas informações para ajudar a organizar a caixa de entrada dos usuários. (PARISER, 2012, p. 23 apud DAVID, 2017, p. 39).

Zanette (2008) aponta que os proponentes do sistema *Tapestry* cunharam o termo ‘filtragem colaborativa’ “visando designar um tipo de sistema específico no qual a filtragem da informação era realizada com o auxílio humano, ou seja, através da colaboração entre os grupos interessados” (p. 16). O autor ainda menciona que posteriormente diversos pesquisadores utilizaram o termo para designar qualquer tipo de sistema de recomendação e, no entanto, Resnick e Varian (1997), em seu trabalho, defendem que o termo “sistema de recomendação” é mais genérico que “filtragem colaborativa”, uma vez que SR podem funcionar independente da colaboração dos usuários.

Bernatt (2008, p. 16) aponta que “sistemas de recomendação computacionais emergiram como forma de suporte, mediação e automação do processo de recuperação da informação”. Para Zanette (2008, p. 25) “[...] é uma tecnologia, de filtragem de informação personalizada, usada para predizer quais ou quantos itens que um usuário poderia se interessar”.

Resnick e Varian (1997) afirmam que muitas vezes é necessário fazer escolhas sem experiência pessoal suficiente das alternativas, de então os sistemas de recomendação surgem com o objetivo de suprir as necessidades do usuário diante da diversidade de possibilidades. Os autores abordam que recomendações podem ser adquiridas de diversas formas como cartas de recomendação, resenhas de filmes, de forma direta (word of mouth), entre outras, e acrescentam que “[...] Os sistemas de recomendação auxiliam e aumentam esse processo social natural.” (p. 56). Bernartt (2008, p. 16) afirma:

A evolução destes sistemas e o fato deles trabalharem com bases grandes de informações, permitiram que recomendações emergentes (não triviais) pudessem ser alcançadas, proporcionando ainda maior credibilidade que uma recomendação humana.

Segundo Morvile e Sindhwano (2010) esses sistemas objetivam gerar recomendações personalizadas a grupos de usuários, de itens e/ou produtos que possam interessá-los.

Zanette (2008) indica a filtragem de informação como solução para a sobrecarga de informações, para que apenas conteúdos interessantes e relevantes sejam indicados ao usuário; e menciona que muitos sistemas baseiam-se na criação do perfil do usuário, com base nos conteúdos acessados, para prever e recomendar informações relevantes. O autor também aponta sistemas que utilizam tecnologias de consulta e indexação (baseadas na análise de conteúdo do item) para a filtragem de informação, e ainda afirma que

Indexar pode ser descrito como o processo de examinar coleções de conteúdos e criar uma estrutura de dados que contenha descrições dos itens avaliados. Uma das questões-chave neste processo está na identificação do que é realmente importante e do que é irrelevante, sendo que retirar o irrelevante é fundamental. A identificação de termos relacionados também é importante para que não ocorra perda de conteúdo dito relevante. (ZANETTE, 2008, p. 26, grifo nosso)

Ricc, Rokach e Shapira (2011) apontam 5 razões para prover o serviço de recomendação em qualquer ambiente, sobretudo em meio comercial:

- Aumento no número de vendas;
- Venda de itens diversificados;

- Aumento da satisfação dos usuários;
- Fidelização dos usuários;
- Melhor entendimento das necessidades dos usuários.

Os autores vão além, mencionando que os SRs podem interessar também aos usuários se suas tarefas e metas forem efetivamente apoiadas pelo sistema, não somente a provedores de *e-service*. Eles afirmam que esses sistemas devem equilibrar suas necessidades e oferecer serviços valiosos para ambos. Ricc, Rokach e Shapira (2011, apud HERLOCKER, 2000) apontam 11 tarefas relacionadas a SRs:

- *Encontrar alguns bons itens*: recomendar uma lista de itens ranqueados com previsões do quanto o usuário gostaria do item, numa escala de 1 a 5, por exemplo;
- *Encontrar todos os bons itens*: recomendar todos os itens que podem satisfazer as necessidades do usuário;
- *Anotação no contexto*: dado uma lista de itens, por exemplo, enfatizar algum deles tendo em vista as preferências a longo prazo do usuário;
- *Recomendar uma sequência*: não concentrar a recomendação em apenas um, mas numa sequência de itens “agradáveis como um todo”;
- *Recomendar um pacote*: recomendar um grupo de itens como um pacote. Ex.: plano de viagem composto de atrações, hospedagem, alimentação etc. “Do ponto de vista do usuário essas várias alternativas podem ser consideradas e selecionadas como um destino de viagem único.” (2011, p. 6, tradução nossa);
- *Navegação*: alternativa para usuário que querem apenas navegar pelo catálogo. A tarefa do recomendador é ajudar o usuário a encontrar itens dentro do seu escopo de interesse na sessão de navegação específica;
- *Encontrar recomendações confiáveis*: tendo em vista que por vezes usuários não dão credibilidade a SRs, “alguns sistemas também podem oferecer funções específicas para permitir que os usuários testem seus comportamentos além daqueles exigidos apenas para obter recomendações.” (2011, p. 6, tradução nossa);
- *Melhorar o perfil*: capacidade do usuário de fornecer informações ao sistema a fim de otimizar as recomendações. “Se o sistema não tiver conhecimento específico sobre o usuário ativo, ele poderá apenas fornecer-lhe as mesmas recomendações que seriam entregues a um usuário ‘médio.’” (2011, p. 6, tradução nossa);
- *Auto expressão*: a recomendação pode não ser o foco do usuário, mas sim a possibilidade de expressar suas opiniões por meio da classificação de itens. “A

satisfação do usuário para essa atividade ainda pode atuar como uma alavanca para segurar o usuário firmemente ao aplicativo (como mencionado acima ao discutir as motivações do provedor de serviços).”(2011, p. 6, tradução nossa);

- *Ajudar outros usuários*: “alguns usuários têm prazer em contribuir com informações, por exemplo, avaliação de itens (ratings), porque eles acreditam que a comunidade beneficia da sua contribuição.” (2011, p. 6, tradução nossa);
- *Influenciar outros usuários*: sobretudo no contexto comercial, o principal objetivo da utilização de SRs é influenciar outros usuários a comprar produtos específicos.

Já Alvarez et al. (2016) vislumbram os sistemas de recomendação como complementares aos resultados obtidos em processos de busca dos usuários e apontam 3 vantagens de sua aplicação em instituições de informação:

- Apresentar recomendações de acordo com os interesses dos usuários;
- Gerar recomendações tendo como base consultas de outros usuários com o mesmo perfil, mostrando uma hierarquização dos resultados de acordo com a quantidade de consulta, bem como colaborar com o princípio da serendipidade, recomendado documentos os quais não foram recuperados em sua busca, mas baseado no histórico de busca de usuários com perfis semelhantes;
- Fornecer recomendações dos documentos mais bem avaliados, bem como comentários, a fim de facilitar “[...] a decisão de consultar primeiro aqueles documentos que especialistas com o mesmo perfil e interesses de pesquisa avaliaram como mais importantes ou valiosos a partir de uma perspectiva científica.” (ALVAREZ; ET AL., 2016, p. 284)

O uso desses sistemas em bibliotecas já é realidade em países europeus e norte-americanos, potencializando o processo de encontrabilidade da informação durante a busca realizada pelo usuário, bem como propiciando serendipidade², e servindo também de mediador entre a informação e o usuário.

² “[...] Anglicismo que se refere às descobertas afortunadas feitas, aparentemente, por acaso.” (WIKIPEDIA). Disponível em: <<https://pt.wikipedia.org/wiki/Serendipidade>>. Acesso em: 21 dez. 2018.

5.1.1 Cenário dos Sistemas de Recomendação

Os Sistemas de Recomendação, atualmente, são peças-chave, não somente no âmbito do comércio eletrônico, como também em *e-business* (ou negócio eletrônico). Em seu trabalho, Lu et.al (2015), reconhecem que os SR oferecem uma série de oportunidades e desafios a diversos domínios. Os autores analisam o cenário de desenvolvimento desses sistemas e agrupam em 8 categorias³, dentre elas: *e-business* e *e-commerce*.

Sobre *e-business* e *e-commerce*, os autores apontam que

Em geral, alguns sistemas enfocam recomendações geradas para clientes individuais, que são sistemas B2C (**business-to-consumer**), enquanto outros buscam fornecer recomendações sobre produtos e serviços para usuários corporativos, que são sistemas **business-to-business** (B2B). Neste estudo, os **sistemas de recomendação de e-business** referem-se a sistemas de recomendação para **aplicativos B2B**. **Sistemas de recomendação de e-commerce / e-shopping** referem-se a sistemas de recomendação para **aplicativos B2C**. (LU; et. al, 2015, p. 17, grifo nosso)

Tem-se como exemplo de SR de *e-business* - recomendação de produtos e serviços - as plataformas de *streaming*: Spotify⁴ (musical) e Netflix⁵ (vídeos).

Figura 1 - Recomendações do Spotify (Daily Mix)



Fonte: *Print* extraído do aplicativo Spotify (2018).

³ E-government, e-business, comércio eletrônico / e-shop, e-learning, e-turismo, e-resource, atividades de e-group e e-library.

⁴ “O **Spotify** é um serviço de música digital que dá acesso a milhões de músicas” (SPOTIFY). Disponível em: <<https://www.spotify.com/br/>>.

⁵ “Netflix é uma provedora global de filmes e séries de televisão via streaming” (WIKIPEDIA). Disponível em: <<https://www.netflix.com/br/>>.

Oliveira (2018, online) informa algumas das características utilizadas e analisadas pelo Spotify para a geração de recomendações, conforme o quadro a seguir:

Quadro 2 - Características utilizadas e analisadas pelo Spotify

CARACTERÍSTICA	DESCRIÇÃO
<i>Acústica</i>	Medida de confiança de 0,0 a 1,0 que define se a música é acústica ou não.
<i>Dançabilidade</i>	Indicador se a música é adequada ou não para dança. O embasamento surge a partir da combinação de elementos musicais, onde é analisado tempo, estabilidade do ritmo, força da batida. Medida entre 0,0 e 1,0.
<i>Duração</i>	Tempo da música em milissegundos.
<i>Energia</i>	Medida entre 0,0 e 1,0 que representa uma medida perceptual de intensidade e atividades. Por exemplo, músicas no estilo metal tem alta energia e pontuação alta, enquanto musicas classicas tem energia baixa nessa pontuação.
<i>Identidade</i>	Número de identificação, criado pelo Spotify, para cada música.
<i>Instrumentalidade</i>	Identifica se uma faixa possui contêm sons vocais ou não
<i>Vivacidade</i>	Algoritmos de recomendação do Spotify que buscam detectar a se existe público na gravação ou não.
<i>Sonoridade</i>	Entendimento do volume total de uma faixa em decibéis (dB). Durante toda a música são calculados valores da sonoridade daquela faixa.
<i>Fala</i>	Através da fonoaudiologia o Spotify consegue identificar se naquela faixa existem mais palavras faladas
<i>Tempo</i>	Estimativa de batidas por minuto (BPM) dentro de uma música.

Fonte: Oliveira (2018, online). Disponível em:

<<https://predictblog.com/sistemas-de-recomendacao/>>.

Oliveira (2018, online) ainda afirma que essas características são chamadas de *features*, e que possuem pesos específicos, possibilitando identificar quais músicas são mais indicadas para cada tipo de perfil de usuário.

Já a Netflix surge em 1997 como um serviço online de locação de filmes. No ano seguinte é lançado o primeiro site de vendas e aluguel de DVDs e, a *posteriori*, é lançado um serviço de assinatura mensal ilimitada de DVDs. Nos anos 2000 a Netflix lança um **sistema de recomendação personalizada de filmes**, que utiliza as classificações de seus próprios assinantes para gerar recomendação. Em 2007 inicia-se o serviço de transmissão online, que permite aos assinantes assistirem séries e filmes instantaneamente no computador, e a partir de 2008 são estabelecidas parcerias para a transmissão de conteúdo online por meio de outros suportes. Atualmente é um dos maiores serviços de streaming de vídeos no mundo; possui mais de 130 milhões de assinantes em mais de 190 países (NETFLIX, online).⁶

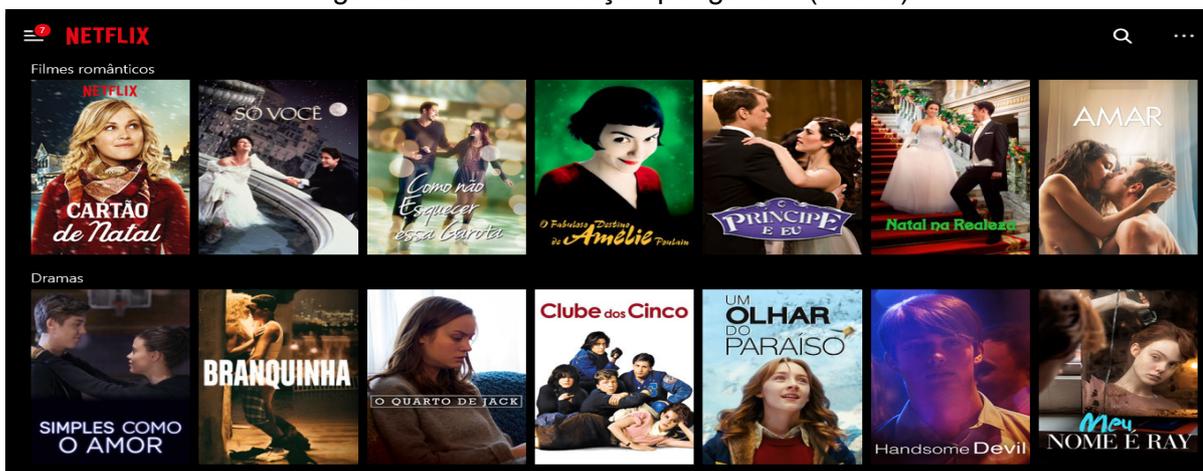
Em 2006 a Netflix, enxergando o potencial dos SR, lança o concurso *Netflix Prize*, uma competição de aprendizado de máquina e mineração de dados para previsão de classificação de filmes, oferecendo um prêmio de US \$ 1 milhão ao grupo que conseguisse otimizar seu sistema de recomendação em 10%.

A personalização começa na nossa página inicial, que consiste em grupos de vídeos organizados em linhas horizontais. Cada linha tem um título que transmite a conexão significativa pretendida entre os vídeos desse grupo. A maioria de nossa personalização é baseada na maneira como selecionamos linhas, como determinamos quais itens incluir nelas e em que ordem colocar esses itens. (AMATRIAN; BASILICO, 2012, online)

Dentre as personalizações mais comuns no serviço de *streaming* em questão, tem-se a coleção de linhas (recomendação) por gênero, que podem ser categorizados de forma genérica, conforme a Figura 2, “Filmes Românticos”, “Dramas”, ou específica, como “Imaginative Time Travel Movies, da década de 1980”.

⁶ NETFLIX. Disponível em: <https://media.netflix.com/pt_br/about-netflix>.

Figura 2 - Recomendação por gênero (Netflix)



Fonte: *print* extraído da Netflix (2018).

A similaridade também é fonte de personalização desse serviço; a semelhança, em questão, pode ser entre filmes/séries, entre usuários, podendo, também, estar presente em diferentes dimensões como metadados, classificações ou visualizações. Amatrian e Basilico (2012, online, grifo nosso) apontam que

[...] essas semelhanças podem ser combinadas e usadas como recursos em outros modelos. A similaridade é usada em vários contextos, por exemplo, em resposta à ação de um membro, como pesquisar ou adicionar um título à fila. Ele também é usado para gerar linhas de "gêneros ad hoc" com base na **semelhança com os títulos com os quais um membro interagiu recentemente** [conforme a figura abaixo].

Figura 3 - Recomendação por similaridade (Netflix)



Fonte: *print* extraído da Netflix (2018).

Como exemplo de SR de *e-commerce*, tem-se o SR da Livraria Saraiva. Borges e Oliveira ([2010]) analisam o sistema de recomendação de três empresas de comércio eletrônico, dentre elas, a Saraiva. Os autores afirmam que a coleta de

informações do usuário ocorre de forma explícita⁷, a partir do momento que o mesmo faz o login na página e acessa os itens; eles ainda apontam que é realizada, também, a coleta de informações da comunidade de usuários, de maneira geral, para serem “[...] computadas avaliações de produtos feitas por outros usuários e a popularidade dos produtos para formar a lista de produtos mais vendidos.” (BORGES; OLIVEIRA, [2010], p. [5]), conforme as Figuras 4, 5 e 6, abaixo.

Figura 4 - Recomendação de itens relacionados aos interesses do usuário (Saraiva)



Fonte: *Print* extraído do site da Saraiva (2018).

Figura 5 - Recomendação de itens geralmente **comprados** por quem procura o item em questão (Saraiva)



Fonte: *Print* extraído do site da Saraiva (2018).

⁷ Comumente chamada de customização.

Figura 6 - Recomendação de itens também **acesados** por quem procura o item em questão (Saraiva)



Fonte: *Print* extraído do site da Saraiva (2018).

Além disso, são utilizadas estratégias de recomendação no site, tais como: lista de recomendação (produtos mais vendidos no site), de acordo com a categoria; avaliação e comentário-como meio de avaliação (positiva/negativa) do produto e opinião; itens semelhantes (produtos relacionados ao item o qual está sendo consultado); associação por conteúdo - por exemplo, ao buscar determinado livro de x autor, são sugeridos outros livros do mesmo autor; e *e-mail* - esta estratégia é utilizada quando o produto está indisponível, o usuário insere seu contato e recebe um *feedback* quando o produto desejado estiver disponível novamente.

5.1.2 Técnicas de Filtragem Aplicadas a Sistemas de Recomendação

Existem diversas técnicas, de pequena ou grande complexidade, as quais permitem que o processo de recomendação aconteça. Ricci, Rokach e Shapira (2011) citam as classificações/avaliações de usuários para itens, descrições ontológicas dos usuários ou itens⁸, e relações sociais e atividades dos usuários.

Os dados utilizados por esses sistemas referem-se a três tipos de objetos: itens, usuários e transações, ou interações, (relações entre itens e usuários). Os **itens** podem ser caracterizados pela sua complexidade, valor ou utilidade. Sendo o

⁸ "As técnicas baseadas em conteúdo têm um limite natural no número e no tipo de **recursos associados**, automática ou manualmente, **aos objetos que eles recomendam. O conhecimento do domínio é frequentemente necessário**, por exemplo, para recomendações de filmes, o sistema precisa conhecer os atores e diretores e, às vezes, também são necessárias **ontologias de domínio.**" (RICCI; ROKACH; SHAPIRA, 2011, p. 101, grifo nosso)

item útil ao usuário, seu valor é positivo, senão, seu valor é negativo. Ainda que o usuário não esteja pagando pela informação recebida, existe um custo cognitivo, além disso, e deve ser considerado o tempo perdido, quando recuperado e/ou recomendado um item cuja utilidade é nula.

Os **usuários** podem ter características e objetivos diferentes, sendo assim, é necessário explorar uma gama de informações desses usuários para a personalização da recomendação, a partir da exploração de modelos de perfil, utilizando de técnicas as quais serão exploradas mais adiante.

As **transações/interações** são as relações entre o usuário e o SR. Ricci, Rokach e Shapira (2011, p. 9) explicitam que

são dados do tipo log que armazenam informações importantes gerados durante a interação humano-computador e que são úteis para o algoritmo de geração de recomendação que o sistema está usando. Por exemplo, um log de transações pode conter uma referência ao item selecionado pelo usuário e uma descrição do contexto (por exemplo, a meta / consulta do usuário) para essa recomendação específica. [...] Na verdade, as classificações são a forma mais popular de dados de transação que um RS coleta. Essas classificações podem ser coletadas de maneira explícita ou implícita.

Ao desenvolver Sistemas de Recomendação é necessário ter em mente seu objetivo e contexto para a escolha da técnica mais adequada. Dentre as mais utilizadas, tem-se:

- **Filtragem baseada em conteúdo:** Na recomendação baseada em conteúdo, recomenda-se ao usuário itens semelhantes aos preferidos por ele no passado. Por exemplo, em um aplicativo de recomendação de filmes, para gerar recomendações a determinado usuário **c**, o SR baseado em conteúdo tenta entender as **semelhanças entre os filmes anteriormente classificados pelo usuário** (atores específicos, diretores, gêneros, assunto etc). Então, apenas filmes com alto grau de semelhança com as preferências do usuário **c**, serão recomendados. (ALDOMAVICIUS; TUZHILIN, 2005) Esses sistemas utilizam dos metadados dos itens para gerar recomendação;

- **Filtragem colaborativa:** Nos sistemas de filtragem colaborativa, um usuário é recomendado com base nas classificações anteriores de todos os usuários [que possuem perfil semelhante] coletivamente. (MELVILLE; SINDHIWANI, 2010). Por exemplo, em um aplicativo de recomendação de filmes, para gerar recomendações a determinado usuário **c**, o SR baseado em filtragem colaborativa tenta encontrar os “pares” do usuário **c**, ou seja, outros usuários com gostos semelhantes em filmes (classificam os filmes de maneira semelhante). Então, apenas os filmes mais apreciados pelos “pares” do usuário **c** seriam recomendados. (ALDOMAVICIUS; TUZHILIN, 2005);
- **Filtragem híbrida:** Esses sistemas geram recomendações a partir da combinação (SR baseado em conteúdo + SR baseado em filtragem colaborativa). (MELVILLE; SINDHIWANI, 2010) (ALDOMAVICIUS; TUZHILIN, 2005).

Quadro 3 – Técnicas de recomendação

Técnicas	Autores
Abordagem baseada em conteúdo	(RICCI; ROKACH; SHAPIRA, 2011), (BERNARTT, 2008); (SAMPAIO; RAMALHO, 2006); (ZANETTE, 2008); (BEL; et al., 2015)
Filtragem colaborativa	(RICCI; ROKACH; SHAPIRA, 2011), (BERNARTT, 2008); (SAMPAIO; RAMALHO, 2006); (ZANETTE, 2008);(BEL; et al., 2015)
Demográfica	(RICCI; ROKACH; SHAPIRA, 2011)
Baseada em conhecimento	(RICCI; ROKACH; SHAPIRA, 2011);
Baseada em comunidade	(RICCI; ROKACH; SHAPIRA, 2011)
Abordagem híbrida	(RICCI; ROKACH; SHAPIRA, 2011); (BERNARTT, 2008); (SAMPAIO; RAMALHO, 2006); (ZANETTE, 2008); (BEL; et al., 2015)
Filtragem econômica	(SAMPAIO; RAMALHO, 2006)
Estereotipagem	(BEL; et al., 2015)

Coocorrência	(BEL; et al., 2015)
Baseado em gráfico	(BEL; et al., 2015)
Relevância global	(BEL; et al., 2015)

Fonte: a autora (2018).

5.2 Filtragem Colaborativa

A Filtragem Colaborativa (FC) é considerada a técnica mais popular e amplamente aplicada em Sistemas de Recomendação (RICCI; ROKACH; SHAPIRA, 2011). Segundo Sampaio e Ramalho (2006, p. 21), essa técnica “baseia-se no fato de que as melhores recomendações para um indivíduo podem ser feitas por pessoas que possuem preferências similares a ele”.

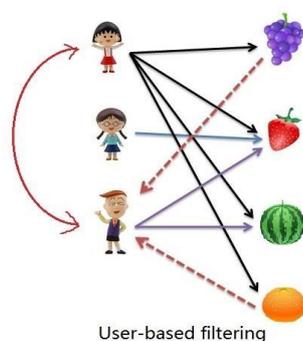
O processo de sugestão gira em torno da similaridade entre os usuários, diferente da filtragem baseada em conteúdo, onde é considerada a similaridade entre os itens. A FC subdivide-se em 2 modelos:

- Baseado em usuário (user-based)⁹: recomendação realizada a partir da similaridade entre usuários;
- Baseado em item (item-based): recomendação realizada a partir dos itens previamente classificados pelo usuário.

A Filtragem Colaborativa baseada em usuário/memória, segundo Madadipouya (2015) e Banati e Metha (2010) fornece recomendações utilizando de técnicas estatísticas, sobretudo para a identificação de usuários semelhantes [também chamados *vizinhos*] cujo comportamento passado é semelhante ao do usuário-alvo.

⁹ Também conhecida como *memory-based* (ALDOMAVICIUS; TUZHILIN, 2005), *heuristic-based* (ZANETTE, 2008) e/ou *neighborhood-based* (RICCI; ROKACH; SHAPIRA, 2011) (MELVILLE; SINDHWANI, 2010)

Figura 7 - Exemplo de filtragem colaborativa baseada em usuário

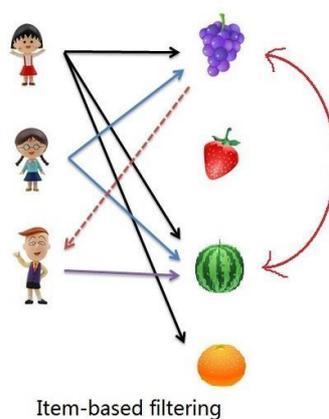


Fonte: Figura extraída do site DevMedia¹⁰

Como exemplo, a partir da Figura 7, é possível perceber a similaridade entre os usuários 1 e 3. Tendo em vista que o usuário 1 classificou positivamente todos itens, o sistema irá recomendar a sua vizinhança [usuário 3] os itens desconhecidos por ele, mas classificados positivamente pelo usuário 1.

Diferente do método baseado em usuário, os algoritmos baseados em item usam a coleção de itens classificados para aprender um *modelo*, o qual é usado para realizar previsões de classificação. (ALDOMAVICIUS; TUZHILIN, 2005).

Figura 8 - Exemplo de filtragem colaborativa baseada em item



Fonte: Figura extraída do site DevMedia²

¹⁰ Disponível em:

<<https://www.devmedia.com.br/apache-spark-como-criar-um-mecanismo-de-sugestao-de-produtos/33459>>.

Como exemplo de FC baseada em item, conforme a Figura 8 demonstra, o sistema entende que os usuários [de perfil semelhante] que classificam positivamente o item 1, classificam positivamente, também, o item 3; sendo assim, será recomendado ao usuário que classificar positivamente o item 1, o item 3, e vice-versa.

Já Queiroz (2003 apud SAMPAIO; RAMALHO, 2006), afirma que a filtragem colaborativa pode ser definida em 3 passos:

- Representação dos dados de entrada:

Segundo Sampaio e Ramalho (2006) a representação dos dados de entrada ocorre quando os usuários expressam suas preferências avaliando itens do sistema. Montainer, López e De La Rosa (2003, p.301, grifo nosso) afirmam que “As duas formas mais comuns de obter feedback relevante são usar informações **explicitamente** ou para obter informações observadas **implicitamente** da interação do usuário”. O *feedback* explícito dá-se em forma de like/dislike, classificações (5 estrelas, escala de classificação numérica, por exemplo), comentários. Já no *feedback* implícito, o próprio sistema infere as preferências dos usuários automaticamente, a partir do monitoramento de suas ações; dá-se a partir do histórico de compras, de navegação, do tempo gasto em uma página específica, análise dos links seguidos pelo usuário.

Quadro 4 - Vantagens e desvantagens (feedback implícito/explicito)

	FEEDBACK IMPLÍCITO	FEEDBACK EXPLÍCITO
VANTAGEM	Não há necessidade de engajamento do usuário com o sistema para a geração de recomendação	Possibilidade de classificar um item tendo em vista seu gosto pessoal; melhor experiência do usuário com o sistema
DESVANTAGEM	Falta de privacidade; possibilidade de obtenção de informações do usuário sem que ele saiba ou autorize	Necessidade de engajamento do usuário com o sistema para a geração de recomendação

Fonte: A autora (2018)

- Formação da vizinhança:

“Para fazer a recomendação o sistema compara o perfil do usuário alvo com os perfis de outros usuários do sistema para encontrar a similaridade entre eles” (SAMPAIO; RAMALHO, 2006, p. 21). Montaner, López e De La Rosa, também pontuam que na FC a recomendação é gerada a partir da combinação de pessoas com perfis semelhantes e que, para isso, são necessárias 3 etapas: a) encontrar usuários similares, b) criar a vizinhança e c) calcular a previsão com base nos vizinhos selecionados.

- a) Encontrar usuários semelhantes - Chan e Jung (2015) mencionam que as preferências do usuário e os itens de recomendação devem ser inseridos em uma matriz.

Usuários com preferências semelhantes são calculados de acordo com as classificações de itens de usuários individuais. Os cálculos de similaridade variam com dados diferentes. [...] Portanto, distâncias mais curtas resultaram em menores valores de distância e maiores semelhanças após a reversão. (CHAN; JUNG, 2015, p. [6])

Montaner, López e De La Rosa (2003) apresentam as técnicas de classificação, clusterização e vizinho mais próximo como mais utilizadas para a realização do cálculo de similaridade entre usuários.

- b) Criação da vizinhança - após o cálculo da similaridade do usuário alvo com os demais usuários do sistema, torna-se necessário quais usuários terão suas avaliações levadas em conta para a predição da relevância dos itens a serem recomendados. Sampaio e Ramalho (2006) discorrem que nos primeiros trabalhos de FC com vizinhos, as avaliações de todos os usuários do sistema eram levadas em conta para a formação da vizinhança, entretanto, hoje é possível utilizar de técnicas para a “adoção de um tamanho fixo de vizinhança, sendo selecionados aqueles usuários de maior similaridade.” (p. 26). A *técnica best-n-neighbour* é apontada por Montaner, López e De La Rosa (2003) como mais adequada, uma vez

que não limita a previsão de cobertura e menciona o trabalho de Herlocker et al. (1999) para formação de vizinhança baseada no centróide¹¹.

O primeiro passo é escolher o usuário mais próximo do usuário alvo e calcular o centróide. Em seguida, outros usuários são incluídos na vizinhança com base na distância até o centróide, que é recalculado sempre que um novo usuário é adicionado. Basicamente, este algoritmo permite os vizinhos mais próximos para afetar a formação do bairro e pode ser benéfico para conjuntos de dados muito esparsos. (MONTANER; LÓPEZ; DE LA ROCHA, 2003, p. 319-320)

c) Calcular a previsão com base nos vizinhos selecionados - Após a seleção da vizinhança, as avaliações dos vizinhos são combinadas a fim de realizar previsões e, posteriormente, escalonar as classificações para uma distribuição comum. Montaner, López e De La Rosa (2003) apontam 3 técnicas: **itens mais frequentes** - analisa-se a vizinhança e os interesses do usuário alvo, extraíndo os itens selecionados com mais frequência [pela vizinhança], recomendando-os então ao usuário-alvo - **regras de associação**. “A recomendação baseada em regras de associação infere regras geradas anteriormente da vizinhança, em vez de usar toda a população de usuários.” (MONTANER; LÓPEZ; DE LA ROCHA, 2003, p. 320). E, por fim, utilizando o cálculo da **média ponderada das classificações**.

- Geração da recomendação:

“[...] com base nas avaliações feitas pelos componentes da vizinhança aos itens de informação, o sistema gera a recomendação para o usuário alvo.” (SAMPAIO; RAMALHO, 2006, p. 21), ou seja, o sistema recomenda a determinado usuário os itens que seus vizinhos mais gostaram.

Há ainda diversas técnicas relacionadas à Filtragem Colaborativa para construção e manutenção de perfis, e geração de recomendação, conforme quadro abaixo:

¹¹ Em geometria, “ponto em que as coordenadas são as médias das coordenadas dos pontos que formam uma figura geométrica; baricentro, centro geométrico”.

Quadro 5 - Classificação das técnicas utilizadas em Sistemas de Recomendação, adaptado de Adomavicius e Tuzhilin (2005).

Recommandation Approach	Recommendation Technique	
	<i>Memory/User/Heuristic/Neighborhood-based</i>	<i>Model/Item-based</i>
<i>Collaborative</i>	<p>Técnicas mais utilizadas:</p> <ul style="list-style-type: none"> • Correlação (cosseno, Pearson, Spearman) • Teoria dos Gráficos <p>Exemplos de Pesquisas:</p> <ul style="list-style-type: none"> • Resnick et al., 1994 • Hill et al., 1995 • Shardanand e Maes, 1995 • Breese et al., 1998 • Nakamura e Abe, 1998 • Aggarwal et al., 1999 • Delgado e Ishii, 1999 • Pennock e Horwitz, 1999 • Sarwar et al., 2001 	<p>Técnicas mais utilizadas:</p> <ul style="list-style-type: none"> • Redes Bayesianas • Agrupamento • Redes Neurais • Regressão Linear • Modelos Probabilísticos <p>Exemplos de Pesquisas:</p> <ul style="list-style-type: none"> • Billsus e Pazzani, 1998 • Breese et al., 1998 • Ungar e Foster, 1998 • Chien e George, 1999 • Getoor e Sahami, 1999 • Pennock e Horwitz, 1999 • Goldberg et al., 2001 • Kumar et al., 2001 • Pavlov & Pennock, 2002 • Shani et al., 2002 • Yu et al., 2004 • Hofmann, 2004 • Marlin, 2003 • Si e Jin, 2003

Fonte: Adaptação de Adomavicius e Tuzhilin (2005, p. 742).

Construir perfis precisos, ou seja, ser capaz de representar de forma precisa os interesses atuais de seus usuários é fundamental para a geração da recomendação, assim como o entendimento das necessidades informacionais dos usuários de uma biblioteca é de extrema importância.

5.2.1 Vantagens e Desvantagens da Filtragem Colaborativa

Em estudo anterior realizado pela autora desta pesquisa, cujo objetivo era identificar as vantagens e desvantagens dos SR baseados em FC aplicados à

recuperação da informação na web, foi possível desenvolver o quadro de vantagens e desvantagens a seguir:

Quadro 6 - Vantagens e desvantagens da filtragem colaborativa

Vantagens	Definição
<i>Incentivo a serendipidade</i>	Uma vez que esses sistemas [FC] permitem ao usuário a descoberta de itens desconhecidos por ele, mas avaliados positivamente por outros usuários de mesmo perfil.
<i>Uso do perfil do usuário para gerar recomendações</i>	Diferente dos SR baseados em conteúdo, os quais utilizam dos metadados dos itens para gerar recomendações, os baseados em FC utilizam as preferências dos usuários.
<i>Possibilidade de classificar um item com base no gosto pessoal</i>	Permitir que o usuário classifique itens com base em seu gosto pessoal, otimiza sua experiência com o sistema.
Desvantagens	Definição
<i>Cold-start (item/user)</i>	Falta de informação sobre o usuário e/ou item; ocorre, normalmente, com itens/usuários novos no sistema.
<i>Necessidade de interação com o sistema</i>	Também pode ser visto como desvantagem, uma vez que exige engajamento do usuário para que haja recomendação.
<i>Esparsidade</i>	Ocorre quando um item tem pouco ou nenhum acesso/avaliação, impossibilitando o processo de recomendação de conteúdos similares.
<i>Super-especialização do usuário</i>	Pode ocorrer em função da recomendação [com base no perfil do usuário] limitá-lo a certo grau de novidade, restringindo o conhecimento de itens de interesse sobre outras perspectivas ao grupo de especialistas.
<i>Escalabilidade</i>	Ocorre devido à complexidade de alguns algoritmos de FC em trabalhar com uma grande quantidade de dados em constante modificação, acarretando em lentidão no sistema.
<i>Privacidade do usuário</i>	Uma vez que, para geração e manutenção do perfil, os dados do usuário podem ser coletados de forma implícita, sem que o mesmo saiba ou autoriza.

	Refere-se, também, a questão da segurança (guarda) do grande volume de dados pessoais coletados e armazenados em ambiente online.
--	---

Fonte: A autora (2018)

Além disso, identificou-se o uso de métodos e técnicas adicionais a fim de potencializar a geração de recomendações e/ou solucionar os problemas/desvantagens desse tipo de sistema, refletindo o cenário de constante desenvolvimento da área da Ciência da Computação.

Quadro 7 - Adaptação do quadro de relação entre as vantagens/desvantagens e técnicas adicionais

Vantagens	Métodos/Técnicas adicionais
a) Incentivo a serendipidade	<ul style="list-style-type: none"> ● Método híbrido (user-based+item-based)
b) Uso do perfil do usuário para gerar recomendações	<ul style="list-style-type: none"> ● Análise de redes sociais
c) Possibilidade de classificar um item com base no gosto pessoal	<ul style="list-style-type: none"> ● Uso de folksonomias
Desvantagens	Definição
a) Cold-start (item/user)	<ul style="list-style-type: none"> ● Abordagem multimodal
b) Necessidade de interação com o sistema c) Super-especialização do usuário d) Esparsidade	<ul style="list-style-type: none"> ● Análise de logs ● Recomendação demográfica ● Abordagem multimodal ● Técnica de co-clusterização
e) Escalabilidade	<ul style="list-style-type: none"> ● Algoritmos evolucionários (algoritmos meméticos e genéticos)
f) Privacidade do usuário	<ul style="list-style-type: none"> ● Computação multipartidária segura

Fonte: Adaptação de (NIEDZEILSKI; GUIMARÃES, 2018, p. 29)

5.3 Estudo de Usuários

Figueiredo (1994) aponta o Estudo de Usuário (EU) como um conjunto de investigações realizadas a fim de descobrir quais as necessidades informacionais dos usuários de determinada biblioteca, centro de informação / documentação, bem como para determinar se suas necessidades informacionais estão, de fato, sendo supridas por esses locais. A autora pontua que “Estes estudos são, assim, canais de comunicação entre a biblioteca e a comunidade a qual ela serve [...] necessários também para ajudar a biblioteca na previsão da demanda ou mudança da demanda de seus produtos e serviços [...]” (1994, p. 7).

Grande parte dos estudos em E.U. foram realizados na segunda metade da década de 40, na conferência da Royal Society (1948), onde foram apresentados trabalhos relacionados ao estudo das necessidades dos usuários. Já em 1958, em Washington, realizou-se a Conferência Internacional de Informação Científica, contribuindo também com trabalhos na temática.

Nice Figueiredo (1994, p.7) aponta ainda que houve uma mudança com relação aos usuários “[...] até então, adotava-se uma atitude passiva, aguardava-se que os usuários aparecessem e soubessem como fazer uso da informação disponível” e que tais mudanças ocorreram com a finalidade de transformar a biblioteca em um ambiente mais ativo e dinâmico, com a criação ou desenvolvimento de serviços já prestados por ela; nesse sentido, os Sistemas de Recomendação, hoje, podem ser considerados como aprimoramento dos serviços prestados por bibliotecas, bem como instrumentos de auxílio no processo de indexação e estudo de usuários.

Dervin e Nilan (1983) caracterizam o estudo de usuários em dois paradigmas, os quais serão tratados como abordagens, neste trabalho: **abordagem tradicional** (com foco no sistema de informação) e **abordagem alternativa** (centrada no usuário), conforme o tópico a seguir.

5.3.1 Abordagens tradicional e alternativa

O estudo de usuários inicialmente teve como foco o levantamento de dados para fins de adequação e aperfeiçoamento dos serviços e produtos de bibliotecas, posteriormente, buscou-se um conhecimento mais aprofundado acerca do processo de obtenção e uso da informação por parte dos usuários.

Figueiredo (1994) categoriza o estudo de usuários em três períodos: o **primeiro** (1948-1965), cujos métodos utilizados (questionários, entrevistas) tinham propósito exploratório, para a obtenção de dados quantitativos relacionados aos hábitos dos usuários com a finalidade de otimizar os sistemas de informação; no **segundo** período (a partir de 1965), eram utilizadas técnicas mais sofisticadas de observação indireta (análise de citações, verificação de compilações estatísticas, análise do uso de coleções) para o estudo de aspectos particulares do comportamento dos usuários; “Começou-se a adquirir um conhecimento mais profundo de como a informação é obtida e usada.” (FIGUEIREDO, 1994, p. 9), no entanto, o foco permanecia ainda nos sistemas de informação; já no **terceiro** período (década de 70) percebe-se a necessidade do estudo de usuários de diferentes áreas que não as ciências puras; os estudos passam a ser amplos e exploratórios, Figueiredo (1994, p. 10) resume esse momento apontando que

A tendência é para estudos de caráter mais restrito nos campos da ciência e tecnologia, dirigidos ao estudo de canais específicos de informação, do ponto de vista do usuário, ou para o esclarecimento de problemas observados em um sistema particular.

Já Dervin e Nilan (1986) traçam uma crítica ao direcionamento dado às pesquisas [à época] no campo de estudo de usuários. Os autores identificam a existência de dois paradigmas dentro do campo: abordagem tradicional e abordagem alternativa. Araújo (2010, p. 17) identifica que na abordagem tradicional tratava-se de

[...] um modelo em que a informação é vista como objetiva e os usuários como processadores de informação; que procura por proposições trans-situacionais sobre a natureza do uso de sistemas de informação; que faz isso enfocando as dimensões externamente observáveis do comportamento.

Araújo (2010, p. 17) afirma que diversas críticas foram apontadas à abordagem tradicional, gerando uma série de teorias e conceitos os quais contribuíram para o surgimento da abordagem alternativa:

- a abordagem sense-making de Dervin, que enfatiza o comportamento informacional em termos das categorias de situação, lacuna e uso;
- abordagem dos valores dos usuários de Taylor, que considera como os diferentes grupos ou contextos criam categorias específicas de valoração da informação;
- o modelo de comportamento informacional de Ellis, que enfatiza as várias categorias de uso de informação correlacionadas com as diferentes atividades desenvolvidas pelos usuários;
- a abordagem baseada em processo de Kuhlthau, que considera as sucessivas etapas do comportamento de busca e uso da informação e as variáveis cognitivas e emocionais que atuam em cada etapa;
- a abordagem do estado anômalo do conhecimento de Belkin, que se centra nos efeitos causados pela ausência de determinado conhecimento pelos usuários.

A abordagem alternativa, difere da anterior [abordagem tradicional] devido a sua visão de *informação* “[...] como algo construído por seres humanos, e os usuários como seres que estão constantemente construindo, como seres que são livres na criação de situações.” (*op. cit.*, 2010, p. 18). O modelo alternativo tem seu foco no uso da informação em situações particulares, centrando-se no usuário, examinando o sistema sob a perspectiva do usuário.

Quadro 8 - Diferenças entre as abordagens tradicional e alternativa

ABORDAGEM TRADICIONAL	ABORDAGEM ALTERNATIVA
Centrada no sistema de informação;	Centrada no usuário;
Busca entender como as bibliotecas e centros de informação são utilizados;	Busca observar o comportamento dos usuários;
Usuários vistos como informantes;	Usuário como ser de conhecimento construtivo e ativo;
Não são verificados os fatores que ocasionam o encontro usuário/sistema de informação;	Busca focar nos aspectos cognitivos envolvidos no encontro usuário/sistema de informação; Busca analisar sistematicamente a

Não considera tarefas relacionadas a interpretação, formulação e aprendizagem envolvidas no processo de busca da informação.	individualidade dos usuários.
--	-------------------------------

Fonte: Dervin e Nilan (1986).

5.4 Busca e Recuperação da Informação

Para este trabalho optou-se por uma abordagem que enfatiza a recuperação da informação englobando o processo de indexação e busca da informação, muito embora o processo de indexação preceda e tenha como objetivo facilitar a busca e recuperação da informação, contribuindo para tornar o sistema de recuperação da informação eficiente e eficaz.

Calvin Mooers (1951, p. 25, tradução nossa) cunhou o termo “Information Retrieval” (Recuperação da Informação), como nova disciplina a qual “[...] trata dos aspectos intelectuais da descrição da informação e sua especificação para busca, e também de qualquer sistema, técnicas ou máquinas empregadas para realizar esta ação”.

Para Saracevic (1996, p. 46),

Tendo se iniciado no começo dos anos 60 [trabalhos em recuperação da informação], prolongando-se até hoje, as questões acerca da natureza, manifestações e efeitos dos fenômenos básicos (a informação, o conhecimento e suas estruturas) e processos (comunicação e uso da informação) tornaram-se os principais problemas propostos pela pesquisa básica em CI. Incluem-se aí, dentre outras, tentativas de se formalizarem as propriedades da informação pela aplicação da teoria da informação, da teoria das decisões e outros construtos da ciência cognitiva, da lógica e/ou da filosofia; várias formas de estudos de uso e de usuários; formulações matemáticas da dinâmica das comunicações (como a teoria epidêmica da comunicação); ricas análises em bibliometria e cienciométrica, pela quantificação das estruturas do conhecimento (como a literatura e a esfera científica) e de seus efeitos (como as redes de citações), etc. Portanto, paralelamente com a aplicação da pesquisa e desenvolvimento, principalmente centrados em torno da recuperação da informação, uma linha básica de pesquisa evoluiu para CI, sendo em alguns casos tão rigorosa, matemática, lógica ou estatisticamente, como qualquer outra pesquisa científica similar.

O autor ainda pontua que, por volta dos anos 70, o paradigma da recuperação da informação volta-se aos usuários e suas interações. Kochen (1974 *apud* SARACEVIC, 1996, p. 46) afirma que

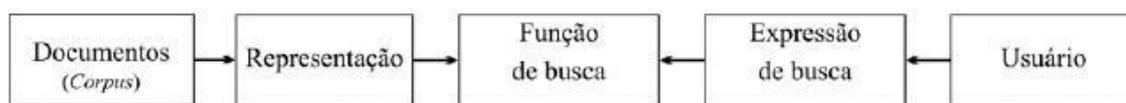
Podemos conceitualizar o sistema de conhecimento, no qual se inscreve a recuperação de informação, como composto por três partes; (a) as pessoas em seu papel de processadores de informações [pessoas com necessidades informacionais]; (b) os documentos em seu papel de suportes de informações; (c) os tópicos como representações [metadados os quais descrevem um item para fins de recuperação].

Portanto, a recuperação da informação constitui-se da comparação entre o solicitado [busca realizada pelo usuário], com o que foi tratado e armazenado; sua limitação corresponde às necessidades informacionais do usuário que faz a busca. Araújo Júnior (2007, p. 65) afirma sobre essa questão que o processo de busca e recuperação da informação pode ser definido pelo lado do usuário que realiza a busca dada uma motivação/necessidade informacional, e da recuperação da informação a qual deve aproximar-se o máximo possível da demanda informacional do usuário.

A base da relação entre as Ciências da Informação e da Computação reside na aplicação dos computadores e da computação na recuperação da informação, sobretudo. Enquanto a Ciência da Informação trata da natureza da informação e de sua comunicação, a Ciência da Computação trata processos que podem ser executados através de um conjunto sequencial de instruções, algoritmos que transformam a informação.

De forma resumida, os sistemas de recuperação da informação devem representar os conteúdos do *corpus dos documentos* e disponibilizá-los aos usuários de forma que seja possível aos mesmos a seleção de itens que satisfaçam suas necessidades informacionais, formalizadas em expressões de busca. A figura abaixo é uma representação simplificada do processo de recuperação da informação.

Figura 9 - Representação do processo de recuperação da informação

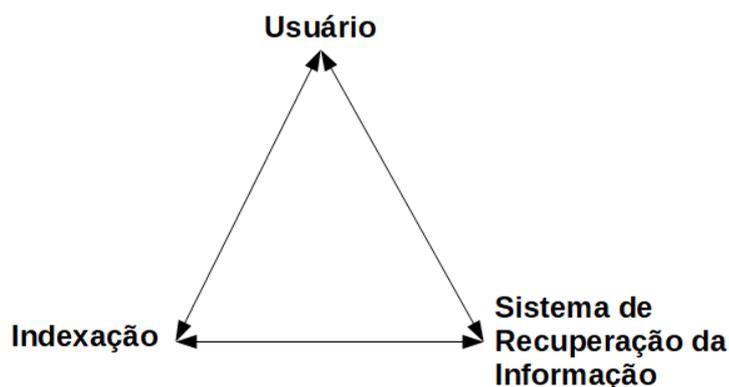


Fonte: Ferneda (2003, p. 15).

Quanto ao trato da informação, é realizada a *representação* do documento (representação documental), constituída da descrição formal dos documentos -descrição física e de elementos de sua identificação (representação descritiva), e da atribuição de assuntos aos documentos, a partir de processos de classificação, indexação e elaboração de resumos (representação temática). Tendo em vista o enfoque deste trabalho a indexação será apresentada mais profundamente na seção 5.5.

A necessidade informacional do *usuário* é representada por meio de sua *expressão de busca* explicitada por meio de linguagem natural ou artificial, e deve resultar na recuperação de documentos que preencham a lacuna informal desse usuário. Ferneda (2003) atenta ao fato de que a principal dificuldade do usuário está em prever através de termos e expressões, por meio da expressão de busca, os termos utilizados na representação dos documentos que satisfarão sua necessidade.

Figura 10 - Representação da ciclicidade do processo de busca e recuperação da informação



Fonte: a autora (2018).

Embora o processo de busca e recuperação da informação seja subjetivo, a responsabilidade de sua eficácia não deve ser em sua totalidade do usuário, mas também do indexador, além de ter também como fator determinante o sistema de recuperação da informação, conforme demonstrado na figura acima.

5.5 Indexação

A indexação pode ser conceituada como a representação do conteúdo temático de um documento seja pela extração de termos do mesmo ou a partir de linguagens documentárias. Trata-se de uma operação a qual representa um documento de acordo com os conceitos nele contidos.

Lancaster (2004) afirma que a redação de resumos e a indexação são atividades intimamente ligadas, uma vez que são partes constituintes da representação do conteúdo temático de um documento, além de ambas atividades terem por finalidade a síntese de seu conteúdo. Sobre a redação de resumos, o autor discorre:

Quanto mais informações são apresentadas, mais claramente a representação revela o alcance do artigo [livro, demais materiais], tornando-se mais provável que venha a indicar para o leitor se esse artigo satisfaz ou não a uma necessidade de informação. (LANCASTER, 2004, p. 7)

A redação de resumos é de extrema importância sobretudo diante da possibilidade do indexador, na atribuição de termos, não representar, de fato, o conteúdo do documento em sua totalidade, mas parcialmente, ou utilizar de termos mais genéricos em detrimento de termos específicos, e vice-versa, interferindo na recuperabilidade de documentos mais precisos em relação à expressão de busca do usuário.

Já Chaumier (1988) acredita que dentre as partes da análise documentária, a indexação é a mais importante. Caracteriza-se em duas etapas: análise conceitual e tradução.

Na **análise conceitual**, conforme a própria nomenclatura, analisa-se e identifica-se o assunto do documento. Ainda tratando da análise de assunto, Lancaster (2004) acredita que esse processo vai além, levando em conta

principalmente o provável interesse de determinado grupo de usuários. O mesmo documento pode ser indexado de forma diferente em diferentes bibliotecas e/ou centros de documentação, e por diferentes indexadores; a variável nesse processo, são os usuários e suas diferentes necessidades informacionais.

Bates (1998, p. 1187 *apud* LANCASTER, 2004, p.10-11) afirma que

[...] o desafio para o indexador é tentar antecipar quais os termos que as pessoas que possuem lacunas de informação de vários tipos procurariam nos casos em que o registros de que dispõem, de fato, fica a meio caminho de satisfazer a necessidade de informação do usuário.

Ao antecipar possíveis termos de forma a preencher lacunas informacionais dos usuários, no contexto dos sistemas de recomendação, torna-se possível a geração de recomendações precisas com base no perfil desses usuários.

É sabido que com o avanço das áreas no decorrer do tempo, novos termos e conceitos são incorporados, tornando necessária a revisão das representações [no âmbito temático], mas, ainda assim, o indexador ao tratar da análise temática de um documento deve realizar o prognóstico quanto ao possível uso do documento na posteridade. Lancaster (2008, p. 14) aponta que os conceitos de atinência e relevância relacionam-se, uma vez que o primeiro refere-se a “[...] relação entre um documento e um enunciado de necessidade de informação/uma consulta”, e que um documento torna-se relevante à medida que se aproxima da necessidade informacional do usuário. Lancaster (2004, p. 17) ainda aponta que

Se aceitarmos que a indexação é mais eficiente quando se orienta para as necessidades de determinado grupo de usuários, a função do indexador será prever os tipos de pedidos para os quais determinado documento será provavelmente uma resposta útil. Talvez isso ainda seja mais difícil do que prever quais os tipos de documentos que têm probabilidade de corresponder de modo útil a determinado pedido, o que constitui, em certo sentido a função de quem faz a busca.

De fato, o trabalho do indexador é de extrema importância, uma vez que essa operação é o início e fim de todo o processo de recuperação da informação-não excluindo a carga de importância do usuário ao determinar a expressão de busca

mais adequada a sua necessidade informacional-. Com relação ao trecho supracitado, mais especificamente à última sentença, no contexto de bibliotecas as quais utilizam da abordagem colaborativa para a geração de recomendações em seu catálogo, o processo da previsão de possíveis documentos correspondentes ao pedido/busca dá-se de forma ainda mais eficiente por meio desse sistema de recomendação.

Já a **tradução**, caracteriza-se como “[...] a conversão da análise conceitual de um documento num determinado **conjunto de termos de indexação.**” (LANCASTER, 2004, p. 18, grifo nosso). No processo de tradução, pode-se fazer a distinção entre indexação por extração (ou derivada) e indexação por atribuição. Na primeira, são selecionados termos em linguagem natural, do próprio documento, já na segunda, conforme a própria nomenclatura, são atribuídos termos de uma fonte que não o documento em questão¹².

Sendo assim, conforme Araújo Júnior (2007, p. 23) “O armazenamento da informação, só pode ser realizado com efetividade, se a indexação for feita de modo satisfatório, ou seja, representando com fidedignidade o conteúdo dos documentos”. A indexação pode ser visualizada sob duas dimensões: da exaustividade e da especificidade.

A exaustividade consiste na representação exaustiva do conteúdo temático do documento, de acordo com dois níveis: a) exaustivo - maior quantidade de termos indexados, de forma a abranger de forma completa o conteúdo temático do documento; b) seletiva - menor quantidade de termos indexados a fim de abranger apenas o conteúdo temático principal do documento. Já o princípio da especificidade, um dos mais importantes na indexação de assuntos, indica que um documento deve ser indexado com o termo mais específico que o represente significativamente.

Tais dimensões interferem diretamente na recuperação da informação quanto à revocação e precisão, e devem ser estabelecidas por meio de uma política, levando em conta principalmente a tipologia da biblioteca e a comunidade usuária. A revocação tende a ser maior quando há exaustividade na indexação, da mesma

¹² Mais frequentemente vocabulários controlados -lista de termos autorizados-, como listas de cabeçalhos de assunto, tesouros e esquemas de classificação bibliográfica.

forma que a precisão tende a ser maior quando indexação seletiva. A indexação seja manual¹³ ou automática, a qual será trabalhada mais profundamente no tópico a seguir, também constitui fator de influência na recuperação da informação.

5.5.1 Indexação automática

A indexação automática pode ser conceituada como qualquer procedimento de identificação e seleção de termos representativos de determinado documento, sem a intervenção direta do homem, realizada apenas com o apoio de computadores. Lancaster (2004) divide a indexação automática em dois tipos: indexação por extração automática e indexação por atribuição automática.

Indexação por extração automática caracteriza-se pela extração de palavras ou expressões com maior frequência no texto; Araújo Júnior (2007, p. 25) afirma que “Nesta tarefa são naturalmente escolhidas palavras com maior frequência ocorrem no texto, sua posição no título, no resumo e o contexto em que aparecem”.

Lancaster (2004) pontua que esses programas podem, ao invés de selecionar palavras e expressões, selecionar radicais; o autor também menciona que os critérios de frequência podem ser complementados com outros critérios, e menciona que

Uma evidente desvantagem do emprego da frequência de palavras simples ou expressões para a seleção de termos está em que, mesmo depois de usar uma lista de palavras proibidas, algumas das palavras que ocorrem frequentemente num documento podem não ser bons discriminantes [...] porque também ocorrem com frequência na base de dados como um todo. (LANCASTER, 2004, p. 287)

A *indexação por atribuição automática* indica que para cada termo atribuído, deve-se criar um ‘perfil’, ou seja, um conjunto de palavras ou expressões que ocorrem frequentemente nos documentos aos quais um indexador humano atribuiria esse termo.

¹³ “A indexação manual ou intelectual consiste na atribuição de termos de indexação ou códigos de indexação realizada por um ser humano. Estes termos serão selecionados e atribuídos por indexadores com base no julgamento subjetivo realizado acerca do conteúdo do documento, ou escolhem termos que tenham probabilidade de virem a ser procurados por um usuário no futuro.” (ROWLEY, 2002 *apud* ARAÚJO JÚNIOR, 2007, p.24)

Araújo Júnior (2007) aponta que a utilização das duas abordagens (humana e automática) seria o ideal para um sistema de recuperação da informação, uma vez que se complementam e tornam o processo de recuperação da informação mais exaustivo e eficiente, propiciando resultados mais relevantes aos usuários. Para Ferneda (2003) a eficiência de um sistema de recuperação da informação está no modelo utilizado. Em seu trabalho o autor indica alguns dos modelos existentes -explicitados no quadro abaixo - e afirma que embora desenvolvidos nas décadas de 60 e 70, e aperfeiçoados na década de 80, muitos deles estão presentes em sistemas de recuperação atuais e mecanismos de busca da web.

Quadro 9 - Modelos quantitativos de recuperação da informação

MODELO BOOLEANO
<p>“Em um sistema booleano, o conteúdo informacional dos documentos é representado por um conjunto de termos de indexação. As buscas são formuladas por meio de uma expressão booleana composta por termos ligados através dos operadores lógicos (AND, OR e NOT). O resultado de uma busca é composto por um conjunto de documentos cuja representação satisfaz às restrições lógicas da expressão de busca.” (SILVA; SANTOS; FERNEDA, 2013, p. 29-30)</p>
MODELO VETORIAL
<p>Modelo baseado na comparação parcial entre a representação dos documentos e a expressão de busca do usuário, por meio da atribuição de pesos tanto aos termos da expressão de busca como aos termos de indexação que representam os documentos. Os pesos variam entre 0 e 1 - quanto mais próximo de 1, maior sua relevância-. “O resultado de uma busca é um conjunto de documentos ordenados pelo grau de similaridade entre a expressão busca do usuário e cada um dos documentos do corpus.” (SILVA; SANTOS; FERNEDA, 2013, p. 31)</p>
MODELO PROBABILÍSTICO
<p>Baseado na teoria matemática das probabilidades. Esse modelo supõe que determinado conjunto de documentos atenda totalmente às necessidades informacionais expressas por meio de expressões de busca no sistema. Sendo assim, “A partir do primeiro conjunto de documentos resultantes de uma busca, o usuário seleciona alguns que considera relevantes para responder à sua necessidade de informação. A expressão de busca, juntamente com os documentos que foram selecionados como relevantes, é submetida novamente ao sistema de informação, procurando refinar a busca e tentando aproximar-se cada vez mais do conjunto ideal de documentos.” (SILVA; SANTOS; FERNEDA, 2013, p. 31)</p>

MODELO FUZZY

Modelo baseado na lógica *fuzzy*. Souza (2006, p. 166 *apud* SILVA; SANTOS; FERNEDA, 2013, p. 32, grifo nosso) afirma sobre o modelo: “[...] busca-se estender o conceito da representação dos documentos por palavras-chave, assumindo que cada *query* [expressão de busca] determina um conjunto difuso e que cada documento possui um grau de pertencimento a esse conjunto, usualmente menor do que 1. O grau de pertencimento pode ser determinado pela ocorrência de palavras expressas na *query*, tal como no modelo booleano, mas pode também utilizar um instrumento – como um **tesauro** – para determinar que termos relacionados semanticamente aos termos índice também confirmam algum grau de pertencimento ao conjunto difuso determinado pela *query*.” (SILVA; SANTOS; FERNEDA, 2013, p.32)

MODELO BOOLEANO ESTENDIDO

“O modelo booleano estendido tenta contornar as limitações do modelo vetorial e do modelo booleano clássico através de uma conceituação matemática mais genérica. [...] Essa generalização é feita através da introdução de dois novos parâmetros em relação ao modelo booleano tradicional: os pesos associados aos termos da expressão de busca e o parâmetro p associado a cada operador booleano.” (FERNEDA, 2003, p. 53)

Fonte: Adaptado de Silva, Santos e Ferneda (2013) e Ferneda (2003).

Além dos modelos quantitativos, Ferneda (2003) discorre, também, acerca dos modelos dinâmicos, os quais têm como principal característica o reconhecimento da importância do usuário na definição das representações dos documentos.

O autor, sobre esse modelo, ainda afirma que

Nessa ótica os usuários interagem e inferem diretamente na representação dos documentos do *corpus*, permitindo uma evolução ou uma adaptação dos documentos aos interesses dos usuários do sistema, percebido através de suas buscas e da atribuição de relevância (e não relevância) aos documentos recuperados (*relevance feedback*). (FERNEDA, 2003, p. 55)

Quadro 10 - Modelos dinâmicos de recuperação da informação

SISTEMAS ESPECIALISTAS NA RI

“[...] o processo de construção automática da base de conhecimento é realizado através da identificação dos principais conceitos contidos nos textos dos documentos do *corpus*. Esses conceitos são identificados utilizando-se cálculos estatísticos de co-ocorrência de pares de palavras. [...] O resultado desse

processo é um conjunto de conceitos representados por grupos de palavras que caracterizam uma ideia contida nos documentos do *corpus*. Esses conceitos são integrados à rede semântica que compõe a base de conhecimento. Essa rede semântica é utilizada para melhorar a eficiência do sistema e auxiliar o usuário na formulação de suas buscas. Para cada novo documento inserido no *corpus* altera-se a configuração da rede semântica.” (FERNEDA, 2003, p. 61)

REDES NEURAIS NA RI

Os modelos baseados em redes neurais têm como objetivo a simulação do sistema nervoso humano em sistemas de recuperação da informação. Silva, Santos e Ferneda (2013, p. 33, grifo nosso) informam que “[...] as redes neurais se utilizam de **padrões** para relacionar as expressões de busca dos usuários com os documentos de um acervo, de modo que cada expressão de busca libera um sinal que ativa os termos do sistema e que se propaga aos documentos relacionados. Tais estímulos retornam os sinais a novos termos, em **interações sucessivas**. As respostas apresentadas ao usuário são definidas por meio desse processo, que podem conter até termos que não foram utilizados na busca, mas que demonstraram ter relação com a expressão pesquisada.”

ALGORITMOS GENÉTICOS NA RI

Os algoritmos genéticos possuem sua base na Biologia -genética-, como uma tentativa de representação matemática da Teoria da Evolução das Espécies. Trata-se do uso de técnicas para a simulação do processo natural da evolução a fim de solucionar determinado problema. “Cada vez que o algoritmo se repete em um processo, são criadas novas estruturas por meio da troca de informações, de modo que as próximas “gerações” sejam cada vez mais aptas a resolver os problemas de uma dada situação.” (SILVA; SANTOS; FERNEDA, 2013, p.33-34). A apresentação de resultados relevantes dá-se em detrimento do nível de interação do usuário com o sistema de busca.

Fonte: Adaptado de Silva, Santos e Ferneda (2013) e Ferneda (2003).

Diante disso, a utilização de sistemas de recomendação, baseados em FC, em bibliotecas propiciaria maior precisão na recuperação da informação, não excluindo a qualidade da indexação, uma vez que recomenda ao usuário itens, de forma personalizada, com base em suas preferências.

5.5.2 Indexação na web

A *web*, ou rede mundial de computadores, pode ser definida como “[...] sistema de documentos em hipermídia que estão interligados e executados na internet.” (WIKIPEDIA)¹⁴. Sobre a *web*, Fernalda (2003, p. 92) menciona que

A *web* é formada por um conjunto de unidades de informação chamadas “páginas”. Uma página é um arquivo de computador cujo tamanho (quantidade de caracteres) pode variar desde o tamanho de uma página de um livro até o tamanho de um livro inteiro.

O autor aponta que dentre as características comuns entre as páginas tem-se a URL (*Universal Resource Locator*) -esquema de endereçamento-, o protocolo *http* (*Hypertext Transfer Protocol*) -protocolo de comunicação base para páginas da *web*-, e a linguagem de marcação HTML (*Hypertext Markup Language*).

Sob a perspectiva da recuperação da informação, segundo Lancaster (2004), os recursos informacionais acessíveis na *web* diferem dos registros bibliográficos de sistemas tradicionais, principalmente por conterem “apontadores” para outros sítios -vínculos de hipertexto¹⁵. No entanto, assemelham-se na medida em que os sítios *web* têm elementos pesquisáveis semelhantes aos presentes em bases de dados, por exemplo.

Traçando um paralelo ao contexto tecnológico atual, os “apontadores” mencionados anteriormente, sob o contexto das bibliotecas, OPACs e registros bibliográficos, podem ser associados às recomendações geradas por SRs; divergindo apenas devido aos apontadores (vínculos de hipertexto) remeterem a conteúdos externos a página de origem, enquanto que SRs, em bibliotecas, remetem a conteúdos “internos”.

Ainda tratando da Rede, o conjunto de informações armazenadas sobre suas páginas formam a base de dados do mecanismo de busca. Lancaster (2004, p. 340) pontua:

Em termos de quais páginas serão realmente recuperadas por uma consulta, a indexação pode até ser mais crítica do que o processo de *crawlers*. O programa de indexação examina as informações armazenadas na base de dados e cria as entradas apropriadas no

¹⁴ Disponível em: <https://pt.wikipedia.org/wiki/World_Wide_Web>. Acesso em: 13 dez. 2018.

¹⁵ “Vínculos de hipertexto (isto é, buscar páginas vinculadas a determinado URL)” (LANCASTER, 2004, p. 342)

índice. Quando se submete uma consulta, é esse índice que é usado a fim de identificar registros bibliográficos.

Os locais de título e URL são vistos como campos de “alto valor” e indexados, no entanto, por vezes não são indexadas as *metatags*, palavras e/ou expressões inseridas em determinada parte do código HTML, as quais têm por fim a descrição do conteúdo da página.

Quanto aos mecanismos de busca e recuperação na *web*, Lancaster aponta 8 fatores a serem observados, conforme o quadro abaixo:

Quadro 11 - Fatores relacionados a busca e recuperação da informação na *web*

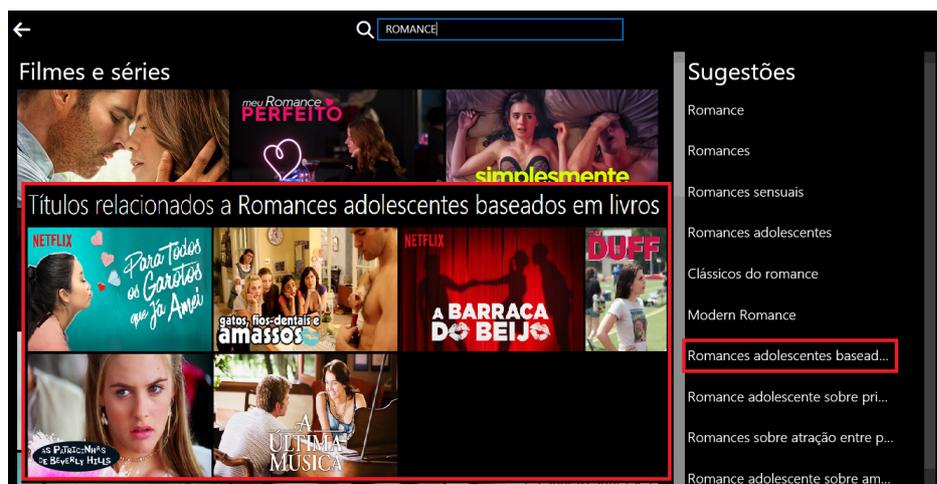
Quanto a busca:	Quanto a recuperação:
Lógica Booleana, inclusive recursos de encaixamento [<i>nesting</i>]	Frequência de ocorrência de termos de busca no registro
Truncamento	Número de coincidência de termos
Buscas com expressões	Localização do termo
Proximidade de palavras	Raridade
Buscas em campos (isto é, poder limitar a busca a um campo especificado no registro, como título ou URL)	Proximidade
Vínculos de Hipertexto (isto é, buscar páginas vinculadas a determinado URL)	Ordem dos termos
Busca em imagens (capacidade de procurar páginas que contenham apenas imagens)	Data
Consulta por exemplo (capacidade de encontrar registros semelhantes a um registro já conhecido como interessante)	Popularidade

Fonte: Extraído e adaptado de Lancaster (2004, p. 341-342)

Tendo em vista a possibilidade de se analisar a entrada (indexação), por meio da saída (recuperação da informação), para melhor representar a indexação em

sistemas de recomendação, realizou-se buscas em duas das plataformas apresentadas anteriormente na seção 5.1.1 (Netflix e Spotify).

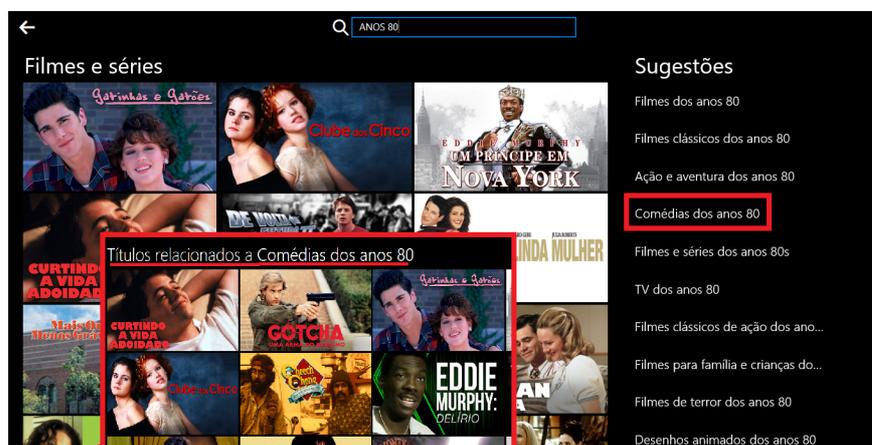
Figura 11 - Busca sob o termo “Romance” (Netflix)



Fonte: a autora (2018).

Ao realizar a busca sob o termo “Romance”, são recuperados títulos relacionados, apresentados em miniatura; no canto direito da tela, são sugeridas subcategorias do gênero “Romance”. Ao selecionar a subcategoria “Romances adolescentes baseados em livros” foram recuperados 6 itens. Para fins de testes, uma nova busca realizada sob o termo “Noah Centineo” - ator principal de um dos 6 títulos, **inseridos na categoria “Romance”**, recuperados anteriormente, “Para todos os garotos que já amei”. Observou-se que não foram recuperados apenas títulos os quais o ator esteve no elenco, mas, também, títulos cuja temática assemelha-se a dos títulos atuados pelo ator.

Figura 12 - Busca sob o termo “Anos 80” (Netflix)

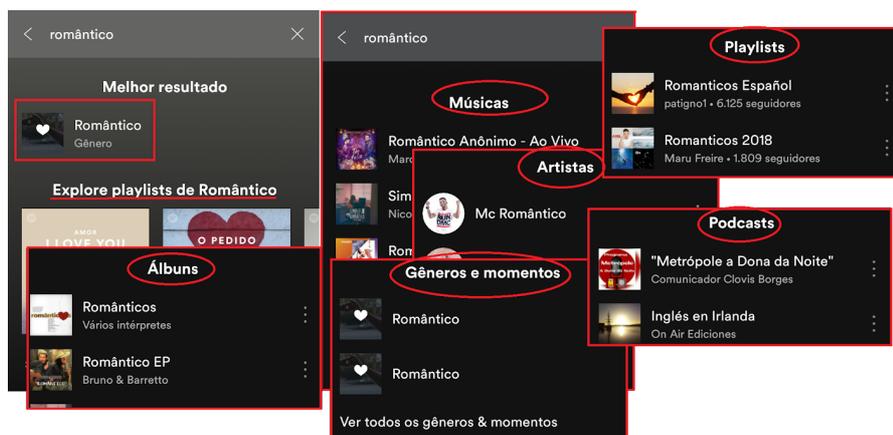


Fonte: a autora (2018).

Uma nova busca foi realizada sob o termo “anos 80”, recuperando uma listagem de itens apresentados em miniatura; novamente, no canto direito da tela foram sugeridas categorias relacionadas aos “Anos 80”. Foi selecionada, então, a categoria “Comédias dos anos 80”, recuperando diversos itens relacionados.

Além da busca por assunto e título, a busca pelo nome do ator e direção também é possível nesse serviço de *streaming*. Além disso, quando realizada busca de um item não incluso no catálogo, recupera-se apenas títulos relacionados.

Figura 13 - Busca sob o termo “Romântico” (Spotify)

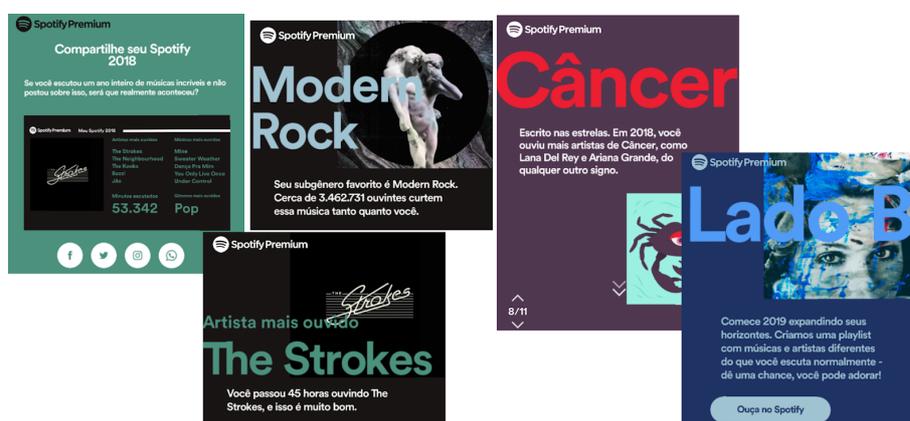


Fonte: a autora (2018).

Já no Spotify, ao realizar a busca sob o termo “Romântico”, recupera-se a categoria gênero *Romântico*, a qual agrupa *playlists*, criadas pelo Spotify,

relacionadas ao gênero. Mais abaixo são agrupadas músicas cujo título, ou diferentes metadados representativos, contenham o termo buscado, artistas cujo nome contenha o termo buscado, o gênero que nos remete às *playlists* relacionadas, podcasts relacionados, álbuns cujo termo buscado ocorre no título, e playlists - aqui, playlists criadas pelos usuários da plataforma, cujo título [da playlist] contenha o termo buscado.

Figura 14 - Estatísticas personalizadas de uso (Spotify).



Fonte: a autora (2018).

Outra questão importante a ser mencionada com relação ao Spotify é o fato de anualmente serem lançadas estatísticas de uso para os usuários que possuem assinatura da plataforma. Na primeira imagem à esquerda são apresentadas as músicas e artistas mais ouvidos no ano de 2018; na segunda imagem a plataforma apresenta o gênero (classe) ou subgênero (subclasse) mais escutado pelo usuário - neste caso, subgênero (“Modern Rock”) do gênero (“Rock”), além de fornecer na descrição a quantidade de ouvintes do subgênero referido-; na terceira imagem, abaixo, é apresentado o artista mais ouvido seguido da quantidade de horas que o usuário o ouviu; na quarta imagem, o sistema utiliza dos dados dos artistas escutados, como por exemplo a data de nascimento, para aferição do signo de maior peso dentre os artistas mais ouvidos pelo usuário; por fim, a quinta imagem representa a recomendação personalizada a cada usuário - da playlist “Lado B” com músicas e artistas diferentes dos escutados normalmente pelo usuário.

5.6 Cenário dos sistemas de recomendação em bibliotecas

Atualmente sistemas de recomendação têm sido utilizados para além de seu contexto inicial, o segmento de *e-commerce*, em bibliotecas (físicas ou digitais).

Com relação às bibliotecas digitais (BDs), Lu et. al (2015, p. 19) mencionam que

são coleções de objetos digitais, juntamente com os serviços associados fornecidos às comunidades de usuários. Os sistemas de recomendação podem ser usados em aplicativos de biblioteca digital para ajudar os usuários a localizar e selecionar fontes de informação e conhecimento.

Lopes, Souto e Oliveira (2006) afirmam que BDs no contexto da Web Semântica propiciam um acesso inteligente e eficiente a documentos digitais disponíveis na *web*. Além disso, os autores apresentam duas vantagens quanto ao uso de metadados para a descrição de objetos de informação: a) maior eficiência durante a coleta de informação para a geração de recomendação; e b) possibilidade de interoperabilidade entre BDs. Em seu trabalho, os autores descrevem um SR de artigos científicos da área da Computação armazenados em bibliotecas digitais, de acordo com os interesses dos usuários identificados a partir de informações do currículo Lattes.

Neste trabalho, os autores utilizam dentre os modelos de indexação, o modelo vetorial. Com relação ao módulo de recomendação são utilizadas informações referentes ao título e palavras-chave contidas no Lattes, bem como a produção bibliográfica do usuário para a determinação do vetor dos termos de busca; quanto às palavras presentes no título, são eliminadas as *stopwords*. A recomendação considera apenas os idiomas em que o usuário possui proficiência e dá preferência às publicações mais recentes e cursos de formação acadêmica em andamento. Com relação à atribuição de pesos: a) termos obtidos em palavras-chave têm peso maior que termos obtidos no título; b) o peso dos termos de acordo com o idioma é atribuído com base no nível de proficiência do usuário; e c) peso maior para publicações mais recentes e cursos de formação acadêmica em andamento.

“Após determinado o vetor de busca, é necessário calcular os pesos dos termos que compõe a expressão de busca em cada um dos documentos que possivelmente serão recomendados, formando assim os vetores dos documentos” (LOPES; SOUTO; OLIVEIRA, 2006, p.[6]). Adota-se no trabalho a abordagem (tf X idf) (*the product of the term frequency and the inverse document frequency*), utilizada para a atribuição automática de pesos aos termos para recuperação de textos; esta abordagem “[...] sugere que os melhores termos para identificação do conteúdo são aqueles capazes de distinguir certos documentos individuais do restante da coleção” (*op.cit*, 2006, p.[6]).

Por fim, foi solicitado a determinado grupo de indivíduos (discentes e docentes da Pós-Graduação do Instituto de Informática da UFRGS) uma avaliação do sistema desenvolvido. Foram geradas 20 recomendações a cada indivíduo. Como resultado, os autores observaram que o sistema identifica adequadamente o perfil dos usuários, uma vez que recomenda ao autor sua própria publicação -contexto de interesse; como ponto negativo, os autores mencionam que a BD utilizada não contempla todas as áreas da Computação, influenciando na qualidade da indexação. Além disso, ele atenta ao fato de muitos autores terem mudado de área, qualificando negativamente, durante a avaliação, publicações que já lhe interessaram a *posteriori*. Outro ponto levantado é a descrição, realizada pelos próprios usuários, em seu currículo Lattes, “[...] mesmo contendo várias palavras-chave utilizadas pelos autores para descreverem suas publicações no Lattes, podem não ter gerado uma boa recomendação se os contextos forem distintos.” (LOPES; SOUTO; OLIVEIRA, 2006, p. [8]). Isso demonstra a importância e necessidade de se levar em conta a qualidade da indexação, seja manual ou automatizada, uma vez que interfere diretamente no processo de recuperação da informação, incluindo a recomendação.

Wakeling, Clough e Sen (2012) traçam um estudo acerca do uso de recomendações em OPACS de bibliotecas. Foram analisados catálogos online de 211 bibliotecas públicas e 118 bibliotecas universitárias no Reino Unido. Os autores concluem que o uso desses serviços em bibliotecas é muito baixo, com apenas 2% das bibliotecas públicas e 11% das bibliotecas universitárias oferecendo esse recurso. Eles ainda afirmam que

Embora as limitações do sistema e as restrições orçamentárias sejam talvez parcialmente responsáveis, sugere-se que os profissionais da biblioteca talvez ainda não tenham sido persuadidos de que o valor das recomendações para os usuários da biblioteca é grande o suficiente para garantir que sua inclusão se torne uma prioridade. (WAKELING; CLOUGH; SEN, 2012, p. 134)

Recursos como marcações, revisões e recomendações oferecem às bibliotecas um meio de engajar-se com seus usuários, proporcionando-os melhor experiência com o sistema, além disso permitem complementar os meios tradicionais de geração de metadados, enriquecendo seu catálogo.

Outros autores cujo trabalho retratou o uso de SR em OPACs de bibliotecas foram Yang e Hofmann (2011 apud WAKELING; CLOUGH; SEN, 2012, p. 138); foram analisadas cerca de 260 bibliotecas acadêmicas na América do Norte. Os autores concluíram que cerca de 34% dos OPACs analisados possuem “[...] alguma forma de “linguagem de recomendação”, embora sua definição de tal linguagem fosse ampla, incluindo funções como ‘encontre mais por este autor’ e ‘itens próximos na prateleira”.

Dentre os exemplos citados por Wakeling, Clough e Sen (2012), tem-se o *Library Thing for Libraries (LTFL)*¹⁶,

O Library Thing é um serviço on-line que permite que os membros cataloguem suas coleções de livros e complementem este catálogo com avaliações, resenhas e tags. Com mais de um milhão de membros, isso representa uma quantidade significativa de feedback explícito para recomendações de filtragem colaborativa. **O LTFL é vendido para bibliotecas como uma sobreposição do OPAC**, usando um pequeno JavaScript para consultar o banco de dados do Library Thing para o ISBN do item que está sendo visualizado no OPAC. (op. cit, 2012, p. 140, grifo nosso).

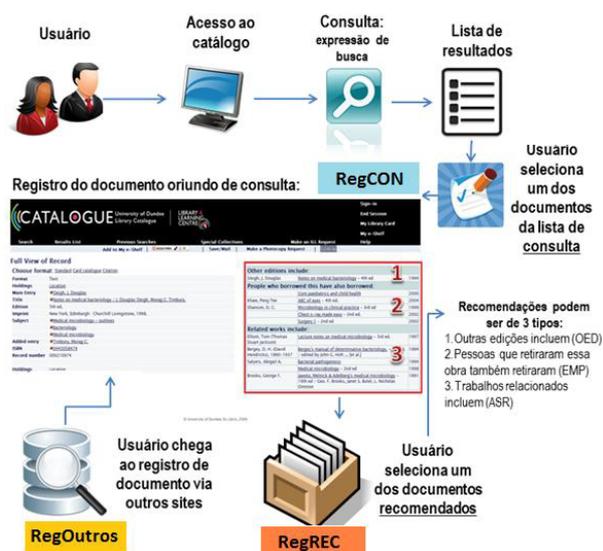
Portanto, trata-se de um serviço de enriquecimento de OPACs fornecendo melhor experiência ao usuário no que tange a navegação, pesquisa e interação com as coleções da biblioteca. Dentre os produtos da LTFL têm-se: a recomendação de livros semelhantes - possui seções adicionais como “mais por este autor”, “autores semelhantes”, “recomendações dos leitores”, *tag* de navegação - permite a pesquisa e navegação entre os itens por meio de *tags*, outras edições e traduções -permite

¹⁶ Disponível em: <<https://www.librarything.com/forlibraries/>>.

vínculo entre edições relacionadas e traduções do mesmo trabalho, comentários, navegação de prateleira - simulação de um mini-navegador de prateleira de livros, *widgets* de exibição de livros - exibição de livros virtuais na página inicial da biblioteca, e *BookPsychic* -“bola de cristal dos clientes”, quanto maior a interação do usuário com os itens, melhores serão as recomendações; integra-se a redes sociais como Facebook e Twitter.

O uso de SR em catálogos online de bibliotecas também foi abordado por Krebs, Rocha e Ribeiro (2017) em seu trabalho sobre a biblioteca da Universidade de Dundee (Escócia), “[...] primeira biblioteca a implementar a ferramenta de recomendação *Related Books in Aleph OPAC*, especialmente criada para funcionar no sistema Aleph.” (op. cit., 2017, p. 155). O estudo deu-se utilizando da análise dos *logs*¹⁷ dos usuários no catálogo da biblioteca. O conjunto de interações realizadas entre o usuário e o catálogo online, em determinado período de tempo, é chamado de sessão; e as informações de sessão são contidas em *posts*.

Figura 15 - Esquema de navegação do usuário no catálogo



Fonte: Krebs, Rocha e Ribeiro (2017, p. 156).

Após consultar o catálogo, o usuário recupera e seleciona o item de interesse; no registro do documento da consulta são apresentados 3 tipos de recomendação

¹⁷ “[...] log de dados é uma expressão utilizada para descrever o processo de registro de eventos relevantes num sistema computacional [...] Para restabelecer o estado original de um sistema ou para que um administrador conheça o seu comportamento no passado.” (WIKIPEDIA). Disponível em: <https://pt.wikipedia.org/wiki/Log_de_dados>. Acesso em: 18 nov. 2018.

(tendo em vista as edições [dos itens], os empréstimos e os assuntos). O usuário também pode chegar a determinado registro de documento por meio de links externos.

Os autores concluem que, em sessões com recomendação, os usuários tendem a visitar mais de um documento por sessão, incentivando a serendipidade. Além disso, eles também mencionam que tanto a recomendação com base no histórico de empréstimo (*“usuários cujo item foi emprestado também retiraram...”*), quanto a recomendação com base no assunto (itens os quais possuem pelo menos três cabeçalhos de assunto ou três números de classificação similares), são igualmente utilizadas no sistema estudado, e finalizam apontando que à medida em que os usuários exigem precisão e personalização em serviços de busca e recuperação da informação, os SR aplicados a OPACs podem tornar-se uma realidade em bibliotecas, satisfazendo as necessidades dos usuários diante do cenário atual.

Já Casagrande (2014), em seu trabalho, busca a implementação de um SR com base na Biblioteca Digital de Literatura Brasileira (BD-LB). Para a descrição dos conteúdos na BD-LB tem-se o conjunto de metadados descritores, a saber: título, autores, pseudônimo, tipo [para além de obra literária, como nota de jornal, publicação na imprensa etc], gênero, ano, localização, descrição, data de inclusão, data de atualização, tipo do documento, idioma e editora.

Através do uso de valores de metadados que descrevem as obras disponíveis, a BD-LB possibilita a construção do perfil de acessos e preferências do usuário [permitindo uma recuperação mais centrada nos interesses dos usuários]. [...] A BD-LB considera que cada conteúdo acessado contribui para a modelagem do perfil do usuário. Com isso, é assumido que o acesso a um conteúdo é um indicativo de interesse por tal conteúdo. Porém, um possível problema se dá pelo acesso precipitado, onde o usuário, após o acesso a determinado conteúdo, percebe que ele não é de seu interesse. Para reduzir este problema, a BD-LB não oferece acesso direto ao conteúdo após o clique, mas sim, uma descrição da obra inicialmente. (CASAGRANDE, 2014, p. 34, 36).

Uma das propostas do trabalho é a diminuição do tempo de processamento dos algoritmos de recomendação [problema da escalabilidade], e os autores utilizam

da técnica de agrupamento ou *clusterização* de itens e/ou usuários, diminuindo a quantidade de informações.

[...] algoritmos de clusterização são realizados no conjunto de itens, e a **similaridade entre itens é usada para classificá-los e inseri-los nos clusters**. Tais características dos itens a serem comparadas são baseadas nas avaliações que foram realizadas aos itens pelos usuários. Com os itens separados em clusters, **cada cluster comporta tipos semelhantes de itens focados num único tópico** [princípio lógico da mútua exclusividade], facilitando a predição de avaliações dos usuários aos itens não avaliados. Uma vez realizada a clusterização, uma técnica tradicional de FC é aplicada para recomendação de itens. Com o escopo da busca por itens similares é reduzida graças à clusterização, a FC aplicada sobre os itens terá uma dimensionalidade reduzida. (CASAGRANDE, 2014, p. 51, grifo nosso)

Sendo assim, o processo de recomendação proposta pelo autor constitui-se de 4 fases: a) a construção dos perfis dos usuários, b) o processo de agrupamento de usuários, c) a determinação dos vizinhos mais próximos em cada grupo, e d) construção da recomendação (lista ordenada de conteúdos recomendados para o usuário).

Quanto à *construção do perfil do usuário*, ao realizar cadastro no sistema é possível identificar suas preferências. Posteriormente, a coleta de dados para manutenção do perfil é realizada de forma *implícita*, a partir da frequência de ocorrência dos valores de um subconjunto dos elementos de metadados que ocorrem nos conteúdos acessados pelos usuários.

Os grupos são identificados e criados por valores de metadados (grupos de determinado autor; grupos de determinado gênero etc), haja vista a frequência de ocorrência do valor de determinado metadado observado no conjunto de conteúdos anteriormente acessados por esse usuário. Como exemplo, se o usuário **X** tem determinado número de acesso em obras de *Machado de Assis*, ele será incluído no cluster *Machado de Assis*, da mesma forma que se o usuário **Y** tem determinado número de acesso em obras do gênero *Romance*, esse usuário será incluído no cluster *Romance*.

Na segunda fase, *processo de agrupamento dos usuários*, o agrupamento ocorre a partir da análise dos perfis dos usuários.

A determinação dos grupos a que um determinado usuário será adicionado depende dos pesos de preferência de determinados valores de MOs. Quando um determinado valor de MO atingir um limiar de peso de preferência, o usuário será incluído no grupo deste valor. Este limiar é chamado de limiar de agrupamento. (CASAGRANDE, 2014, p. 59)

Conforme exposto anteriormente, os usuários serão agrupados em *clusters* haja vista a frequência de ocorrência do valor de determinado metadado observado no conjunto de conteúdos anteriormente acessados ele.

A *formação da vizinhança* se dá a partir da análise dos perfis dos usuários contidos nos mesmos *clusters*, sendo considerados vizinhos os usuários que possuem alto grau de similaridade de acordo com cálculos estatísticos. Depois de identificada a similaridade entre os usuários, segue-se a fase da *construção da recomendação*.

A última fase, em que a recomendação é gerada, ocorre a partir de um conjunto de conteúdos acessados pelo usuário foco; são geradas listas de conteúdos previamente classificados/acessados pela vizinhança, mas desconhecidos pelo usuário foco. A ordenação da lista de recomendações baseia-se no peso de preferência, ou seja, conteúdos que tiverem maior probabilidade de serem relevantes ao usuário foco, devem aparecer primeiro.

6 ANÁLISE E DISCUSSÃO DOS RESULTADOS

Esta seção tem como objetivo discutir os conceitos apresentados acerca dos sistemas de recomendação baseados em filtragem colaborativa em paralelo ao contexto das bibliotecas especializadas. O diferencial dessa tipologia de biblioteca e característica a qual aproxima-a da Filtragem Colaborativa está no perfil especializado de seus usuários.

A abordagem alternativa do estudo de usuários nos remete ao conceito Ranganathiano da espiral do conhecimento, onde “O universo do conhecimento é a soma total no momento do conhecimento acumulado. Ele está sempre em desenvolvimento contínuo”(RANGANATHAN, 1963, p. 359), visto que nessa abordagem as necessidades informacionais dos usuários são observadas sob a perspectiva de sua individualidade, levando em conta o contexto por ele inserido.

Ainda abordando a questão da individualidade, deve-se retomar a necessidade da revisão da representação dos documentos enquanto a sua temática, seja manual ou automática, a fim de acompanhar os avanços no âmbito científico, bem como adequar-se aos seus usuários. O processo de busca e recuperação da informação é cíclico, tendo como peso a tríade (explicitada na figura 10): **usuário <> indexação <> sistema de recuperação da informação**. Ainda que os termos de indexação sejam representativos, caso o usuário não expresse, de fato, sua necessidade informacional na expressão de busca, os resultados obtidos não terão alto grau de relevância; assim como, caso o sistema de recuperação não possua recursos de recuperação, conforme exemplos explicitados no quadro 10, os quais permitam ao usuário uma busca mais precisa, os resultados obtidos podem, também, não ter um alto grau de relevância. Outro caso seria quanto à atribuição de termos não representativos aos documentos, dificultando a recuperação de informação relevante ao usuário, ainda que o sistema possua bons recursos de recuperação e tenham sido utilizados termos representativos na expressão de busca. Sendo assim, é necessário manter o equilíbrio entre essas variáveis da tríade, essenciais para uma recuperação eficiente de informação, de modo a preencher as lacunas informacionais dos usuários.

O exposto tem um peso ainda maior com relação à aplicação em bibliotecas que utilizam SR. Quando o usuário realiza a busca no sistema e avalia, dentre os itens recuperados, o mais relevante, e acessa seu conteúdo, as recomendações são geradas. Portanto, é necessário que o usuário busque e recupere itens para que lhe sejam recomendados outros com base em sua vizinhança; é necessário um SRI que possibilite uma recuperação precisa, e, mais do que isso, é preciso uma indexação de qualidade. Além disso, foi possível perceber o **papel social** da filtragem colaborativa, uma vez que permite ao usuário recuperar informações potencialmente relevantes ainda que o mesmo não tenha "competência informacional"¹⁸ para tal. Outro fator a ser observado é a **participação social** propiciada pela filtragem colaborativa, em razão de sua natureza; a vizinhança é formada de acordo com as mudanças no perfil [gostos/preferências] do usuário, ou seja, quem molda as recomendações é o próprio usuário.

Diferente das bibliotecas digitais que, em sua maioria possibilita a busca por texto completo dos objetos digitais contemplados por ela, a busca em catálogos de bibliotecas físicas restringe-se aos metadados dos itens. Como exemplo, em certas bases de dados é possível realizar a busca dos termos da expressão no próprio *corpus* do documento, recuperando documentos com maior grau de relevância. É nesse contexto que a FC é inserida, e a questão da relevância pode ser solucionada por meio da serendipidade propiciada pela filtragem.

Outro ponto a ser mencionado é o fato de que a lógica classificatória empregada em processos biblioteconômicos é a mesma utilizada em soluções tecnológicas atuais. A principal etapa da Filtragem Colaborativa, também a que difere a FC das demais filtragens, é a formação da vizinhança. A medida em que as informações dos usuários são coletadas, armazenadas e analisadas pelo sistema a fim de gerar recomendações, os usuários são divididos em "grupos" ou "classes" - no âmbito biblioteconômico - a partir de suas semelhanças e diferenças, tal como se caracteriza o processo de classificação.

Sobre a criação e manutenção de perfis de usuários para personalização da recomendação na Filtragem Colaborativa, pode-se pressupor que seja um novo paradigma do estudo de usuário, em se tratando de bibliotecas. E no atual contexto

¹⁸ Optou-se por utilizar o termo "Competência Informacional" como tradução de *Information Literacy*.

tecnológico diferindo da Disseminação Seletiva da Informação, uma vez que engloba todo o acervo para gerar recomendações, além de utilizar de diversas técnicas em constante desenvolvimento para solucionar possíveis problemas ou potencializar resultados.

Saracevic (1996) ainda na década de 90 atentava sobre o papel econômico e social de toda e qualquer atividade de informação como peça chave para o desenvolvimento não somente regional, mas nacional, e progresso social, bem como aos avanços organizacionais e vantagens competitivas. Em se tratando de bibliotecas especializadas, as quais, muitas vezes, estão inseridas em ambientes corporativos, a questão informacional é crucial.

Ranganathan (2004) analisa a biblioteca como um todo e estabelece as 5 leis da Biblioteconomia, as quais serão aqui analisadas haja vista os SR baseados em Filtragem Colaborativa : **1ª lei** *Os livros são para serem usados*: conforme já mencionado, devido à recomendação personalizada, baseada em perfis com gostos similares, o sistema, muitas vezes acaba por recomendar itens desconhecidos pelo usuário, mas de extrema relevância para o mesmo em determinado momento -permitindo serendipidade; com isso, livros “pouco conhecidos”, com pouca circulação, podem ser descobertos e utilizados por usuários. **2ª lei** *Para cada leitor, seu livro* e **3ª lei** *Para cada livro, seu leitor*: explicitam claramente a questão da recomendação personalizada realizada pela filtragem colaborativa, permitindo que o usuário recupere documentos precisos. **4ª lei** *Poupe o tempo do leitor*: sabe-se que, diferente do bibliotecário da antiguidade, conhecedor máximo da biblioteca e de seu acervo, atualmente, diante da sobrecarga informacional, absorver, de fato, informação a fim de gerar conhecimento, não é tarefa fácil. Sendo assim, traçando um paralelo também a 1ª lei, esses sistemas permitem que o usuário descubra itens que supram sua necessidade, mas que talvez não venha a ser recomendado pelo bibliotecário ou recuperado pelo sistema a partir da busca. Além disso, permite praticidade ao usuário, uma vez que não precisaria realizar buscas consecutivas. **5ª lei** *A biblioteca é um organismo em crescimento*: esta lei não deve ser vista apenas da perspectiva física da biblioteca; não se trata apenas do crescimento relacionado ao espaço ou acervo, mas da necessidade das bibliotecas se adequarem e

acompanharem os avanços, sobretudo no âmbito tecnológico. Aprimorar seus produtos e serviços a fim de servir de forma eficiente a sua comunidade.

7 CONSIDERAÇÕES FINAIS

Este estudo buscou contemplar a temática da Filtragem Colaborativa e sua relação com a indexação e o estudo de usuários em bibliotecas especializadas. A Filtragem Colaborativa, além de servir de incentivo a serendipidade, gerar recomendações com base no perfil do usuário e permitir melhor interação e experiência do usuário com o sistema, funciona como ferramenta de auxílio ao estudo de usuário, bem como a indexação.

Atualmente os Sistemas de Recomendação são amplamente utilizados não só no âmbito do *e-commerce* e *e-business*, mas também em bibliotecas, sobretudo em países europeus, EUA e Canadá. Outro fator importante é a comercialização de SR para bibliotecas, como o exemplo mostrado no trabalho sobre a *Library Thing for Libraries (LTFL)*, a nível internacional.

Torna-se cada vez mais claro que os sistemas de recuperação da informação tradicionais frequentemente não conseguem conectar da melhor forma os usuários da biblioteca com o material potencialmente relevante; diferente de modelos hoje utilizados na *web*, os quais utilizam de diversos recursos de recuperação, com o objetivo de proporcionar não somente resultados de busca relevantes, mas uma boa experiência do usuário durante a navegação e/ou busca.

Para além das modificações com relação ao sistema de recuperação da informação (SRI) e todo o ciclo que envolve o processo de recuperação da informação, as bibliotecas devem apropriar-se, sobretudo, da 5ª lei, postulada por Ranganathan - A biblioteca é um organismo em crescimento - e acompanhar os avanços que dizem respeito ao seu fazer.

Os princípios e as técnicas de recuperação da informação estão sendo disseminados para aplicações além da CI, como os próprios sistemas de recomendação que, hoje, são amplamente estudados no campo da Computação, muito embora a temática perpassasse claramente a Ciência da Informação. É necessário que o profissional bibliotecário tenha ciência que seus usuários cada vez mais demandarão precisão e personalização no processo de busca e recuperação da informação, tornando urgente a utilização de ferramentas como os sistemas de recomendação (que não somente otimizam esse processo de busca e recuperação

da informação, como também possibilitam melhor experiência e interação do usuário com a biblioteca), em seu OPAC. Além disso, foi possível perceber que os sistemas de recomendação possibilitam a recuperação de informação potencialmente relevante àqueles usuários que não possuem competência informacional para tal.

Sendo assim, espera-se que este trabalho contribua para a área da Organização do Conhecimento, e que sirva de estímulo para estudos no campo da Ciência da Informação, acerca dos Sistemas de Recomendação e sua forte relação com a Biblioteconomia.

REFERÊNCIAS

- ADOMAVICIUS, G., TUZHILIN, A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. **IEEE Transactions on Knowledge and Data Engineering**, v. 17, n. 6, p. 734–749, 2005. Disponível em: <<http://pages.stern.nyu.edu/~atuzhili/pdf/TKDE-Paper-as-Printed.pdf>>. Acesso em: 30 maio 2018.
- ALVAREZ, E. B. et. al. Os Sistemas de Recomendação, Arquitetura da Informação e a Encontrabilidade da Informação. **Transinformação**. Disponível em: <<http://www.brapci.ufpr.br/brapci/index.php/article/view/0000022026/c931ba39d270aa85801ee5c26235a842>>. Acesso em: 27 ago 2018.
- AMATRIAN, X.; BASILICO, J. **Netflix Recommendations**: Beyond the 5 stars (Part 1). 2012. Disponível em: <<https://medium.com/netflix-techblog/netflix-recommendations-beyond-the-5-stars-part-1-55838468f429>>. Acesso em: 19 nov. 2018.
- ARAÚJO, C. A. A. Abordagem interacionista de estudos de usuários. **Ponto de Acesso**, Salvador, v. 4, n. 2, p. 2-32, 2010. Disponível em: <<https://rigs.ufba.br/index.php/revistaici/article/viewFile/3856/3403>>. Acesso em: 17 out. 2018.
- ARAÚJO, C. A. A. Fundamentos teóricos da classificação. **Enc. Bibli: R. Eletr. Bibl. Ci. Inf.**, Florianópolis, n. 22, 2006. Disponível em: <<https://periodicos.ufsc.br/index.php/eb/article/view/296>>. Acesso em: 15 dez. 2018.
- ARAÚJO JÚNIOR, R. H. de. **Precisão no processo de busca e recuperação da informação**. Brasília: Thesaurus, 2007.
- ARSANT, T.; KÖKSAL, E.; BOSKUS, Z. Comparison of collaborative filtering algorithms with various similarity measures for movie recommendation. **International Journal of Software Engineering & Applications (IJSEA)**. v. 6, n. 3, 2016. Disponível em: <<HTTP://AIRCCONLINE.COM/IJCSEA/V6N3/6316IJCSEA01.PDF>>. Acesso em: 04 jul. 2018.
- BANATI, H.; METHA, S. A multi-perspective evaluation of AM and GA for collaborative filtering recommender system. **International Journal of Computer Science & Technology (IJCSIT)**. v. 2, n. 5, 2010. Disponível em: <<HTTP://AIRCCSE.ORG/JOURNAL/JCSIT/1010IJCSIT08.PDF>>. Acesso em: 04 jul. 2018.
- BELL, J.; et al. Research-paper recommender systems: a literature survey. **ACM/IEEE-CS joint conference on Digital libraries**. Indianapolis, IN, USA: Springer, 2015. Disponível em: <<https://link.springer.com/article/10.1007/s00799-015-0156-0>>. Acesso em: 30 maio 2018.
- BERNARTT, J. L. V. **Um sistema de recomendação baseado em filtragem colaborativa**. 2008. Dissertação (mestrado) - Programa de Pós-Graduação em Engenharia Elétrica, Universidade Federal de Santa Catarina, Santa Catarina, 2008. Disponível em: <<http://repositorio.ufsc.br/xmlui/handle/123456789/90866>>. Acesso em: 30 maio 2018.

- BORGES, D. M.; OLIVEIRA, F. L. de. **Análise e comparação dos sistemas de recomendação de produtos existentes em três empresas de comércio eletrônico (CE)**: Saraiva, Submarino e Amazon. Palmas: Centro Universitário Luterano do Brasil, [2010]. Disponível em: <<http://sistemas-humano-computacionais.wdfiles.com/local--files/capitulo%3Aredes-sociais/Mir2010.pdf>>. Acesso em: 19 nov. 2018.
- CASAGRANDE, M. F. R. **Técnica de recomendação para repositórios digitais baseada em metadados e agrupamento de usuários**. 2014. Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Ciência da Computação. Florianópolis, 2014. Disponível em: <<https://repositorio.ufsc.br/handle/123456789/128758>>. Acesso em: 09 dez. 2018.
- CHANG, W. L.; JUNG, C. F. A hybrid approach for personalized service staff recommendation. **Information systems Frontiers**, 2015, v.19, n.1, p. 149-163. Disponível em: <<https://link.springer.com/article/10.1007%2Fs10796-015-9597-7>>. Acesso em: 03 nov. 2018.
- CHAUMIER, J. Indexação: conceito, etapas, instrumentos. **Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v.21, n.1/2, p. 63-79, jan. Disponível em: <<http://www.brapci.inf.br/index.php/article/view/0000011407/247399c870111947e2009836ea74fb3e>>. Acesso em: 11 dez. 2018.
- CUNHA, M. B. da. **Dicionário de biblioteconomia e arquivologia**. Brasília, DF: Briquet de Lemos, 2008.
- DAVID, N. C. **O uso de sistemas de recomendação para a recuperação da informação em bibliotecas**. 2017. Trabalho de Conclusão de Curso (graduação) - Universidade Federal do Estado do Rio de Janeiro, Escola de Biblioteconomia, 2017.
- DERVIN, B.; NILAN, M. Information needs and uses. **Annual Review of Information Science and Technology (ARIST)**, v. 21, p. 3-33, 1986. Disponível em: <http://www2.hawaii.edu/~donnab/lis670/dervin_nilan.pdf>. Acesso em: 30 maio 2018.
- FERNEDA, E. **Recuperação da informação**: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação. 2003. Doutorado (Tese) - Universidade de São Paulo, Escola de Comunicação e Artes, 2003. Disponível em: <<http://www.teses.usp.br/teses/disponiveis/27/27143/tde-15032004-130230/pt-br.php>>. Acesso em: 27 nov. 2018.
- FERREIRA, S. Novos paradigmas e novos usuários da informação. **Ciência da Informação**, Brasília, v. 25, n. 2, 1995. Disponível em: <<http://revista.ibict.br/ciinf/article/view/660/664>>. Acesso em: 01 nov. 2018.
- FIGUEIREDO, N. M. de. **Avaliação de coleções e estudo de usuários**. Brasília: Associação dos Bibliotecários do Distrito Federal, 1979.
- FIGUEIREDO, N. M. **Estudos de usos e usuários da informação**. Brasília: IBICT, 1994.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. São Paulo: Atlas, 2008.
- GOLDBERG, D. et al. Using collaborative filtering to weave an information tapestry. **Communications of the ACM**, v. 35, n. 12, p. 61-70, 1992. Disponível em: <https://www.ischool.utexas.edu/~i385d/readings/Goldberg_UsingCollaborative_92.pdf>. Acesso em: 26 jun. 2018.

KIM, H. M. et al. Online serendipity: the case for curated recommender systems. **Business Horizon**, v. 60, 2017. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0007681317300630>>. Acesso em: 18 nov. 2018.

LANCASTER, F. W. **Avaliação de bibliotecas e serviços de informação**. Brasília: Briquet de Lemos, 2004.

LANCASTER, F. W. **Indexação e resumos: teoria e prática**. 2. ed. Brasília: Briquet de Lemos Livros, 2004.

LU, J. et. al. Recommender system application developments: a survey. **Decision Support Systems**. Sydney: Springer, 2015. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167923615000627>>. Acesso em: 13 nov. 2018.

LOPES, G. R. et al. Sistema de recomendação para bibliotecas digitais sob a perspectiva da web semântica. **WORKSHOP DE BIBLIOTECAS DIGITAIS**, 2006. Disponível em: <<http://www.lume.ufrgs.br/bitstream/handle/10183/7787/000556995.pdf;sequence=1>>. Acesso em: 30 maio 2018.

MADADIPOUYA, K. A location-based recommender system framework to improve in user based collaborative filtering. *International Journal on Computational Science & Applications (IJCSA)*. v. 5, n. 5, 2015. Disponível em <<HTTPS://WIREILLA.COM/PAPERS/IJCSA/V5N5/5515IJCSA06.PDF>>. Acesso em: 04 jul. 2018.

MELVILLE, P.; SINDHWANI, V. Recommender Systems. In: SAMMUT, C.; WEBB, G.I. (eds). **Encyclopedia of Machine Learning and Data Mining**. Boston, MA: Springer, 2017. Disponível em: <https://link.springer.com/referenceworkentry/10.1007%2F978-1-4899-7687-1_964>. Acesso em: 30 maio 2018.

MONTATER, M.; LÓPEZ, B. L.; DE LA ROSA, J. L. A Taxonomy of Recommender Agents on the Internet. **Artificial Intelligence Review**, v. 19, n. 4, p. 285-330, 2003, Disponível em: <<http://link.springer.com/article/10.1023/A:1022850703159>>. Acesso em: 05 mar. 2018.

NIEDZEILSKI, D.; GUIMARÃES, L. dos S. Sistemas de recomendação baseados em filtragem colaborativa aplicados à recuperação da informação na web. p. 27-30. **XVI Jornada de Iniciação Científica da UNIRIO**. Livro de Resumos. 2018. Disponível em: <<http://www.unirio.br/jic/resumos/2018/livro-de-resumos-3/view>>. Acesso em: 14 nov. 2018.

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

REATEGUI, E. B.; CAZELLA, S. C. Sistemas de Recomendação. In: XXV Congresso da Sociedade Brasileira de Computação, São Leopoldo, 2005. **Anais..** São Leopoldo: UNISINOS, 2005. Disponível em: <<http://www.cin.ufpe.br/~fab/aulas-RI/Sistemas-de-Recomendacao.pdf>>. Acesso em: 30 maio 2018.

- RESNICK, P.; VARIAN, H.R. Recommender systems. **Communications of the ACM**, v.40, n.3, p.55-58, 1997. Disponível em: <http://delivery.acm.org/10.1145/250000/245121/p56-resnick.pdf?ip=200.156.27.19&id=245121&acc=ACTIVE%20SERVICE&key=344E943C9DC262BB%2EE3E361ADC3D62D81%2E4D4702B0C3E38B35%2E4D4702B0C3E38B35&_acm_=1529370469_ad90ad4eb5a1193dc137ec06483d8fe2>. Acesso em: 18 jun. 2018.
- RICCI, F.; ROKACH L.; SHAPIRA, B. Introduction to Recommender Systems Handbook. In: _____. **Recommender Systems Handbook**. Boston, MA: Springer, 2011. Disponível em: <https://link.springer.com/chapter/10.1007/978-0-387-85820-3_1>. Acesso em: 30 maio 2018.
- SAMPAIO, I. A.; RAMALHO, G. L. **Aprendizagem ativa em sistemas de filtragem colaborativa**. 2006. Dissertação (Mestrado). Programa de Pós-Graduação em Ciência da Computação, Universidade Federal de Pernambuco, Recife, 2006. Disponível em: <<https://repositorio.ufpe.br/handle/123456789/2608>>. Acesso em: 30 maio 2018.
- SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**. Belo Horizonte, v. 1. n. 1, p. 41-62, 1996. Disponível em: <<http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/235>>. Acesso em: 25 nov. 2018.
- SILVA, R. G. N. e. **Sistema de Recomendação baseado em conteúdo textual: avaliação e comparação**. Dissertação (Mestrado) – Universidade Estadual de Feira de Santana, Universidade Federal da Bahia, Programa Multi-institucional em Ciência da Computação, 2014. Disponível em: <https://repositorio.ufba.br/ri/bitstream/ri/19281/1/dissertacao_mestrado_ciencia_computacao_rafael_glauber.pdf>. Acesso em: 20 jul 2018.
- SILVA, R. E. da; SANTOS, P. L. V. A. da C.; FERNEDA, E. Modelos de recuperação da informação e web semântica: a questão da relevância. **Inf. Inf.**, Londrina, v. 18, n. 3, p. 27 – 44, 2013. Disponível em: <http://www.uel.br/revistas/uel/index.php/informacao/article/viewFile/12822/pdf_3>. Acesso em: 16 dez. 2018.
- VIEIRA, S. B. Indexação automática e manual: revisão de literatura. **Ci. Inf.** Brasília, v. 17, n. 1, p. 43-57, 1988.
- ZANETTE, L. R. **Sistema de recomendação de itens baseado na rede de confiança do usuário**. Dissertação (Mestrado em Informática) – Universidade Federal do Rio de Janeiro, Instituto de Matemática, Núcleo de Computação Eletrônica, 2008. Disponível em: <http://www.nce.ufrj.br/ginape/publicacoes/dissertacoes/d_2008/d_2008_leonardo_r_osa_zanette.pdf>. Acesso em: 30 maio 2018.