

UNIVERSIDADE DO ESTADO DE SANTA CATARINA – UDESC
CENTRO DE CIÊNCIAS HUMANAS E DA EDUCAÇÃO – FAED
PROGRAMA DE PÓS-GRADUAÇÃO EM GESTÃO DA INFORMAÇÃO

RICARDO DI GIACOMO DONHA

**PROPOSTA DE SISTEMA DE RECOMENDAÇÃO DE ARTIGOS CIENTÍFICOS DE
PERIÓDICOS DE ACESSO ABERTO**

FLORIANOPOLIS

2022

RICARDO DI GIACOMO DONHA

**PROPOSTA DE SISTEMA DE RECOMENDAÇÃO DE ARTIGOS CIENTÍFICOS DE
PERIÓDICOS DE ACESSO ABERTO**

Dissertação apresentada ao Programa de Pós-Graduação em Gestão da Informação, do Centro de Ciências Humanas e da Educação, da Universidade do Estado de Santa Catarina, como requisito parcial para obtenção do título de Mestre em Gestão de Unidades de Informação. Linha de Pesquisa: Gestão de unidades de Informação.

Orientador: Jordan Paulesky Juliani

FLORIANÓPOLIS

2022

**Ficha catalográfica elaborada pelo programa de geração automática da
Biblioteca Setorial do FAED/UDESC,
com os dados fornecidos pelo(a) autor(a)**

Donha, Ricardo Di Giacomo

Proposta de sistema de recomendação de artigos científicos de
acesso aberto / Ricardo Di Giacomo Donha. -- 2022.

111 p.

Orientador: Jordan Paulesky Juliani Juliani

Dissertação (mestrado) -- Universidade do Estado de Santa
Catarina, Centro de Ciências Humanas e da Educação, Programa de
Pós-Graduação Profissional em Gestão de Unidades de Informação,
Florianópolis, 2022.

1. Sistemas de recomendação. 2. Acesso aberto. 3. Artigos
científicos. 4. Periódicos científicos. I. Juliani, Jordan Paulesky
Juliani. II. Universidade do Estado de Santa Catarina, Centro de
Ciências Humanas e da Educação, Programa de Pós-Graduação
Profissional em Gestão de Unidades de Informação. III. Título.

RICARDO DI GIACOMO DONHA

**PROPOSTA DE SISTEMA DE RECOMENDAÇÃO DE ARTIGOS CIENTÍFICOS DE
PERIÓDICOS DE ACESSO ABERTO**

Dissertação apresentada ao Programa de Pós-Graduação em Gestão da Informação, do Centro de Ciências Humanas e da Educação, da Universidade do Estado de Santa Catarina, como requisito parcial para obtenção do título de Mestre em Gestão de Unidades de Informação. Linha de Pesquisa: Gestão de unidades de Informação.

Orientador: Jordan Paulesky Juliani

BANCA EXAMINADORA

Jordan Paulesky Juliani, Doutor em Engenharia e Gestão do Conhecimento

Universidade do Estado de Santa Catarina

Membros:

Jordan Paulesky Juliani, Doutor em Engenharia e Gestão do Conhecimento

Universidade do Estado de Santa Catarina

Jorge Moisés Kroll Prado, Doutor em Ciência da Informação

Universidade do Estado de Santa Catarina

Raphael Winckler de Bettio, Doutor em Engenharia e Gestão do Conhecimento

Universidade Federal de Lavras

Florianópolis, 25 de abril de 2022.

AGRADECIMENTOS

À Universidade Estadual de Santa Catarina, pela oportunidade de poder cursar um Programa de Mestrado gratuito e de qualidade.

Aos docentes do PPGInfo, em especial ao meu orientador, professor Jordan Paulesky Juliani, por todo apoio e direcionamento.

À minha família, por todo o suporte e compreensão em todas as fases.

Aos amigos que sempre incentivaram em todos os momentos.

RESUMO

Os sistemas de recomendação estão embutidos nos diversos tipos de interação que temos com a internet e, sutilmente, recebemos inúmeras recomendações no momento que acessamos a *web*. Além de serem úteis na área comercial, os sistemas de recomendação se destacam também na área acadêmica, ao utilizar ferramentas para analisar os diversos dados produzidos na interação do pesquisador quando esse navega, visita e realiza pesquisas nos sites na web em busca de informações científicas. Paralelo a isso, os sistemas de recomendação são capazes de personalizar o que é exibido ao pesquisador, ou seja, quanto mais se utiliza, se consome ou se pesquisa via internet, mais sugestões pertinentes são dadas. O surgimento e a expansão do movimento do Acesso Aberto, desde o seu surgimento, contribuíram para que os artigos científicos fossem disponibilizados em periódicos dessa natureza, fazendo com que esses documentos científicos fossem ser acessados por um número maior de pessoas. Dentro desse contexto, esta pesquisa possui como objetivo geral apresentar um modelo de Sistema de Recomendação de artigos científicos de Acesso Aberto. Para atingir este resultado, foram traçados objetivos específicos, nos quais, inicialmente, se fez uma breve análise sobre os artigos científicos, os periódicos em Acesso Aberto e o movimento do Acesso Aberto. Em seguida, buscou-se apresentar os sistemas de recomendação, seus conceitos, suas funcionalidades, quais os algoritmos utilizados, os tipos de filtragens, e a existência de sistemas de recomendação de artigos científicos, esses sendo analisados e caracterizados. Após avaliar e comparar os sistemas existentes, foi proposto um modelo de requisitos de Sistema de Recomendação de artigos científicos de periódicos de Acesso Aberto que possua características próprias, atue de forma a entregar informação pertinente e proativamente ao pesquisador.

Palavras-chave: Sistemas de recomendação; Acesso Aberto; Artigos científicos; Periódicos de Acesso Aberto.

ABSTRACT

Recommender systems are embedded in the various types of interaction we have with the internet. We subtly receive numerous recommendations the moment we access the web. In addition to being useful in the commercial area, recommender systems stand out in the academic area. By using tools to analyze the various data produced in the interaction of the researcher when he browses, visits and performs research on websites in search of scientific information, recommender systems are able to customize what is displayed to the researcher. The more it is used, consumed or researched via the internet, the more pertinent suggestions are given. The appearance and growth of the Open Access movement, since its inception, has contributed to making scientific articles available in open access journals, making these scientific documents accessible to an increasing number of people. Within this context, this research has the general objective to present a model of requirements of an open access scientific articles recommender system. To achieve this objective, specific objectives were outlined, where initially, a brief analysis was made on scientific articles, open access journals and the open access movement. Then we sought to present the recommender systems, their concepts, functionalities, algorithms used, types of filtering, and the existence of recommender systems for scientific articles, these were analyzed and characterized. After evaluating and comparing the existing systems, a model of requirements for a system for recommending scientific articles from open access journals was proposed that has its own characteristics, acts in order to deliver relevant and proactive information to the researcher.

Keywords: Recommender systems; Open Access; Scientific papers; Open access journals.

LISTA DE FIGURAS

Figura 1 - Interesses em comum dos usuários	20
Figura 2 - Etapas gerais de um SR	21
Figura 3 - FC para recomendação de artigos	27
Figura 4 - Filtragem baseada em conteúdo – FBC	29
Figura 5 - SR de artigos baseado em FBC.....	32
Figura 6 - Filtragem híbrida.....	34
Figura 7 - SR de artigos baseado em FH	35
Figura 8 - Interface do Google Acadêmico com artigos recomendados	38
Figura 9 - Interface do Google Acadêmico	39
Figura 10 - Funcionalidades premium academia.edu	40
Figura 11 - E-mail com alerta de novos documentos	41
Figura 12 - Tela com acesso restrito.....	42
Figura 13 - Recomendação de artigo na plataforma academia.edu	43
Figura 14 - Interface com opções de download da Academia.edu	44
Figura 15 - Evidências de recomendações no Researchgate	46
Figura 16 - Visão macro da arquitetura do protótipo W-RECMAS.....	48
Figura 17 - Diagrama de casos de uso do sistema	50
Figura 18 - Interface do arXiv com a aba de recomendações	52
Figura 19 - Plugin do SR da CORE.....	54
Figura 20 - Arquitetura básica do sistema SysRecAC	56
Figura 21 - Subdivisões de SR e alguns possíveis algoritmos	58
Figura 22 - Exemplo de classificação KNN	59
Figura 23 - Exemplo de árvore de decisão para frequência de compras de um produto.....	60
Figura 24 - Rede neural	62
Figura 25 - Exemplo de caso de uso.....	68
Figura 26 - Exemplo de índice em um documento de requisito	69
Figura 27 - Detalhamento de requisito	70
Figura 28 - Documento de requisito IEEE 830-1998	71
Figura 29 - Estrutura para “casos de uso” conforme IEEE 830-1998.....	74
Figura 30 - Fluxo de informação nos provedores de dados.....	82
Figura 31 - Fluxo de informação em um Provedor de Serviço	83
Figura 32 - Funcionamento do protocolo OAI-PMH	84
Figura 33 - Visão geral do sistema RECOAR.....	90

LISTA DE QUADROS

Quadro 1 - Objetivos e recursos de estilos de filtragem.....	24
Quadro 2 - Matriz usuário-item.....	26
Quadro 3 - Métodos para FH.....	35
Quadro 4 - Exemplo de avaliações de músicas	63
Quadro 5 - Manifestos sobre o movimento do AA	78
Quadro 6 - Procedimentos metodológicos utilizados	86
Quadro 7 - Quadro Comparativo de SR	89

LISTA DE ABREVIATURAS

AA	Acesso Aberto
BOAI	Iniciativa de Acesso Aberto de Budapeste
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CI	Ciência da Informação
FBC	Filtragem Baseada em Conteúdo
FC	Filtragem Colaborativa
FH	Filtragem Híbrida
KNN	K-vizinhos mais próximos
SR	Sistema de Recomendação
UDESC	Universidade do Estado de Santa Catarina

SUMÁRIO

1	INTRODUÇÃO	12
1.1	OBJETIVOS.....	15
1.1.1	Objetivo Geral	15
1.1.2	Objetivos específicos.....	16
1.2	JUSTIFICATIVA	16
2	REFERENCIAL TEÓRICO	18
2.1	SISTEMAS DE RECOMENDAÇÃO.....	18
2.1.1	Conceitos e propósitos.....	18
2.1.2	Funcionalidades de um SR	19
2.1.3	Tipos de filtragens	25
2.1.3.1	<i>Filtragem Colaborativa.....</i>	<i>25</i>
2.1.3.2	<i>Filtragem Baseada Em Conteúdo</i>	<i>28</i>
2.1.3.3	<i>Filtragem Híbrida</i>	<i>33</i>
2.1.3.4	<i>Filtragem Baseada no Conhecimento</i>	<i>36</i>
2.2	SISTEMAS DE RECOMENDAÇÃO DE ARTIGOS CIENTÍFICOS EXISTENTES	37
2.2.1	Google acadêmico	37
2.2.2	Academia.edu.....	40
2.2.3	Researchgate	44
2.2.4	Protótipo W-RECMAS	47
2.2.5	Protótipo SysPaperAdvice	49
2.2.6	arXiv	51
2.2.7	CORE	53
2.2.8	SisRecAC.....	55
2.3	ALGORITMOS DE SISTEMAS DE RECOMENDAÇÃO	57
2.3.1	K-vizinhos mais próximos.....	58
2.3.2	Arvores de decisão.....	60
2.3.3	Redes Neurais	61
2.3.4	Regressão Linear	63
2.3.5	TF-IDF.....	64
2.4	ANÁLISE E MODELAGEM DE SISTEMAS DE INFORMAÇÃO.....	65
2.4.1	Levantamento de requisitos de informações.....	66

2.4.2	Documento de requisitos.....	69
2.5	PERIÓDICOS CIENTÍFICOS COMO FONTE DE INFORMAÇÃO E CONHECIMENTO	75
2.5.1	Periódicos científicos e o movimento de Acesso Aberto.....	77
2.5.2	A interoperabilidade e os periódicos científicos de Acesso Aberto – o protocolo OAI-PMH.....	80
3	PROCEDIMENTOS METODOLÓGICOS	85
4	RESULTADOS DA PESQUISA	88
4.1	CONSIDERAÇÕES SOBRE ESTRATÉGIAS DE FILTRAGENS PARA UM SR DE ARTIGOS CIENTÍFICOS	88
4.1.1	Análise comparativa dos SR encontrados na pesquisa	89
4.2	ARQUITETURA DO SISTEMA.....	90
4.3	DOCUMENTO DE REQUISITOS.....	92
5	CONSIDERAÇÕES FINAIS.....	102
	REFERÊNCIAS	104

1 INTRODUÇÃO

O surgimento e o desenvolvimento de novas tecnologias contribuíram diretamente para transformações profundas na sociedade, assim como na Ciência. O compartilhamento de informações online facilitou sobremaneira a sua localização, esteja ela contida em trabalhos científicos, em notícias, ou em catálogos de produtos e serviços. Nota-se que diferentes tipos de informação em variados suportes podem ser encontradas com apenas alguns cliques, pois a era digital e a evolução da internet criaram inúmeras facilidades e benefícios para os usuários.

Em virtude do crescimento do uso das tecnologias digitais e do avanço da Internet, ocorreu um aumento do volume de dados e informações disponíveis, existindo uma abundante quantidade de informação para ser processada. Consequentemente, o pesquisador que está em busca de informação relevante, que atenta aos seus propósitos de pesquisa, pode ter dificuldade em identificar aquelas que podem efetivamente contribuir para sua investigação.

A Pesquisa científica segue um conjunto de etapas e metodologias a serem cumpridas de forma sistemática para o atingimento do seu objetivo. Prodanov (2013), explica que a pesquisa considerada como científica, perpassa desde a formulação do problema até a apresentação dos resultados, cumprindo as seguintes fases: preparação da pesquisa (definição e delimitação do problema); coleta de dados; análise e discussão dos dados; e por fim, preparação do relatório da pesquisa.

Conforme Gil (2008), a pesquisa possui como objetivo essencial descobrir respostas para problemas utilizando procedimentos científicos. Durante o processo de pesquisa, a revisão bibliográfica ou fundamentação teórica demanda que o pesquisador realize um levantamento de fontes de informação científicas relevantes.

Portanto, as características que distinguem o conhecimento científico de outros tipos de conhecimento são: possuir uma fundamentação e metodologia a ser seguida, estar baseada em informações classificadas, submetidas à verificação e que oferecem explicações plausíveis sobre o estudo em questão (PRODANOV, 2013).

As fontes de informações se dividem em três categorias, conforme sua natureza, explicadas a seguir por Pereira (2012):

- Primárias – informações novas e originais (livros e artigos de periódicos);
- Secundárias – resultado de análises de fontes primárias, informações filtradas e organizadas de acordo com esquemas determinados (revisões e enciclopédias);
- Terciárias – informações organizadas de forma a se tornarem mais acessíveis, compilação de fontes primárias e secundárias (bibliografias, índices e catálogos).

Brito e Lima (2015) atestam que o periódico científico, impresso ou eletrônico, é atualmente uma das fontes de informação mais utilizadas pela comunidade científica, estabelecendo-se como um meio para expor os resultados de pesquisas, submetendo-os ao julgamento e à contribuição dos pares.

Os periódicos científicos funcionam como veículo para a propagação do conhecimento, desde as suas versões em papel, bem como nas publicações atualmente realizadas em meio eletrônico. Garvey (1979) aponta que a ciência opta por fontes associadas ao domínio formal, com senso de estética e lealdade às informações, atributos do periódico científico.

Lima, Mendes Junior e Pinto (2019) atestam que atualmente é possível acessar a maior parte dos artigos científicos somente mediante à assinatura de bases de dados e/ou de periódicos, que é geralmente financiada pelas bibliotecas das instituições de ensino superior e por centros de pesquisa, ou então custeada individualmente pelos próprios pesquisadores. Porém, existem periódicos que oferecem acesso aos artigos de forma gratuita, enquanto outros não funcionam dessa forma isenta de pagamento, no qual o acesso pode implicar em uma cobrança monetária para o usuário, tornando-se dispendioso e impraticável. Os periódicos oferecidos de forma gratuita são chamados de periódicos de Acesso Aberto, e são instrumentos de comunicação científica que não restringem o acesso.

Um fator importante que impulsiona o crescimento da divulgação de informação científica é o movimento do Acesso Aberto (AA), que busca democratizar o acesso ao conhecimento científico. De acordo com o glossário da Fundação Oswaldo Cruz (FIOCRUZ, 2022, não paginado), AA é definido como a “Disponibilização na Internet de literatura de caráter acadêmico ou científico, permitindo a qualquer usuário ler, baixar (fazer *download*), copiar, distribuir, imprimir, pesquisar ou referenciar (fornecer *link*) o texto integral dos documentos”, portanto, a principal característica do AA é oferecer acesso à informação científica de forma gratuita.

Pode-se dizer que a relação entre o pesquisador e a fonte de informação, em especial os artigos científicos, acontece, geralmente, de uma forma reativa, ou seja, uma vez identificada uma oportunidade de investigação o pesquisador inicia a coleta de dados. Atualmente os pesquisadores também podem contar com um modelo “proativo” no acesso aos artigos científicos, por meio dos sistemas de recomendação (SR), que de alguma forma identificam o perfil do usuário e com base nele recomendam documentos científicos.

Um SR é um sistema que funciona como um filtro de informação e entrega para o usuário itens ou objetos (filmes, livros, artigos, páginas da *web*, entre outros) que provavelmente sejam do seu interesse (SHAFFER, 2001 *apud* MOTTA *et al.*, 2011).

Alguns SR disponíveis na internet foram concebidos para recomendar documentos científicos para pesquisadores, entre eles se destacam o Google Acadêmico, Researchgate, Academia, arXiv entre outros. Também existem protótipos, modelados em ambiente acadêmico, que foram pensados para o mesmo fim, como o W-RECMAS, SysPaperRec, SisRecAC.

Academia.edu, por exemplo, é uma plataforma que funciona como uma rede social para pesquisadores compartilharem seus trabalhos, seus artigos, seus livros e suas pesquisas, possibilitando acesso de forma gratuita. Outra ferramenta muito usada e que funciona de forma diferente da plataforma citada anteriormente é o Google Acadêmico, que, segundo Ciriaco (2015), é um buscador elaborado pela empresa Google que possui em sua plataforma teses, resumos, livros, dissertações, monografias e artigos científicos, auxiliando na busca de referências para trabalhos científicos. A ferramenta indexa conteúdos de bases de Acesso Aberto como Scielo, Altametric e Wiley, ou de materiais que estejam disponíveis no Google Livros, e ela pode definir alertas para receber notificações de novas publicações dentro de um tema que seja do interesse do usuário.

De acordo com Bai *et al.* (2019), sistemas de recomendação de artigos tornaram-se ferramentas indispensáveis no campo acadêmico, nos quais os seus algoritmos são atualizados de forma contínua fazendo com que a precisão das recomendações progrida com o tempo.

Salienta-se que os SR estão presentes nas nossas interações com o ambiente virtual, principalmente quando utilizamos websites e aplicativos de compras, de filmes e de músicas. A relevância dos SR para as organizações comerciais que atuam na internet pode ser mais bem percebida, em 2009, quando a Netflix, provedora de filmes e séries via *streaming*, concedeu um prêmio de US\$ 1 milhão para uma equipe que desenvolveu um algoritmo que aumentou a precisão do mecanismo de recomendação da empresa em 10% apenas (WIRED, 2012).

Grandes empresas do varejo vêm utilizando os SR e obtendo excelentes resultados no incremento das vendas. A varejista norte americana *Amazon* reportou que 35% do que é consumido é diretamente influenciado por recomendações na sua plataforma online e ainda registrou um aumento nas suas vendas na ordem de 29% em relação ao mesmo período do ano anterior. Muito desse crescimento se deve a forma com que a empresa integrou suas recomendações em quase todas as partes do processo de compra (FORTUNE, 2012).

As empresas que dispõem do comércio eletrônico como um meio de realizar suas vendas utilizam os SR para auxiliar os seus clientes e eventuais consumidores a encontrar os itens de sua preferência. A grande quantidade de itens, de categorias, de marcas e de tipos de produtos que são oferecidos com maior facilidade pelo comércio eletrônico podem dificultar ou tornar

cansativa a navegação, dessa forma, os SR entram em ação e coletam informações dos usuários no sentido de mapear as suas preferências e ofertar produtos que sejam relevantes, agregando valor às empresas que os utilizam.

Assim, um SR trata de identificar similaridade entre o comportamento dos usuários e de recomendar itens, textos, vídeos, entre outros que já foram consultados ou utilizados por outros usuários, podendo levar em consideração competências e perfis (COSTA; AGUIAR; MAGALHÃES, 2013).

Observando o valor dos SR de artigos científicos para os pesquisadores e a crescente do movimento do AA, este projeto de pesquisa nasceu com o intuito de modelar um SR que atue de forma proativa, entregando ao pesquisador documentos novos que façam parte do seu interesse de pesquisa e que sejam compatíveis com o Qualis escolhido por ele, a partir de artigos científicos publicados em periódicos de Acesso Aberto. Ao recomendar artigos científicos assertivamente, o sistema poderá gerar novas ideias e inspirações, permitir ao pesquisador perceber novas oportunidades de pesquisa, estabelecer relações com outras áreas do conhecimento a partir de um mesmo eixo temático, entre outros possíveis benefícios esperados.

Em face desse interesse de manter o pesquisador provido de informações que vão ao encontro dos seus interesses é que os SR se apresentam como um suporte, evitando perda de tempo com informações em excesso, filtrando e enviando apenas resultados de acordo com o perfil do usuário que o utiliza.

Considerando a importância de recomendações objetivas e oportunas que vão ao encontro dos interesses do usuário/pesquisador, surge então, a seguinte pergunta de pesquisa: **“Quais são os requisitos funcionais desejáveis em um SR para indicar artigos científicos de periódicos de Acesso Aberto?”**

1.1 OBJETIVOS

A seguir serão definidos o objetivo geral e os objetivos específicos para o presente estudo, com a finalidade de responder o problema de pesquisa apresentado anteriormente.

1.1.1 Objetivo Geral

Modelar os requisitos de um Sistema de Recomendação de artigos científicos de periódicos de Acesso Aberto.

1.1.2 Objetivos específicos

Com a finalidade de atingir o objeto geral foram definidos os seguintes objetivos específicos:

- a) Caracterizar os periódicos científicos e o movimento de Acesso Aberto;
- b) Analisar as características dos sistemas de recomendação de artigos científicos existentes;
- c) Analisar tecnologias/algoritmos para realizar a recomendação.

1.2 JUSTIFICATIVA

Seguindo o conceito de Roesch (2009), um projeto pode ser justificado por meio de sua importância, oportunidade e viabilidade.

De acordo com Lu *et al.* (2015), além de outras áreas, os SR são aplicados em domínios como bibliotecas digitais, áreas educacionais, governo eletrônico e negócios eletrônicos. Assim, no âmbito educacional da graduação e pós-graduação, um SR, voltado para a busca de artigos científicos de periódicos de Acesso Aberto, que mantenha o pesquisador bem-informado em relação ao seu campo de atuação, pode instigá-lo a expandir sua pesquisa, identificando novas oportunidades de investigação.

Esta pesquisa poderá contribuir para o Programa de Pós-Graduação em Gestão da Informação – PPGInfo, da Universidade do Estado de Santa Catarina – UDESC, disponibilizando um resultado de pesquisa em forma de um produto, resultante de uma pesquisa aplicada altamente compatível com a sua proposta do mestrado profissional.

Este trabalho também pode ser considerado como oportuno, uma vez que este pesquisador está cursando o programa de mestrado, e, portanto, convivendo com uma grande quantidade de informação científica e com a necessidade de ter referências bibliográficas que vão ao encontro da pesquisa, seja em forma de livros ou de artigos, para a fundamentação e desenvolvimento de sua pesquisa.

A viabilidade da pesquisa justifica-se pelo fato de ter havido tempo hábil para a realização do trabalho, e de não ter sido necessário o emprego de recursos materiais e/ou financiamento. Ademais, os recursos humanos diretamente envolvidos no desenvolvimento do estudo (orientador e mestrando) possuíam conhecimentos prévios sobre a temática abordada.

Ainda, para a Ciência da Informação (CI), este trabalho mostra-se como uma oportunidade de modelar uma ferramenta que pode ser utilizada por inúmeros usuários,

auxiliando na obtenção de informação e conhecimento científico de acordo com o seu interesse de pesquisa.

Em relação à existência de outros SR de artigos científicos, esta proposta se diferencia dos sistemas encontrados uma vez que propõe a integração com dados provenientes da plataforma Sucupira, desenvolvida pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, em especial o Qualis Periódicos. Conforme Aguiar (2022), o Qualis Periódicos é um conjunto de procedimentos utilizado na classificação da produção científica dos programas de pós-graduação, avaliando a qualidade dos artigos e outros tipos de produção, a partir dos veículos de divulgação, isto é, os periódicos científicos. A consulta ao Qualis é realizada por meio da ferramenta online denominada plataforma Sucupira.

Além disso, a recomendação de artigos no sistema proposto deve somente se basear em periódicos de Acesso Aberto, isso significa que os artigos serão disponibilizados de forma totalmente gratuita. A integração com os dados do sistema Qualis, tornará possível que o usuário estabeleça, na definição do seu perfil, quais áreas do conhecimento e quais estratos Qualis de periódicos lhe interessam, aumentando potencialmente a assertividade das recomendações. Essa integração com os dados do Qualis Periódicos caracteriza a proposta de um Sistema de Recomendação singular, pensado para pesquisadores brasileiros, e todos aqueles vinculados, de alguma forma, com pesquisa científica desenvolvida no Brasil.

2 REFERENCIAL TEÓRICO

Considerando os objetivos propostos nesta pesquisa, foram apresentados nesta seção os conceitos e a aplicabilidade dos SR, assim como os tipos existentes e suas características, suas vantagens, suas desvantagens e seu histórico, no sentido de compreender o seu funcionamento. Além disso, foram abordadas as temáticas dos periódicos científicos e do movimento de AA.

2.1 SISTEMAS DE RECOMENDAÇÃO

Neste tópico serão abordados os conceitos, funcionalidades e tipos de filtragem dos SR.

2.1.1 Conceitos e propósitos

SR são diferentes de sistemas de busca ou sistemas de recuperação de informação, principalmente pelo critério de individualização e dos atributos de interesse e de utilidade específicos para o usuário nas recomendações (BURKE, 2002).

Os SR atualmente fazem parte da vida de todas as pessoas que utilizam meios eletrônicos para acesso à *web*, e acabam por automatizar processos comuns no dia a dia. Um SR indica, por exemplo, qual o próximo artigo a ser lido, a próxima música a ser ouvida, a próxima pessoa para interagir ou para conhecer.

Gatto e Zorzo (2018) explicam que os SR utilizam algoritmos da ciência cognitiva, tais quais: filtragem de informação, mineração de dados, teoria da aproximação, entre outros.

A ideia básica dos SR é utilizar algumas informações fornecidas pelo usuário para deduzir quais são os seus interesses, funcionando como um mapeamento desses para o alcance do objetivo. Usuário é a entidade que recebe a recomendação e o produto dessa recomendação é chamado de item. Considerando esse cenário, um item pode ser considerado qualquer coisa que possa ser recomendada para o usuário. A recomendação é baseada na interação entre usuário e item, e o histórico dessa interação costuma ser um bom indicador para novas recomendações (AGGARWAL, 2016).

De acordo com Ricci, Rokach e Shapira (2010), os SR são ferramentas e técnicas de software que fornecem sugestões de itens que um usuário possa desejar utilizar.

Em comparação com um sistema que procura apenas informações ou que tem simplesmente funcionalidades de filtragem de informação, os SR exigem menos experiência por parte do utilizador e menos esforço para especificar e reter os seus interesses, porque

esses sistemas fornecem recomendações que foram reconhecidas como satisfatórias pelos seus utilizadores, baseados nas suas preferências previamente expressas ou sobre as preferências de outros utilizadores que têm os mesmos interesses (RESNICK; VARIAN, 1997).

Sobre os elementos que constituem parte de um SR, Mota *et al.* (2011, p. 231), afirmam que:

Os principais componentes de um Sistema de Recomendação são cliente e produto. Um produto é um recurso que pode ser de diferentes naturezas, por exemplo: um conteúdo, um arquivo, uma informação, uma pessoa, um objeto. A recomendação é uma função de mapeamento de interesses do cliente para obtenção de um ou mais produtos.

Ainda, a autora explica que, uma recomendação R para a escolha de “ p ” deve ser feita de tal forma que: $R(c,p) = \max F(c, p^i)$. Aqui, considerando que:

“ c ” representa o cliente que utiliza o SR;

P representa um conjunto de produtos que podem ser avaliados;

$p^i \in P$;

F é a função que define a relevância de p^i em relação a c .

Para Aggarwal (2016), um SR possui inúmeros objetivos que compreendem os tópicos abaixo:

- Relevância: um SR deve gerar recomendações que realmente sejam relevantes e tenham seu valor percebido pelo usuário;
- Novidade: no sentido de recomendar itens ainda não conhecidos pelo usuário. Se o SR repetir recomendações, provavelmente não será mais usado;
- Serendipidade: como forma de promover a descoberta de itens inesperados, ao acaso, diferente daquilo que se estava procurando;
- Diversidade: fazendo com que o usuário não fique entediado com a recomendação de itens semelhantes.

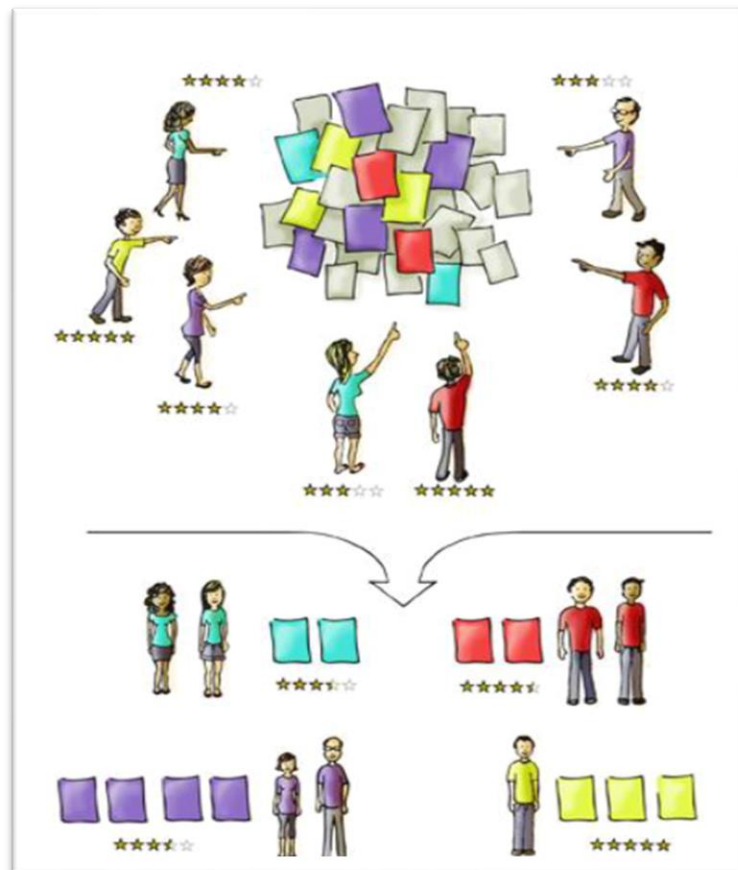
Conforme Cazella, Drumm e Barbosa (2010), os SR permitem personalizar os sistemas web, na medida que são capazes de identificar preferências e sugerir produtos relevantes para cada usuário, por meio da análise do seu comportamento de navegação, produtos comprados ou desejados, entre outras particularidades.

2.1.2 Funcionalidades de um SR

Aggarwal (2016) explica que quando um usuário possui interesse em certo assunto, por exemplo um documento histórico, esse usuário estará mais propenso a se interessar futuramente

por outros documentos ou assuntos históricos, e afirma que o princípio básico das recomendações é que existem dependências significantes entre o usuário e a sua atividade com o item. Isso mostra que alguns tipos de SR irão nos retornar informação relevante de algum usuário que possui um interesse similar ao nosso, os SR possuem diferentes abordagens para recomendar, e uma delas se baseia em avaliações de usuários com perfis em comum.

Figura 1 - Interesses em comum dos usuários



Fonte: Motta *et al.* (2011, p. 232).

Neste caso acima, podemos dizer que um SR funciona de forma colaborativa, uma vez que a recomendação surge da ordenação e da estruturação de avaliações individuais que contribuem para avaliação de produtos que serão acessados por outros indivíduos (MOTTA *et al.*, 2011). Ainda, a autora reitera que os SR se baseiam em perceber que o que é relevante para um usuário também pode ser para algum outro que possua interesse similar, conforme representa a Figura 1.

Para que os SR possam funcionar de forma satisfatória, é necessário identificar o usuário e extrair o máximo de informação possível, além de seus dados pessoais, dos seus interesses, dos seus gostos, de suas preferências, entre outros. Duas formas são consideradas mais

frequentes para essa identificação quando esse acessa algum tipo de sistema, seja ao abrir uma conta em um banco ou ao realizar uma compra em um *website*, segundo Cazela e Reategui (2005):

- Identificação no servidor – funciona na hora de realizar o cadastro, disponibilizando um espaço onde o usuário inclui informações pessoais, além de criar um *login* e uma senha. O servidor armazena essas informações em um banco de dados e, desta forma, consegue identificar o usuário quando este realiza a autenticação utilizando o *login* cadastrado.
- Identificação no cliente – usualmente usa *cookies*, um mecanismo que possibilita identificar o computador que está se conectando a um *website*, assumindo que a máquina é utilizada pela mesma pessoa sempre. Consequentemente, ao identificar a máquina está identificando o usuário, porém torna-se menos confiável quando existe a possibilidade do computador ser utilizado por mais de uma pessoa.

Existem alguns meios para que os SR descubram as informações necessárias sobre o usuário, para que possam trazer recomendações que supostamente interessem a ele. Essas informações podem ser coletadas de forma explícita, quando o próprio usuário cria um perfil no qual informa seus interesses, ou de forma implícita, onde o sistema propõe os gostos baseado nas suas ações (ALVAREZ *et al.*, 2016).

Abaixo, a Figura 2 representa as etapas básicas de um SR:

Figura 2 - Etapas gerais de um SR



Fonte: Elaborada pelo autor (2021).

Portanto, quando é criado um perfil em uma rede social, por exemplo, é fornecido voluntariamente, de forma explícita, além dos dados cadastrais como nomes e endereço,

características sobre nós mesmos, gostos e preferências, que serão utilizadas futuramente para recomendar produtos, serviços, e relacionamentos interpessoais. Por outro lado, nessa mesma rede social citada anteriormente, é fornecida uma grande quantidade de informação sobre nós de forma implícita, pelo histórico de páginas visitadas, pelas ‘curtidas’ em algum tipo de publicação, quando se clica em anúncios de produtos/serviços, quando se avalia algum item, entre outras ações que possam ser usadas para inferir sobre o perfil do usuário. Essas informações ficam registradas no perfil quando o usuário está identificado desde o início da interação, facilitando o mapeamento das suas características. Além disso, mesmo quando o usuário não está identificado, isto é, quando ele não realizou o login em sua conta com nome e senha, ele também gera informações anônimas enquanto navega, essas servem para representar um grupo. Um bom exemplo são as listas de recomendação de itens “mais vendidos” em qualquer categoria, essas listas funcionam como um termômetro para avaliar a aceitação de um produto de forma geral (AGGARWAL, 2016).

No sentido de reforçar o entendimento das formas de coleta de informações dos usuários e utilizar alguns exemplos mais claros, Barcellos *et al.* (2007) salienta que o perfil do usuário pode ser construído a partir de algumas técnicas, como a coleta de dados explícita, implícita e a inferência. Na coleta de dados explícita o próprio usuário informa as suas preferências por meio da inserção de documentos ou pelo preenchimento de um formulário. Na coleta implícita os dados são coletados pelo registro da navegação do usuário no sistema, como por exemplo as páginas acessadas no site ou os produtos adicionados no carrinho de compras em uma loja virtual. A técnica de inferência consiste em definir o perfil do usuário com base no padrão de comportamento de outros usuários.

Em relação ao perfil do usuário, Reategui, Cazella e Osório (2006, p. 5) afirmam que:

A representação do perfil constitui-se no primeiro passo de um projeto de Sistema de Recomendação pois as técnicas que serão aplicadas dependerão desta decisão. Um Sistema de Recomendação não pode iniciar as atividades sem que o perfil do usuário tenha sido previamente definido, pois para iniciar suas atividades o sistema deve saber o máximo possível sobre os usuários. No momento da geração do perfil inicial é importantíssimo ter-se a maior quantidade de informação possível sobre o real interesse e necessidade do usuário.

SR podem atuar de diferentes formas quando interagem com o usuário: enviando sugestões intencionalmente de forma proativa via *e-mail*, por mensagem instantânea ou por outro meio, com consentimento prévio do usuário que é geralmente feito no cadastro para receber as recomendações e também podem atuar sob demanda, de forma que as sugestões apenas são apresentadas ao usuário se ele solicitar, por exemplo na navegação em um site de

compras clicando no *link* indicado como “sugestões para você”. Outra forma seria a atuação passiva, que são sugestões apresentadas dentro de um contexto de navegação, se o usuário estiver navegando para consumir livros, existirá sugestão de outros livros (MOTTA *et al.*, 2011).

Existem várias razões para os provedores de serviços utilizarem os SR, segundo Ricci, Rokach e Shapira (2010):

- Aumentar o número de itens vendidos: provavelmente a função mais importante para um SR comercial;
- Vender mais itens diversos: outra importante função: habilitar o usuário a selecionar itens que seriam difíceis de encontrar sem uma recomendação precisa;
- Ampliar a satisfação do usuário: um SR bem projetado irá melhorar a experiência do usuário durante a navegação;
- Aumentar a fidelidade do usuário: um usuário provavelmente será leal a uma página de internet ou a um aplicativo que o reconheça como freguês e que o trate com importância. Quanto mais ele interagir dentro do sistema, mais refinadas serão as recomendações;
- Entender melhor a necessidade do usuário: função importante do SR que pode ser aproveitada em muitas outras aplicações, descrevendo as preferências coletadas de forma explícita ou implícita.

As motivações acima citadas são focadas nos provedores de serviço, porém usuários também podem ter motivações para quererem utilizar SR, se esses efetivamente atingirem seus desejos. Consequentemente, um SR deve tentar equilibrar as necessidades dos dois lados, oferecendo um serviço excelente para ambos. (RICCI; ROKACH; SHAPIRA, 2010). Portanto, essa é uma função primordial para o sucesso de um SR, fazer com que tanto o usuário, quanto o provedor atinjam os seus objetivos.

Nesse contexto, de acordo com Iskold (2007, não paginado, tradução nossa):

[...] existem duas atividades *online* fundamentais – a procura e a navegação. Quando um consumidor sabe exatamente o que está atrás, ele procura por isso. Mas quando ele não está atrás de nada específico, ele navega. É na navegação que está a oportunidade de ouro para um SR, porque o usuário não está focado em encontrar algo específico, ele está aberto a sugestões.

Para Aggarwal (2016), os modelos básicos de SR trabalham com dois tipos de dados, um baseado em interações do usuário, como o comportamento de compras e avaliações, que

são conhecidos como filtragem colaborativa, e o outro baseado em informações sobre o perfil dos usuários e palavras-chave relevantes, chamados de recomendação baseada em conteúdo.

Parte da literatura pesquisada considera três técnicas de filtragem mais usadas nos SR, a filtragem colaborativa (FC), a filtragem baseada em conteúdo (FBC) e a filtragem híbrida (FH), que serão discutidas a seguir, além de uma quarta técnica chamada de filtragem baseada no conhecimento. Aggarwal (2016) reforça que as abordagens predominantes nos SR são largamente classificadas em três categorias, FBC, FC e FH.

O primeiro SR que se tem notícia, denominado *Tapestry* e desenvolvido em um centro de pesquisa nos Estados Unidos, surgiu da necessidade de filtrar uma quantidade considerável de e-mails que inundavam os usuários e que aumentavam na mesma proporção do acesso à internet, portanto, foi possível utilizar critérios de filtragem como, por exemplo, palavras-chave nesse sistema. Essa filtragem foi chamada de FC, pois era realizada não somente com algoritmos, mas também com o auxílio humano, em colaboração entre os interessados, (GOLDBERG *et al.*, 1992) entretanto, a FC é considerada como uma técnica ou um tipo de SR e, assim como a FBC, possui o mesmo propósito de recomendação, mas ambas utilizam abordagens diferentes, conforme será destacado no transcorrer deste trabalho.

Reforçando a importância dessas técnicas de filtrações citadas, Adomavicius e Tuzhilin (2005) afirmam que didaticamente, a FBC, a FC e a FH são as classificações mais empregadas nos SR. O Quadro 1 apresenta os objetivos e recursos dos três estilos de SR, e o estilo baseado em conhecimento:

Quadro 1 - Objetivos e recursos de estilos de filtragem

Estilo	Objetivos	Recursos
Colaborativo	Recomendações baseadas em abordagem colaborativa utilizando as avaliações de outros usuários.	Avaliações de usuários e avaliações de grupos.
Baseado em conteúdo	Recomendações com base nos atributos de itens favorecidos em avaliações passadas e ações feitas pelo usuário.	Avaliação do usuário e características/atributos dos itens.
Baseado em conhecimento	Retorna recomendações baseadas nas especificações dos atributos desejados em um item.	Especificação do usuário, características/atributo do item e domínio do conhecimento do usuário.
Híbrido	Criar um sistema mais robusto combinando os pontos fortes de dois estilos.	Utiliza os recursos de acordo com os tipos de filtrações combinadas.

Fonte: Adaptado de Aggarwal (2016, p. 16).

Dentre os recursos apresentados no Quadro 1, compreende-se que as avaliações de outros usuários e de grupos são bastante utilizadas para gerar recomendações na FC, enquanto na FBC os atributos dos itens e a avaliação previa do usuário são os principais recursos. Na filtragem baseada em conhecimento, o conjunto de especificação do usuário e a especificação correta do item, somados ao conhecimento do usuário, se fazem necessários para uma recomendação aprimorada. A FH vai utilizar como recurso os pontos positivos de cada filtragem que serão combinadas na criação do sistema para maximizar os seus resultados.

Portanto, para a construção de SR, os dois modelos que mais se aplicam e possuem visibilidade entre os outros são a FC e a FBC, ou seja, um SR pode utilizar qualquer um dos modelos disponíveis de acordo com seu objetivo, ou então utilizar o formato híbrido que engloba mais de um modelo.

Consoante com o que foi observado anteriormente, os tipos de filtrações citados como mais usadas e importantes serão apresentados a seguir, com suas características, seus pontos positivos e negativos.

2.1.3 Tipos de filtrações

Esta subseção não pretende ser exaustiva em relação aos tipos de filtrações existentes, mas sim apresentar e caracterizar, de acordo com autores citados na subseção anterior, os tipos de filtrações que são mais empregados e considerados como principais na construção de SR: a FC e a FBC.

2.1.3.1 Filtragem Colaborativa

Esse tipo de técnica é a mais utilizada entre todas, sendo encontrada em plataformas de *streaming* de filmes por exemplo, na qual as avaliações de usuários considerados de um mesmo grupo servem como base para recomendar itens para essas pessoas, considerando que dentro desse grupo os indivíduos possuem características parecidas, já que as avaliações refletem os gostos e preferências dos usuários.

Burke (2002) assegura que a FC é o tipo de filtragem mais familiar, mais amplamente implementada e mais madura.

Os sistemas baseados em FC não requerem o conhecimento das características e conteúdo dos itens, pois o seu foco são os usuários e suas avaliações, é nessa troca de

experiências entre pessoas em que esses SR se baseiam, e em que os itens são filtrados (CAZELLA; REATEGUI, 2005).

Assim, se dois usuários fizerem avaliações sobre um item em comum, os interesses de ambos serão considerados similares pela FC, e na hipótese de existir algum item que tenha sido bem avaliado por apenas um desses usuários, esse item poderá ser recomendado a outro.

As avaliações de usuários são consideradas essenciais neste tipo de sistema, e para que funcionem de forma regular, isto é, utilizando e confiando no conteúdo dos comentários e avaliações, Cazella e Reategui (2005) reiteram a necessidade de veracidade nas opiniões e avaliações feitas, usando exemplos de sistemas que empregam estímulos que encorajam os usuários a registrar suas opiniões verídicas através do oferecimento de bônus, brindes, descontos, entre outros.

Essa filtragem recomenda ao usuário ativo itens avaliados como bons no passado por outros usuários com gostos similares. Essa similaridade nos gostos de dois usuários é calculada baseada no histórico de avaliações, e por essa razão a FC é referenciada pela ‘relação pessoa para pessoa’, sendo considerada a técnica mais popular e amplamente usada em SR (RICCI; ROKACH; SHAPIRA, 2010).

Isso posto, pode-se dizer que a FC basicamente utiliza as opiniões de usuários para realizar as recomendações, não necessitando possuir informações sobre o conteúdo dos itens que serão recomendados. Portanto, itens sem nenhuma informação sobre o seu conteúdo serão recomendados amparados nas classificações de outros usuários.

Os SR que utilizam a FC, conforme explicam Bai *et al.* (2019), usualmente precisam de uma matriz usuário-item para representar as avaliações e comentários do usuário sobre os itens, destarte, as avaliações podem ser usadas para representar os interesses dos usuários. Após a criação da matriz, o sistema é capaz de calcular a similaridade entre os usuários para encontrar semelhança entre eles, conhecidos como “usuários vizinhos”, e por fim efetuar a recomendação. A matriz usuário-item é mostrada no Quadro 2 onde os elementos da matriz são as avaliações dos usuários.

Quadro 2 - Matriz usuário-item

	Item 1	Item 2	Item 3	Item 4	...	Item X
Usuário 1	0	1	0	1	...	1
Usuário 2	1	1	0	0	...	0
Usuário 3	1	0	1	1	...	0
...
Usuário Y	1	0	0	1	...	1

Fonte: Adaptado de Bai *et al.* (2019, p. 9328).

Na sequência, a Figura 3 mostra uma estrutura de SR de artigos científicos que empregam a FC.

Figura 3 - FC para recomendação de artigos



Fonte: Adaptado de Bai *et al.* (2019, p. 9328).

Dentro desse contexto, os usuários poderiam receber recomendações de itens que não estavam sendo pesquisados de forma ativa ou mesmo itens considerados atípicos, conforme Cazella e Reategui (2005), esse ponto pode ser visto como uma das vantagens desse tipo de filtragem. Além disso, outro benefício existente é a possibilidade de que ocorra a formação de grupos de usuários que se identifiquem por suas preferências e interesses em comum. Diante das vantagens apresentadas, cabe alertar para algumas questões que podem apresentar certos problemas na coleta de informações dos usuários, conforme se demonstra a seguir:

- 1) **Problema do primeiro avaliador:** quando um novo item aparece no banco de dados não existe maneira deste ser recomendado para o usuário até que mais informações sejam obtidas através de outro usuário.
- 2) **Problema de pontuações esparsas:** o objetivo dos sistemas de filtragem colaborativa é ajudar pessoas, focando em documentos lidos ou itens adquiridos. Caso o número de usuários seja pequeno em relação ao volume de informações no sistema existe um grande risco das pontuações tornarem-se muito esparsas.
- 3) **Similaridades:** caso um usuário tenha gostos que variam do normal este terá dificuldades para encontrar outros usuários com gostos similares, sendo assim suas recomendações podem se tornar pobres (CAZELLA; REATEGUI, 2005, p. 318).

Partida a frio ou *cold start*, de acordo com Park e Chu (2009), consiste em um problema nos SR que ocorre quando não existe nenhuma classificação histórica disponível sobre os itens ou sobre o usuário, dificultando que o sistema faça a recomendação.

Em vista disso, pode-se dizer que existem duas categorias que se referem ao problema de partida a frio, uma delas se trata sobre como fazer recomendações para um novo usuário que possui pouca ou nenhuma avaliação prévia sobre itens, dificultando encontrar usuários semelhantes que possam contribuir no processo de filtragem colaborativa. A outra categoria diz respeito aos novos itens que não possuem nenhuma avaliação de usuários, trazendo dificuldade para realizar a sua recomendação.

2.1.3.2 Filtragem Baseada Em Conteúdo

Nos SR baseados em conteúdo, a descrição dos itens e das suas características são usadas para as recomendações, por isso o termo ‘conteúdo’ no nome, referenciando as características e descrições dos itens (AGGARWAL, 2016). Ainda segundo o autor, nesse método as avaliações e comportamentos de compras dos usuários são combinadas com as informações disponíveis dos itens.

Nesse tipo de filtragem, conforme explica Park e Chu (2009), o perfil do usuário é gerado baseado nas descrições de conteúdo dos itens que foram anteriormente classificados por ele, dessa forma é possível recomendar novos itens que ainda não foram avaliados por nenhum outro usuário. No caso de novos usuários, a FBC geralmente pede que eles respondam um questionário que mostra explicitamente as suas preferências para assim poder gerar um novo perfil, e, à medida que esse usuário consome e avalia mais itens, seu perfil é atualizado com o conteúdo dos novos itens consumidos, que terminam por receber um maior peso. Esta técnica busca construir um perfil de usuário baseado em características dos itens em que ele teve alguma interação implícita ou explícita (LÁZARO, 2010). Para realizar a recomendação de um item, essa técnica cria uma conexão entre o item e as preferências do usuário de acordo com o seu perfil.

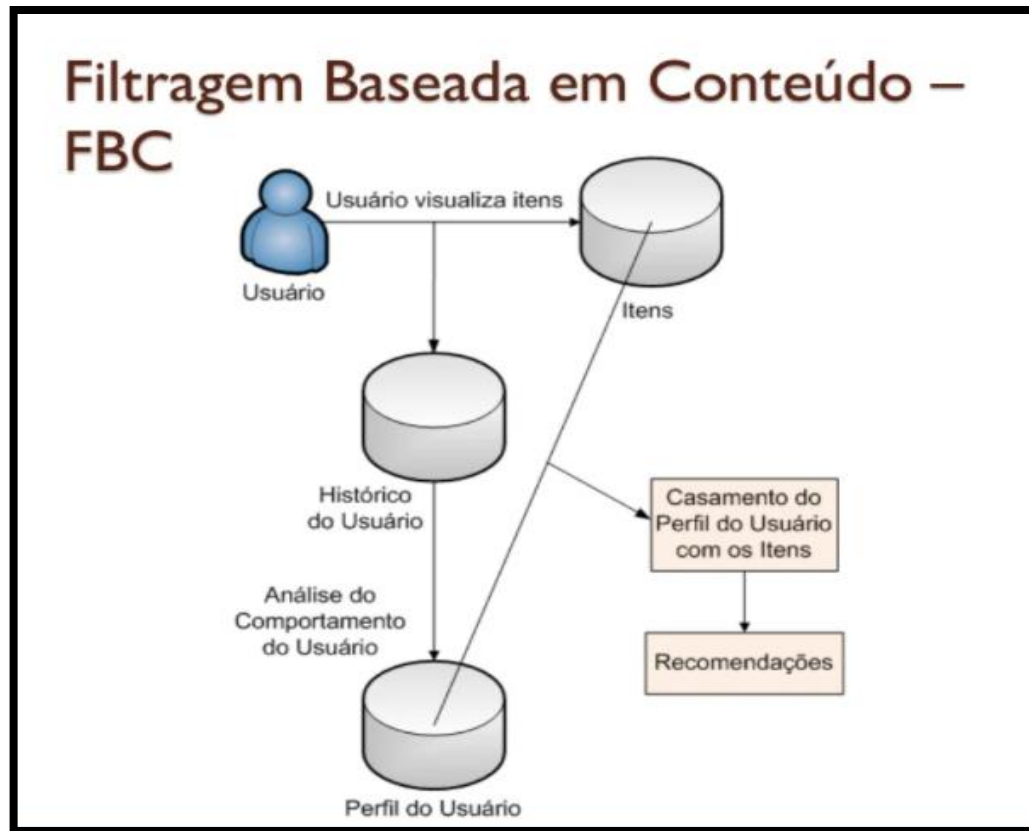
Para Cazella e Reategui (2005, p. 16), esse tipo de abordagem em conteúdo possui vínculos com a área de recuperação de informação, e afirma que:

Devido aos significativos avanços feitos pelas comunidades de filtragem de informação e filtragem de conteúdo, muitos sistemas baseados em filtragem de conteúdo focam nas recomendações de itens com informações textuais, como documentos e *websites*. As melhorias sobre os sistemas tradicionais de recuperação de informação vieram com a utilização do perfil do usuário, que contém suas preferências e necessidades.

Uma das formas para se utilizar a FBC, conforme explicam Motta *et al.* (2011), é demandar ao usuário uma avaliação de itens, indicando seu interesse ou não pelos itens

avaliados. Após a avaliação, o sistema procura itens que reúnam características afins com o que foi avaliado como de interesse, descartando os demais itens avaliados como de não interesse. A Figura 4 apresenta um exemplo de FBC:

Figura 4 - Filtragem baseada em conteúdo – FBC



Fonte: Gatto e Zorzo (2018).

Portanto, esta técnica de filtragem se baseia na construção de um perfil, tanto para um usuário quanto para um item, com cada um desses perfis apresentando informações que versem sobre o seu conteúdo. No caso de um usuário, pode-se conter informações sobre suas preferências de marcas, sua renda e seu local de moradia por exemplo. Já em relação a um item ou produto, informações sobre seu fabricante, sua garantia, seu modelo e seu ano de criação servem como características uteis para compor um perfil.

De acordo com Aggarwal (2016), a FBC possui vantagens nas recomendações para itens novos, ou seja, itens com insuficiência de avaliações, isso porque outros itens que possuam características parecidas podem já ter sido avaliados pelo usuário, fazendo com que a FBC consiga promover essas avaliações junto com as características do item, mesmo quando esse não tenha um histórico de avaliações. Dessa forma, o sistema compreende itens previamente bem avaliados pelo usuário e recomenda itens semelhantes.

Conforme Salton e MacGill (1983 *apud* CAZELLA; DRUMM; BARBOSA, 2010), na hipótese dos itens serem artigos científicos, o procedimento de comparar itens previamente avaliados pelo usuário pode ser favorecido, pois documentos e textos podem ser classificados como semelhantes quando compartilham termos em comum. Dessa forma a FBC mostra-se a técnica mais adequada para a recomendação de itens que possuem textos, pois seu conteúdo é comumente representado por palavras-chave.

Assim, se a efetividade para recomendar itens novos é uma vantagem nessa técnica, não se pode afirmar o mesmo para o caso de novos usuários, isso porque o FBC necessita de um histórico de classificações feitas pelo usuário, e, quanto maior a quantidade de classificações, mais robustas e certeiras serão as recomendações (AGGARWAL, 2016).

Segundo Motta *et al.* (2011), a FBC apresenta algumas vantagens:

- não depende de dados de outros usuários;
- recomenda itens novos ou não comuns;
- não sofre com problemas de Partida a Frio ou de Esparsidade, problemas que ocorrem com itens novos adicionados sem prévias avaliações;
- possibilita recomendações a usuários com gostos peculiares;
- é capaz de fornecer explicações sobre os itens recomendados, listando os atributos do conteúdo que acarretou a recomendação;
- é considerada uma tecnologia madura.

Dessa forma, a FBC geralmente não necessita de avaliações de outros usuários para fazer as recomendações. Aggarwal (2016. p. 140) afirma que:

Os sistemas baseados em conteúdo são largamente utilizados em cenários em que uma quantidade significativa de informação de atributos está disponível. Em muitos casos, estes atributos são palavras-chaves, que são extraídas das descrições dos produtos. Na realidade, a grande maioria dos sistemas baseados em conteúdo extraem atributos textuais dos objetos subjacentes. Sistemas baseados em conteúdo estão, portanto, particularmente bem adaptados para dar recomendações em áreas não estruturadas e ricas em textos.

Assim como todas as técnicas, a FBC também possui desvantagens conforme descrevem Motta *et al.* (2011):

- recomendações estáticas, ignorando a opinião de usuários;
- o conteúdo precisa ser codificado por características inteligíveis, pois muitas vezes é analisado automaticamente para identificar as categorias;
- baixa eficiência quando o conteúdo não é muito informativo;

- para livros técnicos o título é relevante, pois representa uma categoria, mas para outros itens, como filmes, o título pode trazer pouca informação por não possuir relação estreita para indicar a categoria, tornando mais difícil gerá-la.

Visto que os SR baseados em conteúdo funcionam extraíndo os atributos dos itens, eles frequentemente oferecem perspectivas interpretáveis para o processo de recomendação. Em uma recomendação de um filme, por exemplo, seria interessante apresentar ao usuário um motivo pelo qual ele possa gostar do filme sugerido, seja pelo gênero, por um ator/atriz específico, ou apresentar um conjunto de palavras-chave que possam ajudá-lo a decidir se deve assistir ao filme recomendado (AGGARWAL, 2016).

Para Cazella e Reategui (2005), existem algumas limitações na FBC, como a dificuldade na compreensão causada pelo uso de sinônimos, a dificuldade na análise de conteúdos pouco estruturados e a falta de inovação na recomendação, pois o sistema tende a recomendar apenas conteúdos semelhantes ao perfil do usuário.

Observando algumas desvantagens mostradas, conclui-se que esse sistema tende a funcionar de forma mais precisa quando os itens são detalhados de uma forma minuciosa, com o máximo de características explícitas.

Os SR baseados em conteúdo apresentam alguns componentes que, conforme explica Aggarwal (2016), são considerados básicos para o seu funcionamento e também invariáveis. Visto que estes sistemas funcionam com uma grande variedade de descrição de itens e do conhecimento dos usuários, é necessário que todos esses dados não estruturados sejam convertidos em estruturas padronizadas, que geralmente são convertidos em palavras-chaves. Os principais componentes desse sistema incluem a parte de pré-processamento, a parte de aprendizagem (ambas *offline*), e a parte de previsão (*online*). As partes *offline* são utilizadas para a criação de um modelo resumido, que geralmente é um modelo de classificação ou regressão, o qual posteriormente é usado para a geração de recomendações. Os componentes dos SR baseados em conteúdo são:

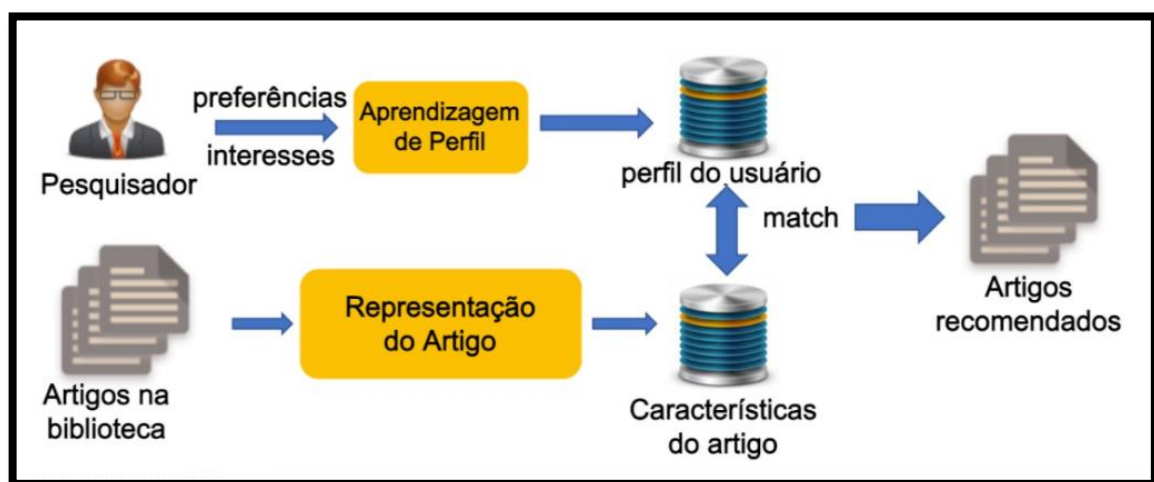
1. Pré-processamento e extração de atributos: responsável por extrair as características dos vários itens/fontes (páginas da web, descrições de produtos, notícias, entre outros) e transformar em uma representação estruturada baseada em palavras-chave. Esse é considerado o primeiro passo em qualquer SR baseado em conteúdo.
2. Aprendizagem baseada em conteúdo de perfis de usuário: neste momento um modelo específico do usuário é construído para prever os interesses do usuário nos itens, com base em seu histórico anterior de compra ou classificação de itens. O *feedback* do usuário, na forma de classificações feitas anteriormente ou pela sua

atividade, é aproveitado. Esses feedbacks são agrupados com os atributos dos itens para construir os dados. Um modelo de aprendizagem é construído com base nesses dados de treinamento. O modelo resultante é conhecido como perfil do usuário, relacionando os interesses dele (classificações) aos atributos do item.

3. Filtragem e recomendação: Nessa fase, o modelo criado anteriormente é utilizado para fazer recomendações para usuários específicos, combinando os perfis desses usuários com as representações dos itens.

Na hipótese de um SR de artigos científicos, Bai *et al.* (2019) explicam que os itens são os artigos em uma biblioteca digital e os usuários são os pesquisadores. As citações dos trabalhos de um pesquisador podem ser usadas para a construção do seu perfil. Ainda, os autores afirmam que existem várias formas para se construir o perfil do pesquisador, entre elas extrair palavras-chave das preferências e interesses do usuário. Além disso, pode-se extrair palavras-chave do título, do resumo e do conteúdo dos artigos para serem usados como representação deles, que são chamados de artigos candidatos e que podem ser recuperados da biblioteca digital. Assim, os SR de artigos calculam a semelhança das palavras-chave do perfil do pesquisador e dos artigos candidatos e os classifica, dessa forma os próximos artigos candidatos com alta similaridade podem ser recomendados para o pesquisador. Abaixo, a Figura 5 mostra um SR de artigos que utiliza FBC, no qual se pode visualizar os três principais passos, que são: a representação do item (artigo), a aprendizagem do perfil e, por último, a recomendação.

Figura 5 - SR de artigos baseado em FBC



Fonte: Adaptada de Bai *et al.* (2019, p. 9326).

Para a representação dos itens, onde os itens (textos) são transformados em representações estruturadas com base em palavras-chave, existem alguns métodos conhecidos

que são utilizados com mais regularidade. O TF-IDF, do inglês *Term frequency-inverse document frequency*, que em português significa ‘frequência do termo - inversa frequência nos documentos’, é uma das medidas estatísticas usadas.

Nesse contexto, Bai *et al.* (2019, p. 9326) explicam que:

O valor TF-IDF é uma medida estatística para avaliar a importância de uma palavra para um documento em uma coleção ou conjunto. A ideia básica do modelo TF-IDF é dividida em dois aspectos. Por um lado, quanto mais vezes a palavra-chave K aparece no documento D, mais importante será K para o documento D. Por outro lado, quanto maior a frequência de K aparecer em diferentes documentos, menor será a importância de K para distinguir os documentos.

Portanto, O TF-IDF vai analisar a frequência de um termo em um documento (TF), e o quão raro é um termo diante dos documentos/textos (IDF).

Para Cazella, Nunes e Reategui (2010, p. 28) “[...] a mineração de textos é amplamente aplicada na técnica de filtragem baseada em conteúdo, uma vez que é necessária a análise do conteúdo que descreve o item a ser recomendado.”

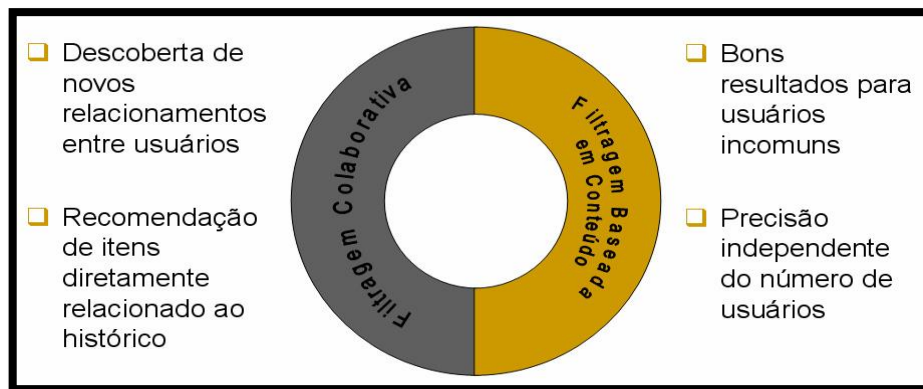
Conforme visto anteriormente, a construção do perfil do usuário aparece na sequência da extração dos atributos dos itens. Utilizando técnicas de recuperação de informação, é possível criar o perfil de um usuário, examinando tudo que já foi avaliado por ele e, ainda, as entradas que foram realizadas pelo próprio usuário com as suas preferências e gostos.

2.1.3.3 Filtragem Híbrida

A filtragem híbrida (FH) utiliza as técnicas citadas anteriormente, de forma a combinar o melhor de cada uma das técnicas existentes, excluindo os fatores limitantes para a criação de um sistema mais robusto. É notável que cada técnica funcione melhor para um tipo de abordagem e que cada uma possua diferentes tipos de entrada e apresentem forças e fraquezas que as diferem. Casos em que há uma grande variedade de entradas, existe a flexibilidade de se usar diferentes tipos de SR para a mesma tarefa, nesses casos tem-se oportunidades para o uso da FH, nas quais as particularidades de diferentes tipos de sistemas combinados podem obter o melhor de cada um, criando assim um sistema mais robusto (AGGARWAL, 2016).

Conforme Cazella e Reategui (2005), essa técnica é fundamentada nos benefícios proporcionados pela FBC e pela FC, reunindo os pontos fortes de cada técnica e excluindo os fatores limitantes (fraquezas), ilustrado pela Figura 6.

Figura 6 - Filtragem híbrida



Fonte: Cazella e Reategui (2005).

Em conformidade com o que foi visto, a Figura 6 mostra que os pontos mais fortes da FC e da FBC são utilizados conjuntamente, no sentido de minimizar os problemas que possuem quando utilizadas isoladamente, para formar a FH.

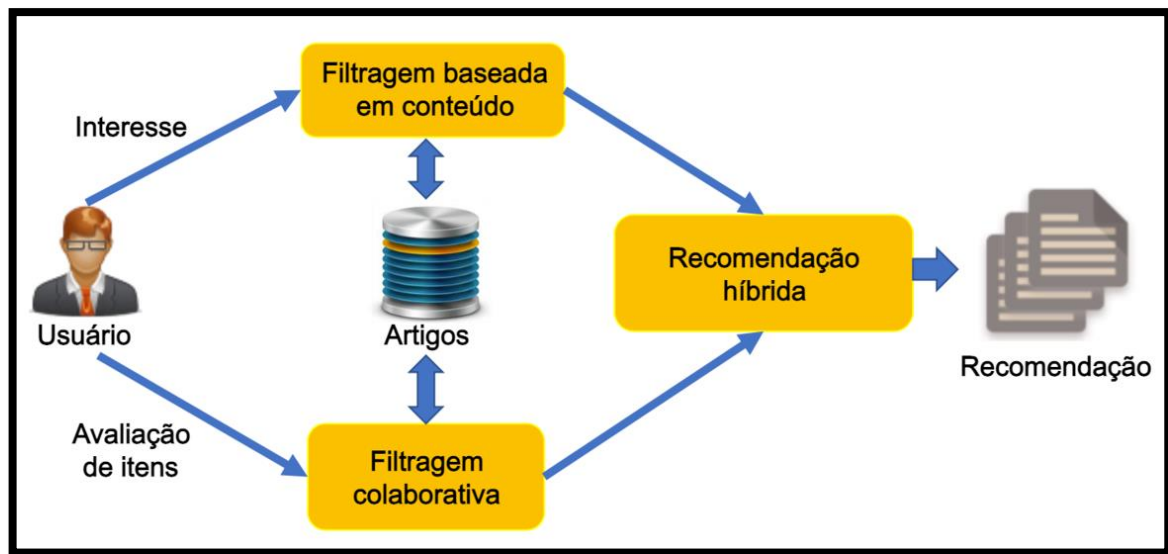
Campos *et al.* (2010) afirmam que a combinação das técnicas utilizadas na FBC e FC estão presentes nos sistemas híbridos de recomendação mais comuns.

Nos SH que utilizam os dois tipos de filtragem citados acima, Adomavicius e Tuzhilin (2005) exemplificam que podem ser classificados em quatro categorias:

- Diferentes recomendadores combinados – nesta abordagem dois sistemas separados são criados e cada um produz uma saída;
- Características da FBC incluídas em modelos de FC – mantem-se a base da recomendação com as técnicas da FC, entretanto são adicionadas informações sobre usuários e itens, que são características da FBC;
- Propriedades da FC adicionadas em modelos de FBC – utilizam-se técnicas de redução de dimensionalidade no conjunto dos usuários de um grupo e de representações de itens, aprimorando a precisão do SR baseado em conteúdo;
- Modelo unificado – neste tipo de abordagem é construído um modelo que utiliza ambas as particularidades das filtragens para gerar a recomendação, e não apenas se adiciona características de uma filtragem sobre a outra.

A Figura 7 mostra um SR híbrido de artigos científicos, que usa a combinação da FBC e FC.

Figura 7 - SR de artigos baseado em FH



Fonte: Adaptada de Bai *et al.* (2019, p. 9331).

Existem alguns métodos mais comuns que têm sido empregados em FH, conforme explica Burke (2002), os quais funcionam utilizando combinações de forma diferentes, conforme o Quadro 3:

Quadro 3 - Métodos para FH

Métodos	Descrição
Ponderado	As pontuações de várias técnicas de recomendação são combinadas para produzir uma única.
Comutação	Sistema alterna entre as técnicas de recomendação conforme a situação.
Misto	Recomendações de diferentes técnicas são apresentadas ao mesmo tempo.
Combinação de Recursos	Recursos de diferentes fontes de dados de recomendação são agrupados em um único algoritmo para recomendar.
Cascata	Um recomendador refina as recomendações feitas pelo outro.
Aumento de Recursos	O resultado de uma das técnicas é usado como recurso para a outra.
Meta-nível	O modelo gerado por um recomendador é usado como entrada para o outro.

Fonte: adaptado de Burke (2002).

Diferentes métodos podem ser utilizados na FH, o Quadro 3 mostra os métodos e apresenta um breve resumo de algumas combinações possíveis, nas quais as técnicas de filtragem se sobrepõem ou se unem, buscando maximizar a recomendação final.

2.1.3.4 Filtragem Baseada no Conhecimento

SR baseados em conhecimento são mais úteis quando os itens que serão recomendados são comprados com muita frequência, ou seja, itens raramente vendidos como imóveis, automóveis ou produtos de luxo possuem especificidades e dificilmente terão um número suficiente de avaliações de outros usuários que possam ser usadas para recomendações (AGGARWAL, 2016).

Assim, como são itens que possuem pouca procura ou são muito específicos, as recomendações baseadas em conhecimento serão realizadas em cima das necessidades particulares que cada um irá informar especificamente.

A filtragem baseada em conhecimento procura identificar conhecimentos específicos do domínio em fontes internas (tendências encontradas nos dados) ou mesmo externas para realizar a recomendação. São consideradas as sugestões e os gostos proporcionados pelos utilizadores com base nas suas interações históricas, ou seja, o modelo deve levar em consideração a base de conhecimento do domínio para gerar novas recomendações de itens para os utilizadores. Diversas informações auxiliares além das avaliações (*ratings*) em diferentes formatos (textos, imagens, vídeos, ...) podem ser exploradas para extrair conhecimento do domínio (AZAMBUJA; MORAIS; FILIPE, 2021).

De acordo com Azambuja, Morais e Filipe (2021) o estilo de filtragem baseada em conhecimento pode trazer ganhos em algumas situações que possuem particularidades e em modelos computacionais que utilizem a FH. Porém esse tipo de filtragem vem sendo utilizada com menos frequência do que a FC e FBC.

Em contrapartida, Ricci, Rokach e Shapira (2010) afirmam que os SR baseados em conhecimento, inicialmente quando implantados, tendem a funcionar melhor do que outros. Porém, caso não estejam equipados com componentes de aprendizado, logo serão superados por outros métodos que exploram a interação humano/computador, como a FC.

2.2 SISTEMAS DE RECOMENDAÇÃO DE ARTIGOS CIENTÍFICOS EXISTENTES

A comunicação entre pesquisadores mostra-se característica intrínseca à comunidade científica, e essa aproximação entre pesquisadores cresce e se aprimora com as chamadas “redes sociais” acadêmicas.

Existem diferentes plataformas categorizadas como redes sociais acadêmicas, que possuem múltiplas aplicações, como divulgação e compartilhamento de documentos científicos, espaço para *upload* de resultados de pesquisas, recomendações de artigos e de perfis de pesquisadores de áreas afins, criação de alertas com palavras-chave, plataformas que permitem que o próprio usuário faça recomendações para seus pares, entre outras funcionalidades.

Devido à essa quantidade de plataformas, neste capítulo são apresentados protótipos e plataformas consideradas como redes sociais acadêmicas que de alguma forma fazem recomendações de artigos científicos.

2.2.1 Google Acadêmico

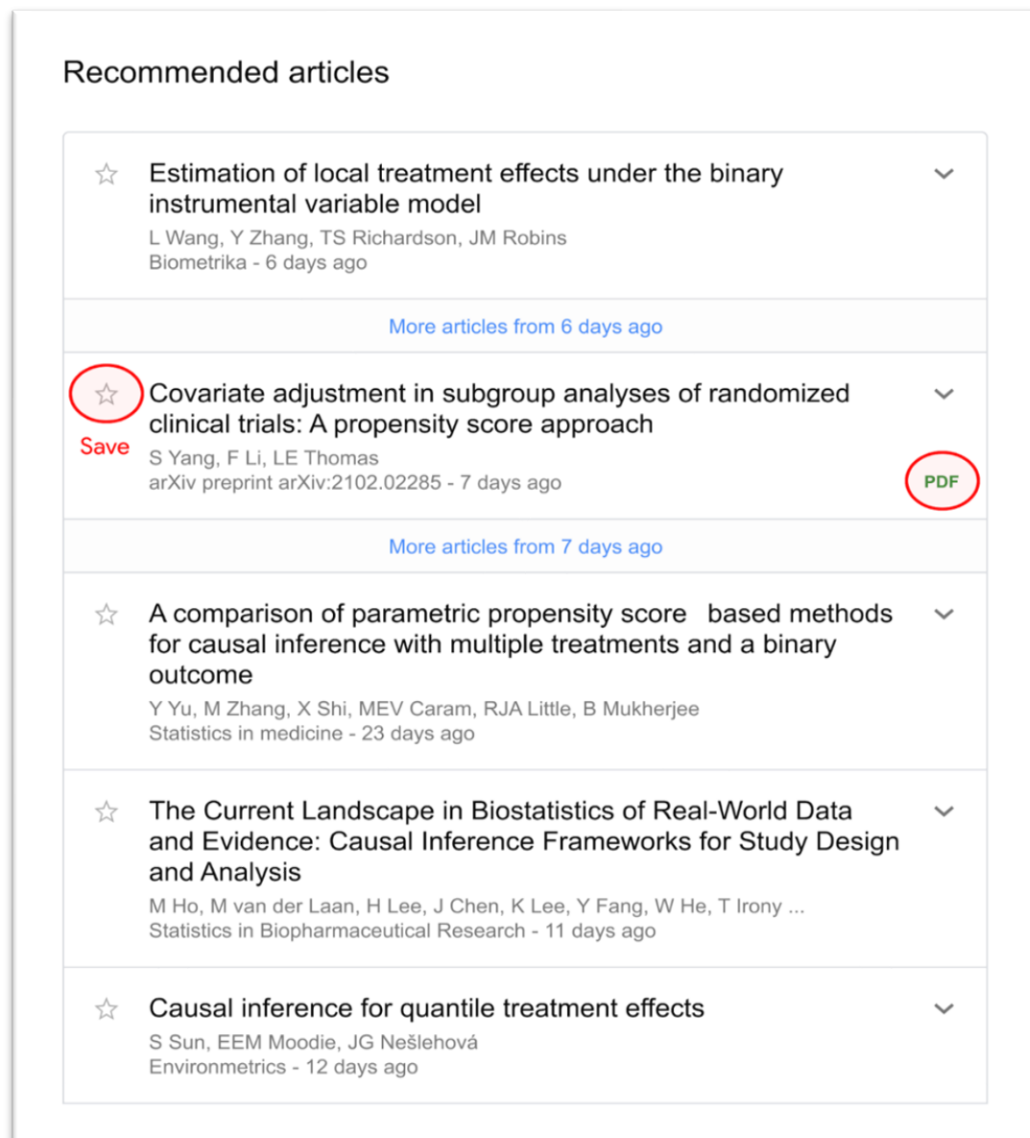
Funcionando de forma parecida com o Google padrão, o Google Acadêmico é uma ferramenta que retorna resultados de busca de acordo com a sua relevância. Para isso, leva em conta detalhes como: o autor, a publicação, a frequência de citação e o texto integral (CIRIACO, 2015).

Apesar de ter surgido como uma ferramenta de recuperação da informação, o buscador acadêmico da Google agrega muitas outras aplicações. Algumas das funções disponíveis:

- ✓ é possível localizar documentos científicos sobre variados assuntos e diversas fontes;
- ✓ ao criar um perfil, o usuário que possui artigos publicados pode acompanhar se houve citações do seu trabalho;
- ✓ é possível definir um alerta utilizando uma palavra-chave sobre algum assunto de interesse, assim, quando um novo artigo relacionado for publicado o usuário recebe um alerta;
- ✓ possui uma biblioteca do usuário na qual é possível salvar publicações e documentos.

Nos casos em que o usuário cadastrado possui um perfil ativo no sistema, isso é, que realize publicações frequentes, Shetty *et al.* (2021) explicam que a página inicial do usuário no Google Acadêmico deverá conter recomendações de artigos que se parecem com a Figura 8 a seguir:

Figura 8 - Interface do Google Acadêmico com artigos recomendados



Fonte: Shetty *et al.* (2021).

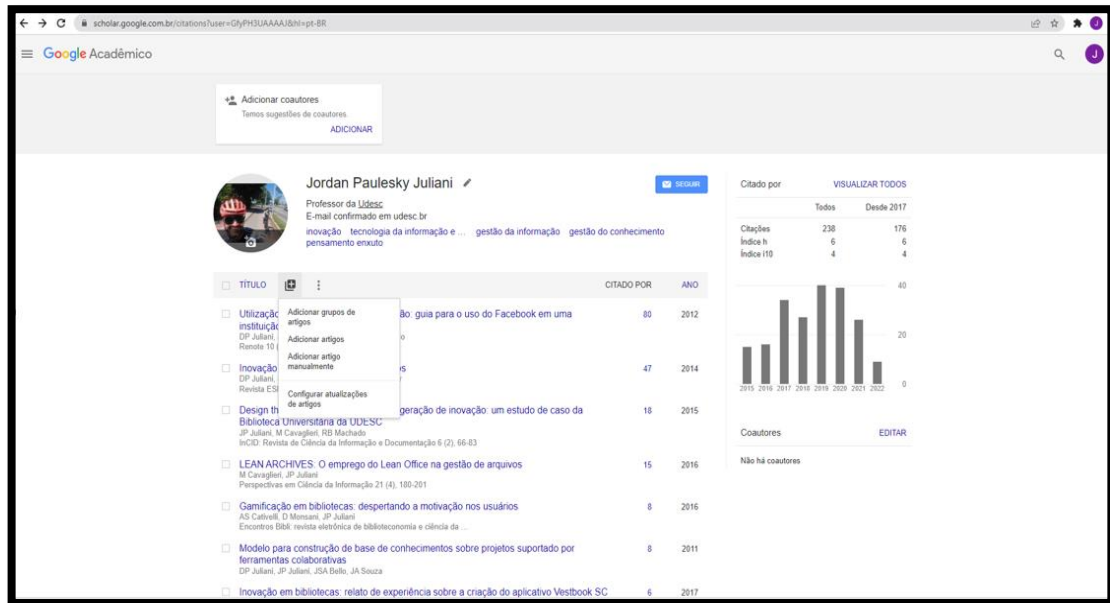
As recomendações mostradas em forma de lista na Figura 8 revelam que os artigos mais recentes são listados no topo. O usuário pode executar funções como: salvar na sua biblioteca para leitura posterior, expandir resumos dos documentos ou ler de forma completa, quando disponíveis.

O Google Acadêmico realiza recomendações para o usuário que possui um perfil acadêmico cadastrado com os trabalhos que escreveu ou fez parte, para que o algoritmo possa efetivar as recomendações, que iniciarão alguns dias após esse processo de criação do perfil. Para gerar as recomendações, os artigos do perfil do usuário são analisados e a sua relevância é calculada com modelos estatísticos que incluem os assuntos dos artigos, os locais de

publicações, as citações utilizadas e os autores que fazem parte da mesma área (SHETTY *et al.*, 2021).

Abaixo, a Figura 9 mostra a interface do Google Acadêmico com algumas opções de configurações para pesquisador incluir as suas produções em sua própria página.

Figura 9 - Interface do Google Acadêmico



Fonte: Elaborada pelo autor (2022).

As opções de configurações que aparecem na Figura 9 sugerem que ao clicar no símbolo ‘+’ ao lado da palavra ‘TÍTULO’ é possível adicionar manualmente os artigos dos quais é autor ou participante. Além de que, de forma automatizada, o próprio sistema pode encontrar os artigos publicados pelo autor. É importante o pesquisador realizar o *upload* de seus trabalhos para que o sistema possa utilizá-las, tornando as recomendações mais personalizadas.

De acordo com Moretti (2020), o Google Acadêmico agrega conteúdos de várias fontes, algumas delas são: Altametric, Biblioteca Digital FGV, SciELO Brasil, Buscalegis UFSC. Por outro lado, existem conteúdos indexados pelo buscador Google Acadêmico que são conteúdos pagos, onde só é possível visualizar resumos e sumários de forma gratuita.

Assim como outros sites e plataformas que são consideradas grande empresas privadas, ou que fazem parte de uma grande corporação, no caso do Google Acadêmico, é extremamente difícil obter de forma clara e explícita o funcionamento de seus algoritmos utilizados para realizar as recomendações, entretanto buscou-se mostrar o seu funcionamento com exemplos que demonstram a sua aplicação.

2.2.2 Academia.edu

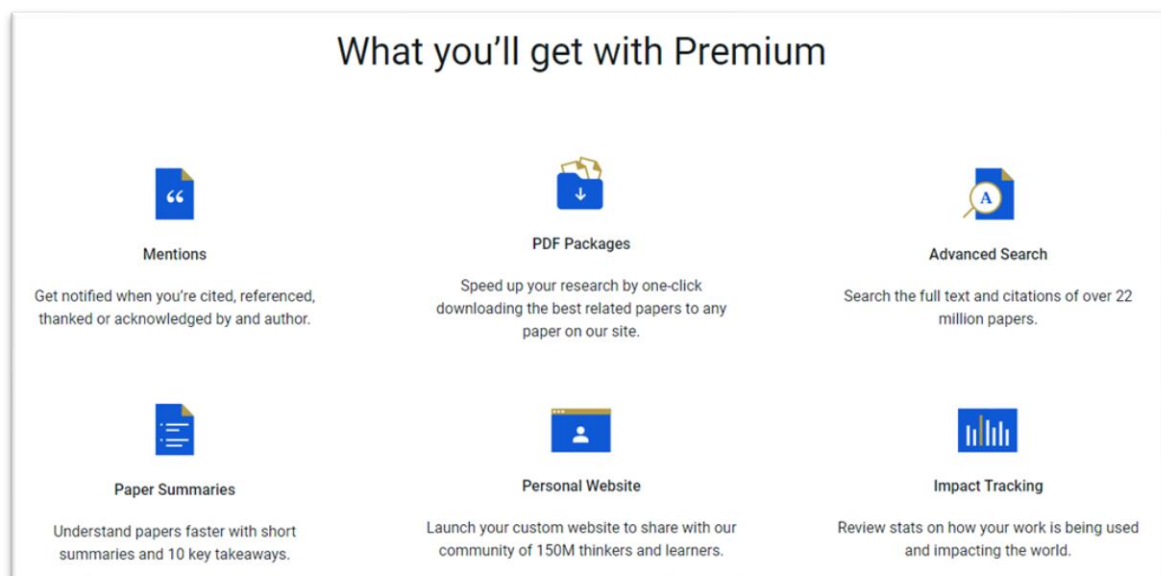
O uso de SR no mundo acadêmico tem aumentado, em parte, por causa de sites de Redes Sociais Acadêmicas, como Academia.edu e Researchgate.net (MONTEIRO-KREBS *et al.*, 2021).

O academia.edu, de acordo com seu próprio sítio eletrônico, tem como missão disponibilizar, na internet e de forma gratuita, todos trabalhos acadêmicos e científicos, além de progredir com as discussões e colaboração acadêmica.

Para que este modelo de negócio que disponibiliza acesso gratuito funcione, a plataforma também possui um serviço pago chamado de *Premium*. Neste formato, os assinantes do Academia *Premium*, que são mais de duzentos e sessenta mil, conseguem garantir o custo para que a plataforma possa abrigar pesquisas gratuitas e cumprir sua missão (ACADEMIA.EDU, 2021).

A Figura 10 mostra as funcionalidades extras disponíveis para os assinantes do Academia *Premium*:

Figura 10 - Funcionalidades premium academia.edu



Fonte: academia.edu (2021)

No sistema *Premium* o usuário poderá obter:

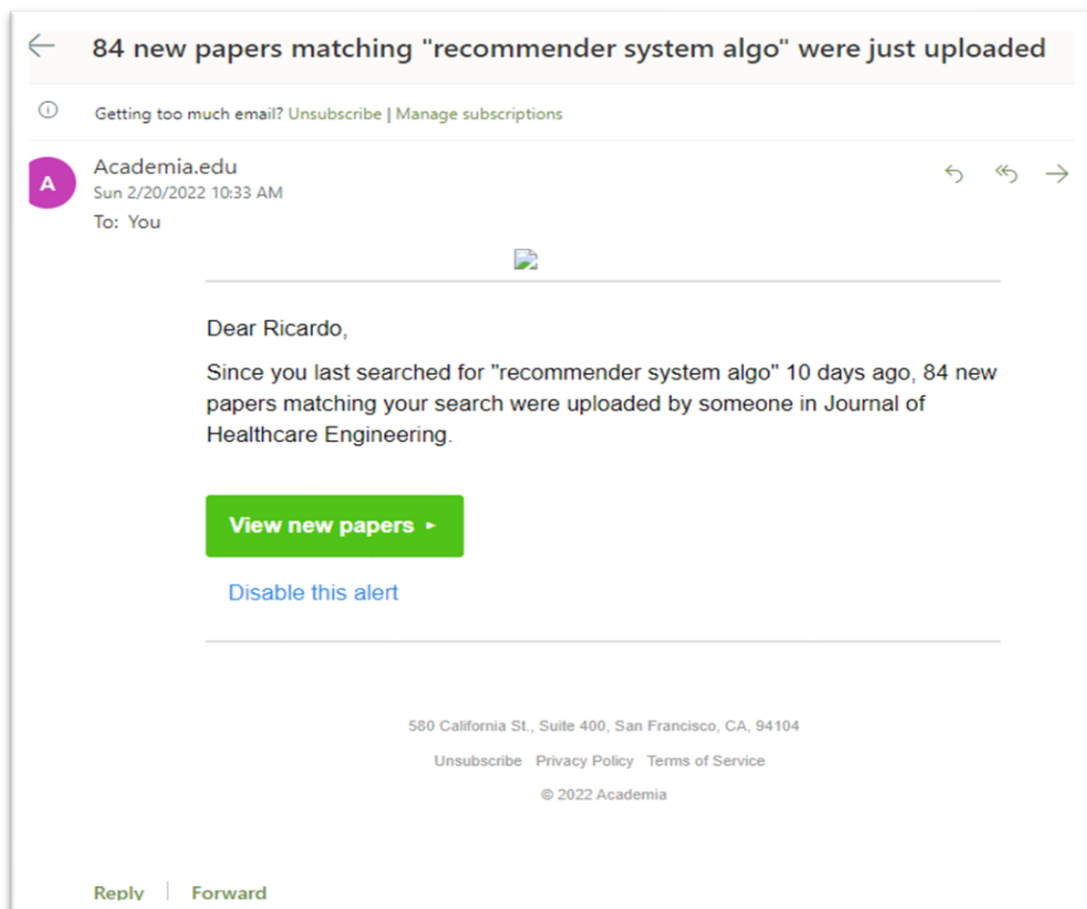
- ✓ Notificações quando for citado, agradecido ou referenciado por um autor;
- ✓ Acelerar sua pesquisa, baixando com apenas um clique os melhores artigos relacionados para qualquer artigo disponível no sítio;
- ✓ Pesquisar textos completos e citações de mais de 22 milhões de artigos;

- ✓ Acesso a resumos curtos e 10 dicas importantes para facilitar o entendimento dos documentos mais rapidamente;
- ✓ Ter o seu próprio sítio na plataforma para compartilhar com a comunidade de 150 milhões de usuários;
- ✓ Revisar estatísticas sobre como seu trabalho está sendo usado.

Portanto, os assinantes que pagam a mensalidade *Premium*, possuem um alcance maior nas suas pesquisas e ao mesmo tempo fazem com que a plataforma continue disponibilizando acesso gratuito para outros usuários.

De acordo com a experiência de uso deste autor, a plataforma Academia.edu faz recomendações de artigos no *feed* do usuário baseado na atividade do pesquisador dentro da plataforma, como as palavras que foram usadas nas buscas recentes, os artigos e os autores buscados. Ademais, o usuário recebe *e-mails* com alertas de novos documentos baseado na sua última pesquisa e que foram adicionados à plataforma recentemente, conforme mostra a Figura 11.

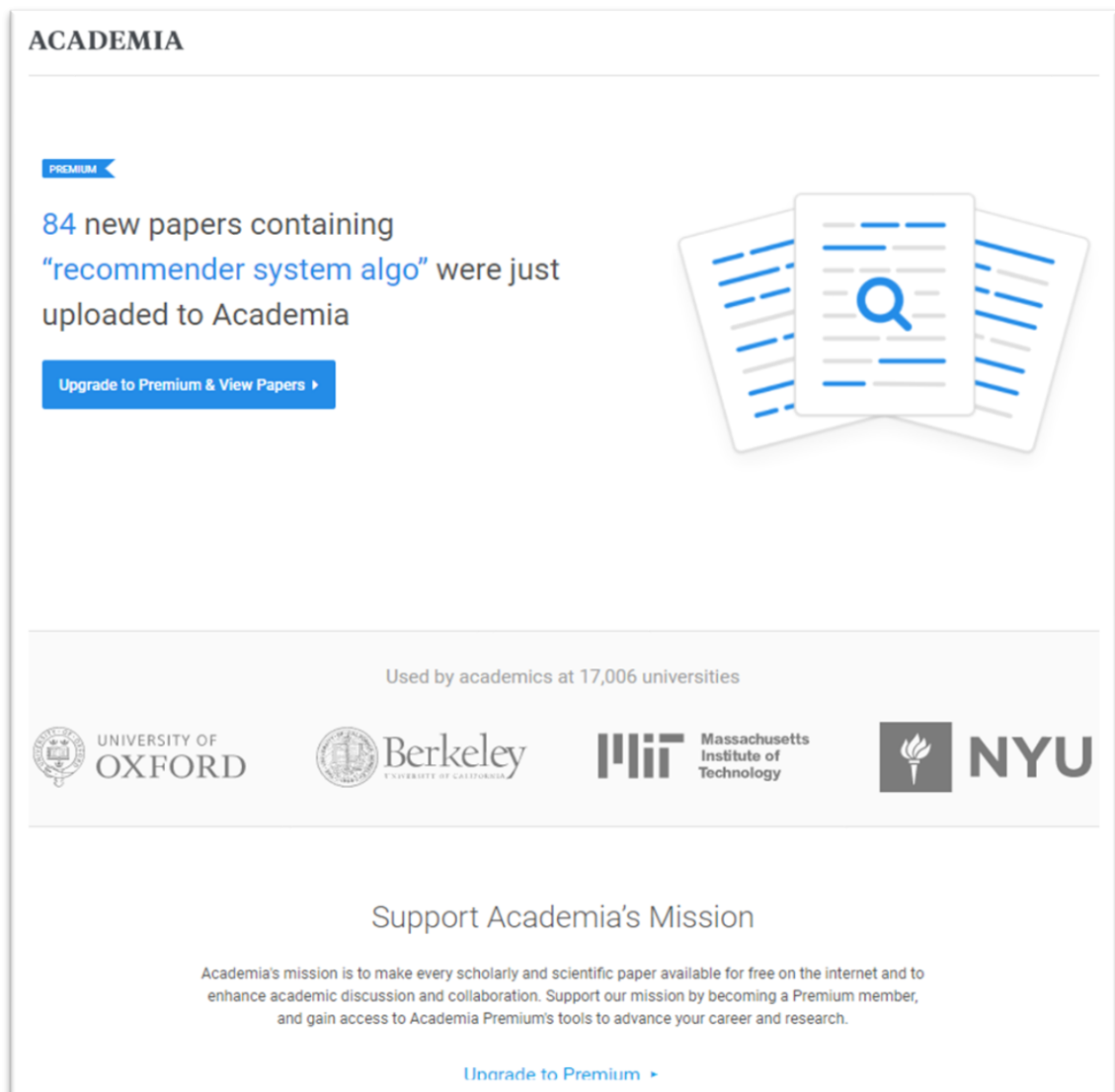
Figura 11 - E-mail com alerta de novos documentos



Fonte: Elaborada pelo autor (2022)

A Figura 11 mostra um *e-mail* enviado pela Academia.edu, o qual informa ao usuário que 84 novos artigos que se referem à última pesquisa feita pelo pesquisador dentro da plataforma foram adicionados. No próprio *e-mail* existe uma opção para clicar (em verde) que direciona o usuário aos artigos e outra na qual é possível desabilitar novos alertas. Não obstante, ao clicar no botão verde o pesquisador é conduzido a uma página em que a plataforma Academia.edu permite que os 84 artigos sugeridos no *e-mail*, só podem ser vistos caso seja feito um *upgrade* para a categoria *Premium*, ou seja, o acesso só será liberado mediante pagamento, conforme mostra a Figura 12:

Figura 12 - Tela com acesso restrito

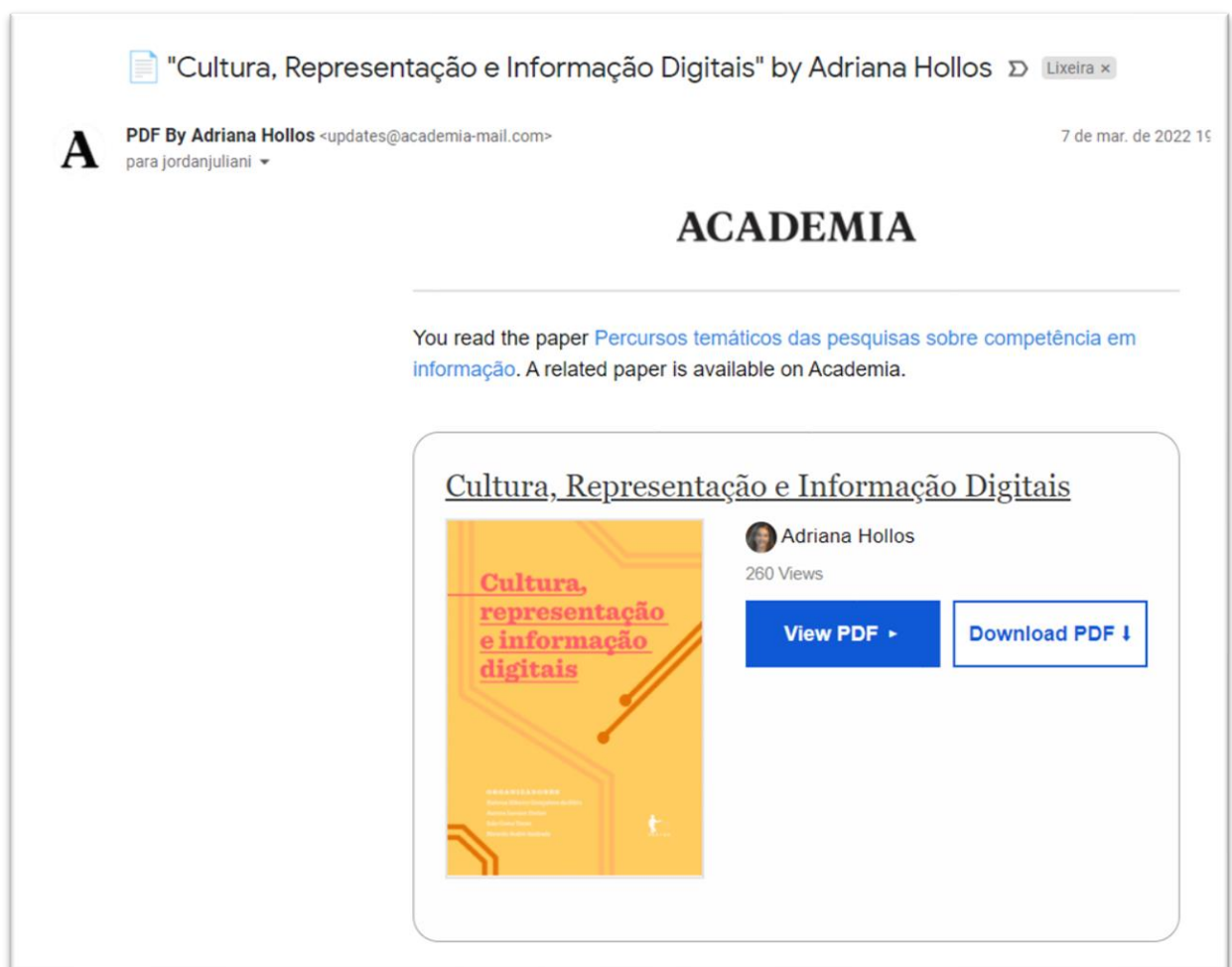


Fonte: Elaborada pelo autor (2022)

Pode-se observar na Figura 12 a tela em que o usuário é direcionado após clicar para acessar os artigos sugeridos via *e-mail*. Nota-se que no campo em azul, na figura, está sugerido que o usuário se torne um membro *Premium* para conseguir o acesso aos documentos.

Em contrapartida, em uma outra situação, a plataforma elaborou uma recomendação de um artigo baseada em uma leitura prévia do usuário e enviou para a caixa de *e-mail*, com a possibilidade do usuário fazer o *download* do artigo em forma de PDF, conforme apresenta a Figura 13:

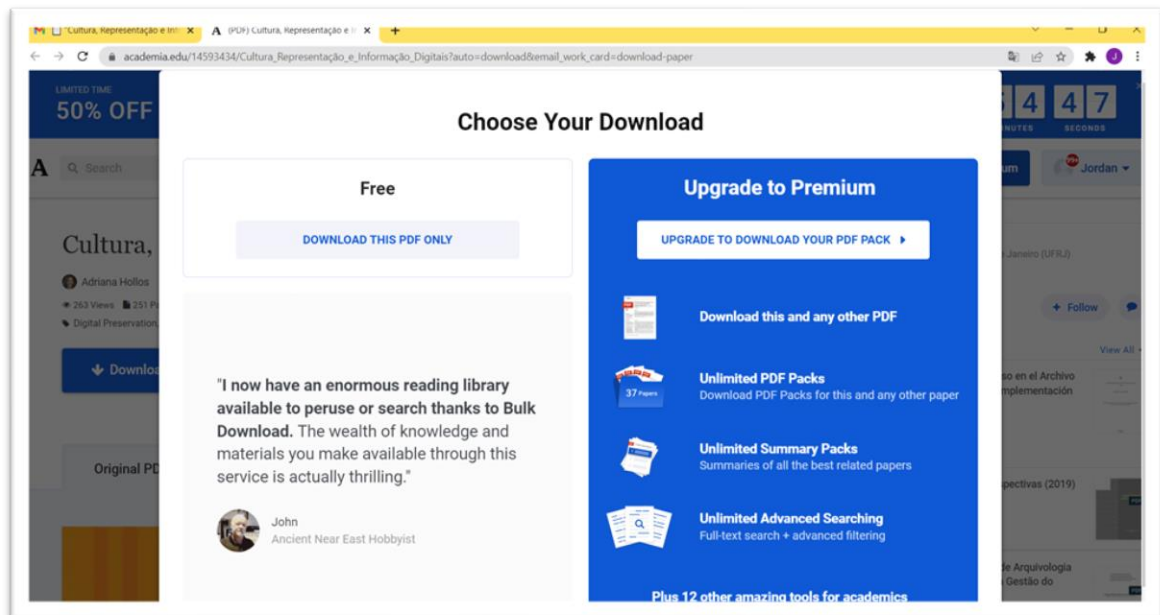
Figura 13 - Recomendação de artigo na plataforma Academia.edu



Fonte: Elaborada pelo autor (2022).

No campo do e-mail recebido com a recomendação na Figura 13 pode-se observar as opções para visualizar o artigo (botão em azul) e ao lado, para baixá-lo. Ao clicar para baixar, o sistema permite ao usuário que realize o *download* do artigo recomendado e, também, oferece um pacote extra de artigos para baixar, porém o usuário só conseguirá obter esse pacote extra caso faça um *upgrade* para se tornar um membro *Premium*, conforme a Figura 14:

Figura 14 - Interface com opções de download da Academia.edu



Fonte: Elaborada pelo autor (2022).

No caso da Figura 14, o sistema oferece ao usuário duas opções de *download*, a primeira de forma gratuita com apenas o artigo que foi recomendado e a segunda com um pacote de artigos e mais benefícios, como pacote de resumos e buscas avançadas, incentivando o usuário a se converter em usuário *Premium*.

Além da experiência do usuário, no espaço de ajuda dentro da plataforma é mostrado que as recomendações são criadas através dos seguintes pontos:

- ✓ Interesses de pesquisa do usuário e de outros acadêmicos que ele segue na rede;
- ✓ Artigos recomendados por outros usuários;
- ✓ Artigos adicionados na sua biblioteca ou artigos populares entre as pessoas que o usuário segue;
- ✓ Artigos que possuem muitas visualizações dentro da sua área de pesquisa.

2.2.3 Researchgate

O Researchgate é uma plataforma muito semelhante à Academia.edu. Ambas possuem as características de rede social com foco na comunidade científica, nas quais os usuários podem interagir e seguir uns aos outros para acompanhar as suas pesquisas e produções científicas.

A plataforma foi criada em 2008 visando solucionar alguns problemas de criação e compartilhamento da ciência. Existem por volta de 20 milhões de pesquisadores inscritos, vindos de diversos campos do conhecimento e diferentes países. Os pesquisadores utilizam o ResearchGate para se conectarem, para colaborarem e para compartilharem seus trabalhos. A missão do Researchgate é interligar o mundo da ciência e tornar a pesquisa aberta a todos (RESEARCHGATE, 2021).

Assim, o pesquisador pode incluir suas publicações para que outros usuários tenham acesso ao conteúdo, além de poder seguir e acompanhar os pesquisadores da sua área recebendo notificações automáticas sobre novas publicações.

No âmbito do Researchgate, de acordo com Monteiro-Krebs *et al.* (2021, p. 20, tradução nossa), “Ao encorajar essa prática (compartilhar o próprio trabalho do pesquisador) a plataforma pode oferecer a publicação gratuita para outros usuários, o que pode aumentar a adesão de novas pessoas.”

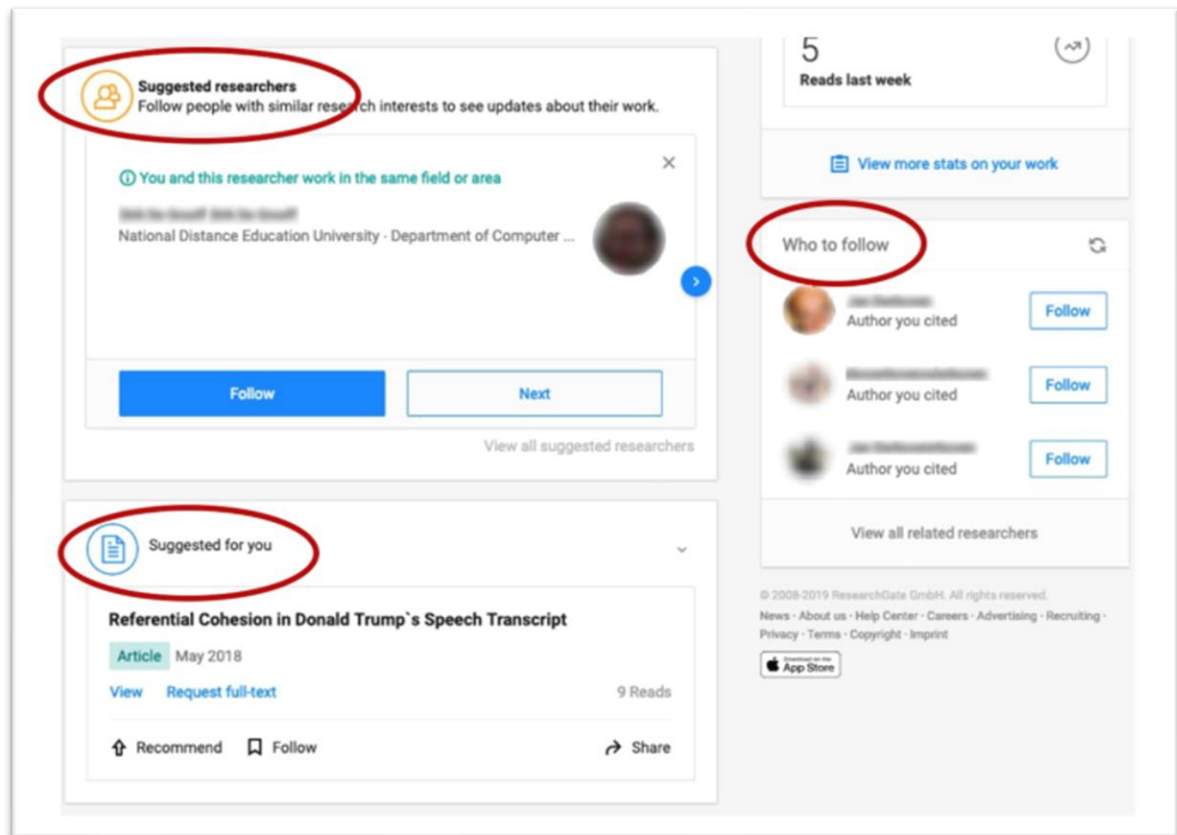
O Researchgate apresenta algumas funcionalidades para se destacar:

- ✓ Possibilidade de acessar e compartilhar publicações;
- ✓ Obter estatísticas sobre publicações (citações, *downloads*);
- ✓ Oportunidade de contato profissional;
- ✓ Encontro e colaboração com pares que atuam no mesmo campo;
- ✓ Existe uma seção para perguntas e respostas na qual é possível solicitar informações sobre pesquisas.

Conforme Monteiro-Krebs *et al.* (2021), após uma consulta ao Researchgate, a qual se questionou com base em quais dados e critérios eles criam as recomendações, a empresa respondeu via *e-mail* que considera e utiliza os metadados que possuem sobre o usuário, tal como nomes de artigos publicados e a interação prévia com a plataforma para apresentar conteúdo que possa ser relevante e interessante. Além disso, a empresa respondeu que parcialmente utiliza *cookies* para fazer isso, e informou um *link* para que a política de *cookies* da empresa pudesse ser visualizada.

Ainda, considerando a experiência deste autor no uso dessa plataforma, um pesquisador que considere algum documento científico como importante e relevante pode recomendá-lo para outros usuários, além disso, pode receber recomendações baseadas em quem o usuário segue na plataforma, assim como nas áreas fins escolhidos na configuração do perfil. Na Figura 15, pode-se visualizar algumas evidências de recomendações feitas pela plataforma.

Figura 15 - Evidências de recomendações no Researchgate



Fonte: Monteiro-Krebs *et al.*, (2021, p. 9).

A Figura 15 mostra indicativos de recomendação ao usuário nos círculos destacados em vermelho. No canto superior esquerdo são sugeridos pesquisadores que possuem o mesmo interesse para serem seguidos. No canto inferior esquerdo estão as sugestões para o usuário (artigos ou outros documentos). No canto direito aparecem sugestões de quem seguir baseadas em autores que foram citados pelo usuário.

Monteiro-Krebs *et al.* (2021), consideraram em seu trabalho, após exaustiva análise da interface da plataforma, que as principais entidades (elementos) envolvidas no processo de recomendação feito pelo Researchgate são: pesquisador, publicação, projeto de pesquisa, instituição, trabalho e perguntas. Isso é, o que autora chama de entidades são as categorias de informação que interagem com a plataforma e que são afetadas pelo algoritmo de recomendação. Os autores ainda frisam que é importante recordar que os algoritmos mudam constantemente e são frequentemente protegidos por sigilo comercial.

2.2.4 Protótipo W-RECMAS

O protótipo de SR W-RECMAS (*Web-RECommender system based on Multi-Agent System for academic paper recommendation*) foi proposto por Silvio Cesar Cazella, em sua tese de doutorado no ano de 2006, no Programa de Pós-Graduação em Computação da Universidade Federal do Rio Grande do Sul.

O objetivo desse sistema é recomendar artigos científicos para pesquisadores e estudantes, utilizando a base de dados do CV-Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e do sistema Qualis da Capes, para coletar as informações necessárias para se calcular o Ranqueamento do Recomendador¹ (métrica proposta para representar a relevância da opinião dos usuários que colaboraram na elaboração de uma recomendação para um usuário alvo). Outrossim, o sistema também cria comunidade de usuários que possuam o mesmo interesse. Esse protótipo funciona com os pesquisadores recebendo as recomendações em conformidade com o seu perfil, e receberão também a informação de qual a relevância da opinião do grupo ou do indivíduo que participou da recomendação (CAZELLA, 2006).

O protótipo W-RECMAS utiliza como diferencial em seu sistema a relevância da opinião do usuário, visando dessa forma quantificá-la nas áreas em que o usuário possui desejo, já que o perfil do usuário estará relacionado a uma ou mais áreas de interesse no sistema.

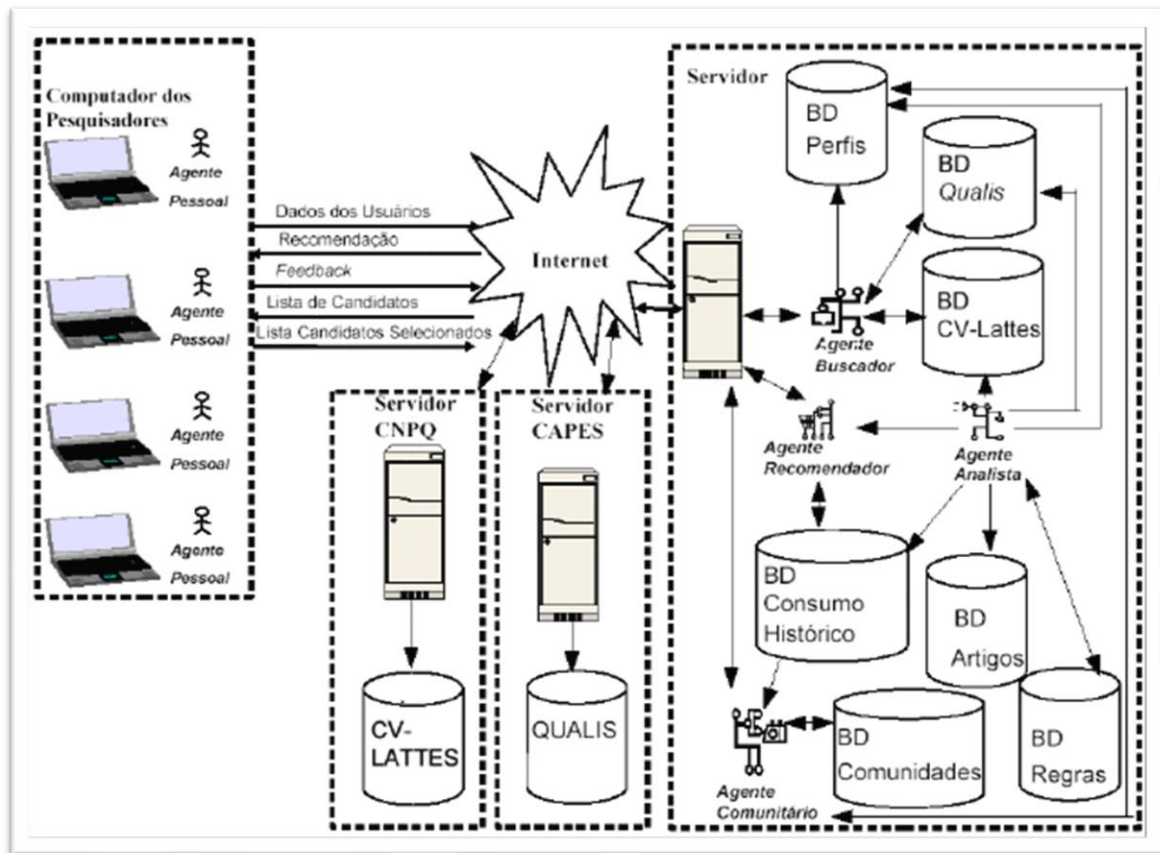
O currículo Lattes do pesquisador, utilizado neste sistema para compor o perfil de cada usuário, é uma base consolidada e uma importante fonte de informação, pois apresenta os documentos publicados, as áreas de interesse e atuação entre outros indicadores.

Para se cadastrar no W-RECMAS, o usuário preenche um formulário online onde informa seus dados como: nome completo, endereço de *e-mail*, nível acadêmico, áreas de interesse e tempo de estudo em cada área. Com esses dados o próprio sistema busca informações no currículo Lattes. No Lattes, o sistema utiliza dados qualitativos como áreas de interesse e dados quantitativos como, por exemplo, as métricas de produção. Com essas informações o sistema pode criar o perfil do usuário e calcular o Ranqueamento do recomendador.

De acordo com Reategui, Cazella e Osório (2006, p. 5) “Um dos pontos fundamentais para trabalhar com sistemas de recomendação está na correta definição do perfil dos usuários do sistema.”

Nesse sistema, os artigos científicos, assim como os usuários, estão relacionados a no mínimo, uma área de interesse. A Figura 16 apresenta uma visão macro da arquitetura do protótipo W-RECMAS:

Figura 16 - Visão macro da arquitetura do protótipo W-RECMAS



Fonte: Cazella (2006, p. 75).

A Figura 16 mostra o sistema realizando suas tarefas específicas. A saber, na construção do perfil do usuário esse entra no sistema com um mínimo dos seus dados, enquanto o sistema busca nas bases descritas anteriormente mais informações sobre esse usuário, tais quais suas produções científicas. Com as informações fornecidas pelo pesquisador, somadas as providas pelas bases, o sistema calcula o Ranqueamento do Recomendador para criar o perfil completo do usuário.

Dentro do W-RECMAS cada parte recebeu um nome e é responsável por um tipo de função específica detalhada a seguir:

- ✓ Agente Buscador – suprir o SR com informações dos usuários no sentido de criar o perfil, utilizando as informações iniciais providas pelo pesquisador e buscando outras no currículo Lattes e lista Qualis;
- ✓ Agente Analista – o mais complexo do sistema por realizar inúmeras tarefas tais quais: iniciar a análise do conteúdo do currículo e lista; calcular a métrica de Ranqueamento do Recomendador; associar o perfil do usuário com os itens

disponíveis, decidindo qual a melhor técnica de filtragem; prever novas áreas de interesse para o usuário;

- ✓ Agente Recomendador – projetado para comunicar a lista de recomendações ao Agente Pessoal, baseado no valor do Ranqueamento do Recomendador do usuário;
- ✓ Agente Pessoal – sua função fundamental é apresentar as recomendações para o usuário alvo;
- ✓ Agente Comunitário – procura criar comunidades de usuários que possuam o mesmo interesse, partindo do princípio de que uma troca direta de informações entre usuários pode ser proveitosa (CAZELLA, 2006).

Conforme visto, em relação ao tipo de filtragens, o protótipo W-RECMAS é um SR que utiliza a FH, empregando a FC e a FBC para realizar as recomendações para seus usuários, definindo o tipo a ser utilizado durante o processo e baseado em cada usuário e as suas particularidades.

2.2.5 Protótipo SysPaperAdvice

O SysPaperAdvice é um protótipo de SR que emprega a FH com a técnica ponderada e entrega as recomendações independentemente da quantidade de dados que o usuário inseriu no sistema. A FH é constituída pela FBC e a FC (BARROS; SILVEIRA; PERTILE, 2016).

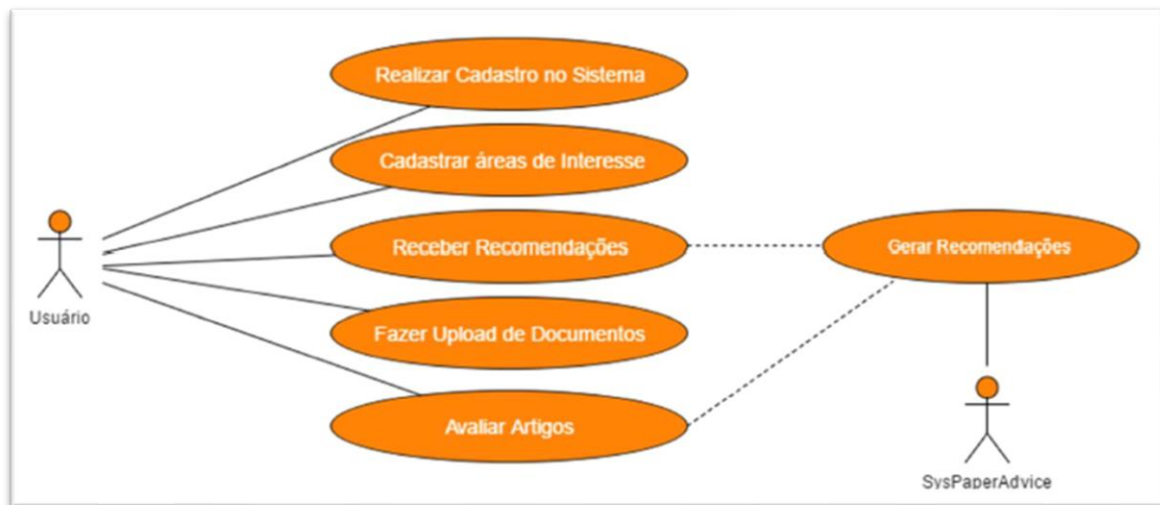
Nesse sistema, foram criados alguns requisitos para o seu desenvolvimento, são eles:

- ✓ o usuário pode realizar *upload* de documentos científicos no sentido de criar um repositório para serem avaliados e recomendados;
- ✓ ao se cadastrar, o usuário poderá definir até 4 palavras-chave sobre assuntos de seu interesse, que serão utilizadas para realizar a FBC;
- ✓ usuários avaliarão uma lista de artigos recebidos, com notas de 1 a 5. Baseado nessa avaliação, o sistema aplicará a correlação de *Pearson*¹ e os artigos que serão recomendados ao usuário serão calculados pela média ponderada. Os cálculos feitos serão aplicados para realizar a FC;
- ✓ as recomendações poderão ser visualizadas distintamente pelo tipo de filtragem utilizada, FBC ou FC.

Na Figura 17 pode-se visualizar as ações que o usuário do sistema pode desempenhar.

¹ Medida que captura a similaridade entre os vetores de avaliação de dois usuários.

Figura 17 - Diagrama de casos de uso do sistema



Fonte: Barros; Silveira e Pertile (2016, p. 15).

A Figura 17 revela as funcionalidades do sistema. Inicialmente o usuário realiza o cadastro, incluindo as suas áreas de interesse que serão utilizadas pela FBC para o envio das recomendações. Após estar cadastrado, o usuário pode fazer o *upload* de documentos que serão colocados à disposição dos outros usuários do sistema, para que esses possam realizar a avaliação sobre os documentos, avaliação essa que será utilizada pela FC. O usuário também poderá avaliar os documentos disponíveis utilizando notas de 1 a 5, que serão usadas pelo sistema para recomendar novos documentos baseados no seu perfil.

O algoritmo utilizado especificamente para fornecer recomendação baseada na FC do SysPaperAdvice, conforme Barros, Silveira e Pertile (2016) explicam, funciona de acordo com os seguintes passos:

1. O usuário quando se cadastra pela primeira vez no sistema recebe uma lista gerada de forma aleatória, com 21 artigos para serem avaliados com notas de 1 a 5;
2. Baseado nas avaliações realizadas, por meio da correlação Pearson, o sistema calcula a proximidade do novo usuário com os antigos que já fazem parte do sistema;
3. Os usuários são agrupados pelos resultados dos cálculos de similaridade, grupos com similaridade igual ou superior a 0.75 formam sua vizinhança;
4. Nos artigos avaliados em comum pela vizinhança é aplicado um cálculo de média ponderada, que servirá para calcular a predição do usuário em relação a todos outros artigos que fazem parte do sistema;
5. O sistema recomendará todos os artigos que tiverem nota de predição igual ou maior do que três (3), em ordem decrescente.

Cada vez em que o usuário acessar a opção “Recomendação FC” no menu do sistema ou realizar uma nova avaliação de artigos, o algoritmo será executado novamente. As tecnologias empregadas no sistema, relacionadas com a linguagem de programação, são PHP e SGBD MySQL.

O algoritmo aplicado no sistema que executa a FBC, no qual o índice *Jaccard* é utilizado, opera de acordo com o andamento mostrado a seguir:

1. Usuário escolhe áreas de interesse em um formulário quando se inscreve no sistema, podendo incluir até quatro palavras afins, que ficam armazenadas;
2. Sistema consulta os documentos cadastrados, que são transformados do formato PDF para o formato texto;
3. Similaridade entre as palavras incluídas pelo usuário e as palavras nos documentos são calculadas pelo índice *Jaccard*;
4. Quando o resultado do índice calculado for maior que zero, o sistema fará a recomendação do documento.

Caso o usuário modifique alguma de suas áreas de interesse ou acesse no menu do sistema a opção “Recomendação FC”, o algoritmo realizará todo o processo novamente (BARROS; SILVEIRA; PERTILE, 2016).

2.2.6 arXiv

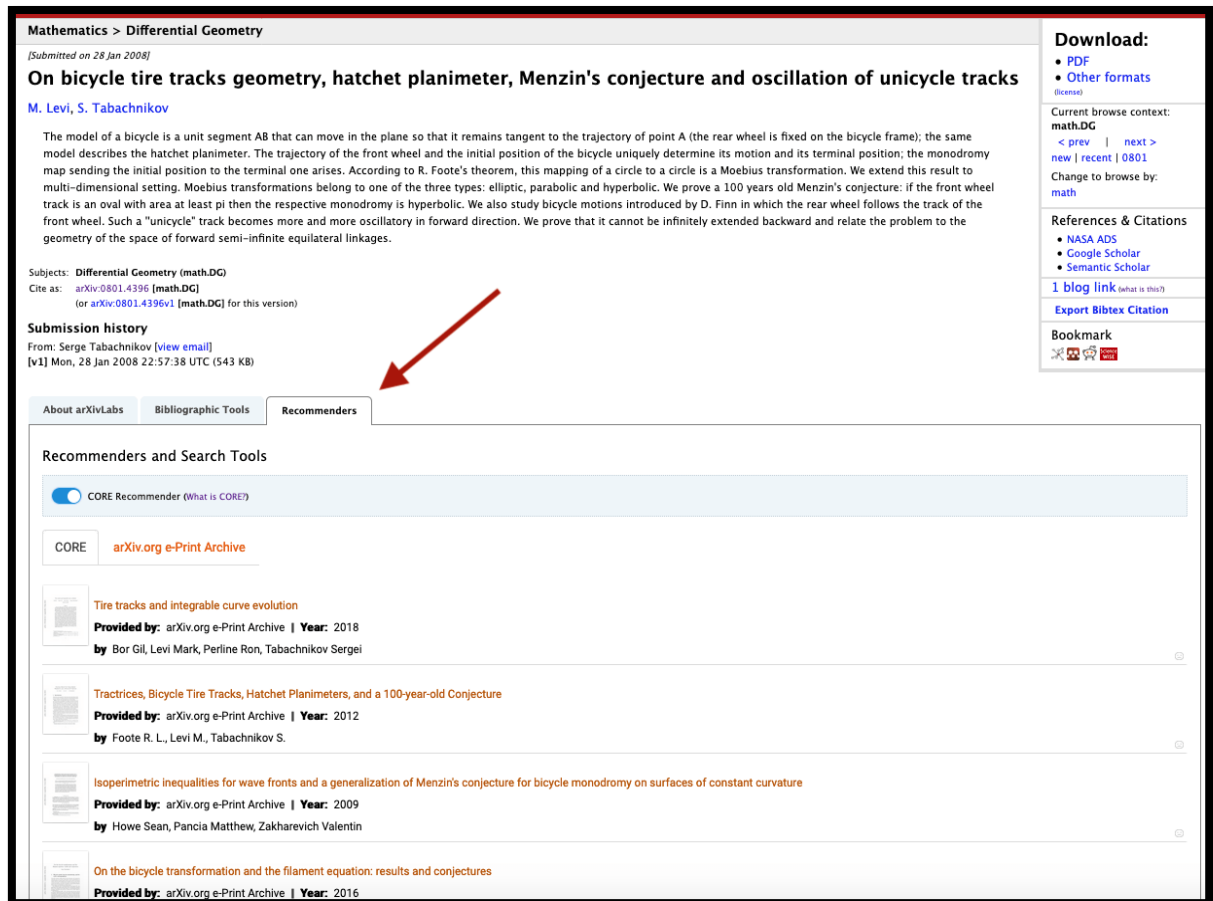
A plataforma arXiv hospeda mais de dois milhões de artigos acadêmicos em oito áreas temáticas, possuindo moderadores voluntários que cuidam da curadoria. Oferecem aos seus usuários diversos serviços como: submissão de artigos, compilação, produção, busca e descoberta, recuperação e distribuição para leitores na *web*. Fundada em 1991, inicialmente como um servidor de pré-prints, hoje é mantida pela Cornell Tech,² que mantém ênfase em Acesso Aberto e colaboração e conhecimento para prosperar, atributos considerados como sua base de sustentação. Usuários da plataforma podem enviar artigos sem custo, que passam pelos moderadores que classificam e hospedam na respectiva área. Não existe revisão por pares, portanto o conteúdo dos artigos enviados é de responsabilidade do remetente (CORNELL UNIVERSITY, 2022).

No ano de 2020, o arXiv e o CORE - maior coleção de artigos de pesquisa de Acesso Aberto do mundo - anunciaram uma parceria, a qual o arXiv irá fornecer recomendações de

² <https://www.tech.cornell.edu/>

artigos de pesquisa para seus usuários. As recomendações serão providas pela interface de programação de aplicações (API) do CORE (BEEL, 2020). A seguir a interface e suas abas na Figura 18:

Figura 18 - Interface do arXiv com a aba de recomendações



Fonte: Beel (2020).

As recomendações são geradas por um algoritmo que se baseia no texto completo e nos metadados dos artigos, isso inclui o título, o resumo e o(s) autor(es). Tanto os arquivos hospedados no arXiv quanto no CORE poderão ser recomendados, num total de cinco artigos, todos com acesso total ao seu conteúdo (CORNELL UNIVERSITY, 2020).

A Cornell University (2022) explica que no modelo de negócio da arXiv, não há cobrança de taxas dos seus usuários para enviarem artigos e utilizarem a plataforma. A plataforma foi fundada para pesquisa e educação, sem fins lucrativos, por isso conta com membros – universidades, bibliotecas, institutos de pesquisa e laboratórios - que financiam e mantém a arXiv como uma grande plataforma de compartilhamento de Acesso Aberto.

2.2.7 CORE

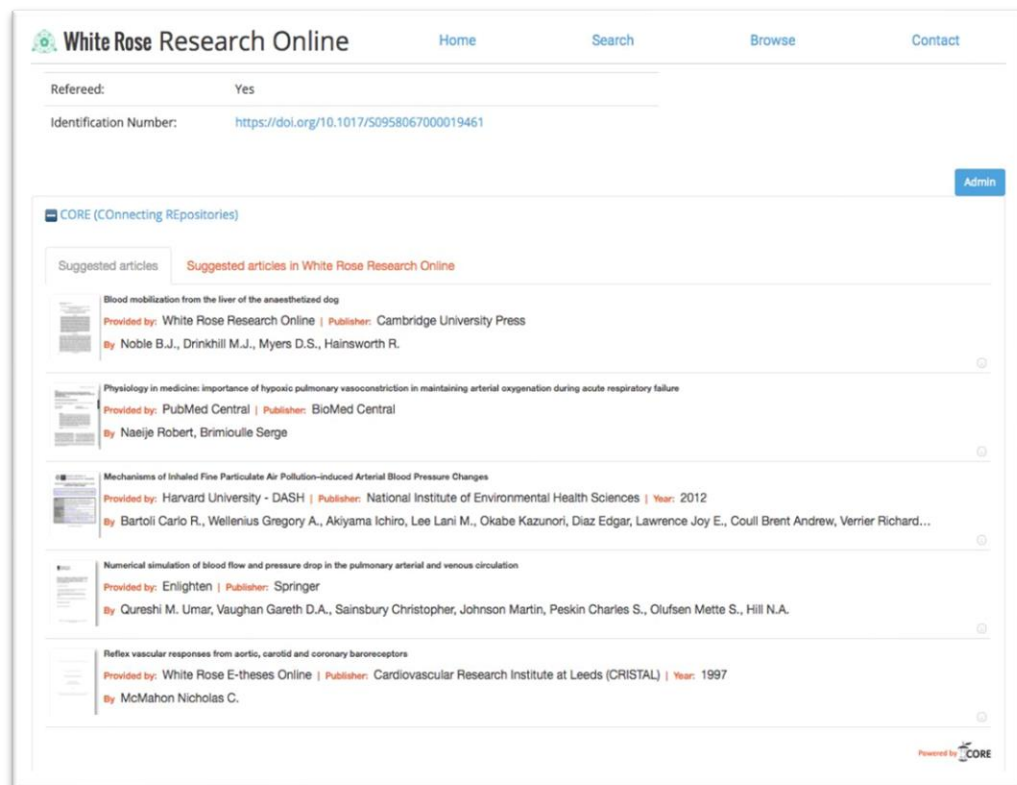
CORE é o maior agregador do mundo de artigos de pesquisa de Acesso Aberto de repositórios e de periódicos. É um serviço sem fins lucrativos dedicado à missão de Acesso Aberto. Entre outras funcionalidades, a plataforma atende a rede global de repositórios e periódicos, aumentando a capacidade de descoberta e reutilização de conteúdo de Acesso Aberto. Além disso, fornece soluções para gerenciamento de conteúdo e para a descoberta e acesso de máquina escalável para pesquisa. Os serviços providos pela CORE apoiam uma ampla gama de partes interessadas, especificamente pesquisadores, público em geral, instituições acadêmicas, desenvolvedores, financiadores e empresas de diversos setores (CORE, 2022).

Além dos serviços citados, a CORE também possui um SR de artigos de científicos, conforme visto no item anterior sobre o arXiv, que possui uma parceria com a CORE, fornecedora do SR.

De acordo com a Core (2022), seu SR funciona como um *plugin* para repositórios, periódicos e interfaces da *web*, provendo sugestões de artigos relevantes que estejam relacionados com os artigos que o usuário está lendo ativamente. Todos os artigos recomendados são gratuitos, valorizando a visibilidade do conteúdo de AA. Os artigos recomendados, mais de 25 milhões de textos completos, fazem parte da base da CORE.

Portanto, após um usuário visualizar uma página de um artigo em um repositório que possua o plugin instalado, essa informação sobre o item visitado, que pode incluir o identificador do item e até seus metadados, é enviada à CORE. A resposta é direcionada de volta ao repositório com uma lista de artigos em duas abas: artigos sugeridos em outros repositórios e artigos sugeridos neste repositório, conforme a Figura 19 (KNOTH *et al.*, 2017).

Figura 19 - Plugin do SR da CORE



Fonte: Knoth *et al.* (2017, p. 3).

O SR da CORE utiliza alguns mecanismos para aumentar a qualidade das recomendações geradas, Knoth *et al.* (2017) esclarecem que apenas são recomendados:

- artigos com texto completo de Acesso Aberto;
- artigos com pelo menos um conjunto mínimo de atributos de metadados;
- artigos com miniatura gerada;
- artigos que não tenham tido *feedback* negativo.

Sabe-se que mesmo com todos os esforços e precauções, algumas recomendações ainda podem ser consideradas irrelevantes para um usuário, dessa forma, no sentido de diminuir esse tipo de ocorrência, é disponibilizado por meio de um botão a possibilidade de reportar à CORE que tal recomendação foi irrelevante, assim o item não será recomendado novamente.

O fundador do CORE, Petr Knoth, detalha alguns processos do algoritmo utilizado. Segundo Knoth *et al.* (2017), um documento qualquer pode ser usado como referência para o algoritmo de recomendação do CORE, que o utiliza como um dado de entrada no sistema. Esse documento é representado por recursos como: título, autor, resumo e ano de publicação, que são obtidos com a utilização do Modelo de Espaço Vetorial. Caso esse documento possua outro documento similar no CORE, no momento da consulta o sistema também utiliza o texto

completo e metadados, que podem ser fornecidos de forma explícita pelos repositórios onde o documento foi obtido. Os metadados utilizados podem incluir, por exemplo, o idioma da fonte e as palavras-chave. Outrossim, o CORE ainda conta com identificadores como o DOI, *downloads* e indicadores de leitura do Mendeley, recursos esses utilizados para que correspondências mais próximas possíveis sejam obtidas na base de artigos do CORE, no momento da consulta. Nem todos os recursos utilizados possuem a mesma importância no sistema, no algoritmo de classificação interna, recursos são reforçados assertivamente ou negativamente para produzir melhores recomendações. Em relação aos recursos numéricos como data de publicação, o algoritmo aplica uma função para promover recomendações de artigos mais recentes.

Um dos desafios para a plataforma é a dificuldade em possibilitar recomendações personalizadas para repositórios, em razão de que para isso é necessário um perfil criado pelo usuário, e os repositórios foram criados sem a necessidade de um login de usuário. Portanto, o ponto chave é como manter repositórios abertos para usuários sem *login* e ao mesmo tempo fornecer personalização para outros com *login*. Pensando nisso, Knoth *et al.* (2017) sugerem que, teoricamente, pesquisadores poderiam fazer o *login* em repositórios via ORCID iD, assim seriam reconhecidos em qualquer outro repositório visitado, possibilitando soluções personalizadas com base nos dados contidos em seus perfis do ORCID como em suas publicações, etc.

O SR da CORE tem características únicas e recomenda artigos completos e não apenas resumos de todos os repositórios globais e periódicos de Acesso Aberto, garantindo Acesso Aberto para o público em geral, mesmo sendo desenvolvido e destinado para a rede de repositórios. Atualmente, o tipo de filtragem usada no SR é a FBC, e pensando em expansão de técnicas de filtrações e em melhorias futuras, existe a vontade de acrescentar a FC para que funcione concomitantemente com a FBC. Adicionalmente, também existe a pretensão de implementar a recomendação de outros itens que fazem parte do processo de pesquisa como oportunidades de financiamento à pesquisa e a eventos de pesquisa (KNOTH *et al.*, 2017).

2.2.8 SisRecAC

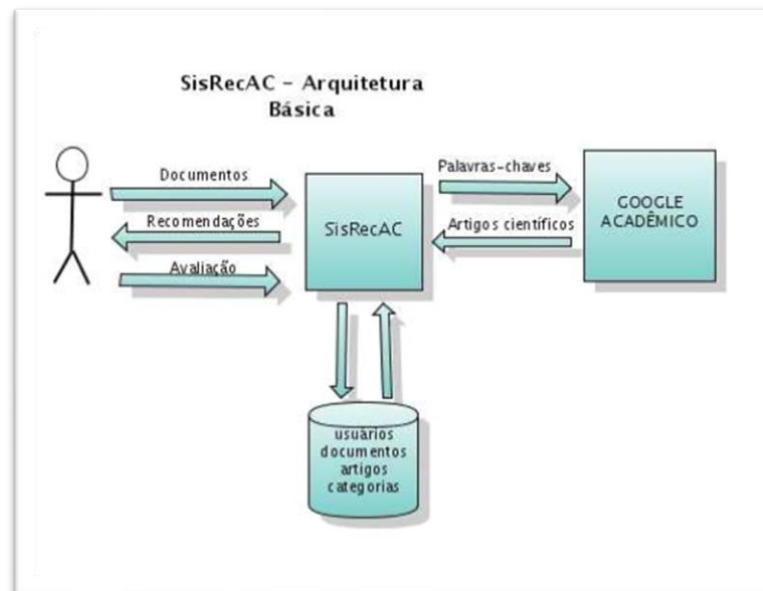
Esse sistema foi desenvolvido pelo Centro de Informática da Universidade Católica de Pelotas – UCPEL para recomendar artigos científicos indexados pelo Google Acadêmico e para analisar os melhores métodos para extração de palavras-chave de documentos.

O SysRecAC se propõe a ser um sistema de metabusca, que utiliza o Google Acadêmico para descobrir artigos científicos baseado em palavras-chave que são extraídas de documentos, que são armazenados no sistema pelos usuários. O objetivo é que os artigos recomendados sejam similares a esses documentos armazenados.

Nesse sistema, a extração das palavras-chave dos documentos ocorre por meio da contagem do número de ocorrências no texto, sendo que as palavras chamadas *stop-words* não são consideradas. Conforme explica Aggarwal (2016), são consideradas *stop-word* palavras comuns que aparecem muitas vezes em um texto, mas não são específicas para o item que se procura, como por exemplo: artigos, preposições, conjunções e pronomes.

O funcionamento do sistema inicia-se quando o usuário realiza o cadastro e, após ser autenticado no sistema, já é possível enviar documentos chamados de documentos base. Esses documentos base serão incluídos em categorias, que podem ser criadas ou já existem, escolhidas por quem os enviou. Ao escolher uma categoria, o sistema mostra ao usuário 5 recomendações de documentos científicos para que, dessa forma, ele possa avaliá-los em relação ao documento base (AVILA, 2006). A Figura 20 mostra a arquitetura básica do sistema:

Figura 20 - Arquitetura básica do sistema SisRecAC



Fonte: Avila (2006, p. 23).

A Figura 20 demonstra a estrutura básica do sistema, na qual o usuário além de receber a recomendação, também avalia o seu conteúdo servindo como *feedback*. O sistema utiliza palavras-chave e as envia para o Google Acadêmico, que retorna artigos científicos que serão recomendados.

De acordo com Avila (2006), os documentos enviados pelos usuários em formato PDF, de onde serão retiradas as palavras-chave, precisam ser convertidos para o formato texto, e essa conversão é efetivada pelo software chamado xpdf,³ via chamadas ao sistema operacional com o comando do PHP.

O SysRecAC pode ser considerado um SR com FH, pois utiliza as técnicas de filtragem FBC, por meio de palavras-chave no texto, e a FC quando recomenda documentos para outras pessoas. Além disso, o sistema utiliza a coleta de dados do usuário de forma explícita.

A linguagem de programação utilizada no SysRecAC foi a PHP 4, o servidor foi Linux RedHat versão 9 e o banco de dados PostgreSQL 8.1.

2.3 ALGORITMOS DE SISTEMAS DE RECOMENDAÇÃO

Azambuja, Morais e Filipe (2021) consideram que os SR são categorias de algoritmos que utilizam inteligência artificial para realizarem recomendações relevantes, empregando técnicas diferentes de filtragem de informação.

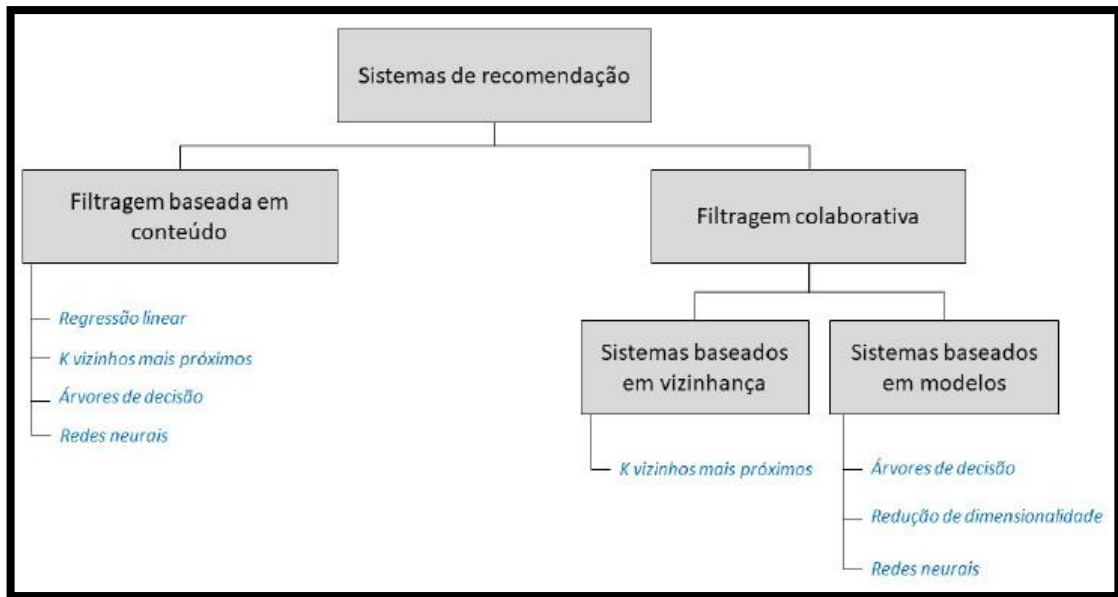
Conforme Gillespie (2018, p. 98) “Os algoritmos são máquinas inertes e sem sentido até serem combinados com bancos de dado para com eles funcionar.”

Um dos pioneiros no uso de SR para *e-commerce* é a plataforma *Amazon*. Seu algoritmo coleta as informações dos usuários de forma explícita e implícita, e as utiliza para a construção do seu sistema. No caso das informações explícitas, os usuários classificam produtos comprados utilizando uma escala de 1 a 5, e por meio implícito, a plataforma reúne informações sobre a navegação do usuário, produtos comprados e categorias visitadas (AGGARWAL, 2016). A *Amazon* possui um forte SR em sua plataforma e utiliza como modelos de recomendação a FH, composta da FBC e FC.

Conforme explica Bonilla (2020), sistemas que utilizam a FC são subdivididos em dois grupos, sistemas baseados em vizinhança e em modelos. No seu trabalho, a autora aborda cinco algoritmos de aprendizado de máquina em SR, a saber: regressão linear; método dos k-vizinhos mais próximos (KNN); redução de dimensionalidade; árvores de decisão e redes neurais artificiais. A Figura 21 apresenta os tipos de filtragem e seus algoritmos.

³ <http://www.xpdfreader.com/>

Figura 21 - Subdivisões de SR e alguns possíveis algoritmos



Fonte: Bonilla (2020, p. 37).

Fica evidenciado na Figura 21 que dos algoritmos citados pela autora em seu trabalho, a maioria deles pode ser utilizada por ambos os tipos de filtragens, enquanto os outros possuem limitações técnicas e são usados em apenas um tipo.

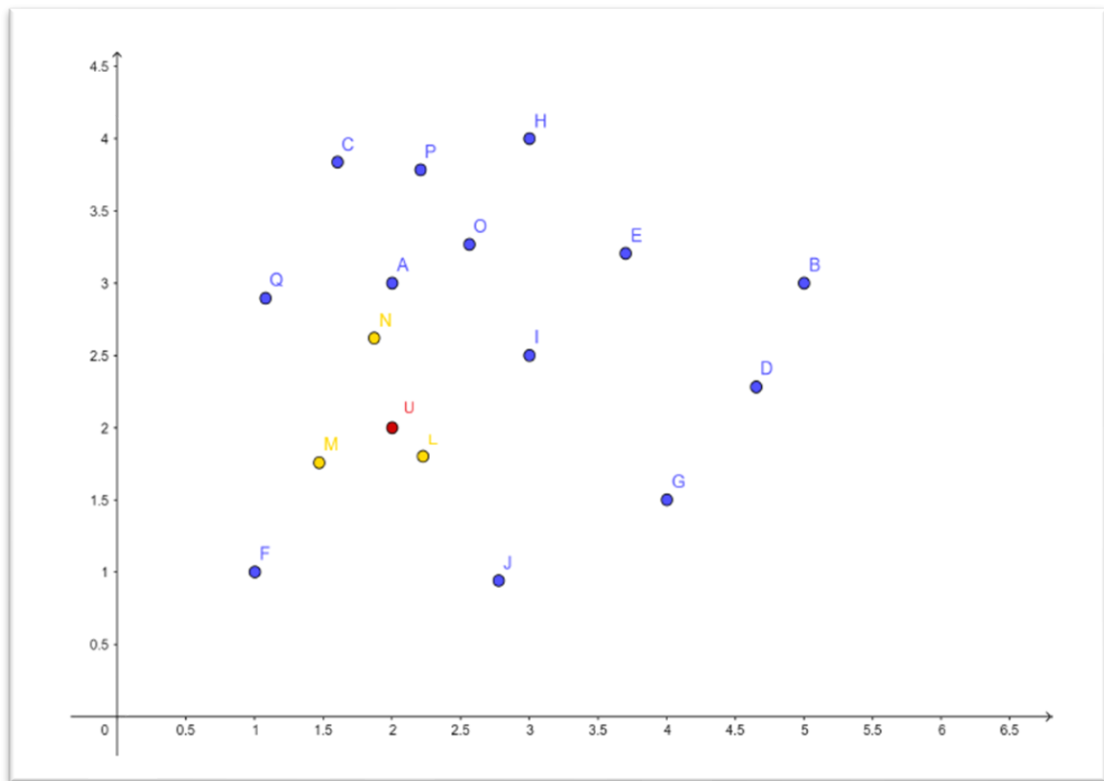
De acordo com Bonilla (2020, p. 32) “De maneira geral, um algoritmo de aprendizado de máquina recebe um conjunto de informações – ao que chamamos de atributos de entrada – e nos retorna um conjunto de respostas, também conhecidas como atributos de saída.”

A seguir, os algoritmos mencionados serão detalhados.

2.3.1 K-vizinhos mais próximos

O algoritmo chamado de K-vizinhos mais próximos ou *K-nearest neighbors*, realiza as predições baseadas em comportamentos de usuários que pertencem a uma vizinhança, que é definida através de critérios de semelhança entre usuários (BONILLA, 2020). Um exemplo de classificação KNN é mostrado a seguir na Figura 22:

Figura 22 - Exemplo de classificação KNN



Fonte: Bonilla (2020, p. 57).

O algoritmo KNN que aparece na Figura 22 pode ser aplicado em sistemas que utilizem FC e FBC. De acordo com Kubat (2017), o fato de cada exemplo poder ser representado por um ponto em um gráfico torna possível calcular a distância geométrica entre qualquer um dos pontos, utilizando por exemplo a distância euclidiana. Portanto, a similaridade será maior quanto mais próximos entre si estiverem os pontos.

Diante disso, Souto Junior e Bezerra (2020, p. 60) explicam o funcionamento do método KNN:

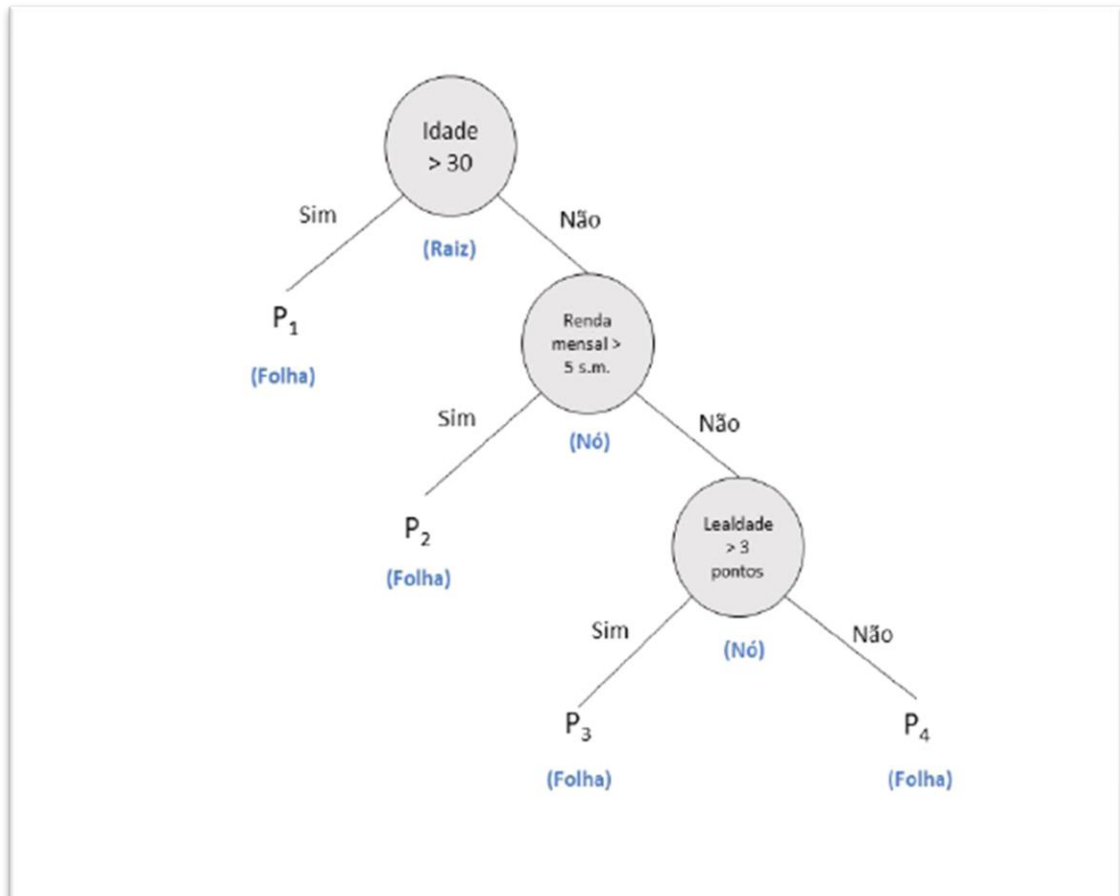
Dada a classificação feita por um grupo de usuários a um conjunto de itens, a tarefa do algoritmo é encontrar os vizinhos mais próximos a partir de algum critério de semelhança, ou em outras palavras, encontrar os usuários mais correlatos com o usuário alvo, e em seguida prever a sua avaliação a partir daquelas de seus semelhantes.

Diversos critérios podem ser utilizados para definir a semelhança entre os usuários, sendo que a correlação de Pearson é considerada como frequentemente usada. O método KNN pode ser aplicado tanto em função do usuário quanto do item.

2.3.2 Árvores de decisão

As árvores de decisão possuem essa denominação devido ao seu formato de árvore, com raiz, nós e folhas conforme a Figura 23 demonstra:

Figura 23 - Exemplo de árvore de decisão para frequência de compras de um produto



Fonte: Bonilla (2020, p. 66).

Conforme representado na Figura 23, o nó com maior hierarquia, isso é, que possui maior importância, é chamado de raiz (idade >30). Abaixo da raiz aparecem os outros nós internos que fazem novas ligações, os nós que não se ligam a nenhum outro são chamados de folha.

Para a construção de uma árvore de decisão, Bonilla (2020) considera as seguintes etapas:

- Estabelecer variáveis para os nós, da raiz até o último nó;
- Escolher as regras para divisão dos dados a partir dos nós;

- Definir regras para quando um ramo deverá formar uma folha ou continuar como um novo nó;
- Fixar o valor previsto para o alvo em cada folha.

Árvores de decisão funcionam como divisões dos dados com o uso de critérios hierárquicos, chamado de critérios de divisão. São regularmente usadas em classificação de dados e projetadas nos casos em que a variável dependente é categórica (AGGARWAL, 2016).

Algumas vantagens da utilização de Árvores de Decisão segundo Lemos, Steiner e Nievola (2005) são:

- a possibilidade de se construir modelos para qualquer função, contanto que o número de exemplos seja suficiente;
- não constituírem distribuição particular para os dados;
- possuírem alto grau de compreensão;
- que os atributos usados podem ser categóricos ou numéricos.

2.3.3 Redes Neurais

Aggarwal (2016) revela que redes neurais funcionam de forma a representar o cérebro humano no seu uso de neurônios, que são conectados uns aos outros por meio de conexões sinápticas. No sistema biológico o aprendizado é realizado alterando a força das conexões sinápticas em resposta a estímulos externos. Em redes neurais artificiais, a unidade básica de computação também é chamada de neurônio, e as forças das conexões sinápticas correspondem a pesos, que definem os parâmetros usados pelo algoritmo.

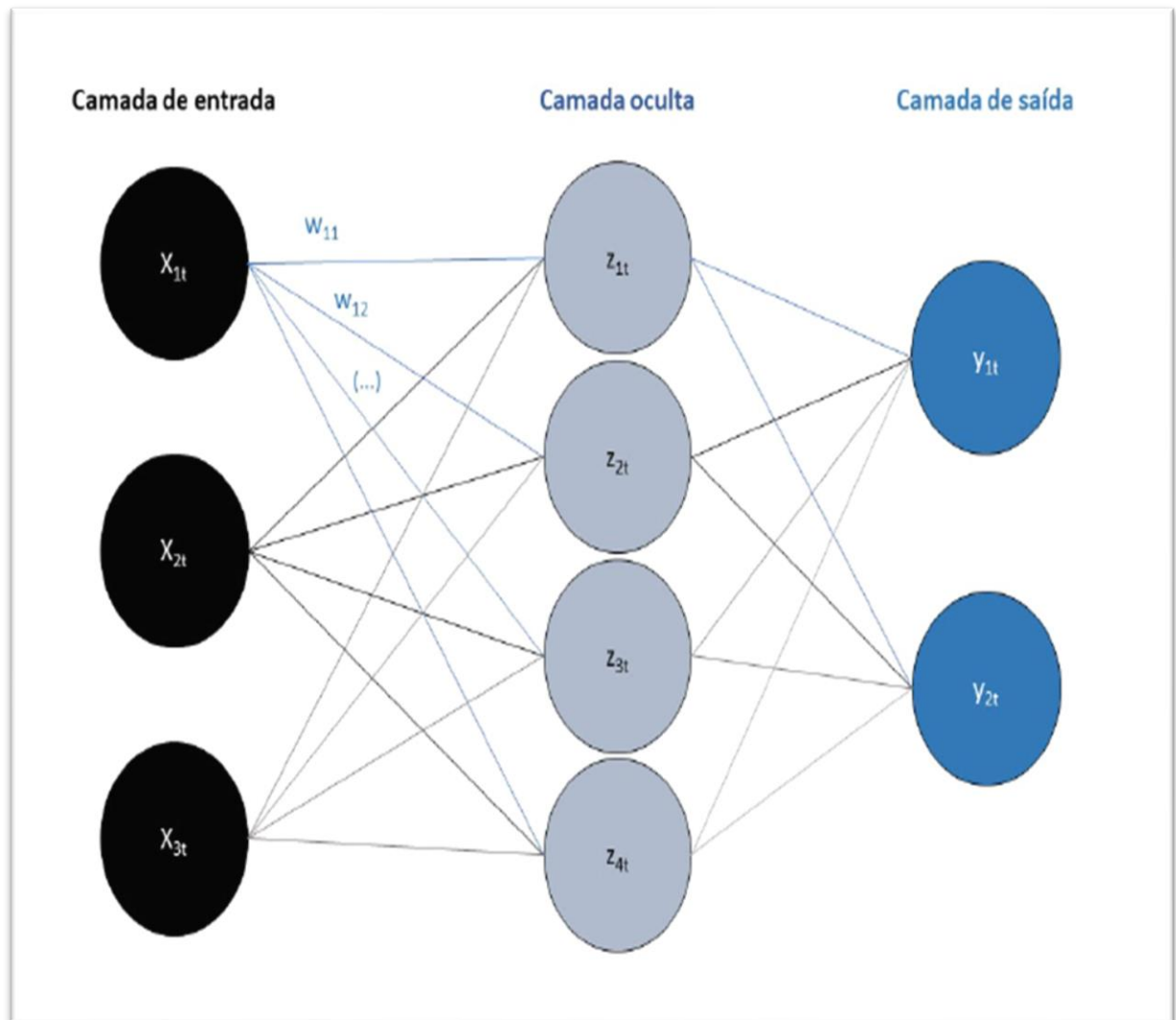
O modelo de rede neural mais utilizado é conhecido como *MultiLayerPerceptron* – MLP. Funciona como redes que retratam uma ou múltiplas camadas de neurônios entre as camadas de entrada de dados e saída de resultados, chamadas camadas intermediárias, e são treinados através do algoritmo de retro propagação (KROSE; VAN DER SMAGT, 1993 *apud* LEMOS; STEINER; NIEVOLA, 2005).

Sobre a organização das redes neurais, Bonilla (2020, p. 85) atesta que:

Os neurônios (também chamados de nós) de uma rede neural são organizados em camadas que podem ser de três tipo: camada de entrada, camadas ocultas e camada de saída. Os nós de cada camada são ligados aos nós da próxima camada através de conexões, sendo que cada conexão possui um peso [...].

A Figura 24 esquematiza uma rede neural onde é possível observar as camadas conforme explicação prévia.

Figura 24 - Rede neural



Fonte: Bonilla (2020, p. 85).

Conforme explica Kubat (2017), enquanto não há comunicação entre os neurônios da mesma camada, neurônios adjacentes estão totalmente interconectados e cada ligação entre eles está associada a um peso. Podemos visualizar na Figura 24 que a ligação entre os neurônios X_{1t} e z_{1t} possui o peso w_{11} , já a ligação entre os neurônios X_{1t} e z_{2t} possui um peso diferente, w_{12} .

Conclui-se que o início do processo está nos dados que alimentam as camadas de entrada, essa camada envia os dados com pesos ponderados para as camadas ocultas e as camadas de saídas recebem esses impulsos das camadas ocultas criando o resultado.

2.3.4 Regressão Linear

“A ideia básica na regressão linear é assumir que as classificações podem ser modeladas como uma função linear das frequências das palavras” (AGGARWAL, 2016, p. 158, tradução nossa).

Bonilla (2020) afirma que a regressão linear tem como objetivo encontrar uma equação linear para explicar a relação entre os atributos de entrada (variáveis de previsão) e os atributos de saída (variável de resposta).

De acordo com Aggarwal (2016), os modelos baseados em regressão têm o mérito de poderem ser usados para vários tipos de classificações, como classificações binárias, classificações baseadas em intervalos ou classificações numéricas. O modelo linear é um exemplo de modelo de regressão adequado para avaliações de valor real. Na prática, as classificações podem ser unárias, binárias, baseadas em intervalos ou categóricas (pequeno número de valores ordinais). Vários modelos lineares foram projetados para diferentes tipos de variáveis de classe de destino.

Esse tipo de algoritmo pode ser utilizado na construção de SR que utilizam FBC. Considerando como exemplo alguns dados onde um usuário realizou avaliações de algumas músicas, conforme a Quadro 4:

Quadro 4 - Exemplo de avaliações de músicas

Código da Música	Estilo	País	Avaliação
01	Pop	USA	2,9
02	Rock	Australia	3,5
03	MPB	Brasil	4
04	Reggae	França	4,2
05	Jazz	USA	4,5

Fonte: autor (2022).

As músicas são representadas por um código na coluna inicial e nas duas colunas seguintes são apresentadas algumas características de cada música (estilo e país). Por fim, na última coluna as avaliações de cada uma delas com notas de 1 a 5, feitas por um usuário. De acordo com Bonilla (2020), a regressão linear, nesse exemplo, buscaria uma forma de encontrar uma equação linear no sentido de relacionar as características das músicas com a avaliação dada pelo usuário. Desse modo, haveria a possibilidade do sistema prever as avaliações que o

usuário daria a músicas que ele ainda não conhece, possibilitando assim, elementos para recomendá-las ou não (BONILLA, 2020).

2.3.5 TF-IDF

O algoritmo TF-IDF, visto na subseção que versa sobre FBC, é bastante utilizado para estabelecer pesos das palavras em um texto. Conforme visto, o TF-IDF possui como objetivo atribuir importância em forma de peso para palavras selecionadas em um grupo de documentos, quanto maior o número de vezes em que uma palavra é encontrada em um documento maior é o seu peso. E caso a mesma palavra apareça em vários documentos, possuirá menos importância.

De acordo com o que foi explicado, o TF-IDF é regularmente utilizado na área de recuperação e mineração de dados, que consiste em uma técnica apropriada para se extrair conhecimento em um banco de dados e segue sendo utilizada em SR, no sentido de preparar os dados que serão utilizados pelos algoritmos.

A mineração de textos (*text mining*), inspirada na mineração de dados, recebeu destaque e significância com o crescimento da internet. Funciona de forma parecida, obtendo informação de qualidade a partir de texto em linguagem natural, onde é possível colher informação relevante em bases de textos não estruturado ou semiestruturado (CAZELLA; NUNES; REATEGUI, 2010).

Um sistema pode obter informações sobre um usuário coletando-as através de itens consumidos ou pesquisas realizadas por ele, utilizando a indexação de frequência de termos, ou seja, TF-IDF. Nesse caso, o algoritmo utiliza vetores com dimensões diferentes para cada palavra na base de dados para descrever as informações dos documentos e das necessidades dos usuários. Cada componente do vetor é a frequência que a própria palavra aparece em um documento ou em uma consulta do usuário, dessa forma, os vetores do documento que estão próximos aos vetores de consulta são considerados os mais relevantes (HERLOCKER *et al.* 2000 *apud* CAZELLA; DRUMM; BARBOSA, 2010).

A frequência do termo (TF), que mede quantas vezes um termo (palavra) aparece em um documento, se dá pela divisão do número de vezes que o termo 'x' aparece no documento dividida pelo número total de termos do documento. Em relação ao inverso da frequência (IDF), que determina a importância geral do termo 'x' no grupo de documentos, pode ser calculada pela divisão da quantidade total de documentos dividida pelo número de documentos com o termo 'x'. Consequentemente, o algoritmo TD-IDF é alcançado com o resultado da

multiplicação das frequências TF e IDF, sendo expresso como: $TF-IDF = TF * IDF$ (AVILA, 2006).

Para que um algoritmo funcione, isso é, que produza uma recomendação como resultado (saída), é necessário que exista algum dado inicial que funcione como entrada. No caso do TF-IDF, as palavras-chave podem ser consideradas como as entradas necessárias para que o sistema produza a recomendação.

2.4 ANÁLISE E MODELAGEM DE SISTEMAS DE INFORMAÇÃO

Para Laudon e Laudon (2004, p. 7) “Um sistema de informação pode ser definido como um conjunto de componentes inter-relacionados que coleta (ou recupera), processa, armazena e distribui informações destinadas a apoiar a tomada de decisões [...]”. Já O’Brien e Marakas (2013, p. 2) conceituam SI da seguinte forma: “[...] pode ser qualquer combinação organizada de pessoas, hardware, software, redes de comunicação, recursos de dados e políticas e procedimentos que armazenam, restauram, transformam e disseminam informações em uma organização.

Um Sistema de Recomendação processa, armazena e distribui informações destinadas a tomada de uma decisão, para isso envolve pessoas, hardware, software, redes, demais recursos de dados, políticas e procedimentos, portanto, entende-se que um SR é um tipo de sistema de informação e, por sua vez, para a sua modelagem e desenvolvimento, deve seguir todo o arcabouço de conhecimento e as boas práticas provenientes da ciência da computação, mais especialmente da área de engenharia de software, que trata da análise/modelagem de sistemas de informação.

No processo de construção de um SI existem metodologias a serem cumpridas para alcançar o seu objetivo com sucesso. Para Martin *et al.* (1994 *apud* VARAJÃO, 2002), a análise de sistema pode ser considerada uma das tarefas mais importantes no desenvolvimento de um SI e o seu êxito resulta de uma definição precisa e completa de requisitos.

Nesse contexto, em relação aos requisitos, Varajão (2002, p. 109) afirma que:

Basicamente, os requisitos de informação de um novo sistema envolvem a análise da estrutura existente e a identificação dos utilizadores potenciais do sistema, de que informação necessitam, onde, quando e como, com vista à definição concreta dos objetivos e funções dos novos sistemas.

Para O’Brien e Marakas (2013), a análise de sistemas procura antever como será o sistema antes do seu funcionamento, construindo um modelo lógico em que o sistema possa ser

visualizado, para assim ter a capacidade de identificar como cada atividade deverá ser realizada e quais as suas necessidades.

Varajão (2002, p. 108) afirma que “No desenvolvimento de um SI, os requisitos que o sistema pretende satisfazer deverão ser cuidadosamente analisados. A atividade responsável pela sua identificação é a Análise de Sistemas.”

Baseado nos conceitos apresentados, pode-se perceber que a análise de sistema é uma atividade que procura mapear todas as particularidades de cada etapa que será seguida para sua criação. A análise se concentra em reconhecer desde os anseios do usuário até capacidade do sistema de processar informações, passando pelas entradas e saídas.

A especificação de requisitos do sistema compreende uma versão simplificada da realidade e é apontada como o principal resultado da etapa de análise. As propriedades do SI para atender as necessidades de informação dos usuários do sistema são definidas pelos requisitos do sistema (EDWARDS *et al.* 1991 *apud* VARAJÃO, 2002).

2.4.1 Levantamento de requisitos de informações

O processo de descobrir todas as funcionalidades que um sistema deve executar e as limitações que agem sobre essas funções é denominado de levantamento de requisitos. Nesse momento, os requisitos devem ser estruturados e detalhados, para que sejam apresentados na fase de elaboração para o desenvolvimento de outros componentes, como casos de uso, classe e interfaces. Essa etapa de levantamento gera como resultado o ‘documento de requisito’, que é considerado o principal elemento do anteprojeto do software (WAZLAWICK, 2011).

Requisitos de informações fazem parte do campo de estudos da Engenharia de Software. Conforme Pressman (2006), a engenharia de requisitos é uma ação da engenharia de software, que se inicia na atividade de comunicação e perdura durante a fase de modelagem.

Requisitos de um sistema são as descrições dos serviços que um sistema deve fornecer e as possíveis restrições às operações. Os requisitos representam as necessidades dos usuários de um sistema, como por exemplo fazer um pedido e encontrar uma informação. A engenharia de requisitos consiste em encontrar, analisar, colocar em documentos e observar o funcionamento do sistema e todas as restrições conhecidas. Para atingir esse resultado, tem-se algumas atividades para se cumprir como: o estudo de viabilidade; a elicitação e análise (encontro dos requisitos); especificação (padronização dos requisitos) e validação (verificação) (SOMMERVILLE, 2011).

Pressman (2006, p. 145) afirma que a análise de requisitos permite ao engenheiro de software:

[...] elaborar requisitos básicos do software estabelecidos durante tarefas anteriores de engenharia de requisitos e construir modelos que descrevam cenários de usuário, atividades funcionais, classes de problemas e seus relacionamentos, comportamento do sistema e das classes, e fluxo dos dados à medida que são transformados. A análise de requisitos fornece ao projetista de software uma representação da informação, função e comportamento, que podem ser traduzidas para os projetos arquitetural, de interfaces e em nível de componentes.

Ainda, sobre os modelos, o autor explica que o foco deve ser em ‘o quê’ o sistema deve executar e não ‘como’ fazer. A modelagem de análise possui quatro tipos de abordagens que visam a criação de diversos modelos que mostram os requisitos de software em relação à função, informação e conhecimento, conforme explica Pressman (2006):

- ✓ Modelo baseado em cenário: representam os requisitos de software de acordo com o ponto de vista do usuário. O caso de uso (descrição narrativa ou orientada por gabarito de uma interação entre ator e o software) é o principal componente da modelagem;
- ✓ Modelo de fluxo: focam no fluxo de objetos de dados na medida em que são transformados por funções de processamento;
- ✓ Modelo baseado em classe: usa informação que provém de elementos de modelagem baseada em cenário e orientada para fluxo para identificar as classes de análise;
- ✓ Modelagem comportamental: mostra o comportamento dinâmico, utilizando elementos baseados em cenário orientado a fluxo como entrada e baseados em classe para representar os estados das classes de análise e do sistema como um todo.

Os casos de uso, utilizados na modelagem por cenários, representam uma visão externa do sistema, sendo considerados ferramentas importantes para captar e entender os requisitos funcionais. Funcionam descrevendo as interações inerentes entre os usuários e o próprio sistema, utilizando uma narrativa de como o sistema interage e é utilizado, de forma a contribuir seu entendimento (FOWLER, 2005).

Requisitos Funcionais, de acordo com Wazlawick (2011), devem conter:

- ✓ descrição de uma função que o sistema irá executar;
- ✓ a origem do requisito (quem solicitou) e/ou quem irá executar a função;
- ✓ informações transmitidas do sistema para o usuário (e vice-versa) quando a função é executada;
- ✓ restrições lógicas ou tecnológicas impostas à função.

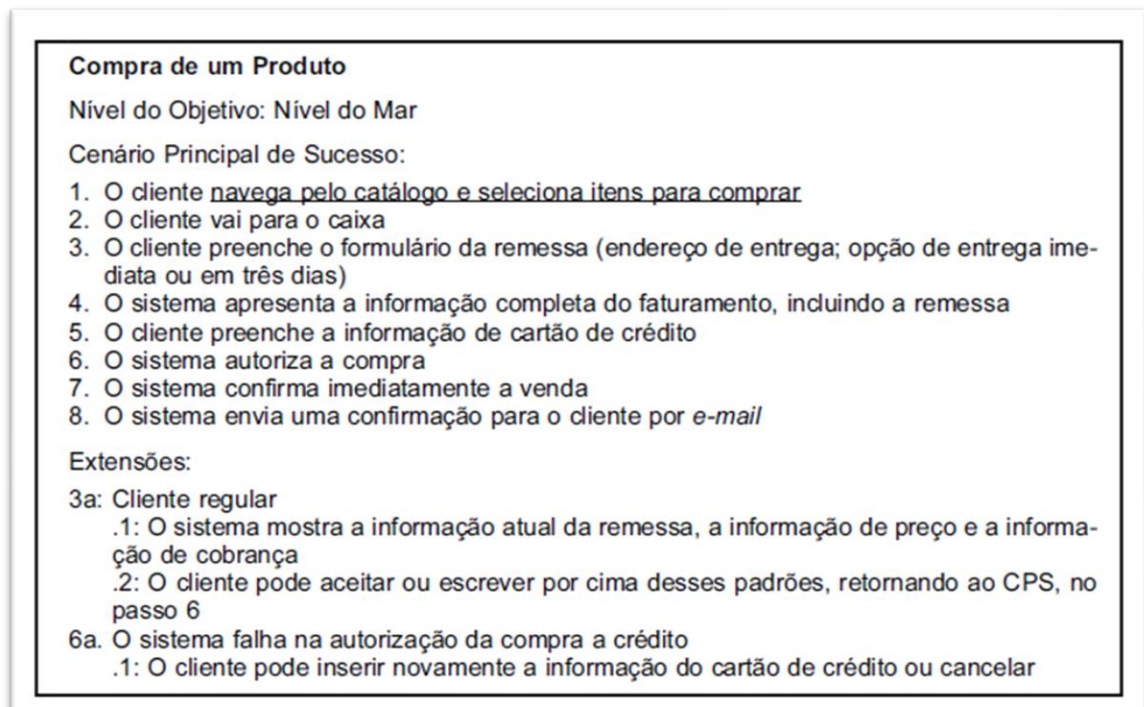
Conforme o autor, os requisitos chamados de não funcionais estão sempre conectados aos requisitos funcionais e podem ser classificados em lógicos (quando fazem parte das regras do negócio) ou tecnológicos (quando dizem respeito à tecnologia usada para a execução de uma função).

Ainda, para Wazlawick (2011, p. 26), “Os requisitos suplementares são todo tipo de restrição tecnológica ou lógica que se aplica ao sistema como um todo e não apenas a funções individuais.”

Por se tratar de um modelo (baseado em cenário) que facilita o entendimento das funcionalidades e interações entre sistema e usuário por meio dos casos de usos, esse modelo será utilizado no desenvolvimento da modelagem de um SR proposta nesta pesquisa.

Um cenário existe quando a sequência de acontecimentos sobre a interação entre o usuário e o sistema é descrita, podendo acontecer cenários diferentes para um mesmo processo que tenha um mesmo objetivo. Para Fowler (2005, p. 104) “Esse objetivo do usuário é a chave dos casos de uso: um caso de uso é um conjunto de cenários amarrados por um objetivo comum de usuário.”

Figura 25 - Exemplo de caso de uso



Fonte: Fowler (2005, p. 105).

Para Fowler (2005), “Cada passo em um caso de uso é um elemento da interação entre um ator e o sistema. Cada passo deve ser uma declaração simples e mostrar claramente quem está executando o passo.”

Conforme visto inicialmente nesta seção, após levantar os requisitos de um sistema utilizando a modelagem por cenário e empregando “casos de uso” como ferramenta, faz-se indispensável documentar o processo por meio um documento de requisitos.

2.4.2 Documento de requisitos

O documento de requisitos aponta todos os itens relativos ao que o sistema deve fazer e em que condições, podendo ser organizado em texto ou em diagrama, subdivididos em subsistemas quando possuírem muitas funções e, de preferência, que contenha um índice com nome da função ou requisito suplementar, além dos requisitos (funcionais e não-funcionais) detalhados no corpo do documento (WAZLAWICK, 2011). Abaixo, a Figura 26 mostra um exemplo de índice de um documento de requisito.

Figura 26 - Exemplo de índice em um documento de requisito

Sistema Livir – Documento de Requisitos	
Requisitos funcionais	
1.	Registrar novos títulos a partir do catálogo das editoras.
2.	Registrar vendas de livros.
3.	Realizar encomendas de livros.
4.	Registrar e autorizar pagamentos com cartão de crédito.
5.	Registrar e aplicar promoções.
6.	Calcular custos de entrega.
7.	Emitir relatório de livros mais vendidos.
8.	Emitir relatório de compradores mais assíduos.
9.	Emitir sugestões de compra para compradores baseadas em compras anteriores.
10.	...
Requisitos suplementares	
1.	O sistema deve operar via interface Web.
2.	Todos os controles de interface devem ter um campo de ajuda associado.
3.	...

Na Figura 26 é possível observar um documento de requisitos com o índice e os requisitos funcionais e suplementares conforme explicado anteriormente. Abaixo, os requisitos devem ser detalhados e mostrados no corpo do documento conforme o exemplo apresentado na Figura 27:

Figura 27 - Detalhamento de requisito

<p>1. Registrar novos títulos a partir do catálogo das editoras</p> <p>Descrição: O gerente seleciona as editoras para as quais pretende fazer a atualização. O processo é automático. O sistema consulta os ISBN disponibilizados e os compara com os existentes na base. Havendo novos ISBN, o sistema atualiza a base com as novas informações.</p> <p>Fontes: Sr. Fulano de Tal (gerente) e manual técnico da interface de catálogo das editoras.</p> <p>Usuário: O próprio gerente.</p> <p>Informações de entrada: O gerente informa quais são as editoras para as quais pretende fazer a atualização a partir de uma lista fornecida pelo sistema.</p> <p>Informações de saída:</p> <ul style="list-style-type: none"> - A lista de editoras (nome). - O relatório de atualizações efetuadas (uma lista contendo: nome da editora, ISBN, título e preço de compra). <p>Restrições lógicas: Não há.</p> <p>Restrições tecnológicas:</p> <ol style="list-style-type: none"> 1. Deve ser usado o sistema de interface com as editoras, de acordo com o manual XYZ.1234. 2. A seleção feita pelo gerente se dá através de uma lista de seleção múltipla, a qual deve ser ordenada de forma alfabética. 3. ...
--

Fonte: Wazlawick (2011, p. 28).

Requisitos detalhados conforme a Figura 27 facilitam o entendimento das funções que serão desempenhadas. Através do detalhamento é possível visualizar as informações que entram e saem do sistema.

Para Sommerville (2011), o documento de requisitos de software (às vezes chamado de “Especificação de requisitos de software”) é uma declaração oficial do que os desenvolvedores

do sistema devem implementar. Pode incluir os requisitos do usuário para um sistema e uma especificação detalhada dos requisitos do sistema. Às vezes, os requisitos do usuário e do sistema são integrados em uma única descrição.

Existe um padrão internacional que baliza a forma como os documentos de requisitos devem ser escritos. Essas são normas que detalham e padronizam os documentos, portanto serão utilizadas como base para esta pesquisa. Entre elas, a norma IEEE 830-1998 que se refere à “Práticas Recomendadas para Especificação de Requisitos de Software”, que trata de regras para escrita e documentação de requisitos de software.

Figura 28 - Documento de requisito IEEE 830-1998

Table of Contents
1. Introduction
1.1 Purpose
1.2 Scope
1.3 Definitions, acronyms, and abbreviations
1.4 References
1.5 Overview
2. Overall description
2.1 Product perspective
2.2 Product functions
2.3 User characteristics
2.4 Constraints
2.5 Assumptions and dependencies
3. Specific requirements (See 5.3.1 through 5.3.8 for explanations of possible specific requirements. See also Annex A for several different ways of organizing this section of the SRS.)
Appendixes
Index

Fonte: IEEE 830 (1998).

As partes do documento são descritas abaixo de acordo com Gonçalves *et al.* (2004).

1. Introdução: Demonstra uma visão geral do documento com base nas suas subseções.

1.1 Propósito: Traçar o propósito e o alvo do documento.

1.2 Âmbito: Identifica o software que será desenvolvido pelo seu nome; explica o que vai fazer e se precisar, o que não vai fazer; descreve a aplicação, os benefícios relevantes e objetivos.

1.3 Definições, Acrónimos e abreviaturas: Fornece definições dos termos usados, acrónimos e abreviaturas necessários para o bom entendimento do documento.

1.4 Referências: Fornece lista completa com outros documentos referenciados, identificando-os pelo título, número de relatório, data e organização de publicação, além de mostrar onde encontrar essas fontes.

1.5 Organização: descreve o conteúdo do documento e a sua organização.

2. Descrição Geral: Descreve os fatores gerais que afetam o produto e suas exigências por meio das seis subseções seguintes.

2.1 Perspectiva do Produto: Descreve os relacionamentos do produto em relação a outros produtos, como outros sistemas, hardware, software, usuário etc. Caso seja um produto independente e autocontido deve ser dito nesta seção.

2.2 Funções de Produto: Resumos das principais funções que serão desempenhadas pelo software, devem ser organizadas de forma compreensível podendo ser utilizados textos ou gráficos.

2.3 Características do Utilizador: deve descrever características gerais dos usuários alvo, incluindo nível de formação, experiência e proficiência técnica.

2.4 Restrições: Fornece uma descrição geral de itens que possam limitar o desenvolvimento, tais como regulamentos; hardware; interface com outras aplicações; operações em paralelo; funções de auditoria e de controle; exigências de alto nível de linguagem e fiabilidade; criticalidade da aplicação e considerações de segurança.

2.5 Assunções e Dependências: deve listar os fatores que afetam as exigências ditas no documento de requisito.

2.6 Divisão e Atribuição de Requisitos: Identifica as exigências que podem ser adiadas para futuras versões do sistema.

3. Requisitos Específicos: Deve conter todas as exigências (requisitos) de software detalhadas, para que permitam que seja feito o desenho de um sistema que satisfaça as exigências e que se façam testes que mostrem que o sistema satisfaz essas mesmas exigências.

3.1 Interfaces Externas: descrição detalhada das entradas e saídas do software, complementando as descrições da seção 2.1 sem repetir as informações.

3.2 Requisitos Funcionais: definem as ações fundamentais que devem ter lugar no software ao aceitar e processar as entradas e ao processar e gerar as saídas. Geralmente são listados utilizando frases na seguinte forma: “*O sistema deve...*”. Pode ser

apropriado dividir as exigências funcionais em subfunções ou subprocessos, não implicando que o desenho do software venha a ser dividido da mesma forma.

3.3 Requisitos de desempenho: deve especificar exigências numéricas estáticas e exigências numéricas dinâmicas impostas ao software ou à interação humana com o software como um todo. Exigências numéricas estáticas podem incluir alguns pontos: o número de terminais a suportar; o número de utilizadores simultâneos a suportar e as quantidades e tipos de informação a processar. Essas exigências são por vezes identificadas sob uma secção separada intitulada 'Capacidade'. As exigências numéricas dinâmicas podem incluir, por exemplo, o número de transações e tarefas e a quantidade de dados a processar num determinado período, em condições de carga normal e carga máxima. Todas estas exigências devem ser definidas em termos quantificáveis. Por exemplo: "95% das transações devem ser processadas em menos de 1 segundo" em vez de "O operador não deve ter de esperar que a transação complete".

3.4 Requisitos lógicos da base de dados: devem ser especificados os requisitos lógicos sobre qualquer informação que deva ser colocada numa base de dados, podendo ser incluídos: tipos de informação usados pelas várias funções; frequência de uso; capacidades de acesso; entidades e relações; restrições de integridade e exigências de retenção dos dados.

3.5 Restrições de desenho: Aqui devem ser especificadas restrições ao desenho que possam ser impostas por outras normas, limitações de hardware etc.

3.6 Atributos do sistema de software: Existe um número de atributos do software que podem servir de exigências. É importante que os atributos requeridos sejam especificados de modo a que o seu cumprimento possa ser objetivamente verificado. Uma lista parcial de exemplos é fornecida: fiabilidade, disponibilidade, segurança, capacidade de manutenção e portabilidade.

3.7 Organização dos requisitos específicos: os requisitos detalhados tendem a ser extensivos quando se trata de um sistema não trivial. Por essa razão recomenda-se uma consideração cuidadosa por forma a organizar os requisitos de uma forma simples de compreender. Não existe uma organização modelo que seja perfeita para todos os sistemas. Diferentes classes de sistemas inclinam-se para diferentes organizações de requisitos na seção 3 de um documento. Por exemplo, o sistema pode se organizar por modo; por classe do utilizador; por objetos; por características, por estímulos ou por resposta. Porém, quando nenhum dos esquemas organizacionais anteriores provam ser

úteis, a funcionalidade total pode ser organizada numa hierarquia de funções (hierarquia funcional) organizadas por entradas comuns, saídas comuns ou dados internos comuns.

3.8 Comentários adicionais: sempre que um documento de requisitos é contemplado, várias das técnicas organizacionais dadas na secção anterior podem ser usadas. Quaisquer exigências adicionais podem ser postas numa secção separada no fim do documento de requisitos;

4. Apêndices: nem sempre são considerados parte do documento de requisitos propriamente dito e não são sempre necessários. Podem incluir: uma amostra de formatos de entrada/saída; descrições de estudos de análise de custos ou resultados de questionários; informação de suporte ou de fundo que possa ajudar os leitores do documento de requisitos; uma descrição dos problemas a serem resolvidos pelo software e instruções de empacotamento especiais para o código e para os suportes de distribuição, por forma a satisfazer exigências de segurança, exportação, carregamento inicial ou outras. Quando apêndices são incluídos, o documento de requisitos deve explicitamente referir se esses devem ser considerados parte integrante das exigências.

A IEE 830-1998 possui uma estrutura que serve como modelo para quando o método de levantamento de requisitos para modelagem de um sistema seja “casos de uso”, conforme mostra a Figura 29:

Figura 29 - Estrutura para “casos de uso” conforme IEEE 830-1998

```

3. Specific requirements
3.1 ...
3.2 Functions Requirements
    3.2.1 Caso de uso 1
        3.2.1.1 Requisito funcional 1.1
        .
        .
        3.2.1.n Requisito funcional 1.n
    3.2.2 Caso de uso 2
        .
        .
    3.2.m Caso de uso m
        3.2.m.1 Requisito funcional m
        .
        .
        3.2.m.n Requisito funcional m.n
    .
    .
3.8 Additional comments

```

Fonte: IEEE 830 (1998).

2.5 PERIÓDICOS CIENTÍFICOS COMO FONTE DE INFORMAÇÃO E CONHECIMENTO

Durante o século XVII, na Europa, ocorreu o surgimento dos primeiros periódicos científicos, primeiramente na França e logo em seguida na Inglaterra. Nesse momento acontecia uma grande mudança no meio científico, ocorria que a dedução não mais seria aceita como principal método de pesquisa, a comunidade científica exigia e procurava por evidências empíricas e baseadas na observação para que os resultados dos trabalhos fossem considerados como científicos. Esses eventos, que caracterizaram o início da ciência moderna, modificaram também a forma da comunicação científica que antes era feita pessoalmente ou por cartas. Com o aparecimento do periódico científico, a comunicação tornou-se rápida, ampla e precisa, permitindo troca de ideias e críticas entre os interessados (MUELLER, 2000).

Um dos primeiros periódicos que surgiu na Inglaterra, *Philosophical Transactions*, fundado por um grupo de filósofos ligados à *Royal Society*, continua sendo publicado até os dias de hoje e, segundo Mueller (2000), esta instituição inglesa considera que são quatro as funções atuais do periódico científico:

- Comunicação formal dos resultados da pesquisa original para a comunidade científica e interessados;
- Preservação do conhecimento registrado;
- Estabelecimento da propriedade intelectual;
- Manutenção do padrão da qualidade na ciência.

Apesar dos periódicos existirem há muitos anos e estarem em constante evolução ao longo do tempo, a função primordial de divulgar os resultados de pesquisas formalmente para a comunidade científica mantém-se inalterada até hoje.

Para que um trabalho científico venha a contribuir para um aumento de conhecimento em qualquer área, ele precisa ser divulgado e difundido. Biojone (2013) atesta que os periódicos seguramente funcionam como a principal ferramenta de propagação científica, e que, quando a publicação do resultado de uma pesquisa por esse meio é lida por seus pares, torna-se legítima. Dessa forma ela poderá ser vista pelos demais pesquisadores, criticada e revisada. Os periódicos são veículos que realizam esse papel de comunicação dentro da comunidade científica, recebendo os trabalhos originais e os encaminhando para avaliadores, para que procedam a revisão do original às cegas e aos pares, garantindo que os artigos, uma vez publicados, atendam aos padrões adequados de qualidade estabelecidos pelo corpo editorial do periódico.

A comunicação científica compreende as atividades relacionadas à produção, disseminação e uso da informação. O processo inicia-se quando a oportunidade de pesquisa é identificada pelo cientista, até que a comunicação dos resultados seja aceita pela comunidade de pesquisadores como parte do conhecimento científico, sendo o periódico o principal meio de comunicação científica (GARVEY, 1979).

Não obstante possuir inúmeros atributos positivos e ser empregado de forma abrangente pela comunidade acadêmica, Mueller (2000, p. 76) afirma que o modelo tradicional de periódico científico apresenta problemas que se acentuam com o rápido desenvolvimento da tecnologia. Dentre alguns, são considerados como principais: a demora na publicação do artigo, que pode levar mais de um ano após o recebimento pelo editor; o alto custo para adquirir e manter as coleções atualizadas; a pouca versatilidade do formato impresso em papel comparado com o formato eletrônico; as dificuldades para o pesquisador perante a grande quantidade de periódicos e os poucos mecanismos de identificação e busca, assim como, a demora para obtenção de cópias de artigos de seu interesse.

O alto custo para manter as coleções atualizadas, somado com o crescente número de títulos assinados, fez com que bibliotecas tradicionais ao redor do mundo na década de 80 cancelassem ou diminuíssem a quantidade de assinaturas de coleções. Logo essa crise chegou ao Brasil no início dos anos 90, intensificada pela conjuntura econômica e política, fazendo com que bibliotecas optassem por certas áreas de ensino em detrimento a outras. Em contrapartida esse processo despertou a busca por alternativas para suprir a crise (MUELLER, 2000).

As modificações que ocorreram e continuam a ocorrer no jeito de se comunicar e a crescente demanda por informação e inovação, seja na forma científica ou não, aceleraram a transição do fluxo de informação para o meio eletrônico, fazendo surgir os periódicos científicos eletrônicos em grande número. A internet possibilita o acesso rápido à informação, além de facilitar a busca e atualização da informação, tornando-se uma importante ferramenta tecnológica na comunicação. Posto isso, foi natural a expansão desta forma eletrônica conforme Veloso e Nascimento (2012, p. 112):

Os periódicos científicos eletrônicos surgiram no início da década de 90, com o advento das tecnologias da informação e da necessidade de divulgar e disseminar o conhecimento científico através de um meio de comunicação, onde o acesso à informação fosse mais rápido e eficiente, ou seja, onde os recursos tecnológicos, em especial as redes de computadores, focalizando a Internet beneficiando uma nova forma de produção eletrônica; consequentemente, possibilitando assim alternativas variadas em termos de estrutura textual e de imagens, além de uma distribuição vasta e ágil para um grande público consumidor.

É no periódico eletrônico que ocorre o início do processo, desde a submissão e editoração, até a distribuição dos artigos científicos, totalmente em meio digital. Oferece, além do texto completo, acesso integral à coleção de artigos, busca em metadados, links, referências cruzadas, estatísticas de citação, acesso, entre outras possibilidades (BIOJONE, 2003).

O Portal Capes de Periódicos é considerado uma biblioteca virtual e oferece acesso, através de instituições participantes, a uma vasta quantidade de documentos científicos completos das mais variadas áreas do conhecimento, nacionais e internacionais. O próprio sítio eletrônico do Portal CAPES explica sobre a sua criação:

Foi criado para reunir material científicos de alta qualidade e disponibilizá-lo à comunidade acadêmica brasileira. Assim, o Portal de Periódicos tem o objetivo de reduzir as assimetrias regionais no acesso à informação científica, cobrindo todo o território nacional. É considerado uma iniciativa única no mundo, pois um grande número de instituições acessa o acervo que é inteiramente financiado pelo Governo federal (CAPES, 2020, não paginado).

As bases de dados que são disponibilizadas dentro do Portal Capes, onde são oferecidos os treinamentos, oferecem conteúdo pago e de AA.

2.5.1 Periódicos científicos e o movimento de Acesso Aberto

Entende-se por AA o acesso gratuito por qualquer pessoa aos resultados de pesquisas científicas, baseado no argumento de que, por ser considerado um bem público, o conhecimento científico deve estar disponível e ao alcance de todos, sendo uma alternativa ao modelo considerado tradicional em que o acesso é cobrado (FAPESP, 2019).

Entre as inúmeras declarações e manifestos de apoio ao Acesso Aberto, em 2002, a *Budapest Open Access Initiative* (BOAI) foi a primeira a articular uma definição pública para o termo AA, além de propor estratégias para o seu atingimento. Essa iniciativa resultou em um documento conhecido como Declaração de Budapest, nele encontra-se a definição sobre AA:

Acesso Aberto” à literatura científica e revisada por pares significa a disponibilidade livre na Internet, permitindo a qualquer usuário ler, fazer *download*, copiar, distribuir, imprimir, pesquisar ou referenciar o texto integral desses artigos, recolhê-los para indexação, introduzi-los como dados em softwares, ou usá-los para outro qualquer fim legal, sem barreiras financeiras, legais ou técnicas que não sejam inseparáveis ao próprio acesso a uma conexão à Internet. As únicas restrições de reprodução ou distribuição e o único papel para o direito autoral neste domínio é dar aos autores o controle sobre a integridade do seu trabalho e o direito de ser devidamente reconhecido e citado (BOAI, 2002, não paginado).

Rios, Lucas e Amorim (2019) acrescentam que a Declaração de Budapeste trouxe na sua composição ideias que possibilitaram desafiar o monopólio de editoras sobre publicações científicas, citando o auto arquivamento (via verde) e a publicação em periódicos de AA (via dourada) como ações para superar essa mediação das editoras comerciais.

Duas estratégias para atingir o Acesso Aberto foram propostas na BOAI, são conhecidas como Via Verde e Via Dourada. A Via Verde pode ser chamada de auto arquivamento, pois os artigos são arquivados pelo próprio autor em repositórios institucionais e próprios e a Via Dourada conhecida por utilizar periódicos financiados por instituições que oferecem material de forma gratuita, como exemplo a SciELO (*Scientific Electronic Library Online*).

Sucessivos manifestos encorparam o apelo em direção ao movimento do AA e cada um deles incrementou de alguma forma esse crescimento. O Quadro 5 apresenta as principais ênfases sobre importantes manifestos ocorridos, de acordo com Rios, Lucas e Amorim (2019):

Quadro 5 - Manifestos sobre o movimento do AA

Manifesto	Destaques
Convenção de Santa Fé (1999)	- Marcou o lançamento do <i>Open Archives Initiative</i> (OAI); - Arquivos <i>e-prints</i> ; - Proposta de novo formato para avaliação de artigos científicos através do auto arquivamento.
Declaração de Budapeste (2002)	- Utiliza o termo Open Access pela primeira vez para definir o acesso livre à informação; - Apresenta o conceito de AA; - Traz duas estratégias para alcançar o AA (auto arquivamento e revista de AA); - Oferece apoio às instituições que possuem interesse em aderir ao movimento AA.
Declaração de Bethesda (2003)	- Apresenta recomendações para os envolvidos com a informação científica: Instituições e agências de financiamento, bibliotecários e editores e sociedade científica.
Declaração de Berlim (2003)	- Promover a Internet como o instrumento funcional ao serviço global e do pensamento humano; - Busca estabelecer o paradigma do AA.
Declaração de Haia (2014)	- Aborda temas como mineração de dados e <i>Big Data</i> ; - Estabelece estratégias práticas para a informação alcançar características de AA; - Recomenda o uso de Orcid, XML e Creative Commons.

Fonte: Adaptado de Rios, Lucas e Amorim (2019, p. 166).

Percebe-se que a cada evento os destaques vão se aprimorando de forma subsequente, ampliando o escopo em busca do movimento de AA. A evolução do aparato tecnológico contribuiu de forma significativa favorecendo as pautas abordadas.

Dentro desse contexto evolutivo do AA, Prado (2019, p. 66) considera que:

[...] o Acesso Aberto, mais que uma espécie de modelo de negócio, também pode ser compreendido como um princípio ideológico que perpassa desde a produção até a disseminação das pesquisas científicas. Valoriza e incentiva que a informação de qualidade seja um bem público que retroalimente novos estudos para o bem comum.

Como benefícios do AA, segundo a Fapesp (2019), pode-se perceber o incremento na visibilidade dos resultados de pesquisas, que possibilita um maior alcance de um público que não teria acesso, potencializa a capacidade de colaboração internacional das atividades de pesquisa e, por último, aumenta o número de citações. Artigos científicos publicados dessa forma recebem, em média, 8% mais citações em relação aos outros e esse número percentual é maior em países como Brasil, Argentina e Rússia, chegando a 25%.

Os repositórios institucionais de Acesso Aberto, coletam e armazenam os trabalhos, dando acesso a toda produção científica realizada nas universidades. Esses repositórios surgem como forma de minimizar a falta de visibilidade da produção científica das instituições. O portal SciELO, considerado um bom exemplo de AA no Brasil, funciona de forma aberta e universal, para indexação e publicação na Web (FAPESP, 2019).

Um dos principais objetivos dos repositórios é disseminar resultados de pesquisa de Acesso Aberto em todo o mundo. Consequentemente, é importante que esses resultados sejam bem acessíveis, ou seja, que estejam conectados de forma satisfatória para que possa oportunizar cada vez mais interações com eles (KNOTH *et al.*, 2017).

Sobre a SciELO, Packer, (2014, p. 15) afirma que:

A criação da SciELO há 15 anos e seu desenvolvimento posterior foram guiados por duas abordagens pioneiras e inovadoras: em primeiro lugar, a indexação de periódicos nacionais de qualidade para complementar os índices internacionais e a publicação de textos completos de Acesso Aberto, na Internet, na modalidade conhecida atualmente como “Via Dourada” (*Golden Road*), surgida por volta de quatro anos antes do lançamento da Declaração de Budapeste, que é reconhecida internacionalmente como o início do movimento de AA; em segundo lugar, a convergência cooperativa de editoras independentes, editores e agências nacionais de pesquisa acerca do objetivo comum de aumentar a qualidade e visibilidade dos periódicos.

Com o crescimento dos recursos digitais e por consequência a maior facilidade para editar e divulgar periódicos, as publicações em AA surgiram como alternativas na área científica. Segundo Laakso *et al.* (2011), o AA, no contexto de publicações acadêmicas, é amplamente utilizado para se referir ao acesso online irrestrito a artigos publicados em periódicos acadêmicos.

De acordo com Rodriguez e Stubert (2015), a ampliação do conceito de AA e das ferramentas editoriais em código aberto viabilizam a criação de títulos de qualidade por editores independentes, que se organizam na busca de credibilidade e visibilidade necessárias para o reconhecimento das revistas.

Entre os recursos e ferramentas existentes que facilitam o alcance em direção ao conceito AA, o Open Journal Systems (OJS) se destaca por oferecer inúmeras funcionalidades. O OJS, recomendado pela CAPES, abarca ações necessárias para a automação de periódicos científicos e o seu processo editorial possibilita aprimoramento na avaliação da qualidade dos periódicos e maior rapidez no fluxo das informações, além de permitir fácil adaptação aos processos de editoração em uso, sendo bem aceito pela comunidade brasileira de editores. Ademais, possibilita que a propagação, a divulgação e a preservação dos conteúdos das revistas brasileiras apresentem um avanço na adoção de padrões internacionais para periódicos on-line 100% eletrônicos.

OJS é um software de código aberto que serve para gerenciar e publicar periódicos acadêmicos, desenvolvido pela PKP⁴ em 2001 para aprimorar o acesso à pesquisa, é a plataforma de publicação de periódico de código aberto mais usada, com mais de 25.000 periódicos utilizados em todo no mundo.

A contar do início das manifestações que surgiram e vão ao encontro do movimento do AA, o número de publicações ou objetos que são resultados das pesquisas científicas divulgados de forma livre ou aberta tem aumentado e com isso ferramentas têm surgido para contabilizar esse aumento. (APPEL, 2019).

2.5.2 A interoperabilidade e os periódicos científicos de Acesso Aberto – o protocolo OAI-PMH

Conforme visto, bases de dados, repositórios e periódicos científicos constituem uma fonte importante de informação para pesquisadores. O grande número de repositórios existentes hoje, seja em universidades, temáticos ou de outro tipo, disponibilizam uma grande quantidade de informação, entretanto, é necessário que essas bases possuam uma capacidade de trocar informações entre si, no sentido de facilitar a pesquisa e o acesso mais preciso, unificando a informação.

⁴ <https://pkp.sfu.ca/>

Entre as ferramentas que proporcionem essa interoperabilidade entre as bases, o chamado protocolo OAI-PMH (*Open Archives Initiative Protocol for Metadata Harvesting*) possibilita disponibilizar os dados dos participantes da iniciativa para que possam ser coletados por outros participantes.

Devido à inexistência de padrões na disponibilização e pesquisa de informações científicas na internet, Garcia e Sunye (2003) explicam que foi criada a Iniciativa *Open Archives* (OAI), e assim, dessa iniciativa originou-se o protocolo OAI-PMH, com o propósito de unificar as consultas as bases de dados existentes, de modo a torná-las mais eficientes. O protocolo OAI-PMH vem se confirmando como base para a interoperabilidade entre repositórios digitais acadêmicos e científicos no mundo todo e através desse protocolo é possível promover visibilidade e integração de informações.

A OAI possui como missão desenvolver e promover padrões de interoperabilidade que buscam facilitar a disseminação eficiente do conteúdo, trazendo nas suas raízes um esforço para melhorar o acesso a arquivos de impressão eletrônica como um meio de aumentar a disponibilidade da comunicação acadêmica (OPEN ARCHIVES, [201-?a]).

No Brasil, o Instituto Brasileiro de Informação em Ciências e Tecnologia – IBICT, apoia a implementação do protocolo OAI-PMH pelas instituições brasileiras. A Biblioteca Digital Brasileira de Teses (BDTD), desenvolvida pelo IBICT, é um provedor de serviço que segue os preceitos da OAI e visa integrar em um único portal as teses e dissertações existente nas Instituições de Ensino Superior (IES) do país, a fim de disponibilizar um catálogo nacional com esses documentos, fazendo com que o acesso a este catálogo seja feito de forma única. A BDTD emprega as tecnologias da OAI e utiliza o modelo baseado em padrões de interoperabilidade, consolidado em uma rede distribuída de bibliotecas digitais de teses e dissertações (IBICT, 2018).

O protocolo OAI-PMH foi um dos que difundiu o formato de provedores de dados e de serviços. Nesse formato, ao utilizarem o protocolo, os integrantes se cadastram escolhendo serem um sistema provedor de dados, provedor de serviços ou os dois. Dessa forma, provedores de dados funcionam como repositórios de documentos digitais que exibem seus metadados, já os provedores de serviços atuam na função de buscar os metadados expostos pelos provedores de dados (GARCIA; SUNYE, 2003).

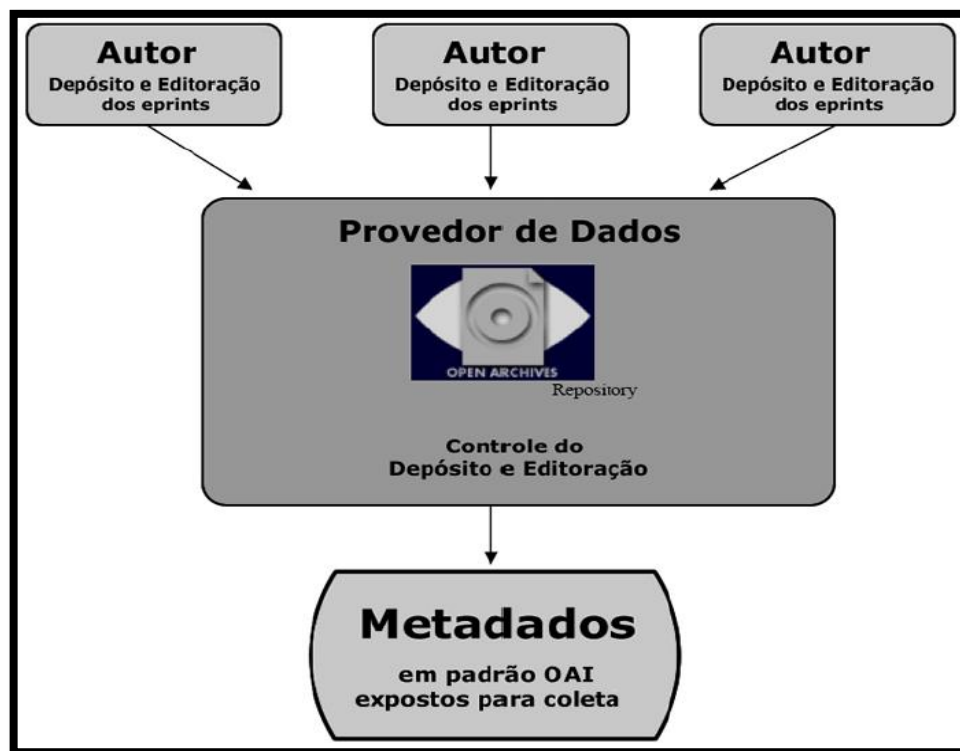
Repositórios podem ser compreendidos, em geral, tal qual plataformas de armazenamento de dados e recursos informacionais. Provedores de dados são exemplos de repositórios que funcionam como depósitos para armazenar objetos digitais, assim como os metadados que os descrevem de forma individual (CUNHA 2021).

Conforme Open Archives [201?b], provedores de dados que suportam o protocolo OAI-PMH tem a possibilidade de listar seu repositório no registro OAI-PMH, com os seguintes objetivos:

- aumentar o acesso e a exposição ao seu trabalho, facilitando o descobrimento do repositório e quais metadados podem ser coletados;
- garantir sua conformidade com a especificação OAI-PMH;
- utilizar o protocolo para compartilhar conhecimento e fornecer informações para a OAI monitorar o uso do protocolo no sentido de planejar melhorias futuras.

De acordo com Carvalho e Oliveira (2009, p. 5) “Os provedores de Dados são responsáveis por gerenciar os objetos, expondo seus metadados para a coleta. Eles são os criadores e mantenedores dos metadados e dos recursos de seus repositórios.” A Figura abaixo ilustra um exemplo do fluxo de informação nos provedores de dados.

Figura 30 - Fluxo de informação nos provedores de dados



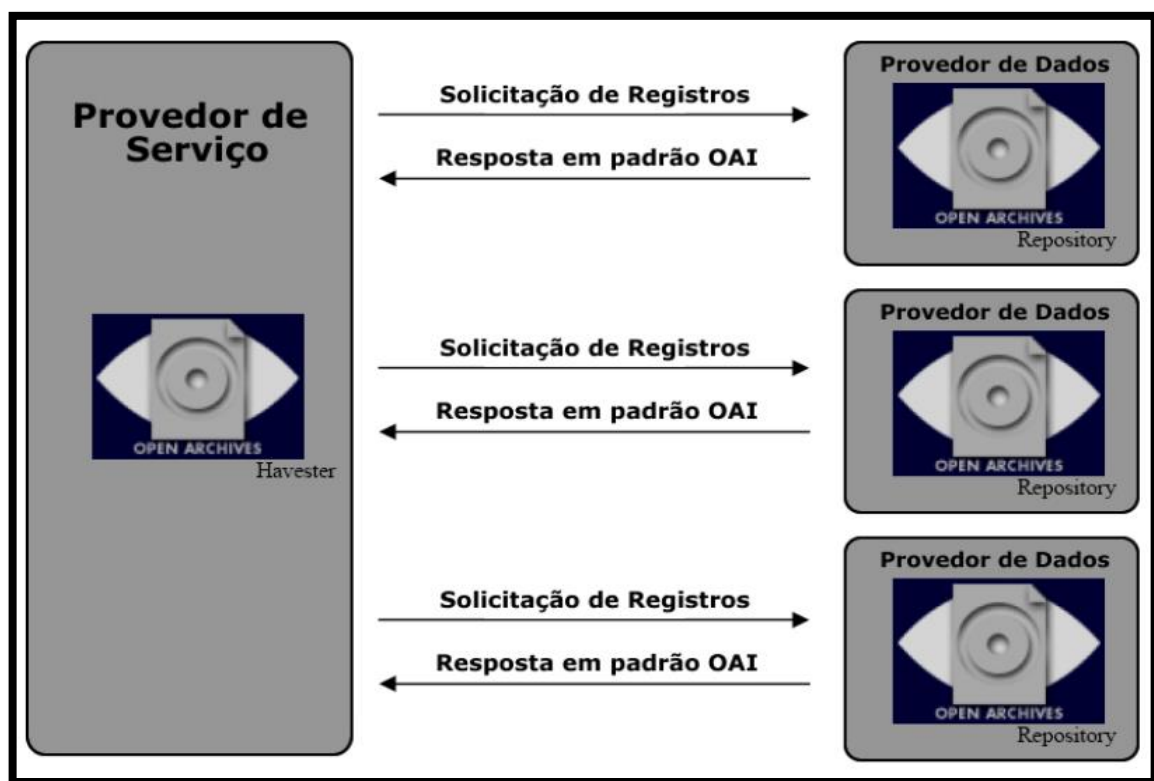
Fonte: Carvalho e Oliveira (2009, p. 6).

A Figura 30 mostra o funcionamento de um fluxo de informações nos provedores de dados, no qual os autores gerenciam seus objetos enquanto os provedores de dados controlam os recursos e exibem os metadados para coleta.

Em relação ao provedor de serviço, Cunha (2021, p. 36, grifo nosso) esclarece que “[...] refere-se aos repositórios que só armazenam metadados dos objetos. Esses objetos **continuam pertencendo aos repositórios originais.**”

A BDTD é um exemplo de provedor de serviço, no qual é possível realizar buscas em um lugar único, retornando como resultados documentos (teses e dissertações) diretamente do seu repositório de origem para o acesso, a impressão ou o *download* de quem realiza a busca. A Figura 31, mostra o fluxo de informação no provedor de serviço:

Figura 31 - Fluxo de informação em um Provedor de Serviço



Fonte: Carvalho e Oliveira (2009, p. 7).

Na Figura 31 é possível visualizar de forma simples o acesso que os Provedores de Serviço têm sobre os metadados dos Provedores de Dados, utilizando sempre o padrão OAI na troca de informações.

A essência da interoperabilidade do protocolo OAI-PMH é transportar e compartilhar os metadados entre os provedores, assim, são usadas três camadas de padrões tecnológicos pré-existentes, que servem como base para a construção do protocolo. As camadas são descritas a seguir:

- Camada de Semântica da Descrição de Dados – foi dada preferência ao padrão *DublinCore* por ser um padrão simples de descrição de metadados, facilitando sua

adoção. Outros padrões de metadados também são aceitos, porém o *DublinCore* é exigido para manter um nível básico de interoperabilidade;

- Camada de Linguagem da Descrição dos Dados – para padronizar a descrição dos metadados, foi definido o padrão XML;
- Camada de Transporte dos Dados – baseado na arquitetura cliente-servidor, foi escolhido o padrão HTTP que serve de base para a Web, para o transporte dos dados entre os provedores de dados e serviços (MARTINS; FERREIRA, 2012).

A seguir, a Figura 32 exibe o funcionamento básico do protocolo OAI_PMH, com as camadas utilizadas como padrões.

Figura 32 - Funcionamento do protocolo OAI-PMH



Fonte: Carvalho e Oliveira (2009, p. 8).

Na Figura 32, o Provedor de Serviço solicita os metadados utilizando a camada de transporte de dados definida como padrão (HTTP) e recebe a resposta codificada em XML, a camada utilizada como padrão para descrever os metadados.

Consoante com o que foi observado neste tópico, acerca do protocolo OAI-PMH, Garcia e Sunye (2003, p. 11) asseguram que “Isto [OAI-PMH] permite que informações de qualidade se tornem disponíveis através de consultas gratuitas através dos provedores de serviço, sem necessidade de pagamento de taxas ou assinatura para acessá-las.”

3 PROCEDIMENTOS METODOLÓGICOS

De acordo com Prodanov e Freitas (2013), a metodologia consiste em aplicar técnicas que devem ser observadas na construção do conhecimento, com o intuito de comprovar a sua validade.

Metodologia pode ser visto como o caminho para se chegar a um determinado fim, e método científico como o conjunto de procedimentos intelectuais e técnicos adotados para se atingir o conhecimento (GIL, 2008).

Para tanto, é necessária a delimitação do tema que, conforme Marconi e Lakatos (2013), é dotada de um sujeito e um objeto, sendo dado como concluído quando é realizada a limitação geográfica e espacial. Assim, o tema proposto versará sobre a recomendação de artigo científicos que sejam exclusivamente de Acesso Aberto. Este trabalho tem como base os passos para modelar um SR proposto.

Em relação aos seus objetivos, a metodologia utilizada se caracterizou como exploratória que, conforme o entendimento de Prodanov e Freitas (2013, p. 51):

[...] tem como finalidade proporcionar mais informações sobre o assunto que vamos investigar, possibilitando sua definição e seu delineamento, isto é, facilitar a delimitação do tema da pesquisa; orientar a fixação dos objetivos e a formulação das hipóteses ou descobrir um novo tipo de enfoque para o assunto. Assume, em geral, as formas de pesquisas bibliográficas e estudos de caso.

Nesta pesquisa, o método de raciocínio utilizado é o indutivo, que parte de algo particular para uma generalização, considerado o oposto do dedutivo. Ambos são métodos que lidam com os ‘procedimentos lógicos’ que devem ser seguidos na pesquisa científica referente aos fatos da natureza e a sociedade (GIL, 2008).

Utilizou-se da abordagem qualitativa para o atendimento dos objetivos propostos. Esse procedimento, segundo Richardson (1999), busca do que é comum e, ao mesmo tempo, do que é específico em um determinado assunto e seus significados múltiplos, além de procurar conhecer um acontecimento ou fenômeno individual em profundidade através de descrições, de comparações e de interpretações.

De acordo como objetivo principal desta pesquisa, de modelar os requisitos de um Sistema de Recomendação de artigos científicos de periódicos de Acesso Aberto, caracteriza-se, do ponto de vista da sua natureza, como aplicada, que utiliza a pesquisa básica somada com tecnologias existentes e possui como objetivo gerar algum benefício prático para solução de um problema concreto. De acordo com Prodanov e Freitas (2013), a pesquisa aplicada tem como

objetivo solucionar problemas específicos com a utilização de uma aplicação prática, envolvendo verdade e interesses locais.

Os dados utilizados são de fontes secundárias, por meio de uma investigação bibliográfica. Segundo Marconi e Lakatos (2013, p. 43-44), esse procedimento “trata-se de levantamento da bibliografia já publicada, em forma de livros, revistas, publicações avulsas e imprensa escrita”. Assim, pretende-se utilizar materiais já elaborados, de preferência revisado pelos pares e atualizados, tais como livros e artigos científicos, publicações em periódicos, dissertações, no intuito de estar em contato com o material sobre o assunto pesquisado.

Conforme Prodanov e Freitas (2013), a pesquisa bibliográfica possui etapas consideradas imprescindíveis para sua consecução, são elas:

- Escolha do tema;
- Levantamento bibliográfico preliminar;
- Formulação do problema;
- Elaboração do plano do assunto;
- Buscas das fontes;
- Leitura do material;
- Fichamento;
- Organização lógica do assunto;
- Redação do texto.

O Quadro a seguir sintetiza os procedimentos metodológicos utilizados na pesquisa:

Quadro 6 - Procedimentos metodológicos utilizados

ASPECTO	ESPECIFICAÇÃO
Delimitação do tema	Artigos de Acesso Aberto
Método de Raciocínio	Indutivo
Natureza da pesquisa	Aplicada
Caracterização da pesquisa	Abordagem qualitativa
Delineamento da pesquisa	Quanto aos objetivos: exploratória
	Quanto aos meios: Investigação bibliográfica e Estudo de Caso
Técnicas e instrumentos de coletas de dados	Dados secundários (Investigação bibliográfica)
Técnicas e instrumentos de Análise de dados	Análise interpretativa

Fonte: autor (2021).

A pesquisa bibliográfica neste trabalho manteve seu foco nos seguintes tópicos: sistemas de recomendação, artigos e periódicos científicos e o movimento do Acesso Aberto. Em relação ao primeiro tópico, inicialmente verificou-se os conceitos, finalidades e abordagens utilizadas nas recomendações, a partir de uma reflexão teórica. Em um segundo momento, foram mapeados alguns SR de artigos científicos existentes, assim como plataformas acadêmicas que realizam recomendações de artigos científicos. Analisou-se suas particularidades e, quando possível, o seu funcionamento na prática, com o propósito de se aproximar dos conceitos vistos e conhecer o funcionamento dos sistemas de recomendação. No que diz respeito ao segundo tópico, buscou-se primeiramente, apresentar o surgimento do artigo e do periódico científico, a sua relação com os pesquisadores e a Ciência, a confiabilidade das informações neles contida, em seguida foi analisada a transição dos periódicos para a forma eletrônica, que oportuniza o acesso de mais pessoas, sendo a forma utilizada no SR proposto. Acerca do terceiro tópico, o movimento do AA, procurou-se conceituar o movimento e apresentar os seus objetivos, para em seguida mostrar os principais manifestos do movimento com seus destaques e crescimento ao longo do tempo. A associação dos periódicos eletrônicos com o movimento do AA trouxe para o trabalho a perspectiva da informação científica alcançável, disponível e sem custo para o pesquisador e usuário.

Assim, o levantamento bibliográfico realizado possibilitou aliar uma estratégia de correlação entre os tópicos pesquisados para a criação do modelo de SR.

4 RESULTADOS DA PESQUISA

4.1 CONSIDERAÇÕES SOBRE ESTRATÉGIAS DE FILTRAGENS PARA UM SR DE ARTIGOS CIENTÍFICOS

Por possuírem um grande conjunto de técnicas e abordagens, diferentes estratégias de recomendações podem ser utilizadas em um SR, variando de acordo com o seu objetivo.

As grandes empresas que atuam no comércio eletrônico disponibilizam em seus sites uma considerável quantidade de itens para venda, atraindo uma enorme base de clientes e buscas por produtos. Por possuírem bastantes clientes, vendas e buscas, a tendência é que os clientes realizem mais avaliações dos itens consumidos, seja por meio de notas aos produtos e comentários, melhorando a reputação do item e facilitando a recomendação de forma colaborativa.

Nesses casos de SR comerciais, o uso da FC torna-se mais viável para implementação, pois, conforme funciona a abordagem colaborativa, os usuários são divididos em grupos de acordo com características semelhantes, e as recomendações que servem para um membro de um grupo, acabam por servirem para todos os outros membros que fazem parte do grupo.

SR de artigos científicos possuem peculiaridades que os diferem dos SR da área comercial. A utilização da FBC é a mais usada nestes sistemas, conforme motivos vistos na subseção 2.1.3.2. Este tipo de filtragem, por ter raízes na área de recuperação de informação, combina bem quando utilizada com arquivos que contém textos.

Por ser um sistema novo e não possuir usuários cadastrados e, conseqüentemente, não possuir dados ou avaliações de outros usuários, o SR proposto neste trabalho vai servir-se da FBC de forma inicial para suas recomendações, entretanto, na medida em que os usuários cadastrados começarem a avaliar os artigos científicos, o sistema já estará preparado para empregar recomendações utilizando a FC. Nesse caso, quanto mais usuários cadastrados e quanto mais avaliações, maior será a acurácia do sistema com a FC.

Tendo em vista que a utilização dos dois tipos de filtragens procura minimizar os pontos fracos de cada um e potencializar os seus pontos fortes, o sistema proposto vai empregar a FH, utilizando ambas as filtragens (FBC+FC) operando concomitantemente, de forma a maximizar a adequação das recomendações com o perfil do usuário, claro que evitando que um mesmo artigo seja recomendado mais de uma vez para o mesmo usuário.

4.1.1 Análise comparativa dos SR encontrados na pesquisa

Com base nas plataformas que foram explanadas nesta pesquisa, sejam as que funcionam somente como SR, sejam protótipos de SR ou as chamadas redes sociais acadêmicas que empregam ferramentas de recomendação, fez-se um quadro comparativo onde são destacados os atributos apresentados por cada uma delas:

Quadro 7 - Quadro Comparativo de SR

	Técnica de filtragem	Tipo de coleta de dados	Perfil	Notificação	Tipo de Documento
Google Acadêmico	Híbrida (FBC+FC)	Explícita e implícita	Formulário + buscas	<i>Email</i>	PDF
Academia.edu	Híbrida (FBC+FC)	Explícita e implícita	Formulário + buscas	<i>Email e plataforma</i>	PDF
Researchgate	Híbrida (FBC+FC)	Explícita e implícita	Formulário + buscas	<i>Email e plataforma</i>	PDF
W-RECMAS	Híbrida (FBC+FC)	Explícita e implícita	Formulário + buscas	Não informa	PDF
SysPaperAdvice	Híbrida (FBC+FC)	Explícita	Formulário	Não informa	PDF
arXiv	Baseada em conteúdo	Implícita	Não	Na plataforma	PDF
CORE	Baseada em conteúdo	Implícita	Não	Na plataforma	PDF
SisRecAC	Híbrida (FBC+FC)	Explícita	Formulário	Não informa	PDF

Fonte: autor (2022).

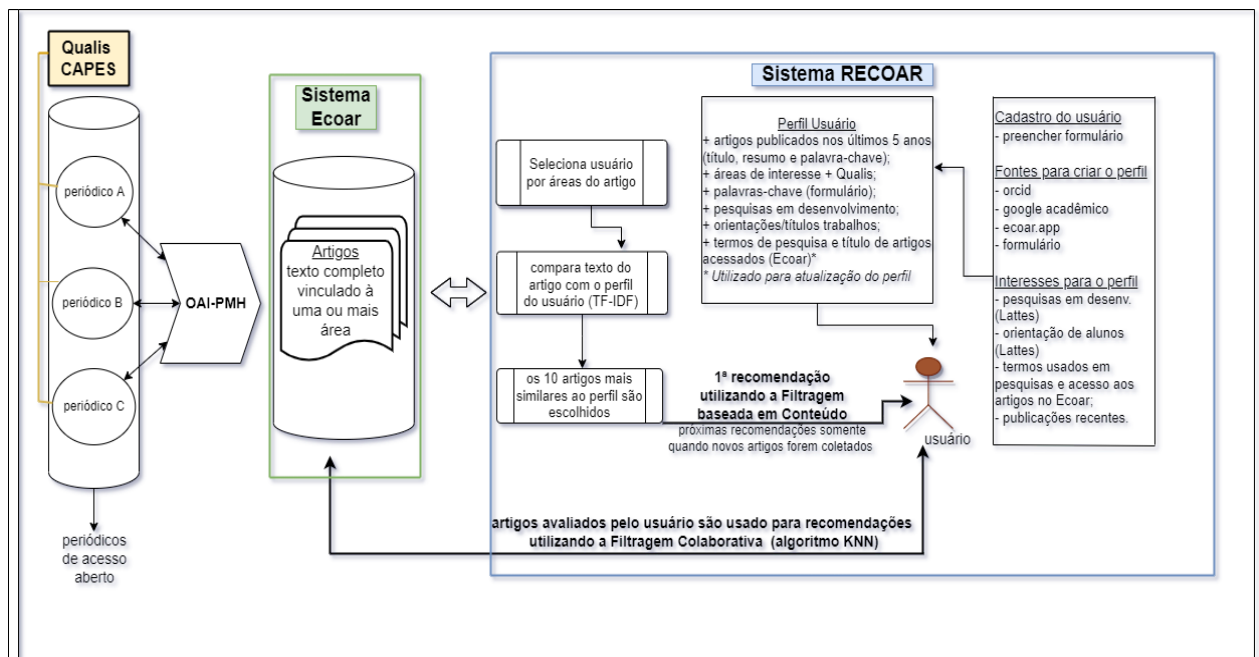
Percebe-se que ao observar o Quadro 7 existem muitas semelhanças entre os SR, com destaque para a utilização da FH como padrão para suas recomendações, revelando que a combinação das filtragens FBC e FC, conforme apresentado no referencial teórico, tende a proporcionar uma maior assertividade nas recomendações. Outra característica comum a maior parte dos sistemas é a coleta de dados feita de forma explícita, ou seja, aplicada no momento do cadastro do usuário no sistema. Ainda, o cadastro é realizado, na maioria dos casos, com o

preenchimento de um formulário no qual o usuário inclui dados pessoais e profissionais, que podem ser usados para o sistema buscar informações para construção do perfil.

4.2 VISÃO GERAL DO SISTEMA PROPOSTO

O sistema proposto possui um esquema de funcionamento conforme aparece na Figura 33, e será detalhado na sequência:

Figura 33 - Visão geral do sistema RECOAR



Fonte: autor (2022).

Observando o sistema proposto na Figura 33, constata-se que o sistema se comunica com a base de dados do sistema Ecoar para coletar os artigos científicos que serão recomendados para os usuários. O sistema Ecoar, hoje, é um buscador de artigos científicos de Acesso Aberto, cuja base foi preenchida com artigos publicados em periódicos listados no Qualis CAPES, estratos A e B, nas 49 áreas de conhecimento. Possui mais de 2 milhões de artigos que podem ser consultados, sendo a busca realizada no texto completo dos artigos, com filtros de área de conhecimento, ano e Qualis. A sua base de artigos é atualizada permanentemente, na medida em que os artigos são publicados nos periódicos, o sistema realiza a sua coleta e indexação.

O sistema Ecoar, por sua vez, interage com a plataforma Sucupira, onde vai buscar os periódicos de Acesso Aberto para poder importar os artigos neles contidos, através do protocolo OAI-PMH. O Ecoar funciona de forma ininterrupta, isto é, quando termina de pesquisar os periódicos de Acesso Aberto na plataforma Sucupira, volta para o início e continua todo o

processo e assim sucessivamente, mantendo a sua base sempre atualizada com artigos, que são importados com textos completos, título, palavras-chave e resumo, em formato PDF e por área de conhecimento, conforme as 49 áreas definidas pela Capes.

O sistema proposto RECOAR também seleciona os seus usuários por áreas de acordo com a Capes, para que assim possa realizar a análise de similaridade entre o perfil do seu usuário com os artigos que fazem parte da base Ecoar. Dessa forma, após comparar o texto do artigo com o perfil do usuário utilizando o algoritmo TF-IDF, o sistema faz a primeira recomendação ao usuário utilizando a FBC, essa recomendação irá conter os 10 artigos que possuírem a maior similaridade com o perfil. As próximas recomendações só ocorrerão quando novos artigos forem coletados e estiverem disponíveis na base de dados do Ecoar.

Quando cadastrado no sistema RECOAR, o pesquisador passa a ter acesso à base de dados do sistema Ecoar, e poderá, caso tenha interesse, realizar buscas nesta base. Ao realizar buscas, o RECOAR grava os termos buscados pelo usuário e os artigos retornados na busca, para futuramente utilizar essas informações para atualizar o perfil do usuário. O usuário também tem a oportunidade de avaliar os artigos que foram retornados pela sua busca, preenchendo um formulário de avaliação. Quando o usuário avalia um artigo, ele é colocado em um grupo dentro da base de dados do RECOAR, esse grupo será formado por usuários que também avaliaram o mesmo artigo, com uma avaliação considerada próxima à sua. Logo, os grupos são formados por semelhança de avaliações, verificadas pelo algoritmo KNN, e serão usados pelo sistema para realizar as recomendações que utilizam a FC. Portanto, quanto mais artigos forem avaliados pelo pesquisador, mais assertiva será a recomendação com base em FC.

Para realizar o cadastro no RECOAR, o usuário preenche um formulário com dados que serão utilizados para criação do seu perfil no sistema. Entre os dados informados, a área de interesse e o estrato Qualis do artigo que deseja receber na recomendação. Além desses dados, o sistema RECOAR vai buscar mais informações nas fontes indicadas pelo usuário no formulário de cadastro, são elas: Orcid, currículo Lattes, Google acadêmico e palavras-chave. Após o cadastro o usuário recebe a primeira recomendação, conforme explicado anteriormente, e caso realize buscas na base Ecoar, os termos utilizados e artigos retornados serão utilizados para atualizar o seu perfil.

Na próxima seção, os casos de uso serão detalhados um a um, com nível de detalhamento maior no documento de requisitos do sistema proposto RECOAR.

4.3 DOCUMENTO DE REQUISITOS

Esta seção contém o principal documento que deve ser utilizado como base para a modelagem do SR proposto no objetivo geral desta pesquisa.

DOCUMENTO DE REQUISITOS DO SISTEMA PROPOSTO – RECOAR

1. INTRODUÇÃO

O modelo de documento de requisitos baseado na norma IEEE 830-1998, utilizado para essa pesquisa, não é prescritivo. Desta forma serão apresentados apenas os elementos do modelo compatíveis com o Sistema de Recomendação proposto, são eles: propósito; âmbito; definições, acrônimos e abreviatura; organização; perspectiva do modelo; funções do produto; características do utilizador; restrições; interfaces externas; requisitos funcionais.

1.1 Propósito: Este sistema tem como propósito recomendar artigos científicos de periódicos de Acesso Aberto, e tem como alvo os pesquisadores científicos.

1.2 Âmbito: O sistema denominado de RECOAR deverá recomendar artigos científicos de periódicos de Acesso Aberto. O sistema irá recomendar, de forma permanente, artigos de periódicos de Acesso Aberto que são compatíveis com o perfil do usuário. Como objetivo, o sistema visa prover ao usuário o recebimento de informações e conhecimentos relevantes por meio de artigos científicos que são da sua área de interesse. Entre os principais benefícios que o sistema pode gerar, pode-se destacar:

- identificar novas oportunidades de pesquisa;
- propiciar maior publicidade aos artigos publicados em periódicos de Acesso Aberto;
- manter o usuário atualizado;
- reduzir o tempo de pesquisa do usuário em busca de novidades científicas;
- identificar pesquisas correlatas;
- estabelecer novos contatos e assim formar novas redes de pesquisadores.

1.3 Definições, acrônimos e abreviatura: Siglas, termos e definições utilizados neste documento de requisitos listadas abaixo.

- Área do conhecimento: uma ou mais áreas de conhecimento entre as 49 áreas definidas pelas CAPES;
- CAPES - Coordenação de Aperfeiçoamento de Pessoal de Nível Superior é uma fundação vinculada ao Ministério da Educação do Brasil que atua na expansão e consolidação da pós-graduação stricto sensu em todos os estados brasileiros;

- CAPTCHA – *Completely Automated Public Turing Test to tell Computers and Humans Apart*: usado como medida de segurança, é um teste de desafio cognitivo utilizado para diferenciar pessoas e robôs;
- Currículo Lattes: currículo onde constam os dados e realizações acadêmicas, faz parte da plataforma Lattes criada pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPQ, destinado a pesquisadores e estudantes;
- DOI – *Digital Object Identifier*: padrão de números e letras que identificam publicações;
- ECOAR - Buscador de artigos científicos de Acesso Aberto que utiliza artigos publicados em periódicos listados no Qualis CAPES, estratos A e B, nas 49 áreas de conhecimento, disponível em <https://ecoar.app>;
- FBC - Filtragem baseada em conteúdo: tipo de filtragem utilizada em sistemas de recomendação que utiliza o conteúdo dos itens do sistema;
- FC - Filtragem colaborativa: tipo de filtragem utilizada em sistemas de recomendação que utiliza colaboração dos usuários de um sistema;
- K-vizinhos ou KNN: algoritmo que realiza previsões baseadas em comportamentos de usuários que pertencem à uma mesma vizinhança ou grupo, que é definida através de critérios de semelhança entre os usuários;
- OAI-PMH - *Open Archives Initiative Protocol for Metadata Harvesting*: Protocolo que define um mecanismo para coleta de metadados em repositórios;
- ORCID – *Open Researcher and Contributor ID*: identificador digital único, tem a função de distinguir pesquisadores, cientistas e acadêmicos em função da ambiguidade e equivalência de nomes de indivíduos;
- Perfil: perfil de um usuário cadastrado no sistema;
- SR - Sistema de Recomendação: sistema que recomenda diversos tipos de itens;
- TF-IDF: algoritmo inspirado em mineração de dados, analisa a frequência que um termo aparece em um documento e pode analisar o quão raro é um termo em um documento.

1.4 Organização: O documento apresenta os requisitos do sistema, sendo que na primeira parte tem-se a introdução com informações preliminares, em seguida na segunda parte as informações gerais, e na terceira com maior aprofundamento, onde os requisitos são apresentados por meio de “casos de uso”.

2. DESCRIÇÃO GERAL

2.1 Perspectiva do produto: O RECOAR é um sistema que trabalha de forma independente, utilizando dados coletados pelo sistema Ecoar, na medida em que utiliza os artigos científicos de Acesso Aberto contidos na sua base para realizar as suas recomendações.

2.2 Funções do Produto: O funcionamento do sistema está baseado nas operações descritas a seguir:

2.2.1 Cadastrar o usuário, por meio do preenchimento de um formulário;

2.2.2 Criar o perfil do usuário, baseado nas informações providas no momento do cadastro;

2.2.3 Selecionar os usuários cadastrados, em função da área de conhecimento dos artigos a serem recomendados;

2.2.4 Realizar a primeira recomendação ao usuário utilizando a FBC, contendo 10 artigos científicos de Acesso Aberto que estejam dentro da(s) mesma(s) área(s) escolhida(s) por ele no momento do seu cadastro. Estes artigos utilizados na primeira recomendação ao usuário, fazem parte da base existente no sistema Ecoar;

2.2.5 Recomendar artigo, a partir da primeira recomendação produzida pelo sistema ao usuário, as próximas recomendações serão feitas à medida que novos artigos cheguem na base do sistema Ecoar;

2.2.6 Classificar artigo científico, o usuário tem a possibilidade de efetuar pesquisas de forma reativa utilizando a base de dados do sistema Ecoar. O usuário poderá classificar os artigos encontrados no sistema Ecoar, preenchendo um formulário de classificação;

2.2.7 Realizar recomendações utilizando a FC, à medida em que o usuário for avaliando os artigos encontrados na base Ecoar;

2.2.8 Atualizar o perfil, sempre que haja algum tipo de mudança feita pelo próprio usuário nas suas áreas de interesse ou quando fizer uma nova publicação. O perfil do usuário também deve ser atualizado com os termos de pesquisa utilizados ao realizar buscas no sistema Ecoar, assim como os títulos dos artigos acessados e a sua classificação também serão informações utilizadas na atualização do perfil, no sentido de aprimorá-lo, para refinar as recomendações;

2.3 Características do utilizador: O usuário alvo do sistema é o pesquisador, graduando, graduado, pós-graduando ou pós-graduado;

2.4 Restrições: Como os interesses de pesquisa do pesquisador podem variar no tempo, o sistema deve permanecer em constante atualização em relação ao perfil do usuário. Os artigos científicos a serem recomendados pelo sistema RECOAR, estão localizados na base de dados do sistema Ecoar, portanto caso este sistema sofra alguma alteração, ajustes poderão ser necessários no sistema RECOAR;

3. REQUISITOS ESPECIFICOS

3.1 Interfaces externas: sistema Ecoar.

3.1 Requisitos funcionais: Neste momento do documento, os requisitos funcionais levantados em função dos “casos de uso” serão apresentados e divididos nos seguintes processos: cadastro do pesquisador; criação do perfil do usuário; seleção dos pesquisadores por área do artigo; importação dos artigos científicos dos periódicos de Acesso Aberto; recomendação de artigos utilizando a filtragem baseada em conteúdo, recomendação de artigos utilizando a filtragem colaborativa e, atualização do perfil do usuário.

3.1.1 Cadastro do pesquisador: Esta subseção relata os processos acima descritos utilizando “Casos de Uso”.

A. CASO DE USO: Cadastrar o usuário/pesquisador.

A1 Escopo: Realizar o cadastro do usuário no sistema.

A2 Descrição Resumida: Este caso de uso relata como o sistema procede ao cadastrar um novo usuário.

A3 Ator Principal: Usuário.

A4 Ator Secundário: Sistema RECOAR.

A5 Pré-condições: o Sistema deve possuir um formulário com campos a serem preenchidos pelo novo usuário.

A6 Fluxo de Eventos: O caso de uso inicia-se com o usuário solicitando novo cadastro e finaliza com o usuário sendo cadastrado na base de dados, após o preenchimento dos campos disponíveis no formulário.

A6.1 Fluxo Básico:

A6.1.1 - O usuário preenche o formulário com os seguintes dados: “Nome Completo”; “Áreas de interesse (Capes)”; “Qualis dos artigos que quer receber”; “Link do perfil no Orcid”; “Link do perfil no Google Acadêmico”; “Link do currículo Lattes”; “Palavras-chave”; “Canal para receber a recomendação”. Caso o usuário não possua algum dos dados solicitados, deve deixar o campo de preenchimento em branco. Os dados preenchidos são gravados no Sistema;

A6.2.1 - Sistema vincula os dados preenchidos no formulário com o usuário e salva os dados na sua base criando um cadastro;

B. CASO DE USO: Criar o perfil do usuário.

B1 Escopo: Construir o perfil do usuário com base nas informações providas no formulário de cadastro.

B2 Descrição Resumida: Este caso de uso permite que o Sistema crie um perfil para o usuário com os dados providos no cadastro, e acessando os links contidos no formulário para buscar informações complementares.

B3 Ator Principal: Sistema RECOAR.

B4 Ator Secundário: Orcid, Google Acadêmico, currículo Lattes, Ecoar e o usuário.

B5 Pré-condições: Estar conectado na internet para poder buscar informações nos links dos sistemas (Orcid, Google Acadêmico, currículo Lattes e Ecoar) onde o usuário possui um perfil, e poder acessar as informações que lá estão.

B6 Fluxo de Eventos: O caso de uso começa com Sistema acessando o formulário preenchido pelo usuário e termina com o Sistema associando os dados encontrados ao usuário, criando assim o seu perfil, e gravando no Sistema. O perfil é criado com base na junção das informações textuais localizadas nos sistemas acessados (Orcid, Google Acadêmico, currículo Lattes e Ecoar).

B6.1 Fluxo Básico

B6.1.1 O caso de uso começa com o Sistema acessando o formulário preenchido pelo usuário onde encontrará os dados sobre “áreas de interesse” e “palavras-chave”, estes dados serão registrados no perfil no usuário;

B6.1.2 O Sistema deve buscar as informações sobre artigos publicados pelo usuário com base no link do Orcid fornecido no cadastro, e resgatará o título dos artigos encontrados, além de utilizar o DOI dos artigos para poder resgatar o resumo e palavras-chave. Após encontrar os dados, eles serão registrados no perfil do usuário;

B6.1.3 O Sistema deve buscar as informações sobre artigos publicados pelo usuário com base no link fornecido do Google Acadêmico, e ao encontrar os artigos, resgatará o título, o resumo e as palavra-chave que serão registrados no perfil do usuário;

B6.1.4 O Sistema deve acessar o currículo Lattes do usuário pelo link fornecido no cadastro, com o objetivo de encontrar informações referentes aos projetos de pesquisa e produções bibliográficas. Ao encontrá-los, deverá resgatar o título dos projetos de pesquisa e das produções bibliográficas, além dos resumos e palavras-chave das produções bibliográficas. O Sistema deve registrar as informações encontradas no perfil do usuário;

B6.1.5 O sistema RECOAR, valendo-se do nome completo fornecido pelo usuário no cadastro, deve comunicar-se com o sistema Ecoar com o propósito de pesquisar nesta base, artigos científicos que possuam o nome usuário em questão como sendo autor ou participante. Quando o nome completo do usuário coincide com o nome do autor ou participante de um ou mais artigos, o(s) artigo(s) é resgatado. Em função da sinonímia, o usuário deve confirmar se o(s)

artigo(s) resgatado(s) é de sua autoria, no caso afirmativo o(s) artigo(s) serão registrados no seu perfil, no caso negativo não haverá registro;

B6.1.6 O Sistema deve eliminar as redundâncias quando elas ocorrerem. Visto que o mesmo artigo pode estar referenciado em mais de uma base, pode aparecer de forma duplicada e, portanto, o sistema deve manter apenas um, eliminando as duplicidades.

B6.1.7 O caso de uso termina com o Sistema registrando no perfil do usuário.

B6.2 Fluxo Alternativo:

B6.2.1 Ao acessar o link do Orcid do usuário, o Sistema deve selecionar o DOI das publicações constantes na plataforma para buscar o resumo e as palavras-chave, entretanto, quando não houver DOI disponível o Sistema utilizará apenas o título da publicação;

B6.2.2 No link do currículo Lattes do usuário, o Sistema pode não ser capaz de acessá-lo devido ao controle de acesso via “*captcha*” que dificulta o acesso computacional, nesse caso os dados do Lattes não serão resgatados.

C. CASO DE USO: Seleção dos pesquisadores/perfis por área do artigo.

C1 Escopo: O Sistema deve selecionar os usuários em função da(s) área(s) de conhecimento cadastrada(s) por ele no sistema RECOAR.

C2 Descrição Resumida: Esse caso de uso permite que o Sistema selecione os usuários cadastrados por área de conhecimento. Esta funcionalidade será executada antes da recomendação de um artigo científico. Um artigo científico de um periódico está vinculado à uma ou mais áreas do conhecimento. Para recomendar um artigo para um usuário, faz-se necessário verificar as áreas do conhecimento que um usuário está vinculado.

C3 Ator principal: Sistema.

C4 Pré-condições: Usuário deve possuir seu cadastro criado no Sistema, e deve existir um artigo a ser recomendado.

C5 Fluxo de Eventos: O Sistema busca na base os usuários com cuja área de conhecimento coincide com a do artigo científico que se pretende recomendar.

C5.1 Fluxo Básico

C5.1.1 O Sistema lista os usuários cadastrados cuja área de conhecimento coincida com as áreas do artigo a ser recomendado;

D. CASO DE USO: Importar artigos científicos de periódicos de Acesso Aberto.

D1 Escopo: Importar os artigos científicos que serão recomendados.

D2 Descrição Resumida: Esse caso de uso possibilita que o Sistema se comunique e utilize os artigos científicos da base de dados Ecoar que foram importados previamente tendo como base os periódicos listados no Qualis periódicos da plataforma Sucupira dentro das 49 áreas do

conhecimento da CAPES. Apesar de não fazer parte do sistema RECOAR, pois essa funcionalidade de importação de artigos é implementada no sistema Ecoar, este caso de uso será detalhado devido à sua importância no processo de recomendação e para melhor entendimento do Sistema de Recomendação proposto.

D3 Ator principal: ECOAR.

D4 Pré-condições: Os metadados dos periódicos no Qualis CAPES devem ter sido previamente importados, e os periódicos devem permitir o *harvesting* (colheita) de metadados dos artigos disponíveis na sua base via protocolo OAI-PMH.

D5 Fluxo de Eventos: O caso de uso inicializa com o sistema Ecoar acessando os metadados dos periódicos para acessar a sua base de dados para importação dos metadados dos artigos e respectivos textos completos (PDF).

D5.1 Fluxo Básico

D5.1.1 Sistema Ecoar deve acessar a lista de periódicos da plataforma Qualis Capes;

D5.1.2 O sistema Ecoar verifica entre os periódicos existentes, por meio do protocolo OAI-PMH, quais os periódicos que possuem Acesso Aberto;

D5.1.3 O sistema Ecoar deve fazer a conexão com o periódico que possui Acesso Aberto por meio do protocolo OAI-PMH, e assim extrair os metadados (título do artigo, resumo, palavras-chave, data de publicação, autores) e o link dos artigos;

D5.1.4 Utilizando o link do artigo, o sistema Ecoar deve buscá-lo no formato PDF para assim poder fazer a indexação. Desta forma o sistema Ecoar deve armazenar na sua base os metadados coletados e os textos completos dos artigos em formato PDF.

E. CASO DE USO: Recomendar os artigos utilizando a filtragem baseada em conteúdo.

E1 Escopo: Recomendar artigos científicos para o usuário.

E2 Descrição Resumida: Esse caso de uso proporciona realizar a recomendação de artigos científicos de Acesso Aberto ao usuário utilizando a FBC.

E3 Ator principal: Sistema RECOAR.

E4 Ator secundário: Sistema Ecoar.

E5 Pré-condições: Sistema deve possuir conexão com a internet.

E6 Fluxo de Eventos: O caso de uso inicializa com o sistema RECOAR selecionando os artigos pertinentes para cada usuário, finaliza recomendando-os ao usuário correspondente pelo canal de comunicação por ele escolhido, e registrando na base de dados do sistema RECOAR, qual artigo foi recomendado para cada usuário.

E6.1 Fluxo Básico

E6.1.1 O sistema Ecoar, a cada novo artigo importado, o envia ao sistema RECOAR para verificar a possibilidade de recomendá-lo aos seus usuários;

E6.1.2 O RECOAR seleciona os pesquisadores cuja(s) áreas de conhecimento e nível do Qualis informadas coincidam com a área do artigo recebido;

E6.1.3 Com os pesquisadores selecionados, o sistema RECOAR deve comparar o texto do artigo (conteúdo do artigo em formato PDF) recebido, com o perfil do pesquisador utilizando o algoritmo IDF-TF, e verificar o grau de similaridade. Neste caso, o sistema pode utilizar a interface Lucene, que é uma biblioteca de mecanismo de pesquisa usada para indexar e pesquisar o texto e funciona de maneira mais eficiente para buscar palavras e metadados em documentos.

E6.1.4 Se o grau de similaridade estiver de acordo com o esperado, o Sistema deve recomendar o(s) artigo(s) ao usuário pelo canal de comunicação por ele escolhido, e registrar na base de dados do sistema RECOAR qual foi o artigo recomendado para o usuário. O sistema deve recomendar apenas artigos que ainda não foram recomendados ao usuário, e não recomendar artigos que tenham o próprio usuário como autor.

E6.2 Fluxo Alternativo (acontece apenas uma vez, logo após o cadastro do usuário)

E6.2.1 Este fluxo acontece apenas quando um novo usuário realiza o cadastro no sistema RECOAR e tem o seu perfil criado, o Sistema deve selecionar o usuário pela(s) área(s) do conhecimento e em seguida pelo nível do Qualis informado;

E6.2.2 O sistema RECOAR busca no sistema Ecoar, dez (10) artigos com maior nível de similaridade encontrado, por meio da comparação do perfil do usuário com o texto completo dos artigos. Somente artigos pertencentes a(s) área(s) de conhecimento do usuário serão recomendados, e com o nível do Qualis escolhido pelo usuário;

E6.2.3 O sistema RECOAR deve recomendar os artigos ao usuário, pelo canal de comunicação por ele indicado.

F. CASO DE USO: Classificar artigos.

F1 Escopo: Classificar os artigos científicos.

F2 Descrição Resumida: Esse caso de uso permite ao usuário realizar a classificação de artigos científicos de Acesso Aberto que lhe foram recomendados e/ou que tenham sido localizados a partir de uma busca na base do sistema Ecoar. A classificação do artigo científico consiste no preenchimento de um formulário indicado, por meio de uma nota de 0 a 5, a sua originalidade, inovatividade, atualidade, rigor metodológico, e, nível de correção, clareza e coerência de linguagem. Ao final será atribuída uma nota ao artigo, que corresponderá a média aritmética das notas atribuídas para cada quesito de classificação.

F3 Ator principal: Usuário.

F4 Ator secundário: sistema Ecoar e RECOAR.

F5 Pré-condições: O usuário deve ter recebido a primeira recomendação feita pelo RECOAR após o seu cadastro, ou ter visualizado artigos a partir de uma busca realizada no sistema Ecoar.

F6 Fluxo de Eventos: O caso de uso inicializa com o usuário recebendo artigo(s) recomendados pelo sistema RECOAR, ou por meio de uma busca no sistema Ecoar, que resulta numa lista de artigos como resultado. Uma vez acessado o artigo, o usuário poderá avaliá-lo.

F6.1 Fluxo Básico

F6.1.1 O usuário recebe artigos recomendados pelo RECOAR ou acessa e busca artigos no sistema Ecoar;

F6.1.2 O usuário deve classificar o artigo, preenchendo o formulário de classificação;

F6.1.3 O sistema deve salvar os dados da classificação e registrá-los.

G. CASO DE USO: Recomendar os artigos utilizando a Filtragem Colaborativa.

G1 Escopo: Recomendar artigos científicos utilizando a FC.

G2 Descrição Resumida: Esse caso de uso permite ao Sistema realizar recomendações de artigos científicos de Acesso Aberto ao usuário utilizando a FC.

G3 Ator principal: RECOAR.

G4 Ator secundário: sistema Ecoar.

G5 Pré-condições: Para receber recomendações de um artigo pelo algoritmo que utiliza a FC, um ou mais usuários precisarão tê-lo classificado previamente, bem como o usuário que irá receber a recomendação.

G6 Fluxo de Eventos: O caso de uso inicializa com o sistema RECOAR acessando a base de dados do sistema Ecoar e captando um ou mais artigos que estejam de acordo com os requisitos da FC e possam ser recomendados, finaliza realizando a recomendação dos artigos ao usuário e registrando na base de dados do sistema RECOAR, qual artigo foi recomendado para cada usuário.

G6.1 Fluxo Básico

G6.1.1 O sistema RECOAR deve acessar o sistema Ecoar.

G6.1.2 O RECOAR deve buscar e selecionar os artigos cuja(s) área(s) do conhecimento e o nível do Qualis desejado, correspondam com a área do usuário que vai receber a recomendação;

G6.1.3 Com os artigos selecionados pela área do conhecimento, o RECOAR deve escolher apenas os artigos que tenham sido classificados pelos usuários cadastrados no Sistema e que pertençam à mesma área do conhecimento e possuam o nível do Qualis escolhido pelo usuário;

G6.1.4 O Sistema deve avaliar o grau de proximidade do usuário que receberá a recomendação com os perfis que fazem parte da mesma área do conhecimento por meio do algoritmo K-vizinhos.

G6.1.5 Os artigos que obtiverem o maior grau de proximidade/similaridade ao comparar a classificação do usuário com os demais usuários, de acordo com o algoritmo KNN, e que tiverem obtido nota superior a três (03), deverão ser recomendados ao usuário via canal de comunicação por ele escolhido, e registrar na base de dados do sistema RECOAR qual artigo foi recomendado para o usuário. O Sistema deve recomendar apenas artigos que ainda não foram recomendados ao usuário, e não recomendar artigos que tenham o próprio usuário como autor.

5 CONSIDERAÇÕES FINAIS

Foi observado que os SR são sistemas que personalizam os interesses e necessidades dos seus usuários para poder oferecer itens que se encaixem com seu perfil. Para que um SR atinja o seu objetivo, é necessária uma maturação do sistema, que geralmente ocorre com esse estando em funcionamento durante um certo período, com uma quantidade suficiente de usuários interagindo com o sistema, recebendo recomendações, classificando os artigos e com um grande acervo de itens disponíveis para serem recomendados.

De acordo com os objetivos específicos propostos, o aprofundamento na teoria mostrou a importância do avanço do movimento do Acesso Aberto e todos os passos que foram dados nos últimos 20 anos, viabilizando o acesso a trabalhos científicos em artigos e periódicos, principalmente. Da mesma forma ocorreu com os periódicos científicos, foi apresentado o seu surgimento, sua caracterização como um veículo que difunde o conhecimento científico, e sua própria evolução ao longo do tempo, acompanhando as novas tecnologias e passando a ser oferecido de forma eletrônica, favorecendo ainda mais o acesso.

Em relação aos SR, foi feita uma análise sobre o seu funcionamento, foram apresentadas as suas características e mostrou-se que os SR vêm sendo utilizados por inúmeros tipos e tamanhos de empresas com os mais diversos fins. Ficou evidente que os SR possuem extrema importância atualmente e muito se deve ao ambiente que hoje vivemos, sempre conectados na *web*. As várias abordagens existentes e empregadas pelos SR, os tipos de filtragens e as suas possíveis combinações revelaram que existe a possibilidade de criação de sistemas que atendam diferentes objetivos e usuários. O ambiente acadêmico tem espaço suficiente para a utilização de SR em vários aspectos e pode se beneficiar com o desenvolvimento de sistemas que agreguem novas possibilidades. Dentro desse ambiente acadêmico, os pesquisadores brasileiros são os principais beneficiados, inclusive no caso da proposta que foi apresentada neste trabalho, no qual o sistema proposto somente recomendará artigos de Acesso Aberto, sem custos ao usuário e com a possibilidade de escolha do nível Qualis do artigo que vai ser recomendado.

Acerca dos algoritmos, existem inúmeros modelos que já são utilizados em SR, conforme observou-se no diagnóstico realizado nos SR de artigos científicos existentes e os protótipos. Após a análise dos principais algoritmos, verificou-se que a sua utilização atende ao que foi proposto, entretanto no campo da computação podem surgir novas ideias e algoritmos que possa ter uma performance superior.

Entende-se que o objetivo geral foi atingido, visto que a modelagem dos requisitos de um Sistema de Recomendação foi criada e a base de dados com os artigos científicos

importados de periódicos de Acesso Aberto (sistema Ecoar) já se encontra em funcionamento com uma vasta quantidade de artigos científicos que podem ser recomendados. Entre os diferenciais do modelo de sistema proposto, destaca-se a recomendação de artigos científicos somente de Acesso Aberto, sem nenhum tipo de cobrança ao usuário e a possibilidade do usuário poder escolher o Qualis dos artigos que serão recomendados, e suas respectivas áreas do conhecimento.

Algumas dificuldades foram encontradas durante o desenvolvimento do trabalho: uma extensa bibliografia sobre os conceitos estudados, com novos termos, especialmente os advindos da área da computação; o tempo insuficiente para internalizar, organizar e associar os conceitos para a modelagem; e a impossível criação de um protótipo que permitisse avaliar o sistema proposto.

Como sugestão para futuros trabalhos, a continuação do sistema com a sua implantação, atraindo a comunidade acadêmica em número suficiente de usuários para que o sistema possa funcionar de forma plena, realizando as recomendações, recebendo o *feedback* dos pesquisadores para poder ser avaliado e melhorado, de forma a manter-se atualizado, aprimorando as recomendações e, conseqüentemente, atraindo mais usuários.

REFERÊNCIAS

- ACADEMIA.EDU. **About**. Disponível em: <https://www.academia.edu/about>. Acesso em: 18 jan. 2021.
- ADOMAVICIUS, G.; TUZHILIN, A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. **IEEE Transactions on Knowledge and Data Engineering**, v. 17, n. 6, p. 734-749, 2005. DOI: <https://doi.org/10.1109/TKDE.2005.99>. Disponível em: <https://ieeexplore.ieee.org/document/1423975>. Acesso em: 5 ago. 2021.
- AGUIA. Agência USP de Gestão da Informação Acadêmica Universidade de São Paulo. **Periódicos Qualis CAPES**. 2015-2022. Disponível em: <https://www.aguia.usp.br/apoio-pesquisador/escrita-publicacao-cientifica/selecao-revistas-publicacao/qualis-periodicos/>. Acesso em: 30 abr. 2022.
- AGGARWAL, Charu C. **Recommender System**: the textbook. Suíça: Springer, Cham. 2016.
- AIVAREZ, E. B. *et al.* Os Sistemas de Recomendação, Arquitetura da Informação e a Encontrabilidade da Informação. **Transinformação**, v. 28, n. 3, p. 275-286, 2016. DOI:10.1590/2318-08892016000300003. Disponível em: <https://brapci.inf.br/index.php/res/v/116887>. Acesso em: 6 dez. 2020.
- APPEL, A. L. **Dimensões tecnopolíticas e econômicas da comunicação científica em transformação**. 2019. Tese (Doutorado em Ciência da Informação) – Escola de Comunicação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2019.
- AVILA, C. M. O. **Desenvolvimento de um Sistema de Recomendação de Artigos Científicos e Avaliações de Métodos de Extração de Palavras-Chave**. Universidade Católica de Pelotas, Pelotas, 2006. Disponível em: <http://ppginf.ucpel.tche.br/TI-arquivos/2006/ChristianoAvila/PPGINF-UCPel-TI-2006-2-01.pdf>. Acesso em: 27 fev. 2022.
- AZAMBUJA, R. X.; MORAIS, A. J.; FILIPE, V. Teoria e Prática em Sistemas de Recomendação. **Revista de Ciências da Computação**, v. 16, n. 16, p. 23-46, 2021. DOI: 10.34627/rcc.v15i0.264. Disponível em: <https://journals.uab.pt/index.php/rcc/article/view/264>. Acesso em: 21 fev. 2022.
- BAI, X. *et al.* Scientific Paper Recommendation: A Survey. **IEEE Access**, v. 7, p. 9324-9339, 2019. DOI: 10.1109/ACCESS.2018.2890388. Disponível em: <https://ieeexplore.ieee.org/document/8598708>. Acesso em: 11 dez. 2021.
- BARCELLOS, C. D. *et al.* Sistema de Recomendação Acadêmico para Apoio a Aprendizagem. **RENOTE**, v. 5, n. 2, 2007. DOI 10.22456/1679-1916.14236. Disponível em: <https://seer.ufrgs.br/renote/article/view/14236>. Acesso em: 5 dez. 2020.
- BARROS, A.; SILVEIRA, S. R.; PERTILE, S. L. **SysPaperAdvice**: desenvolvimento de um protótipo de sistema de recomendação de artigos trabalhos científicos. 2016. Trabalho de Conclusão de Curso (Bacharelado em Sistemas de Informação) – Universidade Federal de Santa Maria, Westphalen, 2016.

BEEL, J. **ArXiv now has a recommender system, provided by CORE**. 2020. Disponível em: <https://recommender-systems.com/news/2020/10/19/arxiv-now-has-a-recommender-system-provided-by-core/>. Acesso em: 28 jan. 2022.

BIOJONE, M. R. **Os periódicos científicos na comunicação da ciência**. São Paulo: Educ/Fapesp, 2003.

BOAI. Budapest Open Access Initiative. **Iniciativa de Budapeste pelo Acesso Aberto**. 2002. Disponível em: <http://www.budapestopenaccessinitiative.org/translations/portuguesetranslation>. Acesso em: 07 nov. 2020.

BONILLA, C. J. C. **Uma introdução aos sistemas de Recomendação**: modelos matemáticos, algoritmos e aplicações. 2020. Dissertação (Mestrado em Matemática Aplicada e Computacional) – Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, Campinas, 2020.

BRITO, G. N.; LIMA, I. F. Periódicos Científicos como fone de Informação: um estudo na Informação & Sociedade e na Biblionline. **Folha de Rosto**, v. 1, n. 2, p. 49-60, 2015. Disponível em: <https://periodicos.ufca.edu.br/ojs/index.php/folhaderosto/article/view/42>. Acesso em: 10 nov. 2020.

BURKE, R. Hybrid recommender systems: Survey and experiments. **User Modeling and User-Adapted Interaction**, v. 12, n. 4, p. 331-370, 2002. DOI: 10.1023/A:1021240730564. Disponível em: <https://link.springer.com/article/10.1023/A:1021240730564>. Acesso em: 10 set. 2021.

CAMPOS, L. M. *et al.* Combining content-based and collaborative recommendations: A hybrid approach based on Bayesian networks. **International Journal of Approximate Reasoning**, v. 51, n. 7, p. 785-799, 2010. DOI: 10.1016/j.ijar.2010.04.001. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0888613X10000460?via%3Dihub> Acesso em: 10 set. 2021.

CAPES. Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Quem somos**. 2020. Disponível em: <http://www-periodicos-capes-gov-br.ezl.periodicos.capes.gov.br/index.php/sobre/quem-somos.html>. Acesso em: 31 nov. 2021.

CARVALHO, C. L.; OLIVEIRA, R. R. **Implementação de Interoperabilidade entre Repositórios Digitais por meio do Protocolo OAI-PMH**. Relatório Técnico, 2009. Disponível em: https://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_003-09.pdf. Acesso em: 20 mar. 2022.

CAZELLA, S. C.; REATEGUI, E. B. Sistemas de Recomendação. *In*: CONGRESSO DA SOCIEDADE BRASILEIRA DE COMPUTAÇÃO, 25. 2005. São Leopoldo. **Sistemas de Recomendação**. São Leopoldo: [s.n.], 2005. p. 306-348.

CAZELLA, S. C. **Aplicando a Relevância da Opinião de Usuários em Sistema de Recomendação para Pesquisadores**. 2006. Tese (Doutorado em Ciência da Computação) – Universidade Federal do Rio Grande do Sul. Porto Alegre, 2006.

CAZELLA, S. C.; DRUMM, J. V.; BARBOSA, J. L. V. Um serviço para recomendação de artigos científicos baseado em filtragem de conteúdo aplicado a dispositivos móveis. **RENOTE-Revista Novas Tecnologias na Educação**, v. 8, n. 3, 2010. Disponível em: https://scholar.google.com.br/scholar?cluster=14316360232910482189&hl=pt-BR&as_sdt=2005. Acesso em: 21 out. 2020.

CAZELLA, S. C.; NUNES, M. A. S. N.; REATEGUI, E. B. A Ciência da Opinião: Estado da arte em Sistemas de Recomendação. In: CARVALHO, André Ponce de Leon F.; KOWALTOWSKI, Tomasz (org.). **Jornada de Atualização de Informática-JAI**. Rio de Janeiro, 2010. p. 161-216.

CIRIACO, D. **O que é e como usar o Google Acadêmico**. 2015. Disponível em: <https://canaltech.com.br/mercado/o-que-e-e-como-usar-o-google-academico/>. Acesso em: 22 out. 2021.

CORE. **About CORE**. Disponível em: <https://core.ac.uk/about>. Acesso em: 10 mar. 2022.

CORNELL UNIVERSITY. **arXiv: About**. 2022. Disponível em: <https://arxiv.org/about>. Acesso em: 28 fev. 2022.

CORNELL UNIVERSITY. **CORE Recommender now supports article discovery on arXiv**. 2020. Disponível em: <https://blog.arxiv.org/2020/10/15/core-recommender-now-supports-article-discovery-on-arxiv/>. Acessado em: 28 fev. 2022.

COSTA, E.; AGUIAR, J.; MAGALHÃES, J. “Sistemas de Recomendação de Recursos Educacionais: conceitos, técnicas e aplicações”. In: **Jornada de Atualização em Informática na Educação - JAIE**. 2013.

CUNHA, B. M. C. **Interoperabilidade em Provedores de Dados e Provedores de Serviços: uma análise dos metadados e protocolos OAI-PMH e OAI-ORE**. 2021. Dissertação (Mestrado em Ciência da Informação) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Marília, 2021.

FAPESP (São Paulo). **Acesso Aberto**. 2019. Disponível em: https://www.aguia.usp.br/apoio-pesquisador/acesso-aberto-usp/entenda-o-que-e-acesso-aberto/?doing_wp_cron=1603999780.6702959537506103515625. Acesso em: 27 nov. 2020.

FIOCRUZ. **Glossário da Ciência Aberta**. 2022. Disponível em: <https://portal.fiocruz.br/glossario-da-ciencia-aberta>. Acesso em: 11 fev. 2022.

FOWLER, M. **UML essencial: um breve guia para linguagem-padrão de modelagem de objetos**. 3. ed. Porta Alegre: Bookman, 2005.

FORTUNE. **Amazon's recommendation secret**. Fortune, 2012. Disponível em: <https://fortune.com/2012/07/30/amazons-recommendation-secret/>. Acesso em: 11 out. 2021.

GARCIA, P.A. B.; SUNYE, M. S. O protocolo OAI-PMH para interoperabilidade em bibliotecas digitais. In: CONGRESSO DE TECNOLOGIA PARA GESTÃO DE DADOS E METADADOS DO CONE SUL, 1., 2003, Ponta Grossa. **Anais [...]**. Ponta Grossa: UEPG,

2003. Disponível em: <https://docplayer.com.br/14283209-O-protocolo-oai-pmh-para-interoperabilidade-em-bibliotecas-digitais.html>. Acesso em: 25 jan. 2022.

GARVEY, W. D. **Communication: the Essence of Science**. Oxford: Pergamon Press, 1979.
GATTO, E. C.; ZORZO, S. D. **Personalização: O que é?** 15 mar. 2018. Apresentação em Slideshare. 50 slides, color. Disponível em: <https://pt.slideshare.net/elainececiliagatto/personalizao-o-que>. Acesso em: 21 nov. 2021.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GILLESPIE, T. A relevância dos algoritmos, **Parágrafo**, v. 6, n. 1, p. 95-121, 2018.
Disponível em: <http://revistaseletronicas.fiamfaam.br/index.php/recicofi/article/view/722/0>.
Acesso em: 08 jan. 2022.

GOLDBERG, D. *et al.* Using collaborative filtering to weave an information tapestry. **Communications of the ACM**, v. 35, n. 12, p. 61-70. 1992.

GONÇALVES, A. *et al.* **IEEE Std 830 Prática Recomendada Para Especificações de Exigências de Software**. Lisboa, 2004. Disponível em: <http://professor.pucgoias.edu.br/sitedocente/admin/arquivosUpload/17785/material/IEEE830.pdf>. Acesso em: 20 fev. 2022.

IBICT. **Biblioteca Digital Brasileira e Dissertações**. 2018. Disponível em: <https://antigo.ibict.br/informacao-para-a-pesquisa/bdtd#apresentacao>. Acesso em: 21 fev. 2022.

IEEE Recommended Practice for Software Requirements Specifications. *In: IEEE Std 830-1998*. Nova Iorque. p. 1-40, 1998. DOI: 10.1109/IEEESTD.1998.88286. Disponível em: <http://www.math.uaa.alaska.edu/~afkjm/cs401/IEEE830.pdf>. Acesso em: 11 fev. 2022.

ISKOLD, A. The Art, Science and Business of Recommendation Engines. **Readwrite**, 16 jan. 2007. Disponível em: https://readwrite.com/recommendation_engines/. Acesso em 20 out. 2020.

KNOTH, P. *et al.* **Towards effective research recommender systems for repositories**. 2017. Disponível em: <https://arxiv.org/abs/1705.00578>. Acesso em: 25 fev. 2022.

KUBAT, M. **An Introduction to Machine Learning**. Berlim: Springer: 2017.

LAAKSO, M. *et al.* The development of Open Access Journal Publishinh from 1993 to 2009. **PLoS ONE**. v. 6, 2011. DOI: 10.1371/journal.pone.0020961. Disponível em: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0020961>. Acesso em: 21 out. 2021.

LAUDON, C. K.; LAUDON, J. D. **Sistemas de Informação Gerenciais**. 5. ed. São Paulo: Pearson Prentice Hall, 2004.

LÁZARO, A. S. **Análise e seleção de algoritmos de filtragem de informação para solução do problema cold-start item**. 2010. Monografia (Graduação em em Sistemas de Informação) - Universidade Federal de Lavras, Minas Gerais, 2010.

LEMOS, E. P.; STEINER, M. T. A.; NIEVOLA, J. C. Análise de crédito bancário por meio de redes neurais e árvores de decisão: uma aplicação simples de data mining. **Revista de Administração: RAUSP**, São Paulo, v. 40, n. 3, p. 225-234, 2005. Disponível em: <https://www.redalyc.org/articulo.oa?id=223417392002>. Acesso em: 20 fev. 2022.

LIMA, K. C. R.; MENDES JUNIOR, R.; PINTO, J. S. P. Artigos científicos publicados em acesso aberto pelos docentes dos programas de pós-graduação stricto sensu no Setor de Ciências Sociais Aplicadas na UFPR. **AtoZ: novas práticas em informação e conhecimento**, v. 8, n. 2, p. 106-113, out. 2019. Disponível em: <https://revistas.ufpr.br/atoz/article/view/76295>. Acesso em: 16 nov. 2021.

LU, J. *et al.* Recommender system application developments: A survey. **Decision Support Systems**, v. 74, p. 12-32, 2015. DOI: 10.1016/j.dss.2015.03.008. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0167923615000627?via%3Dihub>. Acesso em: 27 jan. 2021

MAES, P.; SHARDANAND, U. Social information filtering: Algorithms for automating “word of mouth”. *In: Human Factors in Computing Systems*. Proceedings, 1995, p. 210-217.

MARCONI, M. A.; LAKATOS, E. M. **Metodologia do Trabalho Científico**: procedimentos básicos, pesquisa bibliográfica, projeto e relatório, publicações e trabalhos científicos. 7. ed. São Paulo: Atlas, 2013.

MARTINS, D. L.; FERREIRA, S. M. S. P. Protocolo OAI-PMH e Sistemas Federados de Informação: fundamentos de arquitetura da informação para análise de dados do portal de produção científica da área de Ciências da Comunicação Univerciencia.org. **Liinc em Revista**, Rio de Janeiro, v. 8, n. 2, p. 431-447, set. 2012. DOI: 10.18617/liinc.v8i2.479. Disponível em: <http://revista.ibict.br/liinc/article/view/3359/2965>. Acesso em: 10 fev. 2022.

MONTEIRO-KREBS, L. *et al.* Depicting Recommendations in Academia: How ResearchGate Communicates with Its Users (via Design or upon Request) About Recommender Algorithms. *In: ÁLVAREZ, E. B. Data and Information in Online Environments*. Berlim: Springer, 2021. p. 3-25.

MORETTI, Isabella. **Google Acadêmico**: o que é e como achar artigos científicos. 2020. Disponível em: https://viacarreira.com/google-academico/#Outras_funcionalidades_do_Google_Scholar. Acesso em: 30 jan. 2022.

MOTTA, C. *et al.* Sistemas de Recomendação. *In: PIMENTEL, M., FUKS, H. Sistemas Colaborativos*. Rio de Janeiro: Elsevier, 2011. p. 230-244.

MUELLER, S. P. M. O periódico científico. *In: CAMPELLO, B. S.; CENDON, B. V.; KREMER, J. M. Fontes de informação para pesquisadores e profissionais*. Belo Horizonte: UFMG, 2000. p. 73-94.

O'BRIEN, J. A.; MARAKAS, G. M. **Administração de sistemas de informação**. 15. ed. Porto Alegre: McGraw-Hill, 2013.

OPEN ARCHIVES. Open Archives Initiative: organization. **Open Archives**, [201-?a].

Disponível em: <https://www.openarchives.org/organization/>. Acesso em 25 fev. 2022.

OPEN ARCHIVES. Open Archives Initiative: Data Provider Registration. **Open Archives**, [201-?b]. Disponível em: https://www.openarchives.org/pmh/register_data_provider. Acesso em 25 fev. 2022.

PACKER, A. L. *et al.* **15 Anos de Acesso Aberto**: um estudo analítico sobre Acesso Aberto e comunicação científica. Paris: UNESCO, 2014. E-book. Disponível em: <http://www.scielo15.org/livro-scielo-15/> Acesso em: 25 out. 2020.

PARK, S. T.; CHU, W. Pairwise preference regression for cold-start recommendation. *In: RecSys '09: Proceedings of the third ACM conference on Recommender systems*. Nova Iorque, p. 21-28, 2009. DOI: 10.1145/1639714.1639720. Disponível em: <https://dl.acm.org/doi/pdf/10.1145/1639714.1639720>. Acesso em: 11 nov. 2021.

PEREIRA, D. M. R. **Fonte de informação científica**. 19 de mar. 2012. Apresentação em slideshare. 68 slides, color. Disponível em: <https://pt.slideshare.net/bsararangua/fontes-de-informao-cientfica>. Acesso em: 5 fev. 2022.

PRADO, J. M. K. **Evidências teóricas para um marketing da Ciência em Acesso Aberto**. 2019. 139 f. Tese (Doutorado em Ciência da Informação) – Universidade Federal de Santa Catarina, Florianópolis, SC, 2019. Disponível em: <https://repositorio.ufsc.br/handle/123456789/215017>. Acesso em: 28 nov. 2021.

PRESSMAN, R. S. **Engenharia de software**. 6. ed. São Paulo: McGraw-Hill, 2006.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do Trabalho Científico**: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico. 2. ed. Novo Hamburgo: Freevale, 2013.

REATEGUI, E. B; CAZELLA, S. C.; OSÓRIO, F. S. **Personalização de Páginas Web através dos Sistemas de Recomendação**. 2006. Disponível em: <http://osorio.wait4.org/publications/Reategui-et-al-IHC2006.pdf>. Acesso em: 12 nov. 2021.

RESEARCHGATE. **About**. 2021. Disponível em: <https://www.researchgate.net/about>. Acesso em: jan. 2022.

RESNICK, P.; VARIAN, H. R. Recommender Systems. **Communications of the ACM**, Nova Iorque, v. 40, n. 3, p. 55-58, 1997. DOI: 10.1145/245108.245121. Disponível em: <https://dl.acm.org/doi/10.1145/245108.245121>. Acesso em: 31 out. 2020.

RICCI, F., ROKACH, L., SHAPIRA, B. Introduction to recommender systems handbook. *In: Ricci et al. Recommender systems handbook*. Berlim: Springer, 2010. p. 1-35.

RICHARDSON, R. J. **Pesquisa social**: métodos e técnicas. 3. ed. São Paulo: Atlas, 1999.

RIOS, F. P.; LUCAS, E. R. O.; AMORIM, I. S. Manifestos do movimento de acesso aberto: análise de domínio a partir de periódicos brasileiros. **RBBB - Revista Brasileira de Biblioteconomia e Documentação**, São Paulo, v. 15, n. 1, p. 148-169, 2019. Disponível em: <https://rbbd.febab.org.br/rbbd/article/view/1152/1103>. Acesso em: 20 set. 2021.

RODRIGUEZ, R. S.; STUBERT, D. Periódicos Científicos da Ciência da Informação: os títulos indexados na WoS. **Revista Española de Documentación Científica**, v. 38, n. 3, p. 1-15, 2015. DOI:10.3989/redc.2015.3.1197. Disponível em: <https://redc.revistas.csic.es/index.php/redc/article/view/897>. Acesso em: 5 out. 2021.

ROESCH, S. M. A. **Projetos de estágio e de pesquisa em administração**. 3. ed. São Paulo: Atlas, 2009.

SHETTY, N. *et al.* **Scholar Recommendations Reloaded! Fresher, More Relevant, Easier**. 2021. Disponível em: <https://scholar.googleblog.com/2021/02/scholar-recommendations-reloaded.html>. Acesso em: 28 fev. 2022.

SOMMERVILLE, I. **Engenharia de Software**. 9. ed. São Paulo: Pearson Prentice Hall, 2011.

SOUTO JUNIOR, M.; BEZERRA, B. Uma Investigação Sobre o Efeito da Combinação de Algoritmos em Sistemas de Recomendação. **Revista de Engenharia e Pesquisa Aplicada**, v. 5, n. 1, p. 58-66, 2020. DOI: 10.25286/rep.v5i1.1199. Disponível em: <http://revistas.poli.br/index.php/rep/article/view/1199>. Acesso em: 11 jan. 2022.

VARAJÃO, J. E. G. A. S. **Função de Sistemas de Informação**: Contributos para a melhoria do sucesso da adopção de tecnologias de informação e desenvolvimento de sistemas de informação nas organizações. Tese (Doutorado em Sistemas de Informação) - Universidade do Minho, Guimarães (Portugal), 2002. Disponível em: <http://repositorium.sdum.uminho.pt/bitstream/1822/25960/1/TesePhdJoaoVarajao.pdf>. Acesso em: 14 mar. 2022.

VELOSO, E. C. L. M.; NASCIMENTO, G. B. Uso dos periódicos biblionline pelos alunos pré-concluintes do curso de Biblioteconomia da Universidade Federal da Paraíba do período 2011. **Biblionline**, João Pessoa, v. 8, n. esp., p. 109-121, 2012. Disponível em: <https://periodicos.ufpb.br/ojs/index.php/biblio/article/view/14195/8102>. Acesso em: 17 set. 2021.

WAZLAWICK, R. S. **Análise e projeto de sistemas de informação orientados a objetos**. 2. ed. Rio de Janeiro: Elsevier, 2011.

WIRED. **Netflix never used its \$1 million algorithm due to engineering costs**. Wired, 2012. Disponível em: <https://www.wired.com/2012/04/netflix-prize-costs/>. Acesso em: 10 out. 2021.