

UNIVERSIDADE FEDERAL DO MARANHÃO
CENTRO DE CIÊNCIAS SOCIAIS
CURSO DE BIBLIOTECONOMIA

JULIENE LOBATO DA SILVA

**SISTEMAS DE RECOMENDAÇÃO COMO FERRAMENTA PARA INDICAÇÕES
DE LIVROS: UM ESTUDO DE CASO NA PLATAFORMA GOODREADS**

São Luís
2021

JULIENE LOBATO DA SILVA

**SISTEMAS DE RECOMENDAÇÃO COMO FERRAMENTA PARA INDICAÇÕES
DE LIVROS: UM ESTUDO DE CASO NA PLATAFORMA GOODREADS**

Monografia submetida à coordenação do curso de Biblioteconomia, pela Universidade Federal do Maranhão (UFMA), como requisito para obtenção do grau de bacharel em Biblioteconomia.

Orientador: Prof. Dr. Roosevelt Lins Silva

São Luís

2021

JULIENE LOBATO DA SILVA

**SISTEMAS DE RECOMENDAÇÃO COMO FERRAMENTA PARA INDICAÇÕES
DE LIVROS: UM ESTUDO DE CASO NA PLATAFORMA GOODREADS**

Monografia submetida à coordenação do curso de Biblioteconomia, pela Universidade Federal do Maranhão (UFMA), como requisito para obtenção do bacharelado em Biblioteconomia.

Orientador: Prof. Dr. Roosevelt Lins Silva

Aprovada em 24 de setembro de 2021

BANCA EXAMINADORA

Prof. Dr. Roosevelt Lins Silva

Profa. Dra. Jaciara Januário da Silva

Prof. Dr. Márcio Ferreira da Silva

São Luís

2021

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

SILVA, JULIENE LOBATO DA.

SISTEMAS DE RECOMENDAÇÃO COMO FERRAMENTA PARA
INDICAÇÕES DE LIVROS : UM ESTUDO DE CASO NA PLATAFORMA
GOODREADS / JULIENE LOBATO DA SILVA. - 2021.

75 f.

Orientador(a): Roosevelt Lins Silva.

Monografia (Graduação) - Curso de Biblioteconomia,
Universidade Federal do Maranhão, São Luís, 2021.

1. Disseminação Seletiva da Informação. 2. Filtragem
da Informação. 3. Recuperação da Informação. 4. Sistemas
de Recomendação. I. Silva, Roosevelt Lins. II. Título.

AGRADECIMENTOS

Não posso iniciar este texto sem agradecer a maior figura de sustentação da minha vida: Deus. Quando acreditei que não teria mais neurônios suficientes para trabalharem e construir o que agora vejo como resultado final deste trabalho, foi dEle que tirei forças. Agradeço também à minha mãe que, em meio à pandemia, emprego, afazeres domésticos e responsabilidade com um filho de pouco mais de um ano, deu tudo de si para pedir o mínimo da minha ajuda e, assim, eu poder me focar nas linhas que aqui estão escritas. À minha avó que mesmo distante em todas as chamadas de ligação e vídeo me dizia como acreditava em mim. Aos meus amigos queridos Igor Danilo, Kaio Yukio, Laís Costa, Lucelia Almeida e Sarah Sheyevinna por sempre me oferecerem palavras de carinho e puxões de orelha quando estava procrastinando ou querendo desistir mais uma vez. Agradeço também ao meu orientador, o dr. Roosevelt Lins, pois imagino a dificuldade em orientar uma aluna tão confusa como eu; obrigada por sempre me fazer sentir mais inteligente e responsável do que eu julgava ser. Também não poderia deixar de mencionar as professoras Aldinar e Jaciara por terem me oferecido palavras doces e alguns nomes de autores que discutirei nas seções posteriores. Gostaria de agradecer especialmente ao meu trio acolhedor, às “cocotinhas”, como a professora nos apelidou na aula de Linguística uma vez: Thamyres Pontes, Caroline dos Anjos e Raynara Mendes. Vocês foram um refúgio, um ombro amigo, um riso no final das aulas, um conselho, uma ligação que atravessou os muros da UFMA e alcançou profundamente meu coração. Por último, e nem por isso menos importante, queria agradecer a alguém que provavelmente nunca saberá destas letras, alguém que fora meu mar e minha rocha desde sempre: meu avô. A José Ferreira Lobato agradeço por ter sido um pai, amigo e confidente, e também dedico a ele todo o esforço que impus nesta pesquisa. Queria poder agradecer pessoalmente. Sei que me abraçaria com força e diria um curto “eu sabia”, então beijaria o topo da minha cabeça e voltaria para cozinha em busca de uma xícara de café com leite desnatado. Queria muito dizer te amo mais uma vez.

Informação é poder, porém se tens tal domínio e não o divulgas, torna-te responsável pela ignorância alheia.

Ivan Teorilang

RESUMO

Desde a explosão de informações na web, importantes discussões começaram a ser levantadas sobre a recuperação de informações pertinentes para o usuário. Levando em consideração a lógica dos Sistemas de Recuperação da Informação muitas ferramentas surgiram para tentar levar itens relevantes e de acordo com os interesses dos usuários, com foco na Disseminação Seletiva da Informação, nascida em centros de pesquisa, e os Sistemas de Recomendação, criados graças a necessidade de conquista de clientes no comércio eletrônico. Neste trabalho propõe-se um estudo de caso com um voluntário utilizando a rede social literária Goodreads, cuja principal característica é a recomendação de livros, em busca de um entendimento como essas recomendações são indicadas e recebidas por um usuário brasileiro com baixo índice de leitura. A metodologia se deu através de uma pesquisa exploratória-descritiva, através da coleta de dados provenientes de levantamentos bibliográficos, observação do voluntário conforme cria seu perfil na rede social, questionário fechado e uma entrevista semiestruturada. Os resultados demonstraram como as recomendações tendem a levar em consideração o número de avaliações e a média positiva associada aos itens antes de recomendá-los, fazendo uma previsão de quais livros o voluntário mais se agradaria com base nos históricos de ações e nas prateleiras “lidos” e “quero ler”. Embora tenha alguns obstáculos de idioma, o site se mostrou bem eficaz em sua meta de indicar leituras novas e diferentes aos seus usuários brasileiros, ajudando-os a encontrar novas leituras.

Palavras-chave: Sistemas de Recomendação. Recuperação da Informação. Disseminação Seletiva da Informação. Filtragem da Informação.

ABSTRACT

Since the explosion of information on the web, important discussions have started to arise about the retrieval of pertinent information for the user. Taking into account the logic of Information Retrieval Systems, many tools have emerged to try to take relevant items and according to the interests of users, with a focus on Selective Dissemination of Information, born in research centers, and Recommendation Systems, created thanks to the need to win over customers in e-commerce. This work proposes a case study with a volunteer using the literary social network Goodreads, whose main feature is the recommendation of books, in search of an understanding of how these recommendations are indicated and received by a Brazilian user with a low reading rate. The methodology was carried out through an exploratory-descriptive research, through the collection of data from bibliographic surveys, observation of the volunteer as he creates his profile on the social network, a closed questionnaire and a semi-structured interview. The results showed how recommendations tend to take into account the number of ratings and the positive mean associated with items before recommending them, predicting which books the volunteer would most like based on stock histories and "to read" shelves " and "want to read". Although it has some language obstacles, the site has proven to be quite effective in its goal of providing new and different readings to its Brazilian users, helping them to find new readings.

Keywords: Recommender Systems. Information Retrieval. Selective Dissemination of Information. Information Filtering.

LISTA DE FIGURAS

Figura 1 — Avaliação de sistemas de recuperação da informação.....	17
Figura 2 — Esquema de funcionamento dos Sistemas de Recuperação da Informação	18
Figura 3 — Estrutura do Sistema de Recomendação	23
Figura 4 — Exemplo de coleta explícita de dados através das avaliações dos usuários do site da Amazon	27
Figura 5 — Seleção de gêneros literários na construção do perfil no Goodreads	29
Figura 6 — Lista de recomendação do Goodreads, baseada nos livros lidos	31
Figura 7 — Esquema de funcionalidade da Filtragem Colaborativa.....	32
Figura 8 — Top10 de filmes mais vistos na Netflix.....	35
Figura 9 — Lista de livros a serem avaliados dentro dos gêneros favoritos.....	39
Figura 10 — Página inicial do Goodreads traduzida pelo Google Chrome.....	40
Figura 11 — Lista de lançamentos de livros no Goodreads	42
Figura 12 — Lista de livros mais bem avaliados no Goodreads Choice Awards 2020 por categoria	43
Figura 13 — Escolha dos gêneros literários favoritos pelo voluntário	48
Figura 14 — Ranking de recomendações baseada nos livros lidos e avaliados .	54
Figura 15 — Recomendação de livros baseados no universo da saga Divergente, de Veronica Roth.....	57
Figura 16 — Lista de recomendações baseada nos livros adicionados em “quero ler”	58
Figura 17 — Justificativa de recomendação.....	59

LISTA DE TABELAS

Quadro 1 — Plataformas que se destacam por seus sistemas de recomendação ..	25
Quadro 2 — Atividades que fornecem dados dos usuários do Goodreads	40
Quadro 3 — divisão de etapas metodológicas da pesquisa	44
Quadro 4 — Perfil resultado do questionário realizado durante a pesquisa	47
Quadro 5 — Avaliação e classificação de 20 livros lidos pelo Voluntário no Goodreads	49
Quadro 6 — Livros indicados com desejos de leitura.....	53
Quadro 7 — Lista de livros recomendados de acordo com a prateleira “lidos”	55
Quadro 8 — Lista de livros recomendados de acordo com a prateleira “desejo ler”	59

LISTA DE SIGLAS

DSI	Disseminação Seletiva da Informação
FBC	Filtragem Baseada em Conteúdo
FC	Filtragem Colaborativa
GCA	<i>Goodreads Choice Awards</i>
KNN	K <i>Nearest Neighbors</i> , algoritmo de baseado na semelhança entre vizinhos
RI	Recuperação da Informação
SR	Sistema(s) de Recomendação
SRI	Sistema de Recuperação da Informação
SVD	Decomposição por Valor Singular
TIC	Tecnologia(s) de Informação e Comunicação

SUMÁRIO

1	INTRODUÇÃO	13
2	RECUPERAÇÃO DA INFORMAÇÃO	15
2.1	Disseminação Seletiva da Informação	19
2.2	Sistemas de Recomendação	22
2.2.1	Algoritmos baseados em SVD e KNN	33
2.2.2	Sistemas de Recomendação não-personalizados.....	35
3	GOODREADS: PLATAFORMA DE RECOMENDAÇÃO DE LIVROS ...	37
4	METODOLOGIA	44
5	COLETA E ANÁLISE DOS DADOS	47
5.1	Perfil do voluntário	47
5.2	Recomendações	54
5.3	Entrevista com o voluntário	61
6	CONSIDERAÇÕES FINAIS	64
	REFERÊNCIAS	67
	APÊNDICE A	73
	APÊNDICE B	75

1 INTRODUÇÃO

Desde o advento da internet no final do século XX, os avanços tecnológicos têm proporcionado mudanças significativas nos meios de acesso a informação. Pela primeira vez na história, uma grande escala de usuários é capaz de interagir com milhares de informações, absorvendo-as, criando novas e transformando aquelas disponíveis. Segundo Breitman (2014), a internet já é capaz de conectar mais de 7 bilhões de pessoas a 200 bilhões de objetos pelo mundo, produzindo mais de 2 exabytes (10^{18} bytes) de dados por dia. Como foi reforçado por Santini (2020, p. 98):

Os progressos recentes na área da Informática e da Computação e o advento das redes digitais tornam possível o amplo (e quase infinito) estoque de conteúdo em uma escala inédita. Essas tecnologias contribuem para centralização da informação e possibilitam paralelamente o acesso aos conteúdos digitais de modo mais rápido, flexível, descentralizado e de forma mais individualizada.

Com tanto conteúdo disponível em rede, fica claro que nunca se leu tanto quanto no século XXI, mas dizer que essas informações estão produzindo conhecimento aos seus receptores seria ir longe demais. Por isso, surgem profissionais que buscam utilizar as ferramentas disponíveis para levar conhecimento às pessoas, recuperando e disseminando informações que chamem atenção e tragam alguma consciência, entendimento aos leitores.

É tentando melhorar a recuperação das informações, documentos e demais itens oportunos aos usuários que surgem sistemas como os de Disseminação Seletiva da Informação, mais comuns em centros informacionais (LUNH, 1961; EIRÃO, 2009), e os de Recomendações (SCHAFER; KONSTAN; RIEDL, 1999; SANTINI, 2020), utilizados com mais frequência no comércio eletrônico; esses sistemas buscam levar informações relevantes, de acordo com o perfil de cada usuário, respeitando suas necessidades informacionais e gostos subjetivos. Analisando as oportunidades adquiridas com esses sistemas, encontrou-se a rede social literária *Goodreads*, na qual se destaca o serviço de recomendação de livros, baseado na personalização do perfil do usuário; capaz de ajudar a encontrar novas leituras, e ainda contribuir para que não-leitores encontrem conteúdos que aforem seu gosto por literatura. Pensando nisso, levantou-se o seguinte problema: como funcionam as recomendações de livros no *Goodreads* para novos usuários com baixo índice de leitura?

Para responder essa questão, é preciso observar e entender como opera as recomendações de livros feitas pelo *Goodreads* em um perfil recém-criado por um usuário com baixo índice de leitura. Para alcançar esse objetivo geral, alguns objetivos específicos devem ser cumpridos, são eles: a) discutir os conceitos de Sistemas de Recomendação e quais suas estratégias de filtragem da informação; b) apontar qual(ais) técnica(s) e estratégias o *Goodreads* utiliza para construir as recomendações; e por fim, c) investigar como as listas de recomendações personalizadas são construídas para um novo usuário com baixo índice de leitura.

Como mencionado no objetivo geral, esse trabalho buscou examinar como o site oferece as recomendações, principalmente no primeiro contato com o consumidor. Por isso, a pesquisa foi dividida em discussões principais como: Recuperação da Informação, DSI e Sistemas de Recomendações; a contextualização da plataforma *Goodreads*, como esta surgiu e suas características que chamam atenção no mercado. Depois, encontra-se a seção de metodologia, em que estão explicados todos os métodos utilizados nesse estudo; em seguida, a seção de coleta e análise dos dados traz os resultados obtidos e a discussão sobre como eles se relacionam com os objetivos propostos. A última seção trata da conclusão geral sobre esse estudo, confirmando algumas hipóteses levantadas no decorrer da pesquisa e oferecendo novos rumos de estudo relacionados a Biblioteconomia e Ciência da Informação.

2 RECUPERAÇÃO DA INFORMAÇÃO

Na Era Digital, em que apenas 1% dos dados são de fato analisados antes de se transformarem em informações (BREITMAN, 2014), imagina-se como as técnicas de recuperação da informação se tornam essenciais para obter o conhecimento desejado.

Kent, ainda na década de 1970, garantia que o desenvolvimento tecnológico nos mais variados setores da sociedade culminaria na produção exacerbada do conhecimento, quando as unidades de informação, como bibliotecas e arquivos, não conseguiriam acompanhar sua produção e nem teriam espaço físico para manter todo o conhecimento registrado (KENT, 1972). O que o autor talvez não enxergasse na época era que o desenvolvimento tecnológico possibilitaria um novo meio de suporte para informação: o digital. É por meio dele que barreiras físicas são vencidas e o acúmulo de dados online se torna um problema quando o usuário não consegue recuperá-los de acordo com sua necessidade informacional.

Baeza-Yates e Ribeiro-Neto (1999) e Krebs (2014) afirmam que o *boom* no acesso à informação na web ocorreu por três características principais:

1. O baixo custo do acesso: a possibilidade de encontrar várias fontes de informação, muitas vezes de forma gratuita, em vários sítios da internet, até mesmo por meio da mediação de centros de informação e pesquisa;
2. A maior acesso e rapidez: possibilidade de alcançar informações publicadas em outras partes do mundo, em questão de segundos;
3. A liberdade de publicação: a internet oferece diversas formas de criação e reprodução do conhecimento.

Essas características, possibilidades, melhor dizendo, foram a principal razão para o crescimento de usuários online; e também influenciou para a mesma migração das unidades de informação — agora presentes em ambientes físicos e digitais.

Contudo, o ambiente digital também alterou os comportamentos e sentimentos dos usuários (RIBEIRO, 2014), isso porque a massa informacional disponível em rede facilita a insatisfação dos usuários quando estes não conseguem encontrar as informações que desejam na velocidade prometida pela internet; o que ocorre também em ambientes físicos, quando os centros informacionais já não são mais capazes de suprir a enorme demanda com a mesma rapidez e precisão que os

usuários desejam (EIRÃO, 2009). Recuperar a informação desejada se tornou tarefa difícil, nas últimas décadas, então, aumentou a preocupação em resolver os obstáculos e falhas de recuperação para melhorar as respostas oferecidas.

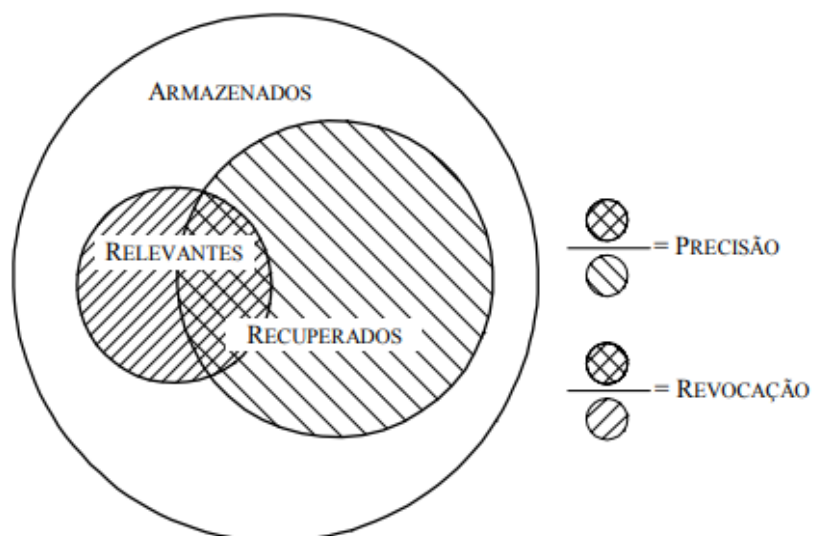
Levando em consideração o conceito de Mooers (1951 *apud* SERACEVIC, 1996, p. 44), sobre a recuperação da informação (RI) se apresentar como um conjunto de “aspectos intelectuais da descrição de informações e suas especificidades de buscas, além de quaisquer sistemas, técnicas ou máquinas empregadas para o desempenho da operação”; ou o esclarecimento de Nascimento (2004) ao dizer que a recuperação da informação — ou *information retrieval* — se foca na criação de índices e no processamento de consultas do usuário, utilizando-se de técnicas de modelagem, visualização de dados, algoritmos de ordenação (*ranking*), filtragem, indexação e linguagens documentárias (BAEZA-YATES; RIBEIRO-NETO, 1999; NASCIMENTO, 2004); entende-se o processo de RI como qualquer método de pesquisa, seja manual, mecânico ou computacional, que consiga recuperar itens pertinentes que estejam de acordo com o desejo do usuário.

A partir do entendimento do que seria a RI e seu objetivo — a disseminação do conhecimento — criou-se também os Sistemas de Recuperação da Informação (SRI), onde, de acordo com Baeza-Yates e Ribeiro-Neto (1999), são utilizadas máquinas e sistemas computacionais capazes de recuperar informações de uma base de dados.

Reforçando a ideia de Ribeiro (2014), Santini (2020, p. 105) explica que

A Recuperação da Informação inclui a noção de precisão (*precision*) e revocação (*recall*). Segundo os critérios que normalmente definem a performance de um sistema de RI, o poder de revocação é a habilidade do sistema de descobrir documentos relevantes e corresponde ao atributo mais importante para o usuário. Em contrapartida, é possível supor que a obtenção de um máximo de revocação pode ocorrer com a recuperação de todos os documentos indexados.

Figura 1 — Avaliação de sistemas de recuperação da informação

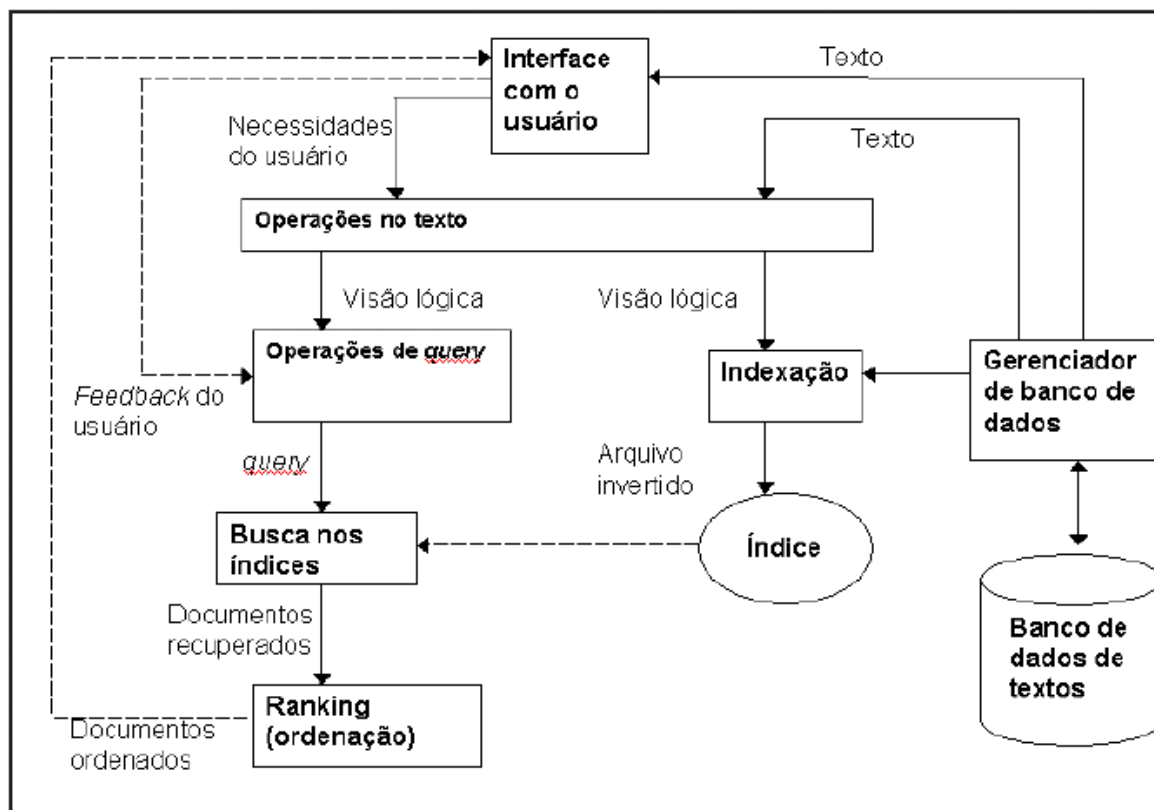


Fonte: Nascimento (2004)

Como ilustrado na figura 1 e explicado pela fala anterior de Santini, a precisão se refere ao número de documentos relevantes encontrados dentro da comunidade de informações recuperadas; enquanto a revocação é a capacidade de recuperar todos os itens relevantes, ou seja, é a porcentagem em cima informações relevantes que foram recuperadas.

As técnicas e estratégias para recuperar essas informações precisam ser bem definidas e criteriosas, selecionando palavras-chave que melhor descrevem o documento, restringindo ou ampliando os resultados de acordo com a demanda (LOPES, 2002), definindo a melhor maneira de exibir os resultados de documentos, através de um *ranking* de relevância (NASCIMENTO, 2004). Contudo, a recuperação não se atém apenas no “encontro” de uma informação, todo o processo é relevante e crucial para que a recuperação final obtenha sucesso.

Figura 2 — Esquema de funcionamento dos Sistemas de Recuperação da Informação



Fonte: Baeza-Yates; Ribeiro-Neto (1999, 2013).

Para que haja uma recuperação computacional, como a expressa na figura anterior, é preciso que todos os dados, ou características, do documento sejam registrados em um **banco de dados de texto** — onde todas as características são armazenadas em caracteres alfanuméricos (texto e número) —, a ser manipulado por um software SGBD (**Sistema Gerenciador de Bancos de Dados**) (NASCIMENTO, 2004; BAEZA-YATES E RIBEIRO-NETO, 2013, grifo nosso). Depois do registro, cada item passará pelo processo de **indexação**, ou seja, termos que identifiquem o assunto do item serão atribuídos a ele; o final dessa primeira fase resultará em um **índice**, uma lista de todos os assuntos de todos os documentos presentes no banco de dados do sistema.

Assim, quando um usuário precisar recuperar alguma informação, ele precisará interagir com sistema de recuperação, em que sua **interface com o usuário** conduzirá quais ações cumprir para fazer o processo de busca. Em um processo completo de recuperação (BAEZA-YATES; RIBEIRO-NETO, 2013, grifo nosso), comum em bibliotecas, com o auxílio de um bibliotecário de referência, o usuário

precisa construir uma consulta (**query**), preenchendo formulários e entrevistas até conseguir elucidar sua necessidade informacional e, assim, escolher as palavras-chave que mais se encaixem nos itens a ser recuperados. Quando os termos são escolhidos, o sistema **busca nos índices** produzidos na primeira fase; a respeito dos documentos que são encontrados a partir daí, forma-se uma lista ordenada (**ranking**) de itens, em ordem decrescente de relevância. Agora basta o usuário decidir qual será mais interessante para si.

O importante de todo esse processo é mencionar que a partir dessa lógica de recuperação da informação foi possível desenvolver inúmeras aplicações bem-sucedidas como produtos, sistemas, redes e serviços (SERACEVIC, 1996) que objetivam associar itens/informações relevantes a usuários/clientes; sistemas de recomendação e disseminação seletiva da informação, são bons exemplos disso.

2.2 Disseminação Seletiva de Informação (DSI)

O alto número de informações disponíveis aliado às atividades oferecidas pela internet culminou na dificuldade de encontrar itens relevantes na velocidade esperada pelos usuários. Recuperar informações não é tarefa simples, se carece entender a necessidade informacional da demanda e associá-la aos documentos disponíveis por meio de palavras-chave. O resultado é a disseminação da informação. Souto (2010, p. 11) fala de disseminação seletiva como

[...] aquele processo que a partir do perfil individual ou de grupo, identifica [...], encaminha, exhibe e/ou disponibiliza, aos usuários, um pacote informacional, resultante da seleção [...] a partir da comparação dos perfis dos usuários com os recursos informacionais disponíveis.

No dicionário Aurélio (2010), “disseminar” significa “difundir(-se)” e “propagar(-se)”, ou seja, transmitir, espalhar ou divulgar alguma coisa. Dentro da Biblioteconomia, a disseminação da informação surge como um dos pilares fundamentais que formam o objetivo profissional do bibliotecário: levar o conhecimento, fazer com que os usuários encontrem informações de seu interesse e em tempo hábil.

Sampaio e Moreschi (1990, p. 38) associaram disseminação com

[...] o sentido de semear, espalhar a informação, ou seja, o ato de levar ao conhecimento do usuário os documentos novos recebidos pela biblioteca, ou, ainda, num sentido mais amplo, divulgar entre os leitores as publicações mais relevantes e atuais para que possam

através da atualização constante desenvolver suas pesquisas e projetos.

Tomando a lógica da recuperação da informação, entendeu-se que para disseminar conhecimento era preciso mais que do levar informações aos usuários, era preciso encontrar conteúdo que os interessasse.

Quando as primeiras universidades surgiram, ainda na Idade Média, o processo de encontrar livros específicos de acordo com a demanda dos acadêmicos, foi um retrato simples de como começou o processo de seleção de informação. Contudo, é importante salientar, o serviço de disseminação seletiva propriamente dito surge em bibliotecas e centros documentais entre as décadas de 1950 e 1960, com Hans Peter Luhn, da IBM Corporation, no intuito de melhorar os serviços de alertas e notificação corrente em bibliotecas especializadas (KREBS, 2013; EIRÃO, 2009). Segundo seu próprio criador, a DSI é um “[...] serviço de informação, vindos de quaisquer fontes, para aqueles pontos onde a probabilidade de utilização, em conexão com o interesse corrente do usuário, seja alta” (LUNH, 1961, p. 132). O objetivo era chegar ao conhecimento dos usuários os documentos recebidos pela biblioteca de acordo com seu perfil de interesse, previamente criado.

Confirmando isso, Mauerhoff (1974) afirma que a DSI é uma extensão/mecanização dos serviços de alertas tradicionais que, manualmente, produziam listas selecionadas de títulos, artigos de periódico e resumos segundo a linha de pesquisa do usuário. Freund e Tomita (1978), dois anos depois, concordaram que se tratava de um processamento de perfis de usuários feito por computador para gerar uma base de dados de referências bibliográficas, que está em constante atualização pela realimentação do acervo e pela geração de relatórios estatísticos para ajustes dos perfis de interesse.

É possível notar como o entendimento sobre DSI se transforma com o passar das décadas, justamente por conta da popularização da computação no final do século XX; antes, década de 1950, as disseminações eram serviços de alertas manuais feitos por bibliotecários; já na década de 1970, a mecanização contribuiu para que esses serviços se tornassem mais rápidos e dinâmicos. Souto (2010) divide o serviço de disseminação em três gerações: a primeira foi a manual, com as fichas; a segunda, o serviço automatizado, quando se adicionou os computadores aos trabalhos de DSI; e por último a geração *internet*, quando, por conta do grande volume de informações, a disseminação seletiva passou a integrar outros ambientes, não

apenas os educacionais, como universidades e demais centros de pesquisa (SOUTO, 2010, grifo nosso); as atualizações também ofereceram uma disseminação que ia além dos materiais tradicionais (livros e artigos científicos), podendo indicar também noticiários, jornais, rádio e televisão, mercado da bolsa de valores, filmes e séries, entre outros produtos informacionais (EIRÃO, 2009).

Desde que o computador conseguiu cruzar informações de uma base de dados, a DSI obteve um crescimento significativo, tornando-se um importante instrumento do processo de gestão do conhecimento, ao oferecer praticidade, agilidade e maior ligação entre usuário e o centro informacional (EIRÃO, 2009). Para o desenvolvimento e execução da base de dados, Nocetti (1980, p. 47) lista seis estágios primordiais:

- I – Levantamento do perfil de interesse dos usuários: descrição detalhada da qualificação, especialidade, necessidades e interesses dos usuários;
- II – Análise e tradução dos perfis: atribuição de descritores, palavras-chave e códigos legíveis pelo sistema, que representam os temas serem recuperados;
- III – Arquivamento dos perfis: armazenamento no sistema dos perfis dos usuários, para processamento automatizado;
- IV – Recuperação da informação: realizada por computador, pelo confronto dos perfis dos usuários com a base de dados;
- V – Controle de qualidade: verificação realizada para os resultados, a fim de identificar possíveis erros de estratégia e de linguagem;
- VI – Expedição aos usuários: envio das listagens e ficha de avaliação, após os controles de expedição;

É após o cumprimento desses estágios que a unidade informacional vai poder concluir quais informações o usuário quer e quais não. Quase três décadas depois, Campos (2007) demarca as fases do processo de disseminação seletiva como:

- a) Definição de perfis de usuário: descrição de interesses e histórico de buscas, pesquisas e leituras de um usuário ou comunidades de usuários com perfis em comum;
- b) Acesso periódico às fontes de informação: busca de referências para itens novos ou alterados e, quando for o caso, armazena indexações efetuadas em um repositório de dados;
- c) Casamento de descritores armazenados com os perfis: aqui se faz um cruzamento de palavras-chave entre os itens e os interesses dos usuários;

- d) Disseminação dos recursos ou das referências para os usuários selecionados: faz-se chegar ao usuário a lista de itens de seu interesse.

Santini (2020, p.106) reafirma que essas etapas “[...] permitem a recuperação da informação conforme as perspectivas e necessidades individuais, e prioriza os modos de percepção de um grupo ou indivíduo específico”. Claro que, por ter surgido dentro de bibliotecas especializadas e centros documentais, voltada para pesquisadores e especialistas, é natural o serviço de DSI ainda ser associado e utilizado para fins de pesquisa. Mas, ainda que seja proveniente de um serviço tradicional da Biblioteconomia, as novas tecnologias permitem que a disseminação seletiva vá além e recupere notícias, boletins, partituras etc., toda e quaisquer informações de interesse, dentro e fora dos centros informacionais.

Baseado nisso, criou-se a tecnologia RSS (*Rich Site Summary*, ou a versão mais atualizada: *Really Simple Syndication*) que, segundo Cunha e Eirão (2012), foi criada para notificar automaticamente os usuários sobre novos conteúdos disponibilizados na internet por meio de *feeds* de linguagem estruturada. Nesse contexto, o usuário passaria de buscar a informação para que ela fosse entregue quase que instantaneamente através de *tags* de interesse, selecionadas previamente por ele.

2.3 Sistemas de Recomendação

Na mesma perspectiva de acesso a informações personalizadas, surge ideia dos sistemas de recomendação (SR) dentro do comércio eletrônico ao oferecer indicações de serviços e produtos que se encaixem nos desejos dos seus clientes. Pereira Filho (2010, p. 30) considera os SR

[...] ferramentas que podem ser utilizadas passivamente pelos usuários para explorar as possibilidades oferecidas pelo mercado, não havendo a necessidade de qualquer ação explícita dos mesmos para receber as recomendações. Estes sistemas analisam o comportamento de usuários para tentar entender o gosto dos mesmos, analisam os itens disponíveis e recomendam a estes usuários itens que provavelmente irão agradá-los. Tais sistemas podem, portanto, ajudar os usuários a encontrar produtos de nicho específico e melhorar as vendas de lojas virtuais.

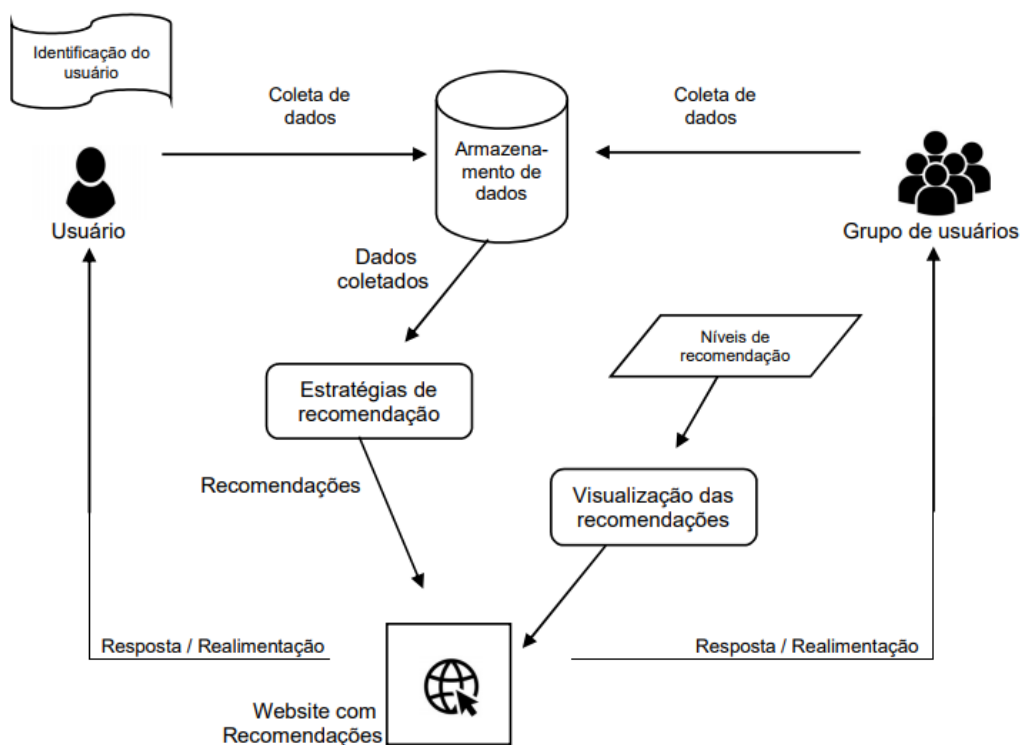
Os sistemas de recomendação, diferente dos métodos dos SRI e da DSI, o usuário não precisa necessariamente solicitar o serviço ou preencher um formulário para criar o seu perfil de interesses, boa parte dos SR utilizam algoritmos que

interpretam as ações do usuário em rede para fazer previsões sobre suas inclinações pessoais.

Silva (2014) explica que a origem dos sistemas de recomendação, datada em meados dos anos 1990, como consequência da tentativa de resolver a sobrecarga de informações advinda de um serviço de email. A ideia era criar listas de discussões automáticas, nas quais o usuário receberia preferencialmente mensagens que se adequassem com seu perfil de interesse, tendo como base os e-mails recebidos e enviados por eles e avaliações realizadas (“bom” ou “ruim”) sobre o conteúdo das mensagens lidas (GOLDBERG *et al.*, 1992). Esse primeiro “experimento” ao tentar inferir e prever os interesses de um grupo de indivíduos teve suas falhas, já que muitas informações pertinentes ficaram de fora das listas finais recomendadas (SILVA, 2014), no entanto, ainda foi o suficiente para se entender que, sim, era possível fazer um mapeamento de preferências de pessoas com base nos dados coletados de maneira explícita, como foi realizado no serviço de e-mail; ou implicitamente, como foi desenvolvido anos depois.

Para entender a lógica de recomendação dos SR, é preciso saber como sua estrutura de coleta, armazenamento e disseminação dos dados resultantes funciona, demonstrado na Figura 3.

Figura 3 — Estrutura do Sistema de Recomendação



Fonte: Schafer; Konstan; Riedl (1999).

A gênese do funcionamento dos SR, como dito por Schafer, Konstan e Riedl (1999) e demonstrado na figura acima, sugere a reunião (coleta) dos dados de um usuário ou de um grupo de usuários em rede. Esses dados podem ser o mais simples como nome, idade, religião, endereço etc.; ou mais complexos como histórico de curtidas em uma rede social, comentários em um jornal eletrônico ou vídeos assistidos no Youtube. Cada sistema de recomendação escolhe a técnica de coleta de dados que seja mais útil para alcançar o seu objetivo. Após obter os dados pertinentes, deve se escolher as estratégias e níveis de recomendação, sejam elas mais simples ou mais específicas. Um site com esse serviço por oferecer recomendações a partir dos itens mais bem avaliados (recomendação não-personalizada) ou indicar item com base no perfil de interesse calculado pelos algoritmos (recomendação personalizada).

Mais de vinte anos depois de Schafer, Konstan e Riedl, a internet alcançou mais de 4 bilhões de usuários (ISTOÉ, 2021); esse aumento contribuiu para a migração contundente das lojas físicas para o mercado de vendas online, e obviamente as possibilidades de indicações de produtos oferecidas pelos SR se tornou mais atraente, o que explica o grande investimento em tecnologias para aperfeiçoar as técnicas de recomendação no *e-commerce*.

Andrade (2018) relaciona principalmente a crescente massa informacional eletrônica à dificuldade dos clientes escolherem produtos dentro da variedade de plataformas digitais; sendo assim, foi preciso melhorar os mecanismos capazes de filtrar as opções disponíveis nos sites e fazer com que o produto certo chegasse à vista do indivíduo interessado. Conseqüentemente, a maneira mais eficiente de alcançar e satisfazer esses clientes é investir na melhoria de sistemas de recomendações para que os produtos de relevância para os clientes possam ser encontrados com facilidade, facilitando a realização da compra e/ou do consumo.

Os sistemas de sugestões se tornaram muito importantes na indústria online de vendas (Amazon, Submarino) e entretenimento (Youtube, Netflix) por causa da sua capacidade de influenciar as escolhas dos usuários. A Netflix acredita que 75% da atividade da audiência é direcionado pelo sistema de recomendações. É mais provável que um usuário padrão siga as sugestões que ele recebe, seja de amigos ou da própria plataforma do que procure por conteúdo por conta própria, (MIRANDA, 2017, p. 6).

As empresas que hoje mais se destacam no mercado online são as mesmas que investem na melhoria dos seus sistemas de recomendação, a exemplo o Quadro 1:

Quadro 1 — Plataformas que se destacam por seus sistemas de recomendação

Categoria	Sistema	Item recomendado
<i>E-commerce</i>	Amazon, Americanas, Mercado Livre	Produtos variados
<i>Streaming de vídeo</i>	Netflix, Youtube, Movielens	Filmes, vídeos
<i>Streaming de música</i>	Spotify, Deezer	Música, podcast
Rede social	Facebook, Pinterest	Fotos, posts, perfis pessoais e empresas
Serviços de turismo	Booking, Airbnb	Hotéis, restaurantes e outros estabelecimentos

Fonte: Andrade (2018).

Atualizando a tabela de Andrade, tem-se os destaques recentes como o Tiktok (rede social voltada à postagem de vídeos), que cresceu exponencialmente desde a quarentena ocasionada pelo covid-19, e a plataforma de recomendação de livros Goodreads, que em 2013 foi vendida à Amazon graças ao seu alto número de usuários oferecendo a possibilidade de impulsionar as vendas de livros da maior companhia de comércio eletrônico do mundo (G1, 2013). O fato é que os SR são uma operação constitutiva do sistema capitalista (SANTINI, 2020), no entanto, não apenas de lucros comerciais sobrevive essas ferramentas. Elas também possuem uma participação especial dentro da comunidade acadêmica. Na área da educação, por exemplo, os SR podem ajudar professores e alunos na escolha de materiais de aprendizagem, disciplinas, linhas de pesquisa para estudo, entre outros; em bibliotecas (físicas e digitais), podem oferecer documentos, informações personalizadas e pontuais aos seus usuários (ALVAREZ, 2017).

Essa característica de recuperação de informações personalizadas pode fazer muitos relacionarem os SR à DSI, que também utiliza técnicas para construir perfis de interesse dos usuários em bibliotecas para ajudar na disseminação informacional da organização. No entanto, embora suas origens sejam consequência

da explosão informacional iniciada em meados do século XX, os SR surgem da demanda capitalista de melhorar as vendas com base nas indicações de produtos/serviços específicos a clientes de interesse; enquanto a DSI surgiu em bibliotecas especializadas para facilitar a recuperação de informação para estudantes e pesquisadores, oferecendo a eles menos tempo com avaliação e seleção de conteúdo (SANTINI, 2020). Além disso, as técnicas para personalização dentro dos sistemas de recomendação e disseminação seletiva da informação são distintas. A principal delas é como se realiza a coleta dos dados pertinentes para formação dos perfis de interesse.

Dentro dos SR, segundo Bezerra (2002), para desenvolver uma personalização adequada, precisa-se pensar além da coleta de dados, focando-se também na exploração adequada desses dados.

A coleta pode ser providenciada a partir de levantamentos explícitos, através de feedback voluntário do usuário; e/ou implícitos, com o monitoramento da atividade do usuário dentro do sistema (CAVALCANTI JÚNIOR, 2017; SANTINI, 2020). Como demonstrado na Figura 4, tem-se uma avaliação de um produto — o livro “Rainha Vermelha”, da escritora Margaret Beaufort —, no qual é possível associar uma nota a ele e fazer um comentário, mencionando como foi sua experiência durante o consumo do produto. Essa coleta é considerada explícita, pois o site deixou claro o interesse em saber como foi a experiência do cliente com o produto, se este está satisfeito com o que foi vendido e entregue, e nesse caso, saber se o conteúdo do livro alcançou as expectativas do leitor que o adquiriu.

Figura 4 — Exemplo de coleta explícita de dados através das avaliações dos usuários do site da Amazon



Fonte: Amazon¹ (2021).

Sobre esse tipo de coleta de dados, Alvarez (2017, p. 122) anunciou que

[...] a medida em que evoluem os SR e as pesquisas sobre estes se desenvolvem, nota-se uma clara tendência à coleta implícita de dados em contraponto à explícita, pois, a primeira, permite captar de forma mais eficiente a informação sobre os hábitos dos usuários, permitindo assim, aumentar a precisão das recomendações.

Em contrapartida, a coleta implícita aparece mais frequentemente no *e-commerce* também, o que garante a ela os maiores investimentos por melhorias. Ela se dá através de coletas de dados de ações “escondidas”, ou seja, que refletem o interesse do cliente sem que ele precise dizer claramente que se trata de um interesse seu. Por outro lado, nem sempre essa aquisição se prova eficiente. Bezerra (2002, p. 4) oferece o seguinte exemplo:

Considere por exemplo um sítio de recomendação de filmes. Assumindo que tal sistema não comercializa filmes, não seria possível aplicar o modelo de aquisição implícita de forma adequada. Sendo assim, o sistema necessita uma avaliação do usuário com relação aos filmes apresentados, dando origem a um perfil de usuário constituído dos filmes avaliados e suas respectivas notas atribuídas por esse usuário.

¹ <https://www.amazon.com.br/rainha-vermelha-Vol-2>

Explicando o posicionamento do autor, entende-se que se esse sítio não vende filmes nem oferece ferramentas para que o cliente assista nenhum conteúdo online, então não seria possível fazer uma coleta implícita de dados a partir do histórico de compras ou visualizações (se soubessem o que o cliente tem consumido anteriormente, seria possível fazer uma relação de filmes semelhantes para serem recomendados), sendo assim, a única maneira a se construir os perfis de recomendações é oferecer formulários de coletas explícitas, perguntando diretamente quais filmes o usuário já viu e gostou.

Obviamente, dependendo da situação e necessidade da empresa ou organização que vai usufruir dos sistemas de recomendação, pode-se escolher a melhor forma de coletar os dados ou incluir uma versão combinada e alavancar a proveniência dos dados coletados.

Ambas as técnicas, implícitas e explícitas, dependem da maneira como essas informações serão tratadas, ou seja, qual abordagem, filtragem ou estratégia será escolhida para modelar o perfil do usuário.

[...] recomendadores, independentemente das técnicas utilizadas, baseiam-se no princípio da personalização da recuperação de informação de modo a agregar valor na experiência do usuário na web. Este princípio da personalização tem forte dependência em como modela e identifica as preferências do usuário, pois é deste modelo que se infere as recomendações (SANTANA, 2014, p. 8-9).

Tal estágio de tratamento informacional se encaminha para o que Bezerra (2002) chamou de exploração adequada de dados. Algumas dessas soluções se estruturam na relação entre o conteúdo dos itens com o perfil do usuário, elaborando um *ranking* de respostas (recomendações), de maior e menor relevância (BEZERRA, 2002; SANTINI, 2020). São essas abordagens que ajudarão os SR a predizer itens que mais e menos lhe interessam, a partir da análise de suas preferências, características dos itens ou avaliações e atributos de usuários semelhantes (PEREIRA, 2016); as técnicas mais comuns de filtragem da informação em SR são as: filtragem baseada em conteúdo, filtragem colaborativa e filtragem híbrida; que serão abordadas a seguir.

Filtragem baseada em conteúdo

A filtragem baseada em conteúdo (FBC) é caracterizada pela descrição de itens, utilizando-se de técnicas de similaridade entre documentos (SAMPAIO, 2006). O princípio é simples, basta correlacionar o conteúdo do item aos interesses do

usuário, alcançados por meio do seu perfil (LÁZARO, 2010). Em uma biblioteca, por exemplo, a FBC pode investigar os documentos já pesquisados e emprestados ao usuário, buscando outros documentos com o mesmo conteúdo — aqui com ajuda da classificação e indexação dos documentos no acervo — e oferecer recomendações de informações semelhantes e relevantes ao seu interesse. De fato, a FBC nasce na área da recuperação de informações e utiliza muitas de suas ferramentas (PEREIRA FILHO, 2010).

Pereira Filho (2010, p. 21) sugere que, em FBC,

[...] normalmente os usuários precisam preencher um perfil inicial para indicar ao sistema o seu gosto pessoal. Desta forma, os usuários destes sistemas possuem perfis e os itens disponíveis possuem descrições. Os perfis dos usuários são confrontados com os itens, e aqueles que possuem maior similaridade com o perfil de um determinado usuário são recomendados.

Algo semelhante ocorre na criação de *login* no Goodreads (Figura 5), quando o sistema solicita a escolha de pelo menos um gênero literário de interesse, como um dos passos iniciais na personalização do usuário. Mais tarde, o algoritmo dessa plataforma fará um cruzamento entre livros com os mesmos gêneros literários.

Figura 5 — Seleção de gêneros literários na construção do perfil no Goodreads

goodreads Começando Encontre amigos Determine uma meta **Classifique livros** Ver recomendações

Em seguida, selecione seus gêneros favoritos.

Usamos seus gêneros favoritos para fazer recomendações melhores de livros e personalizar o que você vê em seu feed de atualizações.

<input type="checkbox"/> Arte	<input type="checkbox"/> Biografia	<input type="checkbox"/> O negócio	<input type="checkbox"/> Chick Lit	<input type="checkbox"/> Infantil
<input type="checkbox"/> cristão	<input type="checkbox"/> Clássicos	<input type="checkbox"/> Histórias em quadr...	<input type="checkbox"/> Contemporâneo	<input type="checkbox"/> Livros de receitas
<input type="checkbox"/> Crime	<input type="checkbox"/> Ebooks	<input type="checkbox"/> Fantasia	<input type="checkbox"/> Ficção	<input type="checkbox"/> Gay e lésbica
<input type="checkbox"/> Novelas gráficas	<input type="checkbox"/> Ficção histórica	<input type="checkbox"/> História	<input type="checkbox"/> Horror	<input type="checkbox"/> Humor e Comédia
<input type="checkbox"/> Mangá	<input type="checkbox"/> Memórias	<input type="checkbox"/> Música	<input type="checkbox"/> Mistério	<input type="checkbox"/> Não-ficção
<input type="checkbox"/> Paranormal	<input type="checkbox"/> Filosofia	<input type="checkbox"/> Poesia	<input type="checkbox"/> Psicologia	<input type="checkbox"/> Religião
<input type="checkbox"/> Romance	<input type="checkbox"/> Ciência	<input type="checkbox"/> Ficção científica	<input type="checkbox"/> Autoajuda	<input type="checkbox"/> Suspense
<input type="checkbox"/> Espiritualidade	<input type="checkbox"/> Esportes	<input type="checkbox"/> Filme de ação	<input type="checkbox"/> Viajar por	<input type="checkbox"/> Jovem adulto

[Não encontra seus gêneros favoritos aqui?](#)

Selecione pelo menos um gênero para continuar

Fonte: Goodreads (2021).

Pereira (2016) ainda exemplifica a aplicação da FBC com um sistema de recomendação de livros: através de algoritmos de aprendizagem de máquina, faz-se

uma análise dos livros já lidos e buscar novos conteúdos a partir de atributos em comum (autor, gênero, título e etc.). Sampaio (2006, p. 21) já tinha mencionado algo parecido ao dizer que:

Os sistemas de recomendação que utilizam filtragem baseada em conteúdo fazem a sugestão de itens que sejam semelhantes aos que usuário demonstrou interesse no passado. A aquisição das preferências neste tipo de filtragem apoia-se fortemente na descrição do conteúdo dos itens de informação (e.g. um filme poderia ser descrito pelo conjunto de atributos: nome, diretor, atores, ano, gênero). Com base nas opiniões que o usuário demonstra sobre itens da base e na descrição desses itens, o sistema deve ser capaz de criar um perfil que reflita os tipos de informação nos quais o usuário se interessa.

Uma das limitações dessa abordagem é a necessidade de processar uma boa quantidade de itens para possibilitar o preenchimento das preferências do usuário. Por outro lado, a vantagem em sistemas de recomendação que utilizam a FBC é não necessitar de uma grande quantidade de usuários envolvidos, a verificação de similaridade é feita entre itens compatíveis com o perfil do usuário e novos itens ainda não avaliados (PEREIRA, 2016, p. 17).

Filtragem colaborativa

A Filtragem Colaborativa (FC) está voltada a interação entre usuários, ou seja, “[...] sistemas visando designar um tipo de software específico no qual a filtragem de informação é realizada com o auxílio humano — ou seja, com colaboração entre grupos de usuários e interessados” (SANTINI, 2020, p. 103). Tomando o Goodreads novamente como exemplo (Figura 6), tem-se a lista de recomendações de novos itens com base nos livros classificados em “lidos” e bem avaliados por um perfil específico. O algoritmo analisou as avaliações e correlacionou com perfis que tiveram um histórico de ações semelhantes para gerar essa lista final.

Figura 6 — Lista de recomendação do Goodreads, baseada nos livros lidos

Aqui estão alguns livros que achamos que você vai gostar, com base nos livros que você adicionou a esta estante. Outros leitores com interesses semelhantes têm gostado deles. Como melhorar suas recomendações ...
atualizado: 4 minutos atrás

Ver: capas | Lista

Recomendações por prateleira

visão global

leitura 50

ler 50

Dica: Crie uma nova estante de livros relacionados que você gostou ou deseja ler. Você receberá recomendações com base nesses livros!

Adicionar uma prateleira:

adicionar

Recomendações por gênero

Você ainda não tem recomendações baseadas em seus gêneros favoritos.

Mais ações

Recomendações de usuários

Dê recomendações

Peça recomendações

Livros marcados como 'Não tenho interesse'

FROM BLOOD AND ASH
JENNIFER L. ARMENTROUT

Cinder
MARIA NAZARENE PEREIRA

THE WAGE OF WINNIPEG and Me
MARIANA ZAPATA

SHATTER ME
TAMERESH MATTI

UGLY LOVE
COLLEEN HOOVER

Quero ler
★★★★★
Não interessado

Quero ler
★★★★★
Não interessado

Quero ler
★★★★★
Não interessado

Quero ler
★★★★★
Não interessado

Quero ler
★★★★★
Não interessado

Quero ler
★★★★★
Não interessado

WAIT
JENNIFER L. ARMENTROUT

PUNK 57
PENELOPE DOUGLAS

JULIE GARWOOD
The Bride

Paris Pleasure
Julie Anne Long

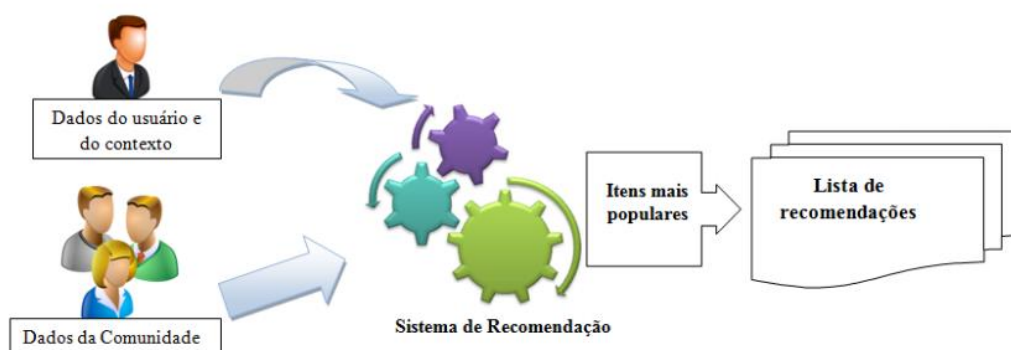
VICIOUS
L. J. SHEN

Fonte: Goodreads (2021).

Pereira Filho (2020, p. 25) explica: “[...] sistemas em que usuários podem avaliar explicitamente os diversos itens disponíveis, o algoritmo analisa as avaliações realizadas e, então, faz a predições das notas que usuários dariam a itens ainda não avaliados”; assim forma-se um “*ranking*” de itens mais e menos agradáveis aos interesses do usuário. Embora para sistemas com FC seja melhor uma aquisição explícita de dados, é possível que os algoritmos ajam implicitamente com informações do histórico de navegação, dados demográficos (idade, sexo, localidade, idioma, etc.) para reunir uma comunidade de usuários com características similares, nomeados “vizinhos” (ALVAREZ, 2017).

A principal diferença aqui é a presença de colaboração entre usuários; assim como na FBC, a FC também utiliza métodos de similaridade entre itens, porém o conteúdo deles não é levado em consideração (PEREIRA, 2016). Sampaio (2006) explica que as técnicas colaborativas não precisam da descrição de conteúdo, pois se beneficiam apenas do julgamento e interação humana.

Figura 7 — Esquema de funcionalidade da Filtragem Colaborativa



Fonte: Alvarez (2017).

Como exposto pelo esquema de Alvarez, os itens recomendados pela FC tendem a ser os mais populares no banco de dados, ou seja, aqueles itens que estão sendo mais bem avaliados, vistos ou consumidos no sistema, o que impede que aquisições recentes não sejam recomendadas até alcançar um número significativo de visualizações/avaliações. Este é descrito por Reategui e Cazella (2005) como **problema do primeiro avaliador** ou “*cold start*” uma das limitações dos SR que utilizam a filtragem colaborativa.

Tanto novos itens podem não ser recomendados quanto nenhum usuário os avaliar, quanto novos usuários podem obter recomendações de péssima qualidade, já que o sistema não possui informações suficientes sobre os mesmos para lhes fazer recomendações personalizadas. (PEREIRA FILHO, 2010, p. 35).

Em sítios com um número de usuários muito inferior ao número de informações a serem recomendadas, a FC limita-se mais uma vez no **problema de pontuações esparsas**, já que nessas situações poucos itens são avaliados e existe maior dificuldade em formação de comunidades de vizinhos. Outra dificuldade surge com usuários com gostos variados, fugindo dos padrões e encontrando obstáculos em encaixá-los em alguma comunidade; este é o **problema da similaridade** (REAGATEGUI; CAZELLA, 2005; PEREIRA FILHO, 2010; SANTINI, 2020, grifo nosso), tais problemas podem oferecer recomendações pouco relevantes e/ou repetitivas.

Filtragem híbrida

Os sistemas híbridos de recomendação buscam reunir técnicas das filtragens colaborativas e baseadas em conteúdo para alcançar recomendações mais precisas (ANDRADE, 2018).

Segundo Santini (2020), existem várias maneiras de combinar os métodos de filtragem da informação e formar uma abordagem híbrida. Entre elas, a autora põe em evidência as seguintes:

1. Implementação de métodos 'colaborativos' e 'baseado em conteúdo' separadamente, porém combinando suas predições;
2. Incorporação de algumas características da abordagem baseada em conteúdo no método colaborativo;
3. Incorporação de algumas características do método colaborativo na abordagem baseada em conteúdo;
4. Construção de um modelo geral unificado, ambos com características da abordagem colaborativa e baseada em conteúdo. (SANTINI, 2020, p. 120).

O intuito da abordagem híbrida é tentar vencer as fraquezas da FC e FBC e formar um sistema de recomendação mais completo e eficiente.

2.3.1 Algoritmos baseados em SVD e KNN

As recomendações se baseiam basicamente em dois fatores: usuário e item (levando em consideração ou não o seu conteúdo), são eles os principais fatores que oferecem informações para o mapeamento das recomendações. Contudo, como já foi discutido na seção anterior (ver 2.3), as abordagens das recomendações, sejam elas focados no usuário ou no item, possuem seus limites e falhas². É tentando resolver esses problemas que surgem algumas técnicas algorítmicas que são utilizadas com frequência.

Uma das mais populares é a técnica algébrica de fatoração de matrizes, Decomposição por Valor Singular ou *Singular Value Decomposition* (SVD) (PEREIRA FILHO, 2010), capaz de descobrir características “escondidas” dos dados.

Trata-se de uma técnica de redução de dimensionalidade que busca encontrar os fatores latentes mais importantes em uma matriz de avaliações (Usuário x Item x Avaliação). Tais fatores podem ser interpretados como conceitos semânticos e a relevância de cada conceito pode ser calculada. (CAVALCANTI JÚNIOR, 2017, p. 19-20).

Lázaro (2010) explica que o SVD é capaz de descobrir por meio da matriz usuário-item interesses dos usuários a partir de aspectos dos itens, como por exemplo as características dos livros (gênero, autor, série, assuntos, etc.) fazendo uma

² São os casos mencionados na seção secundária 2.3: necessidade de processamento de uma grande quantidade de itens (limitação da FBC); *cold start* ou problema do primeiro avaliador, problema da esparsidade ou esparsidade e problema da similaridade (limitações da FC).

correlação entre itens e sendo capaz de prever quais produtos serão mais bem-vistos pelo usuário.

Em outras palavras, essa técnica busca entender porque usuários avaliam positivamente uns itens e negativos outros (PEREIRA FILHO, 2010), são características como gênero, participação de autores ou atores, e até mesmo características mais complexas que fazem mais sentido dentro de cálculos matemáticos que ajudam a fazer um *ranking* de itens relevantes.

Outra técnica importante e muito comum em muitos sistemas de recomendação (LÁZARO, 2010), é *K-Vizinhos Mais Próximos* ou *K-Nearest Neighbor* (KNN). Trata-se de um algoritmo baseado em memória³, que deu origem a outros algoritmos com técnicas parecidas de previsões avaliativas (BEZERRA, 2002; CAVALCANTI JÚNIOR, 2010; SILVA, 2014). Ou seja, o KNN se foca na análise do histórico de ações, fazendo associações com outros usuários ou itens semelhantes, através de comunidades de interesse em comum. Sendo mais específico,

O algoritmo *K Nearest Neighbors* (KNN) é um método de análise de vizinhança que utiliza uma técnica de aprendizado supervisionado não-paramétrico. O aprendizado é supervisionado porque é feita uma análise sobre os dados de treinamento e o que se deseja aprender e conhecido. O KNN é considerado um algoritmo não-paramétrico porque não supõe qualquer distribuição de probabilidade. (PEREIRA FILHO, 2010, p. 63).

Esse algoritmo seleciona os “melhores” vizinhos, de acordo com a similaridade calculada, a partir do princípio que apenas usuários com grande número de preferências em comum são capazes de produzir indicações satisfatórias (SILVA, 2014). Por criar relações de interesses entre usuários com gostos parecidos, denominados de vizinhos, o KNN é comum em sistemas que utilizam filtragem colaborativa, por se basear na vizinhança, tentando prever classificação de um item específico, através de similaridades entre usuários ou itens (CAVALCANTI JÚNIOR, 2010), ou seja, se, por exemplo, um usuário avaliou positivamente pratos típicos como vatapá e arroz de cuxá (ambos nordestinos), é possível que ele avalie — goste — de outro prato regional, então um algoritmo de recomendação faria indicações de comidas presentes no nordeste.

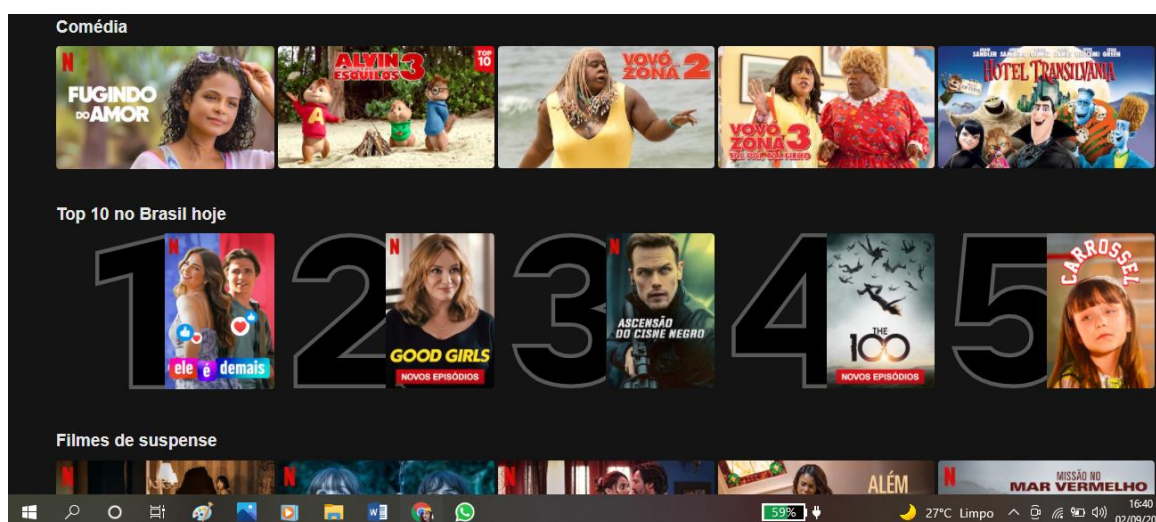
³ Estes algoritmos são assim denominados por sua característica de manter em memória os interesses do usuário para então realizar o cálculo de previsão da avaliação do usuário para um item (SILVA, 2014).

No entanto, quando o algoritmo KNN é aplicado à FC não se leva em consideração o conteúdo dos itens, como no caso do exemplo citado, não vai ser levado em conta os ingredientes necessários para preparar a comida. Segundo Pereira Filho (2010), a técnica criada pelo algoritmo KNN se foca em fazer previsões para reconhecer padrões, processamento digital de imagem e afins.

2.3.2 Sistemas de recomendação não-personalizados

Uma forma mais simples de recomendar novos itens é abrindo mão da personalização dos usuários. Sampaio (2006) relata que os sistemas de recomendação não-personalizados não são capazes de prever as preferências do usuário, muito menos recomendar automaticamente com base nas suas necessidades. Normalmente, essas recomendações são escolhidas dentro da comunidade itens mais bem avaliados e/ou visualizados dentro da plataforma.

Figura 8 — Top10 de filmes mais vistos na Netflix



Fonte: Netflix (2021)⁴.

A figura acima, por exemplo, mostra o top10 de filmes mais vistos na famosa plataforma de *streaming*, Netflix. Mesmo que utilize técnicas de recomendação personalizadas, indicações como o “top10”, “originais Netflix” e “lançamentos” oferecem recomendações mais simples e independentes do perfil do usuário; muitas empresas modernas oferecem as recomendações personalizadas e não-personalizadas para oferecer um número maior e mais variado de indicações, além de ajudar na publicidade dos conteúdos.

⁴ <https://www.netflix.com/browse>

A seção a seguir disserta sobre o caso de uma importante livraria online, cuja característica de recomendação personalizada e não-personalizada de livros vem se destacando nos últimos anos.

3 GOODREADS: PLATAFORMA DE RECOMENDAÇÃO DE LIVROS

O Goodreads foi criado ainda em 2007 por Otis Chandler e Elizabeth Khuri com o objetivo de ser uma rede social de interação entre leitores, cuja principal característica é a possibilidade de recomendar novos livros. De acordo com Chung (2011), o principal objetivo da plataforma é deixar as pessoas mais entusiasmadas, impulsioná-las a ler mais e com mais frequência. Para Otis Chandler, a missão é ajudar seus usuários a encontrar e compartilhar livros que amam (GOODREADS, 20-).

Sua característica mais significativa é a possibilidade de recomendação a partir de classificação de livros e interação entre leitores.

O Goodreads Recommendation Engine combina vários algoritmos proprietários que analisam 20 bilhões de pontos de dados para prever melhor quais livros as pessoas vão querer ler a seguir. Ele mapeia as conexões entre os livros, observando quantas vezes eles aparecem nas mesmas estantes e se foram apreciados pelas mesmas pessoas. Em média, os membros do Goodreads têm 140 livros em suas prateleiras. Com essas informações, o mecanismo aprende como seus gostos são semelhantes ou diferentes dos gostos de outros membros do Goodreads. (CHUNG, 2011, não paginado, tradução nossa)⁵.

Em 2011, após o sucesso entre os leitores, o Goodreads chegou a ter mais de 20 milhões de acessos mensais, possuindo incríveis 30 mil clubes de leitura e uma média de quatro livros sendo adicionados por segundo nas prateleiras “quero ler” (LEE, 2012; G1, 2013). Com esse número extraordinário de acessos e dados sendo criados, Otis Chandler precisou pensar em como melhorar as atividades propostas pelo site. A compra da empresa Discovereads.com (FOLHA DE SÃO PAULO, 2011), então, veio como uma tentativa de aperfeiçoar a personalização das recomendações através de algoritmos de aprendizagem de máquina.

Com sua profunda tecnologia de recomendação de livros algorítmica [referindo-se ao Discovereads], seremos capazes de sondar nosso banco de dados de 100 milhões de classificações de livros de 4,6 milhões de usuários para encontrar padrões gerais dos tipos de livros

⁵ “The Goodreads Recommendation Engine combines multiple proprietary algorithms which analyze 20 billion data points to better predict which books people will want to read next. It maps out the connections between books by looking at how often they appear on the same bookshelves and whether they were enjoyed by the same people. On average, Goodreads members have 140 books on their shelves. With this information, the engine learns how your tastes are similar to or different from the tastes of other Goodreads members.”.

que as pessoas lêem e para gerar recomendações personalizadas de alta qualidade. (CHANDLER, 2011, não paginado, tradução nossa)⁶.

Em 2013, a Amazon adquiriu o Goodreads em uma transação sigilosa que não revelou o valor da empresa criada por Chandler e Elizabeth (G1, 2013). O intuito da Amazon foi impedir que o Goodreads se tornasse uma livraria online concorrente ou ainda redirecionar vendas de livros a outros sites (FLOOD, 2013). Em palavras mais diretas, a maior plataforma de comércio eletrônico — Amazon —, cuja principal característica é venda de livros e do leitor digital Kindle, tornou-se dona do maior site de recomendação de livros — Goodreads.

Quando anunciou a venda, Chandler (2013) se mostrou muito entusiasmado com as possibilidades que a Amazon poderia oferecer ao Goodreads, principalmente em relação à leitura online de livros. Agora, os usuários da Amazon e do Goodreads podem compartilhar o mesmo *login* e oferecer uma gama mais completa de dados, que formarão uma personalização mais fidedigna do cliente.

Atualmente, a plataforma de recomendação de livros é em inglês e não apresenta versões oficiais em outros idiomas. Apesar de o Google Chrome oferecer a ferramenta de tradução automática, algumas falhas de tradução e falta de alguns livros em outras línguas podem dificultar a experiência de alguns usuários.

Para ter acesso ao site e suas funcionalidades, é bem simples. Em primeiro lugar, basta criar um perfil no Goodreads, vinculando com o login da Amazon, Facebook ou Apple, ou até mesmo criando um perfil novo com um endereço de e-mail. O nível mais simples de recomendação se inicia após a avaliação de 20 livros; esta estratégia, segundo Pereira Filho (2010) busca vencer o *cold start*, obrigando cada novo usuário a ter um histórico de leituras, embora não resolva plenamente o problema de itens recém-adicionados.

⁶ “*With their deep algorithmic book recommendation technology, we’re going to be able plumb our database of 100 million book ratings from 4.6 million users to find general patterns of the kinds of books people read and to generate high-quality personalized recommendations*”.

Figura 9 — Lista de livros a serem avaliados dentro dos gêneros favoritos

The image shows the Goodreads website interface. At the top, there's a navigation bar with the Goodreads logo and buttons for 'Começando', 'Encontre amigos', 'Determine uma meta', 'Classifique livros', and 'Ver recomendações'. Below this, a green banner reads 'Avalie os livros que você leu' and 'Eu terminei a classificação'. A sidebar on the left shows a list of genres, with 'Clássicos' highlighted in red. The main content area displays a grid of book covers with 'Avalie:' and 'Quero ler' buttons below each. The books shown include 'Pride and Prejudice', 'To Kill a Mockingbird', 'George Orwell Animal Farm', 'The Great Gatsby', 'Charlotte Brontë Jane Eyre', 'The Alchemist', '1984', 'Wuthering Heights', 'The Catcher in the Rye', and 'Little Women'.

Fonte: Goodreads (2021).

Em destaque em vermelho na figura acima, mostra os gêneros escolhidos por um perfil no Goodreads, e é partir desses gêneros que o site oferece uma lista de livros, baseada nos itens mais visualizados na plataforma, que podem ser avaliados (entre uma e cinco estrelas) e adicionados às prateleiras “lidos” e “quero ler”. Caso o usuário não tenha lido nenhum livro oferecido ou queira escolher títulos além das indicações, logo acima da lista de gêneros há uma caixa de pesquisa de livros e autores — destacado de verde.

Por também ser uma estante virtual, o Goodreads oferece a possibilidade para que seus usuários organizem suas “prateleiras” de livros de acordo com os “lidos”, “quero ler” e “estou lendo”. As avaliações daqueles já lidos vão fazer entender quais livros tem a maior compatibilidade a partir do gênero, do autor e do histórico de leitura e avaliação de perfis semelhantes. Melgão (2015) adiciona a essas ações a possibilidade compartilhamento de listas de leitura, discussão de livros entre os usuários, participação em clubes de leitura, adicionar amigos e receber e enviar mensagens privadas. Todas essas possibilidades facilitam e contribuem para uma personalização mais fiel, cujas recomendações terão mais utilidade.

Figura 10 — Página inicial do Goodreads traduzida pelo Google Chrome

The screenshot shows the Goodreads homepage with several sections highlighted by colored boxes:

- Blue box:** "LENDO ATUALMENTE" (Reading Currently) featuring the book "Corte de gelo e estrelas (Corte de Espi..." by Sarah J. Maas.
- Green box:** "2021 DESAFIO DE LEITURA" (2021 Reading Challenge) showing 0 books completed and 2 books behind.
- Red box:** "QUERO LER" (Books I Want to Read) showing a list of books including "Of One's Own" and "Finding It".
- Yellow box:** "ATUALIZAÇÕES" (Updates) showing a post from Cameron leu about "Rend (Apocalypse Necrótico # 3)".
- Cyan box:** "NOTÍCIAS E ENTREVISTAS" (News and Interviews) featuring a post about 15 new travel romances.
- Purple box:** "RECOMENDAÇÕES" (Recommendations) featuring a post about "Cercos e tempestades (Trilogia Shadow and Bone, # 2)".

Legend below the screenshot:

- Blue box: Prateleira "lendo atualmente"
- Green box: Desafio/meta de leitura
- Red box: Prateleira "quero ler"
- Yellow box: Feed
- Cyan box: Notícias e entrevistas
- Purple box: Recomendações

Fonte: Goodreads⁷ (2021).

A Figura 10 mostra a página inicial do Goodreads para um perfil já pronto. As principais informações disponíveis são: a prateleira de leitura atual, na figura representada pelo livro Corte de Gelo e Estrelas; o desenvolvimento da meta de leitura que o usuário indicou para aquele ano; a prateleira de desejos de leitura; o feed de ações de usuários com perfis parecidos (vizinhos); notícias e entrevistas sobre autores, editoras e livros escolhidos pela plataforma (não baseados no perfil); e por últimos, mas não menos importante, as recomendações de livros.

Todas as informações da página principal foram organizadas com base nos dados coletados na formação do perfil e nas ações subsequentes do usuário dentro da plataforma. Como já explicado por Bezerra (2002) na seção 2.3, entende-se que o Goodreads utiliza a combinação de coletado de informações implícita e explícita.

Quadro 2 — Atividades que fornecem dados dos usuários do Goodreads

Dados explícitos	Dados implícitos
Seleção de gêneros preferidos	Resenhas de livros
Avaliação dos livros já lidos	Participação de fóruns de discussão e grupos de leitura

⁷ Goodreads.com/

Votação no Goodreads Choice Awards	Organização das prateleiras “lidos”, “quero ler” e “lendo atualmente”
	Histórico de buscas e compras de livros na Amazon
	Interação entre usuários (comentários, lista de amigos etc.)

Fonte: a autora (2021).

Segundo Hopkins (2011), a plataforma utiliza um algoritmo de SVD e KNN para solucionar os problemas de espacialidade e similaridade⁸ da base de dados. Em entrevista, Chandler garantiu que a capacidade dos algoritmos alcança “[...] 100 milhões de avaliações de livros de 4,6 milhões de usuários para encontrar padrões gerais dos tipos de livros que as pessoas leem e para gerar recomendações personalizadas de alta qualidade” (HOPKINS, 2011, não paginado).

O Goodreads também utiliza técnicas estatísticas dos SR não-personalizados nas áreas de “Lançamentos” e “Goodreads Choice Awards”; ou seja,

[...] listas de ‘itens recomendados’ são exibidas ao usuário cujo cálculo efetuado na geração do ranking (pontuação que define o destaque a ser dado ao item na representação da recomendação) é uma simples agregação estatística (itens ‘mais vistos’, ‘mais compartilhados’, ‘mais procurados’, ‘quantidade de marcações como favoritos’, etc.) (CAVALCANTI JÚNIOR, 2017, p. 4).

⁸ (REAGATEGUI; CAZELLA, 2005; PEREIRA FILHO, 2010; LÁZARO, 2010; SANTINI, 2020). Ver também a seção secundária 2.3 deste trabalho na discussão sobre Filtragem Colaborativa.

Figura 11 — Lista de lançamentos de livros no Goodreads

The image shows a screenshot of the Goodreads website interface. At the top, there is a search bar with the text "Pesquise livros" and navigation links for "Lar", "Meus livros", "Navegue", and "Comunidade". Below the search bar, the main heading is "Novos lançamentos para agosto de 2021". There are two tabs: "Meus autores" and "Por gênero", with "Por gênero" selected. The page is divided into several sections:

- ficção:** A row of five book covers including "We Are the Brenners" by Trudy Liebet, "The People We Keep" by Jessi G. Giddens, "The Reading List" by Sarah Dessen, "Adaptation" by Lisa Klein, and "Once There Were Wolves" by Jennifer L. Holm.
- não-ficção:** A row of five book covers including "Paradise" by Lisa Klein, "Never Say You Can't Survive" by Matt Siegel, "The Secret History of Food" by Matt Siegel, "Flourish" by Carol Dweck, and "Flourish of the Anarchy" by Holly Jackson.
- jovem adulto:** A row of five book covers including "Lesson in Vengeance" by Holly Jackson, "The Sad Girl" by Jessi G. Giddens, "How We Fall Apart" by Jessi G. Giddens, "The Mysterious Girl" by Jessi G. Giddens, and "Witch Haven" by Jessi G. Giddens.
- infantil:** A row of five book covers including "Poison at Breakfast" by Jessi G. Giddens, "The Mysterious Girl" by Jessi G. Giddens, "Turkey Goes to School" by Jessi G. Giddens, "Dead Wednesday" by Jessi G. Giddens, and "The Mysterious Girl" by Jessi G. Giddens.
- história e biografia:** A row of five book covers including "Hero of Two Worlds" by Mike Duncan, "The Girls Who Stepped Out of Line" by Jessi G. Giddens, "Maiden Voyages" by Jessi G. Giddens, "The Real Valkyrie" by Jessi G. Giddens, and "Viking Heart" by Jessi G. Giddens.
- memória e autobiografia:** A row of five book covers including "Seeing Ghosts" by Jessi G. Giddens, "Sometimes I Feel Like Happy" by Jessi G. Giddens, "All in" by Jessi G. Giddens, "Made in China" by Jessi G. Giddens, and "Boy to Live" by Jessi G. Giddens.

On the right side, there is a section for "Preferências do mailer dos novos lançamentos da Juliene:" with options to "Receber novos lançamentos por e-mail" (checked) and "Receba apenas dos meus autores". Below this is a "Enviar" button and links to "editar preferências de gênero" and "editar preferências gerais de e-mail".

At the bottom right, there is a section titled "NOVOS LANÇAMENTOS POPULARES" with a grid of book covers including "The Girl on the Train" by Lisa Klein, "The People We Keep" by Jessi G. Giddens, "A Slow Fire Burning" by Paula Hawkins, "Lesson in Vengeance" by Holly Jackson, "Blood as Blood" by Holly Jackson, "Stephen King" by Stephen King, "The People We Keep" by Jessi G. Giddens, "My Heart is a Chisel" by Jessi G. Giddens, "Chevy Stevens Dark Roads" by Jessi G. Giddens, "All's Well" by Jessi G. Giddens, and "Did I Say You Could Go" by Jessi G. Giddens. A link "mais novos lançamentos" is at the bottom.

Fonte: Goodreads (2021).

Na plataforma, como mostra a Figura 11, os lançamentos estão separados em “lançamentos mais populares”, em que estão os livros mais bem avaliados e que foram lançados nos últimos meses; e os “lançamentos por gênero”, em que estão listados os livros lançados no mês da busca, separados em 14 categorias.

É neste ambiente que o Goodreads pode favorecer editoras e escritores para que apareçam em destaque através de pacotes de anúncios oferecidos.

A categoria Goodreads Choice Awards também oferece oportunidades de visibilidade para histórias menos conhecidas. Chandler cita, em 2012, em uma entrevista com Ellen Lee, como esse processo impulsionou as vendas de Cinquenta Tons de Cinza, como exemplo.

No início, o livro foi escolhido por um dos maiores clubes de livros de romance do Goodreads, os quais são grupos de discussão virtuais na plataforma. Depois que os grupos maiores o leram, houve um pico de avaliações. Então nosso editor percebeu (por volta de outubro).

Mesmo tendo apenas mil avaliações, foi extremamente bem avaliado, então foi incluído como uma indicação para o Goodreads Choice Awards. Mesmo sendo relativamente desconhecido, acabou vencendo a segunda colocação. EL James veio ao site e conversou

com alguns de nossos outros clubes do livro. A mídia conseguiu (a história) em março, e ela explodiu a partir daí. (LEE, 2012, não paginada, tradução nossa).

Figura 12 — Lista de livros mais bem avaliados no Goodreads Choice Awards 2020 por categoria

The image shows a screenshot of the Goodreads website's 'Best Books of 2020' page. At the top, there is a banner with the text 'goodreads CHOICE AWARDS 2020' and 'Anunciando os vencedores do 12º Prêmio Goodreads Choice Awards, o único grande prêmio de livros decidido pelos leitores. Parabéns aos melhores livros do ano!'. Below the banner is a grid of book categories and their winners. The categories include Ficção, Mistério e suspense, Ficção histórica, Fantasia, Romance, Ficção científica, Horror, Humor, Não-ficção, Memórias e autobiografia, História e Biografia, and Tecnologia científica. Each category features a book cover, the author's name, and a 'Quero ler' button. On the left side of the page, there is a sidebar with various links and information, including 'NOMEADOS EM SUAS PRATELEIRAS', 'CHOICE CHATTER', 'CALENDÁRIO DE VOTAÇÃO', and 'ANOS ANTERIORES'.

CATEGORIA	LIVRO VENCEDOR	AUTOR
Ficção	Midnight Library	Matt Haig
Mistério e suspense	The Guest List	Lucy Folley
Ficção histórica	The Vanishing Half	Brit Bennett
Fantasia	Crescent City	Sarah J. Maas
Romance	From Blood and Ash	Jennifer L. Armentrout
Ficção científica	To Sleep in a Sea of Stars	Paolini
Horror	Mexican Gothic	Silvia Moreno-Garcia
Humor	Stranger Planet	Stephan Lee
Não-ficção	Stamped: Racism, Resistance, and Hope	Jason Reynolds, Ibram X. Kendi
Memórias e autobiografia	A Promised Land	Barack Obama
História e Biografia	Caste: The Origins of Our Discontents	Isabel Wilkerson
Tecnologia científica	A Life on Our Planet	David Attenborough

Fonte: Goodreads (2021).

Com mais de 5 milhões de votos contabilizados (GOODREADS, 2020), o GCA 2020 oferece na página inicial a lista de recomendações com base em lançamentos do último ano que obtiveram algum destaque dentro da plataforma, segundo Chandler. A lista é formada e basta aos usuários votarem nos seus livros preferidos, também dentro das 14 categorias oferecidas pelo site. O resultado é uma lista dos “melhores” livros lançados naquele ano. Em destaque há apenas o livro mais votado, porém, em cada categoria, há uma seleção dos 20 livros mais bem avaliados.

4 METODOLOGIA

Com grande influência no positivismo de Augusto Comte, as ciências sociais desenvolveram métodos de estudo e pesquisas científicas baseadas na observação e experimentação para identificar problemas e fenômenos sociais e compreender melhor como se desenvolve a sociedade (GIL, 2012). Baseado nessa percepção, iniciou-se os percursos metodológicos desta pesquisa, como indicados no quadro abaixo:

Quadro 3 — divisão de etapas metodológicas da pesquisa

ETAPAS DO TRABALHO	ATIVIDADES
Pesquisa exploratória	<ol style="list-style-type: none"> 1. Pesquisa bibliográfica; 2. Delimitação do tema Sistemas de Recomendação;
Estudo de caso	<ol style="list-style-type: none"> 1. Escolha do voluntário; 2. Formulação das perguntas fechadas para o questionário inicial; 3. Observação da criação do perfil; 4. Análise dos resultados das recomendações.
Pesquisa descritiva	<ol style="list-style-type: none"> 1. Entrevista semiestruturada com o voluntário; 2. Descrição dos dados observados; 3. Tabulação e discussão dos resultados.

Fonte: a autora (2021).

Como a discussão principal desta pesquisa são os Sistemas de Recomendação — o qual ainda é pouco estudado nas Ciências Sociais —, classificou-se esse trabalho como proveniente de uma pesquisa exploratória-descritiva. Exploratória, segundo os parâmetros de Gil (2012, 2021), por ter encontrado como finalidade o desenvolvimento e esclarecimento de ideias sobre SR e suas técnicas de recomendações e filtragem da informação. Também seguindo as indicações de Gil (2012) e Severino (2007) buscou-se um levantamento bibliográfico em repositórios institucionais, revistas científicas e portais de notícia sobre os estudos voltados a recuperação da informação, disseminação seletiva e sistemas de recomendações; sempre lembrando que as

Pesquisas exploratórias são desenvolvidas com o objetivo de proporcionar visão geral, de tipo aproximativo, acerca de determinado fato. [...] habitualmente envolvem levantamento bibliográfico e documental, entrevistas não-padronizadas e estudos de caso. (GIL, 2012, p. 27).

O intuito da pesquisa foi entender quais os resultados de recomendações de livros são sugeridos para um usuário que tenha tido pouco contato com o mundo literário; então, como objeto de estudo encontrou-se a rede social para leitores, Goodreads, famosa por suas recomendações de livros.

O passo seguinte foi desenvolver o estudo de caso com um voluntário cujos critérios de escolha foram: não pertencer a área de Biblioteconomia, já ter lido pelo menos vinte livros — exigência mínima do site para construção do perfil —, não conhecer o Goodreads e não ter lido mais de cinco livros nos últimos vinte e quatro meses. Através de um questionário preliminar (apêndice A) que mensurou o perfil leitor, divulgado nas redes sociais, encontrou-se aquele que mais se adequava nos requisitos propostos. Também foi necessário que o voluntário se dispusesse a construir um perfil no Goodreads e usufruir de suas ferramentas durante uma semana antes de ser submetido à entrevista (apêndice B), quando foi possível entender a perspectiva do usuário perante a rede social literária.

Durante o processo de criação do perfil no site, o voluntário foi observado através do compartilhamento de tela via Google Meet⁹ no intuito de observar a interação do voluntário no primeiro contato com a rede social. Nesse momento também foi possível observar os resultados de recomendação que o voluntário obteve.

O momento descritivo da pesquisa se iniciou com os dados obtidos da observação, focando-se principalmente na análise dos resultados de recomendação, relacionando-os com as técnicas algorítmicas, citadas nas seções 2.3 e 3, utilizadas pelo Goodreads para construir o *ranking* de livros indicados.

Após entendido como as listas de recomendações foram construídas, a entrevista foi realizada — também via Google Meet. Todo o momento foi gravado, as respostas, transcritas para que, a posteriori, fossem analisadas as respostas e assim se concluísse a análise dos dados. Todos os dados provenientes do voluntário foram obtidos através de um questionário fechado, estudo de caso e entrevista

⁹ Serviço de comunicação por vídeo.

semiestruturada; em seguida esses dados foram interpretados e relacionados com o levantamento bibliográfico para alcançar as conclusões finais deste trabalho.

Com base nos objetivos descritos e a questão-problema levantada na introdução — ver seção 1 —, a coleta de dados e, principalmente, a entrevista realizada com o voluntário, foi possível ter um vislumbre de como as recomendações são ofertadas para usuários brasileiros com baixos índices de leitura, ainda entendendo os atrativos e obstáculos que o Goodreads oferece para essa comunidade de clientes.

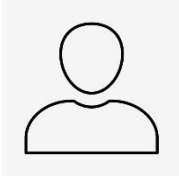
5 COLETA E ANÁLISE DOS DADOS

Esta seção reúne os percursos realizados para coleta de dados do voluntário escolhido e das recomendações oferecidas a ele. A análise se foca em três aspectos: o perfil inicial adquirido pelo questionário; as recomendações oferecidas, provenientes do perfil criado pelo Goodreads; e as impressões do voluntário sobre o site, resultado da entrevista presente no Apêndice B.

5.1 Perfil do voluntário

Para entender as recomendações oferecidas pelo Goodreads, foram descritas aqui informações provenientes do questionário e do perfil criado no site pelo voluntário. A reunião de dados sobre o número de livros lidos, gênero de interesse, idade, escolaridade, entre outros, ajudou a entender seus comportamentos de leitura e suas impressões finais — recolhidas na entrevista — perante o site.

Quadro 4 — Perfil resultado do questionário realizado durante a pesquisa

 (Nunca leu livro em outro idioma)	VOLUNTÁRIO										
	<p><i>Gênero:</i> masculino;</p> <p><i>Idade:</i> 23 anos</p> <p><i>Escolaridade:</i> Ensino Superior Completo (área: Ciências Exatas);</p> <p><i>Nível de leitura e compreensão em língua inglesa:</i> intermediário;</p> <p><i>Número de livros lidos nos últimos dois anos:</i> entre 4 e 6 livros;</p> <p><i>Redes sociais literárias:</i> não tem contato com nenhuma rede social voltada ao mundo literário;</p> <p><i>Gêneros literários de interesse:</i></p> <table style="width: 100%; text-align: center;"> <tr> <td>Fantasia</td> <td>Ficção-científica</td> <td>Mistério</td> <td>Ficção-histórica</td> </tr> <tr> <td></td> <td>Contos</td> <td>Infanto-juvenil</td> <td></td> </tr> </table>				Fantasia	Ficção-científica	Mistério	Ficção-histórica		Contos	Infanto-juvenil
Fantasia	Ficção-científica	Mistério	Ficção-histórica								
	Contos	Infanto-juvenil									

Fonte: a autora (2021).

É possível notar que o Voluntário, apesar de ter muitos gêneros literários de interesse, o que indica uma inclinação ao mundo dos livros, possui um número

baixo de leitura nos últimos dois anos. Ele também não tem contato com nenhuma rede social literária listada, o que inclui o Goodreads.

Seu nível de compreensão em língua inglesa — o que resultou ser intermediário — contribuiu nas suas impressões diante do site, já que este possui apenas a versão oficial em língua inglesa; no entanto, o voluntário declarou nunca ter lido nenhum livro em outro idioma. Resumidamente, é possível inferir que, embora o Voluntário tenha contato com o mundo literário, seu nível de leitura está abaixo da média nacional de leitura¹⁰. Logo após essa compreensão, e sabendo que esses dados corresponderam os critérios estabelecidos pela metodologia para a escolha do voluntário a ser observado, deu-se início ao registro na plataforma Goodreads.

Como primeiro passo para criar seu login no site, o voluntário precisou escolher entre criar uma conta a partir de um e-mail ou vincular alguma conta já criada nas plataformas Facebook, Amazon ou Apple. O voluntário optou por vincular à sua conta do Facebook. Em seguida, ele foi encaminhado para página de escolha dos gêneros literários:

Figura 13 — Escolha dos gêneros literários favoritos pelo voluntário

Next, select your favorite genres.

We use your favorite genres to make better book recommendations and tailor what you see in your Updates feed.

<input type="checkbox"/> Art	<input type="checkbox"/> Biography	<input type="checkbox"/> Business	<input type="checkbox"/> Chick Lit	<input type="checkbox"/> Children's
<input type="checkbox"/> Christian	<input type="checkbox"/> Classics	<input type="checkbox"/> Comics	<input type="checkbox"/> Contemporary	<input type="checkbox"/> Cookbooks
<input checked="" type="checkbox"/> Crime	<input type="checkbox"/> Ebooks	<input checked="" type="checkbox"/> Fantasy	<input type="checkbox"/> Fiction	<input type="checkbox"/> Gay and Lesbian
<input type="checkbox"/> Graphic Novels	<input checked="" type="checkbox"/> Historical Fiction	<input type="checkbox"/> History	<input type="checkbox"/> Horror	<input type="checkbox"/> Humor and Comedy
<input type="checkbox"/> Manga	<input type="checkbox"/> Memoir	<input type="checkbox"/> Music	<input checked="" type="checkbox"/> Mystery	<input type="checkbox"/> Nonfiction
<input type="checkbox"/> Paranormal	<input type="checkbox"/> Philosophy	<input type="checkbox"/> Poetry	<input type="checkbox"/> Psychology	<input type="checkbox"/> Religion
<input type="checkbox"/> Romance	<input type="checkbox"/> Science	<input checked="" type="checkbox"/> Science Fiction	<input type="checkbox"/> Self Help	<input type="checkbox"/> Suspense
<input type="checkbox"/> Spirituality	<input type="checkbox"/> Sports	<input type="checkbox"/> Thriller	<input type="checkbox"/> Travel	<input checked="" type="checkbox"/> Young Adult

Don't see your favorite genres here?

Continue

Fonte: Goodreads (2021).

Ele escolheu os gêneros Crime, Ficção-histórica, Fantasia, Ficção-científica, Mistério e *Young Adult*, destoando dos gêneros escolhidos anteriormente pelo questionário (apêndice A) apenas o gênero “crime”, que foi adicionado à lista do

¹⁰ De acordo com Torkania (2020) a média brasileira é de cinco livros lidos por ano.

Goodreads. O gênero “Young Adult” (Jovem Adulto), voltada para jovens entre 13 e 18 anos¹¹, foi encarado aqui como equivalente ao gênero Infanto-juvenil por possuírem públicos da mesma idade.

Entendido essa associação, demonstra-se (quadro 5) o próximo passo: a escolha de, no mínimo, vinte livros para avaliação — o que só é possível dentro da categoria de livro lidos — que resultaram em dados processados pelos algoritmos da plataforma e assim formaram um perfil de recomendação.

Os escolhidos foram:

Quadro 5 — Avaliação e classificação de 20 livros lidos pelo Voluntário no Goodreads

LIVROS LIDOS	AVALIAÇÃO DO VOLUNTÁRIO	CLASSIFICAÇÃO OFERECIDA PELO GOODREADS
Jogos Vorazes (livro 1 – Jogos Vorazes), de Suzanne Collins	★★★★★	Young Adult , ficção, fantasia , romance, ficção-científica , aventura, distopia, ação, adolescente, pós-apocalíptico, áudio-livro
<i>Catching Fire</i> (livro 2 – Jogos Vorazes), de Suzanne Collins	★★★★★	Ficção, fantasia , romance, não-ficção, young adult , ficção científica , aventura, fantasia urbana, distopia, ação, adolescente, adulto, romance m/m, filme, pós-apocalíptico, mídia
<i>Monkingjay</i> (livro 3 – Jogos Vorazes), de Suzanne Collins	★★★★★	Ficção, fantasia , romance, Young Adult , ficção-científica , aventura, distopia, adolescente, pós-apocalíptico, áudio-livro
A Casa Torta, de Agatha Christie	★★★★★	Ficção, clássicos, mistério , thriller, crime , thriller de mistério, detetive, áudio-livro, literatura britânica, mistério de assino
<i>The Lightning Thief</i> (livro 1 – Percy Jackson e os Olimpianos), de Rick Riordan	★★★★★	Fantasia , Young Adult , mitologia, ficção, aventura, infantil, mitologia grega, fantasia urbana, magia
<i>The Sea of Monsters</i> (livro 2 – Percy Jackson e os Olimpianos), de Rick Riordan	★★★★★	Ficção, fantasia , Young Adult , comics, aventura, romance gráfico, infantil, fantasia-urbana, magia, mitologia, história em quadrinho, mitologia grega

¹¹ <https://www.goodreads.com/genres/young-adult>

<i>The Titan's Curse</i> (livro 3 – Percy Jackson e os Olimpianos), de Rick Riordan	★★★★★	Fantasia, Young Adult , mitologia, ficção, aventura, infantil, mitologia grega, fantasia urbana, magia
<i>The Battle of the Labyrinth</i> (livro 4 – Percy Jackson e os Olimpianos), de Rick Riordan	★★★★★	Fantasia, Young Adult , mitologia, ficção, aventura, infantil, mitologia grega, fantasia urbana, magia
<i>The Last Olympian</i> (livro 5 – Percy Jackson e os Olimpianos), de Rick Riordan	★★★★★	Fantasia, Young Adult , mitologia, ficção, aventura, infantil, mitologia grega, fantasia urbana, magia
Quarto do Despejo, de Carolina Maria de Jesus	★★★	Não-ficção, clássicos, história, biografia, memórias, escola, raça, mulheres, Brasil, literatura portuguesa
Os Melhores Contos de Lygia Fagundes Telles, de Lygia Fagundes Telles	★★★★★	Ficção, clássicos, contos, escola, Brasil
Alguma Poesia, de Carlos Drummond de Andrade	★★★	Clássico, poesia, escola
Me Chame Pelo Seu Nome, de André Aciman	★★★	Ficção, romance, Young Adult , contemporâneo, adulto, LGBT, áudio-livro, queer, gay
<i>The Selection</i> (livro 1 – A Seleção), de Kiera Cass	★★★★★	Young Adult , romance, distopia, fantasia , ficção, ficção-científica , juvenil, chick lit, fantasia para jovens adultos, áudio-livro
<i>Wonder</i> (livro 1 – Extraordinário), de R. J. Palacio	★	Young Adult , ficção, contemporâneo, ficção realista, infantil, escola, família, romance
<i>Divergent</i> (livro 1 – Divergente), de Veronica Roth	★★★	Young Adult , distopia, ficção, fantasia , ficção-científica , romance, aventura, juvenil, pós-apocalíptico, ação
<i>The body in the Library</i> , de Agatha Christie	★★★★★	Mistério , ficção, crime , clássico, thriller de mistério, detetive, literatura britânica, filme de ação, mistério de assassinato, áudio-livro
<i>Death on the Nile</i> , de Agatha Christie	★★★★★	Mistério , ficção, crime , clássico, thriller de mistério, detetive, literatura britânica, filme de ação, mistério de assassinato, áudio-livro

<p><i>A Game of Thrones</i> (livro 1 – Crônicas de Gelo e Fogo), de George R. R. Martin</p>		<p>Fantasia, ficção, fantasia épica, adulto, fantasia de ficção-científica, aventura, épico, dragões, áudio-livro</p>
---	---	--

Fonte: a autora, a partir de informações fornecidas pelo Goodreads (2021).

As classificações de assuntos oferecidas pelo site estão baseadas nos gêneros literários presente nos livros e nas principais prateleiras de classificação criadas por outros usuários da plataforma. No quadro 5, foi destacado em negrito as classificações em comum com os gêneros escolhidos anteriormente pelo voluntário.

Um dado interessante a ser analisado também é a série de livros de Jogos Vorazes, da Suzanne Collins, que foi lida completamente e obteve uma boa avaliação (cinco estrelas); se observar a lista de classificação de assuntos dessa série, é possível perceber que ela aborda três gêneros em comum com aqueles escolhidos anteriormente como favoritos, ou seja, a metade dos gêneros totais de interesse. Em contrapartida, o livro *Divergent*, da série Divergente, da Veronica Roth, embora também incluía os mesmos gêneros de interesse e tenha em comum outras classificações com Jogos Vorazes (“distopia”, “pós-apocalíptico”, “aventura”, por exemplo), apresentou uma avaliação inferior, obtendo uma classificação de três estrelas, também não foi indicada a continuação de leitura dos outros livros da série Divergente. Tal fato representa que mesmo livros com enredos muito semelhantes, um usuário pode apresentar interesse por um e desinteresse por outro.

Essa previsão do motivo do usuário preferir um em vez do outro é calculada pela técnica algébrica SVD do algoritmo do Goodreads, como mencionado na seção 2.3.1 e 3. Obviamente, é muito difícil o algoritmo de recomendação adivinhar se o usuário vai ou não gostar de um livro mesmo que seja semelhante a outros que ele já leu e gostou, o SVD apenas busca indicações que possam justificar a agradabilidade em um item específico e a falta dela em outro item. Dessa forma, são coletados dados da plataforma como número e média de avaliações de ambos os livros, comentários e resenhas dos usuários, entre outros dados que o algoritmo considere pertinente.

Observando os níveis de classificação média dos dois livros indicados pelo voluntário, notou-se que, enquanto Jogos Vorazes possui uma classificação de 4.32 em um universo de quase 7 milhões de avaliações¹², *Divergent* alcança 4.17 pontos

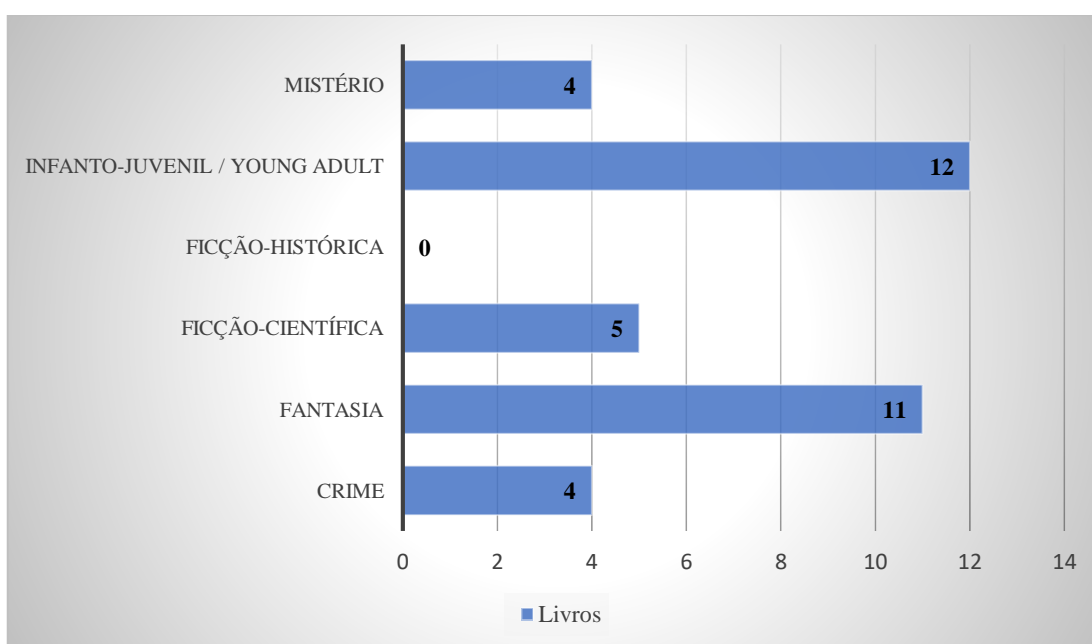
¹² <https://www.goodreads.com/book/show/8518234-jogos-vorazes>

dentro de uma comunidade um pouco maior a três milhões¹³. Assim, entendeu-se que Jogos Vorazes, além de possuir uma popularidade superior, sua média de avaliação também é maior em comparação a *Divergent*; esses dados podem ser considerados motivos pertinentes para o voluntário ter avaliado o primeiro livro com um número superior ao do segundo.

O SVD também levou em consideração que a própria avaliação mediana do voluntário em relação ao livro *Divergent* e a falta de avaliações dos livros seguintes — o que indica a descontinuidade na leitura da série de livros da Veronica Roth. Por isso, espera-se que os livros seguintes dessa série — *Insurgent*, *Allegiant* e *Four* —, apareça em últimos lugares no ranking final de recomendações, se aparecer.

Vencida a expectativa decorrente da avaliação desses dois livros, a atenção do estudo voltou-se aos gêneros favoritos presentes nos livros selecionados; obtendo-se o seguinte resultado:

Gráfico 1 — Presença de gêneros literários favoritos nos livros avaliados



Fonte: a autora (2021).

Analisando o gráfico, gêneros como *Young Adult* e Fantasia se destacam não apenas no número de ocorrência — estão presentes em 60% e 55% dos livros, respectivamente, e na maioria dos itens selecionados, ambos gêneros aparecem juntos —, mas também pelo número de boas avaliações feitas pelo voluntário,

¹³ <https://www.goodreads.com/book/show/13335037-divergent>

segundo o Quadro 5. Por esses fatores, espera-se que o algoritmo entenda que esses dois gêneros se destaquem e as primeiras recomendações se foquem neles.

Em contrapartida, o gênero “ficção-histórica”, apesar de ter sido selecionado como favorito, não esteve presente em nenhum dos livros avaliados; com isso, mesmo que ocorra recomendações desse gênero, dificilmente ele estará em posições superiores no ranking, a não ser por livros que estejam indexados com os demais gêneros de interesse.

Um ponto interessante a ser observado também é o fato de o voluntário ter avaliado muitos títulos em inglês, apesar de nunca ter lido nenhum livro em outro idioma. Isso ocorreu devido ao fato do Goodreads ser uma plataforma inteiramente em língua inglesa, oferecendo uma versão traduzida das informações através de uma ferramenta oferecida pelo Google Chrome. No entanto, os principais livros listados na primeira página de avaliação são todos livros em inglês, fazendo com que o usuário acreditasse que os livros em suas versões americanas eram os únicos disponíveis no site — fato que foi confirmado durante a entrevista (ver seção 5.3).

Após uma pesquisa sobre alguns livros nacionais, ele notou que havia, sim, edições brasileiras de livros estrangeiros. Contudo, as escolhas de livros em outro idioma influenciaram o algoritmo a indicar outros livros em língua inglesa.

Outro ponto de interesse para o algoritmo foram os livros indicados como futuras leituras, dentro da prateleira “quero ler”. As escolhas do voluntário foram:

Quadro 6 — Livros indicados com desejos de leitura

LIVRO ESCOLHIDO	CLASSIFICAÇÃO OFERECIDA PELO GOODREADS
Como Ler Livros, de Mortimer J. Adler e Charles Van Doren	Não-ficção, autoajuda, livros sobre livros, escrita, referência, desenvolvimento pessoal, filosofia, clássicos, produtividade
A Sutil Arte de Ligar o F*da-se, de Mark Manson	Não-ficção, autoajuda, psicologia, desenvolvimento pessoal, filosofia, humor, adulto
O Poder do Hábito, de Charles Duhigg	Não-ficção, psicologia, autoajuda, desenvolvimento pessoal, produtividade, ciência, liderança, sociologia
<i>Dune</i> , de Frank Herbert	Ficção-científica , ficção, fantasia , fantasia de ficção-científica, romance, aventura, <i>space opera</i> (ópera espacial)

Fonte: a autora, a partir de informações fornecidas pelo Goodreads (2021).

Aqui, o voluntário já estava ciente da possibilidade de escolher livros em suas edições nacionais, o que justifica a escolha dos três primeiros livros estarem em português. No entanto, voltando a página principal que ainda oferecia a listagem de possibilidades dos vinte livros de avaliação inicial, o voluntário encontrou o livro *Dune*, que, embora publicado no Brasil em uma versão traduzida, o voluntário achou mais prático escolher a versão original por já estar ali disponível. Um ponto interessante é que a maioria dos livros adicionados a essa prateleira não estão indexados com os gêneros de interesse do voluntário; isso pode resultar em recomendações fora dos gêneros favoritos, mas em posições inferiores no ranking final.

Após essas escolhas, o perfil final do voluntário foi processado e ele pôde usufruir finalmente das possibilidades do Goodreads.

5.2 Recomendações da plataforma

Embora os algoritmos da plataforma estejam sempre coletando dados implícitos e explícitos conforme os usuários utilizam o site para melhorar a lista de itens recomendados, é importante lembrar que as recomendações observadas são provenientes de um perfil mais simples, em seu primeiro acesso ao site.

As recomendações oferecidas para o voluntário foram retiradas das indicações dos livros já lidos — além das avaliações oferecidas a eles — e aqueles adicionados nos desejos de leitura. Os primeiros livros recomendados foram:

Figura 14 — Ranking de recomendações baseada nos livros lidos e avaliados

The screenshot shows the Goodreads interface. At the top, there's a navigation bar with 'Lar', 'Meus livros', 'Navegue', 'Comunidade', and a search bar. Below that, the page title is 'Recomendações > Ler prateleira'. A text block says: 'Aqui estão alguns livros que achamos que você vai gostar, com base nos livros que você adicionou a esta estante. Outros leitores com interesses semelhantes têm gostado deles. Como melhorar suas recomendações ... atualizado: 23 horas, 9 minutos atrás'. Below this is a grid of 10 book covers: Harry Potter (two), Game of Thrones, Brandon Sanderson's Elantris, The Song of Achilles, Patrick Rothfuss's The Windup Girl, Fablehaven, Dune: The Great Dune Desert, The Fellowship of the Ring, and The Help. Each book has a 'Quero ler' button and a star rating (all are 4 stars). On the right, a sidebar titled 'Recomendações por prateleira' has a dropdown menu with 'ler' selected (48 items). Below it, there's a tip: 'Dica: Crie uma nova estante de livros relacionados que você gostou ou deseja ler. Você receberá recomendações com base nestes livros!'. There's also a field to 'Adicionar uma prateleira:'. Further down, 'Recomendações por gênero' says 'Você ainda não tem recomendações baseadas em seus gêneros favoritos.' and 'Mais ações' includes 'Recomendações de usuários', 'Dê recomendações', 'Peça recomendações', and 'Livros marcados como "Não tenho interesse"'.

Fonte: Goodreads (2021).

A figura 14 exhibe apenas os 10 primeiros livros recomendados, contudo, é importante mencionar, que o site ofereceu 48 sugestões nas recomendações dessa prateleira.

Como dito na seção de metodologia, foram observados apenas os cinco primeiros livros no ranking, e são eles:

Quadro 7 — Lista de livros recomendados de acordo com a prateleira “lidos”

LIVROS RECOMENDADOS	MÉDIA DE AVALIAÇÃO DA PLATAFORMA	CLASSIFICAÇÃO OFERECIDA NO GOODREADS
<i>Harry Potter and the Chamber of Secrets: Sheet Music for Flute with cd</i> , de John Williams	4,62 106.200 avaliações	Fantasia , ficção, Young Adult , magia, música, infantil, romance m/m, escola
<i>Harry Potter and the Order of the Phoenix</i> (livro 5 – Saga Harry Potter), de J. K. Rowling	4,62 146.364 avaliações	Fantasia , ficção, Young Adult , magia, infantil, aventura, mistério
<i>A Game of Thrones: the Graphic Novel</i> , de Daniel Abraham, George R. R. Martin e Tommy Patterson (ilustrador)	4,61 68.402 avaliações	Fantasia , histórias em quadrinhos, ficção, romance gráfico, fantasia épica, fantasia de ficção-científica
<i>Elantris</i> , de Brandon Sanderson	4,17 199.429 avaliações	Fantasia , ficção, alta fantasia, adulto, magia, fantasia com ficção-científica, romance, Young Adult
<i>The Song of Achilles</i> , de Madeline Miller (autora do Goodreads) ¹⁴	4,41 451.263 avaliações	Ficção-histórica , fantasia, ficção, mitologia, romance, LGBT, histórico, mitologia grega, queer

Fonte: a autora, a partir de informações do Goodreads (2021).

Os primeiros livros recomendados foram todos em língua inglesa, visto que é a língua de origem do site, e porque 65% dos livros avaliados pelo voluntário durante a criação do perfil estavam em inglês, então o algoritmo entendeu que essa língua é comum na leitura do usuário. Mesmo que as recomendações estejam em outra língua, não é um problema já que os livros, com exceção do primeiro, possuem edições nacionais; com isso, a maioria das recomendações é válida para o usuário. Todos os livros oferecem pelo menos um gênero de interesse do voluntário, e como esperado os primeiros lugares nas recomendações estão livros que apresentam os dois gêneros

¹⁴ https://www.goodreads.com/author/show/176372.Madeline_Miller

de maior persistência nas avaliações anteriores: fantasia e *young adult*. Mas por que os livros vieram nessa ordem?

Obviamente, os algoritmos do Goodreads possuem a sua própria lógica de avaliação de relevância de acordo com o perfil de interesse calculado para o voluntário, mas isso não impede de buscar justificativas, baseadas em tudo que já foi discutido sobre as atividades realizadas pelos SVD e KNN (ver seção 2.3.1), dos motivos que levaram os algoritmos a escolher esses itens, nessa ordem.

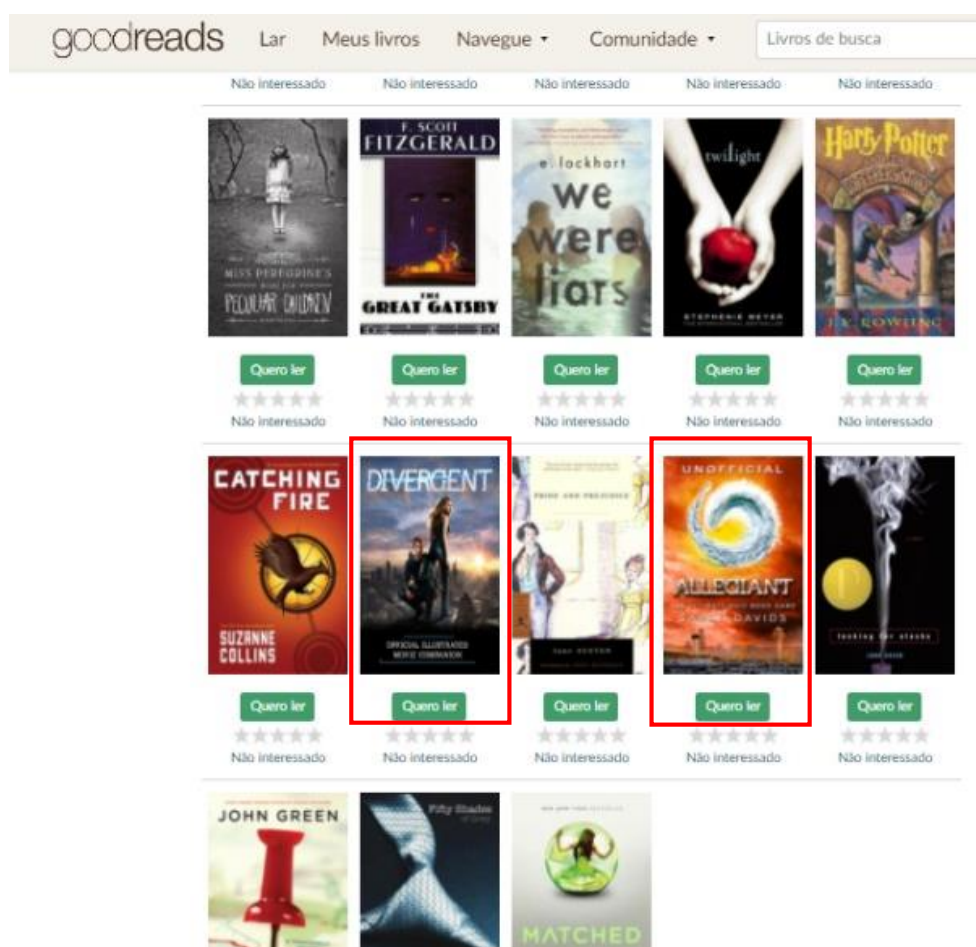
O universo de Harry Potter está na primeira e segunda posição, o que não é surpresa, entendendo que esses livros são bem conhecidos no mundo da fantasia infanto-juvenil, tendo ganhado até mesmo adaptações para o cinema. O que resultou curioso é a primeira recomendação: o livro de partitura da trilha sonora do segundo filme, baseado no livro Harry Potter e a Câmara Secreta. É bom lembrar que o objetivo do site é encontrar novas leituras, então possivelmente a escolha desse item tenha se baseado nisso: um livro com temática bem conhecida (os filmes sobre Harry Potter), mas voltado para trilha sonora do filme, que seria a inovação prometida pelo site. A alta classificação do livro recomendado (4.62 estrelas) também deve ter influenciado na escolha desse item em específico.

Os demais itens alocados nas cinco primeiras posições contam mais de um gênero favorito, incluindo as classificações de subgênero que fazem parte do universo de interesse do voluntário. A exemplo temos *Elentris*, que dentro do controle de vocabulário dos gêneros indicados pelo usuário, inclui apenas “fantasia”, mas a indicação do subgênero “fantasia de ficção-científica” aborda um segundo gênero de interesse, a “ficção-científica”.

Os primeiros cinco livros recomendados são do gênero fantasia, a segunda maior ocorrência de avaliação do usuário; seguido *young adult*, o gênero que esteve presente em 60% dos livros lidos pelo usuário. Ficção-histórica, apesar de não ter ocorrência alguma nos livros avaliados, aparece em quinto lugar na recomendação de *The Song of Achilles* por dois motivos primordiais: o primeiro se refere a classificação “fantasia” (que ocorreu diversas vezes em outros livros lidos) e também está presente nesse livro; o segundo motivo está associado aos demais gêneros indexados, em especial o de “mitologia”, que esteve presente 25% dos livros avaliados na construção do perfil. A junção desses três gêneros — ficção-histórica, fantasia e mitologia — foram interpretados como de grande interesse do usuário, o que resultou na alta posição do livro no *ranking*.

Relembrando também a discussão sobre a série Divergente na seção 5.1, quando foi entendido que talvez o usuário não tivesse interesse em continuar lendo a saga de livros — esperou-se que o algoritmo chegasse a uma conclusão parecida —, foi concebida a ideia de que, caso os demais livros da série fossem recomendados, ocupariam as últimas posições na lista, o que resultou está correto de acordo com a Figura 15:

Figura 15 — Recomendação de livros baseados no universo da saga Divergente, de Veronica Roth



Fonte: Goodreads (2021).

Respectivamente, dois livros foram recomendados: o *Divergent Oficial Illustrated Movie Companion*, de Kate Egan, na 42ª posição, que não pertence a mesma autora da série principal, mas faz menção ao universo de Veronica Roth, trazendo ilustrações e curiosidades sobre a adaptação que foi para os cinemas — situação parecida à recomendação do livro de partituras de Harry Potter. Em 44º lugar se apresentou outro livro do universo de Divergente, o *Allegiant: the Ultimate*

Figura 17 — Justificativa de recomendação

The screenshot displays the Goodreads interface. At the top, there's a navigation bar with 'goodreads', 'Lar', 'Meus livros', 'Navegue', 'Comunidade', and a search bar. Below this, the page title is 'Recomendações > Para ler a prateleira'. A text block explains that recommendations are based on books added to the shelf and similar interests of other readers. The main content area features a grid of book covers. A central pop-up window provides details for 'O Senhor dos Anéis: A Arte da Sociedade do Anel' by Gary Russell, including a 4.62 average rating from 71,562 reviews and a 'Quero ler' button. To the right, a sidebar titled 'Recomendações por prateleira' shows a list of books with counts (48 and 50) and a 'Dica' section. Below that, 'Recomendações por gênero' shows no recommendations for the user's favorite genres. At the bottom, there are 'Mais ações' and 'Recomendações de usuários' sections.

Fonte: Goodreads (2021).

Comprovando Hopkins (2011) quando mencionou que uma das técnicas algorítmicas utilizadas pelo Goodreads ser o KNN, a figura 17 traz a justificativa da escolha do livro *The Lord of the Rings: The Art of the Fellowship of the Ring*, de Gary Russel está relacionada a escolha anterior — o algoritmo guardou na memória esse dado — pelo livro *Dune* na categoria “quero ler”. Essa mesma justificativa ocorreu nas demais primeiras posições de recomendação. Esse fato indica uma priorização dos livros que estão indexados de acordo com os gêneros de interesse, visto que os demais livros adicionados à prateleira “quero ler”, como visto no quadro 6. Nenhum dos demais livros adicionados nessa prateleira durante a construção do perfil não resultou em recomendação por não apresentarem nenhum gênero em comum com os gêneros listados como favoritos.

Tabulando os cinco primeiros livros recomendados na figura 17, tem-se:

Quadro 8 — Lista de livros recomendados de acordo com a prateleira “desejo ler”

LIVROS RECOMENDADOS	MÉDIA DE AVALIAÇÃO NA PLATAFORMA	CLASSIFICAÇÃO OFERECIDA NO GOODREADS
<i>The Lord of The Rings: The Two Towers</i> , de J. R. R. Tolkien	4,45 780.050 avaliações	Fantasia , ficção, clássicos, aventura, alta fantasia, fantasia de ficção científica, fantasia épica, romance, literatura

<i>The Lord of the Rings: The Art of the Fellowship of the Ring</i> , de Gary Russel	4,62 71.774 avaliações	Fantasia , arte, ficção, clássicos, filme, aventura, alta fantasia, épico, referência
<i>A Game of Thrones: the Graphic Novel</i> , de Daniel Abraham, George R. R. Martin e Tommy Patterson (ilustrador)	4,61 68.402 avaliações	Fantasia , histórias em quadrinhos, ficção, revista em quadrinhos, fantasia épica, fantasia de ficção-científica
<i>Elantris</i> , de Brandon Sanderson	4,17 199.429 avaliações	Fantasia , ficção, alta fantasia, adulto, magia, fantasia com ficção-científica, romance, Young Adult
<i>Enders' Game: Battle School</i> , de Orson Scott Card	4,41 42.798 avaliações	Ficção-científica , revista em quadrinhos, ficção, Young Adult , Fantasia , distopia, fantasia de ficção-científica, aventura

Fonte: a autora, a partir de informações fornecidas pelo Goodreads (2021).

A lógica das recomendações desse *ranking* é a mesma da lista de recomendações baseada na prateleira “livros lidos”, ou seja, o algoritmo armazenou o histórico de livros indicados e os associou com outros livros presentes nos bancos de dados do Goodreads.

Um fato curioso observado aqui é as indicações de itens sobre a série O Senhor dos Anéis (*The Lord of the Rings*), não seguindo uma ordem cronológica da publicação dos livros, já que a primeira recomendação oferecida é o segundo livro da série (*The Two Towers*) e em nenhum momento indicou-se o primeiro livro da série, o que esta pesquisadora não conseguiu entender o porquê.

Na segunda posição do ranking de recomendações, trouxe um livro de ilustrações das criaturas do mundo fantástico de Tolkien, além de curiosidades sobre o filme que adaptou o segundo livro da saga. Sabendo que *Dune* foi o principal motivo dessas indicações, e observando os comentários que esses livros possuem no Goodreads, descobriu-se que *Dune* também será adaptado para o cinema — semelhança que justificaria as indicações a livros que relatam curiosidades sobre o universo cinematográfico em torno das obras de Tolkien.

Todas as cinco primeiras indicações aqui também seguem a linha de fantasia, sendo, dessa vez, apenas duas delas também dentro do gênero *young adult*.

Também chamou a atenção as indicações dos livros *A Game of Thrones: the Graphic Novel*, de Daniel Abraham, George R. R. Martin e Tommy Patterson (ilustrador) e *Elantris*, de Brandon Sanderson, que já estavam presentes nas mesmas

posições na lista anterior (ver quadro 7 e figura 14) de recomendações. Mesmo que seja justificável suas recomendações, parece um erro que estejam presentes em duas listas distintas, de maneira que recomendações duplicatas podem ocupar o lugar de outros itens pertinentes à indicação.

De modo geral, à luz dos teóricos que discutem Sistemas de Recomendação e algoritmos de filtragem da informação, todas as indicações oferecidas pelo Goodreads ao voluntário desta pesquisa estão de acordo com o proposto pelo site: recomendações conforme o perfil leitor dos usuários. Embora seja lógica o resultado oferecido pela plataforma e os livros recomendados estejam dentro do universo de gêneros literários que interessam ao voluntário, é muito difícil garantir que esses livros agradarão o usuário.

Para compreender o outro lado da questão, foi preciso submeter o voluntário a uma entrevista. O principal objetivo com ela foi descobrir se ele sentiu realmente satisfeito com as indicações de livros oferecidas pelo Goodreads.

5.3 Entrevista com o voluntário

O Goodreads oferece muitos serviços dentro do mundo literário. Desde organização de leituras por meios de estantes virtuais até participação de fóruns de discussão dos mais variados assuntos, sua principal característica continua sendo as recomendações de livros. São essas indicações, sendo personalizadas ou não, que influenciam novas leituras e atraem boa parte do público.

Mas, para saber com certeza se o site promete aquilo que oferece — recomendações capazes de ajudar a escolher próximas leituras —, é necessário ouvir a opinião de seus usuários. Nesta seção será esboçada as principais impressões do voluntário observado, por meio da entrevista semiestrutura, de acordo com o apêndice B, a qual foi vocalizada por meio de um encontro através do Google Meet e gravada para que pudesse ser transcrita; assim temos os seguintes resultados:

Em primeiro lugar, você teve alguma dificuldade ao fazer o login inicial na plataforma?

Resposta = *não, foi tudo muito fácil de fazer. Achei o site bem intuitivo nessa parte de criar um perfil.*

Boa parte dos livros que você organizou nas suas prateleiras foram livros em inglês, por que fez essa escolha?

Resposta = *Eu não sabia que tinha livros em português (voluntário ri). Quer dizer, como o site é todo em inglês, pensei que os livros também eram todos em inglês. Só depois que me dei conta.*

Você indicou seis gêneros literários de interesse no questionário e outros seis durante a construção no perfil do Goodreads. Apesar de ter mencionado no decorrer desse estudo que leu quatro livros nos últimos dois anos, você se considera uma pessoa interessada no mundo literário?

Resposta = *Eu já fui um leitor mais voraz, mas com o tempo, e por causa da correria do dia-a-dia, não consegui manter o mesmo ritmo de leitura de quando eu era mais novo. Mas eu gosto do mundo literário.*

Sobre a aba “recomendações”, você se interessou por algum dos livros indicados? Algum chamou mais atenção? Por quê?

Resposta = *Todas as recomendações estavam super de acordo com o tipo de livro que eu gosto de ler, mas eu não tive muitas surpresas com algumas recomendações. O Senhor dos Anéis e Harry Potter, por exemplo. São o tipo de livro que eu acho que gostaria de ler, eu até já assisti os filmes inspirados nos livros, e gostei bastante. Mas acho que não leria nenhum deles agora. Os outros livros [recomendados] pareceram bem legais também, gostei da parte dos melhores livros do ano [Goodreads Choice Awards]. Vi uns livros novos bem bacanas.*

As abas “Lançamentos” e “Goodreads Choice Awards” indicaram algum livro que você não conhecia? Algum deles chamou sua atenção?

Resposta = *Pois é, como eu disse, gostei muito do Goodreads Choice Awards, achei uma iniciativa bem legal, e tem uns livros diferentes que parecem ser bem interessantes. Os lançamentos eu não vi muito bem, mas acho que segue no mesmo*

rumo do Melhores do Ano. São livros diferentes, alguns ainda nem foram lançados no Brasil.

Você considera que as recomendações estão de acordo com o seu gosto literário? Houve alguma recomendação que não esteve do seu agrado?

Resposta = Sim, todos os livros são bem parecidos com livros que já li antes. Acho que só os livros de Divergente [referindo-se a série de livros resultante do livro Divergente] que me indicaram eu não gostei, mas isso foi porque eu não gostei do primeiro livro [o voluntário ri novamente], não penso em ler mais nada sobre essa saga.

De forma geral, você se sentiu satisfeito com o Goodreads? Você acha que voltará a utilizá-lo?

Resposta = Gostei bastante do site, fora as recomendações, ele tem umas ferramentas bem legais. Provavelmente eu vou usar de novo os quizz literários que tem no site, achei eles bem divertidos.

Quais aspectos da plataforma você acredita que possa melhorar?

Resposta = Acho que não ter uma versão em português atrapalhou um pouco. Eu sei um pouco de inglês, então tive como usar o site na versão original sem grandes problemas, mas quando precisei organizar aquelas estantes tive traduzir a página pra entender como colocar os livros no lugar certo, e ainda assim teve uma traduções erradas que o Google fez. Acho que, de longe, isso precisa melhorar, porque fiquei bem confuso.

Apesar de não entender muito bem como o Goodreads funcionava, o voluntário não teve grandes problemas em usar as ferramentas oferecidas, além de ter sua curiosidade despertada através de recomendações diferentes das quais ele esperava, contribuindo para que ele tenha vontade de utilizar a plataforma novamente.

6 CONSIDERAÇÕES FINAIS

Sistemas de Recomendação são quase imprescindíveis no comércio eletrônico, sabendo que esses sistemas são capazes de oferecer produtos e serviços de acordo com os gostos pessoais de cada cliente. As recomendações não oferecem apenas um serviço que facilita a recuperação de informações personalizadas; diferentemente do SRI e da DSI que oferecem informações com base em interesses explicitamente declarados, os sistemas de recomendação oferecem predições com base na coleta de dados explícita e/ou implícita, capaz de fazer o usuário encontrar itens que talvez nunca pudesse imaginar que fosse do seu interesse. Sendo personalizadas (baseadas em um perfil de interesse) ou não (baseadas em boas avaliações ou novas aquisições), os sistemas de recomendação foram divididos em três grandes grupos de estratégia de indicação: os sistemas baseados em conteúdo, quando o foco da recomendação se dá através do conteúdo descrito do item; baseados em colaboração, que recomenda através das ações semelhantes dos usuários; e sistemas híbridos, que reúnem estratégias dos dois anteriores para oferecer recomendações de maior qualidade.

Favorecendo-se das possibilidades oferecidas pelos SR, o Goodreads surgiu com uma ideia inovadora para época: oferecer indicações de novos livros com base no histórico de leitura e usuários com gostos similares, ou seja, um sistema colaborativo. Desde 2007, o site só cresceu, chegando a chamar atenção da maior livraria online do mundo: a Amazon. Agora juntos, o Goodreads conseguiu adquirir algoritmos baseados em SVD e KNN, oferecendo recomendações mais especializadas aos usuários.

Notou-se uma tendência do site em recomendar livros baseados em universos famosos dos gêneros Fantasia e *Young Adult* voltados às suas adaptações cinematográficas, como por exemplo os livros sobre os filmes de Harry Potter, Divergente e O Senhor dos Anéis.

No caso específico da série Divergente, mesmo que o algoritmo tenha entendido que o enredo do primeiro livro não tenha agradado o usuário, ainda assim preferiu indicar dois livros do universo, só que em posições inferiores no *ranking* final. Isso se deve à técnica SVD, que calculou os possíveis motivos do usuário não ter oferecido uma nota maior ao livro *Divergent*, como o caso do livro Jogos Vorazes, que possui muitas características semelhantes, mas com uma avaliação superior — pelo

usuário e pela comunidade em geral do Goodreads. Após fazer essas análises, mesmo que considerasse pertinente a recomendação de outros itens do universo de *Divergente*, sua posição ficou comprometida.

O caso da técnica do KNN é ainda mais fácil de ser entendida, já que é baseada na memória, foca-se totalmente nas ações do usuário dentro da plataforma e na construção de grupos de usuários semelhantes. Até mesmo o Goodreads expõe isso ao associar indicações de livros com outros que já foram lidos ou postos na lista de desejos de leituras. Contudo, essa técnica vai além, chegando a analisar comentários, avaliações, participação em enquetes, entre outras ações dentro do site que possam ser associadas a itens literários de interesse.

Observou-se que mesmo com um perfil simples, com poucas avaliações/ações, como foi o caso do voluntário deste estudo, o Goodreads conseguiu calcular recomendações pertinentes, ainda que algumas delas tenham falhas, como o caso da duplicação na recomendação de dois itens (ver seção 5.2). No entanto, é preciso lembrar que a simplicidade do perfil processado pode ter sido o principal motivo dessas falhas. Quanto mais ações o usuário fizer no site, o que também inclui a retirada de itens da lista de recomendações, mais dados serão processados e interpretados pelos algoritmos e assim construir um perfil mais específico.

Àqueles brasileiros que utilizam o Goodreads, dois obstáculos precisam ser vencidos: a escassez dos livros em edições nacionais e a falta de uma versão brasileira do site; esses fatores são observados desde a escolha inicial de avaliação dos livros lidos, pois todos os itens de destaque são em língua inglesa, levando boa parte dos usuários a escolher livros internacionais, como ocorreu com o voluntário desse estudo. Ambos problemas dificultam a possibilidade de usufruir completamente o site, visto que a maioria das recomendações é em língua inglesa. Falhas nas traduções automáticas do navegador ou até mesmo a falta de publicação do livro recomendado na língua nacional, podem frustrar os usuários brasileiros. Ainda assim, o Goodreads é uma ferramenta muito pertinente a todos que têm alguma dificuldade para decidir qual a próxima leitura ou precisa de indicações de novos livros dentro de suas preferências literárias. Sua possibilidade de organização de leituras, fóruns de discussões, enquetes e jogos interativos sobre livros e gêneros literários diversos, até mesmo a possibilidade de seguir vários autores de prestígio no site atrai muitos ávidos pela literatura e podem chamar atenção daqueles que estejam em busca de incentivo para ler com mais frequência.

Inclusive, as estratégias de recomendação do site podem ser muito interessantes para bibliotecas, por exemplo, como forma de incentivar a leitura na comunidade. Além disso, as classificações encontradas nos livros disponíveis no Goodreads, através de classificações dos usuários (folksonomia) podem ajudar na indexação de suas edições em unidades de informação que buscam relacionar mais palavras-chave com os livros disponíveis em seu acervo, contribuindo para uma análise temática mais ampla e favorecendo ainda mais a recuperação da informação. Outra possibilidade é analisar a interação entre o usuário com a plataforma, em uma perspectiva voltado ao UX (*user experience*); já que se trata de um site voltado para o público leitor, entender como os usuários interagem e experienciam o site pode abrir possibilidades às bibliotecas digitais para melhorar seu ambiente web e até mesmo atrair novos leitores ao consumo dos acervos online.

REFERÊNCIAS

ANDRADE, Leanjoelson Souza. **Padrões de design de interação para sistemas de recomendação**. Dissertação (mestrado) — Universidade Federal do Maranhão, Programa de Pós-graduação em Design. São Luís, 2018. Disponível em: <https://tedebc.ufma.br/jspui/bitstream/tede/2333/2/Leanjoelson%20Souza.pdf>. Acesso em: 29 jul. 2021.

ALVAREZ, Edgar Bisset. **Sistemas de recomendação para bibliotecas universitárias**: um aporte teórico da arquitetura da informação. Tese (doutorado em Ciência da Informação) — Universidade Estadual Paulista, Faculdade de Filosofia e Ciências. Marília, 2017. Disponível em: <https://repositorio.unesp.br/handle/11449/151685>. Acesso em: 29 jul. 2021.

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. **Recuperação da informação**: conceitos e tecnologias das máquinas de busca. Tradução de Leandro Krug Wieves e Viviane Pereira Moreira. 2. ed. Minas Gerais: Bookman, 2013.

_____. **Modern Information Retrieval**. New York: ACM Press, 1999.

BEZERRA, Byron L. Dantas. **Estudos de algoritmos de filtragem de informação baseados em conteúdo**. Trabalho de Conclusão de Curso (graduação) — Universidade Federal de Pernambuco, curso de Inteligência Artificial. Recife, 2002.

BREITMAN, Karin. Big Data Seen from the Clouds. Palestra apresentada no **2nd. EMC Summer School on Big Data. EMC/NCE/UFRJ**. Rio de Janeiro. 2014. Disponível em: http://2014.emcbigdataschool.nce.ufrj.br/images/presentations/_Big_Data_Summer_School_Karin.pdf. Acesso em: 14 mar. 2021.

CAMPOS, Luiz Fernando de B. Web 2.0, biblioteca 2.0 e ciência da informação: um protótipo para disseminação seletiva da informação na web utilizando *mashups* e *feeds* rss. In: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 8., 2007, Salvador. **Anais...** Salvador: ENANCIB, 2007. Disponível em: <http://200.20.0.78/repositorios/bitstream/handle/123456789/336/GT2--232.pdf>. Acesso em: 17 set. 2021.

CAVALCANTI JÚNIOR, Flávio de Holanda. **Avaliação de técnicas de filtragem colaborativa para sistemas de recomendação**. Trabalho de Conclusão de Curso — Universidade Federal de Pernambuco, curso de Ciência em Computação. Recife, 2017. Disponível em: <https://www.cin.ufpe.br/~tg/2017-1/fhcj2-tg.pdf>. Acesso em: 13 jul. 2021.

CHANDLER, Otis. Recommendations and discovering good reads. **Goodreads**, 10 mar. 2011. Disponível em: <https://www.goodreads.com/blog/show/271-recommendations-and-discovering-good-reads>. Acesso em: 22 jul. 2021.

_____. Exciting News About Goodreads: We're Joining the Amazon Family!. **Goodreads**, 28 mar. 2013. Disponível em: <https://www.goodreads.com/blog/show/413-exciting-news-about-goodreads-we-re-joining-the-amazon-family>. Acesso em: 30 jul. 2021.

CHUNG, Kyusic. Announcing goodreads personalized recommendations. **Goodreads**, 15 sep. 2011. Disponível em: <https://www.goodreads.com/blog/show/303-announcing-goodreads-personalized-recommendations>. Acesso em: 22 jul. 2021.

CUNHA, Murilo Bastos da.; EIRÃO, Thiago Gomes. A atualidade e utilidade da disseminação seletiva da informação e da tecnologia rss. **Enc. Bibli: R. Eletr. Bib. Ci. Inf.**, Florianópolis, v. 17, n. 33, p. 59-78, jan./abr., 2012. Disponível em: <https://brapci.inf.br/index.php/res/download/52740>. Acesso em: 12 maio, 2021.

EIRÃO, Thiago Gomes. Disseminação seletiva da informação: uma abordagem. **Revista digital de Biblioteconomia e Ciência da Informação**, Campinas, v. 7, n. 1, p. 20-29, jul./dez. 2009. Disponível em: https://www.brapci.inf.br/_repositorio/2011/09/pdf_d55acad50a_0018786.pdf. Acesso em: 15 ago. 2021.

FERREIRA, Aurélio Buarque de Holanda. **Mini Aurélio**: o dicionário da língua portuguesa. 8. ed. Curitiba: Positivo, 2010.

FREUND, Jorge E.; TOMITA, Mari. Comunicação científica e tecnológica: a disseminação seletiva de informações. **R. Bibliotecon. Brasília**, v. 6, n. 2, p. 155-170. jul./dez. 1978. Disponível em: <https://periodicos.unb.br/index.php/rbbsb/article/view/29184/25024>. Acesso em: 01 maio. 2021.

FLOOD, Alison. Amazon purchase of Goodreads stuns book industry. **The Guardian**, 2 abr. 2013. Disponível em: <https://www.theguardian.com/books/2013/apr/02/amazon-purchase-goodreads-stuns-book-industry>. Acesso em: 30 jul. 2021.

FOLHA DE SÃO PAULO. Rede traz serviço de recomendação de livros. **Folha de S. Paulo Tec**, 16 mar. 2011. Disponível em: <https://www1.folha.uol.com.br/fsp/tec/tc1603201125.htm>. Acesso em: 22 jul. 2021.

G1. Amazon compra site de recomendação de livros Goodreads. **G1**: economia, 28 mar. 2013. Disponível em: <http://g1.globo.com/economia/noticia/2013/03/amazon-compra-site-de-recomendacao-de-livros-goodreads.html>. Acesso em: 30 jul. 2021.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2012.

_____. **Como elaborar projetos de pesquisa**. 6. ed. São Paulo: Atlas, 2021.

GOLDBERG, David; NICHOLS, David; OKI, Brian M.; TERRY, Douglas. Using collaborative filtering to weave an information tapestry. **Communications of the ACM**, [s.l.], v. 35, n. 12, p. 61-70, dec. 1992.

GOODREADS. About Goodreads. **Goodreads**, San Francisco, [20-?]. Disponível em: <https://www.goodreads.com/about/us>. Acesso em: 22 jul. 2021.

HOPKINS, Curt. Goodreads buys recommendation service discovered. **Redwrite**, 10 mar. 2011. Disponível em: https://readwrite.com/2011/03/10/goodreads_buys_recommendation_service_discovered/?__cf_chl_jschl_tk__=pmd_6cbe32971749485901399252a884eff786d7d517-1626980144-0-gqNtZGzNAmKjcnBszQti. Acesso em: 22 jul. 2021.

ISTOÉ. Número de usuários na internet chega aos 4,66 bilhões. **ISTOÉ**: dinheiro, n. 1232, mar. 2021. Disponível em: <https://www.istoedinheiro.com.br/numero-de-usuarios-de-internet-no-mundo-chega-aos-466-bilhoes/>. Acesso em: 29 jul. 2021.

KENT, Allen. **Manual da recuperação mecânica da informação**. Tradução de Beatriz Berrini. São Paulo: Polígono, 1972.

LÁZARO, Alexsandra da Silva. **Análise e seleção de algoritmos de filtragem de informação para solução do problema cold-start item**. Monografia (graduação) — Universidade Federal de Lavras, Curso de Sistemas de Informação. Lavras, 2010. Disponível em: http://repositorio.ufla.br/bitstream/1/5316/1/MONOGRRAFIA_Analise_e_selecao_de_algoritmos_de_filtragem_de_informacao_para_solucao_do_problema_cold-start_item.pdf. Acesso em: 29 jul. 2021.

LEE, Ellen. Goodreads' Otis Chandler reviews growth. **SFGATE**, 22 jul. 2012. Disponível em: <https://www.sfgate.com/default/article/Goodreads-Otis-Chandler-reviews-growth-3725030.php>. Acesso em: 22 jul. 2021.

LOPES, Ilza L. Estratégia de busca na recuperação da informação: revisão da literatura. **Ci. Inf.**, Brasília, v. 31, n. 2, p. 60-71, maio/ago. 2002. Disponível em: <https://www.scielo.br/j/ci/a/VPbDtPqgnFs5cm8GSLhtsyM/?format=pdf&lang=pt>. Acesso em: 04 ago. 2021.

LUNH, H. P. Selective dissemination of new scientific information with the AID of electronic processing equipment. **American Documentation**, v. 12, p. 131-138, 1961.

MAUERHOFF, George. R. Selective dissemination of Information. **Advances in Librarianship**, v. 4, p. 25-62, 1974a.

MELGÃO, Laura Margarida B. Goodreads. *In: _____*. **A experiência da leitura e participação digital dentro do grupo de leitores portugueses no goodreads**. Dissertação (Mestrado) — Instituto Universitário de Lisboa, Curso Comunicação, Cultura e Tecnologias da Informação. Lisboa, 2015. Disponível em: https://repositorio.iscte-iul.pt/bitstream/10071/10376/1/Laura_Melgao_disserta%c3%a7%c3%a3o.pdf. Acesso em: 22 jul. 2021.

MIRANDA, Wagner Rodrigues. Netflix: big data e os algoritmos de recomendação. *In: Congresso de Ciências da Comunicação na Região Sudeste*, 12. Anais da Sociedade Brasileira de Estudos Interdisciplinares da Comunicação, Volta Redonda, 2017. Disponível em: <https://portalintercom.org.br/anais/sudeste2017/resumos/R58-0517-1.pdf>. Acesso em: 13 mar. 2021.

NASCIMENTO, Luiz Antonio do. **Proposta de um sistema de recuperação de informação para extranet de projeto**. Dissertação (mestrado) — Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Construção Civil. São Paulo, 2004. Disponível em: <https://teses.usp.br/teses/disponiveis/3/3146/tde-13052004-140558/publico/NascimentoLA2004.pdf>. Acesso em: 18 ago. 2021.

PEREIRA FILHO, João Bosco A. **Um algoritmo de filtragem colaborativa baseado em SVD**. Dissertação (Mestrado) – Universidade Federal de Santa Catarina, Pós-graduação em Ciências da Computação. Florianópolis, 2010. Disponível em: <https://repositorio.ufsc.br/xmlui/bitstream/handle/123456789/94342/286638.pdf>. Acesso em: 13 jul. 2021.

PEREIRA, Alyson Bispo. **Sistemas de recomendação baseados em contexto físico e social**. Dissertação (mestrado) — Universidade Federal de Pernambuco, curso de Ciência da Computação. Recife, 2016. Disponível em: <https://repositorio.ufpe.br/bitstream/123456789/19521/1/risethesis.pdf>. Acesso em: 30 jul. 2021.

REATEGUI, Eliseo B.; CAZELLA, Sílvio C. Sistemas de Recomendação. *In*: CONGRESSO DA SOCIEDADE BRASILEIRA DA COMPUTAÇÃO, 25., 2005, São Leopoldo-RS. **Anais...** UNISINOS: São Leopoldo-RS, 2005. p. 306-348. Disponível em: https://scholar.google.com/citations?view_op=view_citation&hl=pt-BR&user=AxhyoRwAAAAJ&citation_for_view=AxhyoRwAAAAJ:u5HHmVD_uO8C. Acesso em: 03 ago. 2021.

RIBEIRO, Claudio J. Silva. Big Data: os novos desafios para o profissional da informação. **Informação & Tecnologia (ITEC)**, João Pessoa/Marília, v. 1, n. 1, jan./jun. 2014. p. 96-105. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/itec/article/view/19380/11156>. Acesso em: 14 mar. 2021.

SAMPAIO, Maria Imaculada Cardoso; MORESCHI, Erica Beatriz Pinto. DSI — Disseminação seletiva da informação: uma abordagem teórica. **R. Bras. Bibliotec. e Doc.**, São Paulo, v. 23, n. ¼, p. 38-57, jan./dez. 1990. Disponível em: https://brapci.inf.br/_repositorio/2011/09/pdf_d55acad50a_0018786.pdf. Acesso em: 29 abr. 2021.

SAMPAIO, Igor Azevedo. **Aprendizagem de filtragem colaborativa**. Dissertação (mestrado) — Universidade Federal de Pernambuco, curso de Ciência da Computação. Recife, 2006. Disponível em: https://repositorio.ufpe.br/bitstream/123456789/2608/1/arquivo5346_1.pdf. Acesso em: 02 ago. 2021.

SANTANA, Veronica Cardoso de. **Da disseminação da informação à disseminação seletiva da informação**: contribuições para o serviço de referência em bibliotecas. Monografia (graduação) — Universidade Federal de Sergipe, curso de Biblioteconomia e Documentação. São Cristóvão, 2014. Disponível em: <https://ri.ufs.br/bitstream/riufs/6701/2/Veronica%20Cardoso%20de%20Santana.pdf>. Acesso em: 29 abr. 2021.

SANTINI, Rose Marie. História e desenvolvimento dos sistemas de recomendação. *In*: _____. **O algoritmo do gosto**: os sistemas de recomendação on-line e seus impactos no mercado cultural. Curitiba: Appris, 2020.

SCHAFER, J. Ben; KONSTAN, Joseph; RIEDL, John. Recommender Systems in e-commerce. *In*: Conference on Electronic Commerce, 1999, Minneapolis. **Proceedings of the first ACM on Eletronic Commerce**, pp. 158-166. Disponível em: <https://dl.acm.org/doi/abs/10.1145/336992.337035>. Acesso em: 31 jul. 2021.

SCHWALBE, Will. Elizabeth Khuri Chandler tells origin story of goodreads. **But That's Another Story**, 3 dec. 2018. Disponível em: <https://lithub.com/elizabeth-khuri-chandler-tells-the-origin-story-of-goodreads/>. Acesso em: 22 jul. 2021.

SILVA, Guilherme Melo da. **Avaliação entre algoritmos de filtragem colaborativa baseada em vizinhança e transferência de conhecimento em CD-CARS.**

Monografia (graduação) — Universidade Federal Rural de Pernambuco, curso de Ciência da Computação. Recife, 2019. Disponível em:

https://webcache.googleusercontent.com/search?q=cache:ECFOxnQyblQJ:https://repositorio.ufrpe.br/bitstream/123456789/1873/1/tcc_guilhermemelodasilva.pdf+&cd=1&hl=pt-BR&ct=clnk&gl=br. Acesso em: 28 jul. 2021.

SILVA, Rafael Glauber Nascimento e. **Sistemas 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. Feira de Santana, 2014. Disponível em: <https://repositorio.ufba.br/ri/handle/ri/19281>. Acesso em: 29 jul. 2021.

SOUTO, Leonardo Fernandes. Disseminação seletiva da informação: discussões de modelos eletrônicos. **Enc. Bibli:** R. Eletr. Bibliotecon. Ci. Inf., Florianópolis, n. esp., 1º sem. 2006. Disponível em:

<https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2006v11nesp1p60/386>. Acesso em: 07. Set. 2021.

_____. **Informação seletiva, mediação e tecnologia:** a evolução dos serviços de disseminação seletiva da informação. Rio de Janeiro: Interciência, 2010.

TOKARNIA, Mariana. Brasil perde 4,6 milhões de leitores em quatro anos. **Agência Brasil:** educação, 20 set. 2020. Disponível em:

<https://agenciabrasil.ebc.com.br/educacao/noticia/2020-09/brasil-perde-46-milhoes-de-leitores-em-quatro-anos>. Acesso em: 04 ago. 2021.

APÊNCIDE A

(questionário submetido aos possíveis voluntários)

1) Gênero:

- Feminino Masculino Fluído Não-binário Outro
- Não quero informar

2) Idade: _____

3) Escolaridade:

- Fundamental completo Ensino Médio incompleto E. M. completo
- Ensino Superior completo Ensino Superior incompleto

Caso esteja no Ensino superior ou tenha-o concluído, qual a área do seu curso?

4) Quantos livros você leu nos últimos dois anos ano:

- Nenhum Até 2 livros Entre 3 e 5 livros
- Entre 6 e 10 livros + de 10 livros

5) Quais seus gêneros literários preferidos?

- Infantil Infanto-juvenil Romance Fantasia
- Distopia Suspense Hot Policial
- Não-ficção Ficção-histórica

Quer mencionar mais um? Diga qual: _____

6) Você utiliza alguma dessas redes sociais literárias? Se sim, marque-as.

- Universidade Falada Spirit Fanfiction Goodreads

Skoob Many Books Wattpad

Não utilizo nenhuma dessas redes

Alguma outra? _____

7) Você já leu algum livro em outro idioma?

Sim Não

8) Qual seu nível de leitura e compreensão em inglês?

Baixo Alto Intermediário Nulo, não sei nada de inglês

APÊNDICE B

(entrevista semiestruturada)

Esta entrevista será gravada apenas para fins acadêmicos, mantendo o nome dos voluntários em anonimato. Após a tabulação de dados, a gravação será excluída e os únicos dados sobre constarão apenas na monografia intitulada “**SISTEMAS DE RECOMENDAÇÃO COMO FERRAMENTA PARA INDICAÇÕES DE LIVROS: UM ESTUDO DE CASO NA PLATAFORMA GOODREADS**”.

1. Em primeiro lugar, você teve alguma dificuldade ao fazer o login inicial na plataforma?
2. Boa parte dos livros que você organizou nas suas prateleiras foram livros em inglês, por fez essa escolha?
3. Você indicou seis gêneros literários de interesse no questionário e outros seis durante a construção no perfil do Goodreads. Apesar de sua média de leitura ser dois livros por ano, você se considera uma pessoa interessado no mundo literário?
4. Sobre a aba “recomendações”, você se interessou por algum dos livros indicados? Algum chamou mais atenção? Por quê?
5. As abas “Lançamentos” e “Goodreads Choice Awards” indicaram algum livro que você não conhecia? Algum deles chamou sua atenção?
6. Você considera que as recomendações estão de acordo com o seu gosto literário? Houve alguma recomendação que não esteve do seu agrado?
7. De forma geral, você se sentiu satisfeito com o Goodreads? Você acha que voltará a utilizá-lo?
8. Quais aspectos da plataforma você acredita que possa melhorar?