

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE COMPUTAÇÃO

IGOR THIAGO CORREA FILHO
Orientador: Prof. Dr. Guilherme Tavares de Assis

**PROPOSTA E DESENVOLVIMENTO DE UM SERVIÇO DE
RECOMENDAÇÃO PARA A BIBLIOTECA DIGITAL DO MESTRADO
PROFISSIONAL EM ENSINO DE CIÊNCIAS DA UNIVERSIDADE
FEDERAL DE OURO PRETO**

Ouro Preto, MG
2021

UNIVERSIDADE FEDERAL DE OURO PRETO
INSTITUTO DE CIÊNCIAS EXATAS E BIOLÓGICAS
DEPARTAMENTO DE COMPUTAÇÃO

IGOR THIAGO CORREA FILHO

**PROPOSTA E DESENVOLVIMENTO DE UM SERVIÇO DE RECOMENDAÇÃO
PARA A BIBLIOTECA DIGITAL DO MESTRADO PROFISSIONAL EM ENSINO DE
CIÊNCIAS DA UNIVERSIDADE FEDERAL DE OURO PRETO**

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como parte dos requisitos necessários para a obtenção do grau de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Guilherme Tavares de Assis

Ouro Preto, MG
2021



FOLHA DE APROVAÇÃO

Igor Thiago Correa Filho

Proposta e desenvolvimento de um serviço de recomendação para a Biblioteca Digital do Mestrado Profissional em Ensino de Ciências da Universidade Federal de Ouro Preto

Monografia apresentada ao Curso de Ciência da Computação da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Ciência da Computação

Aprovada em 26 de Agosto de 2021.

Membros da banca

Guilherme Tavares de Assis (Orientador) - Doutor - Universidade Federal de Ouro Preto
Dayanne Gouveia Coelho (Examinadora) - Doutora - Universidade Federal de Ouro Preto
Michele Hidemi Ueno Guimarães (Examinadora) - Doutora - Universidade Federal de Ouro Preto

Guilherme Tavares de Assis, Orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 26/08/2021.



Documento assinado eletronicamente por **Guilherme Tavares de Assis, PROFESSOR DE MAGISTERIO SUPERIOR**, em 26/08/2021, às 17:10, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0210293** e o código CRC **69298B39**.

Dedico este trabalho a todos que contribuíram para sua realização.

Agradecimentos

Agradeço a Deus pela minha vida; aos meus pais pelo amor, ensinamentos e oportunidades; ao meu irmão pelo companheirismo de sempre.

Agradeço a UFOP e a todos os professores do DECOM pela oportunidade e ensinamentos, especialmente ao professor Guilherme Tavares que me acompanhou e me ensinou muito durante esse tempo.

Agradeço aos meus irmãos da república Molotov “amizade verdadeira, marca que ninguém remove”; e a todos os grandes amigos que fiz em Ouro Preto.

Resumo

Serviços de recomendação fornecem sugestões de itens a serem recomendados para um usuário. Com o surgimento das bibliotecas digitais e com a grande quantidade de objetos digitais armazenados em seus acervos, a tarefa de encontrar os objetos digitais mais relevantes a um determinado usuário pode se tornar bem árdua e ineficiente. Visando melhorar a satisfação de usuários ao utilizarem um biblioteca digital, é necessário o desenvolvimento de serviços que auxiliam na recuperação de objetos digitais que sejam de interesse dos mesmos. Assim, considerando a Biblioteca Digital de Mestrado Profissional em Ensino de Ciências (BDMPEC) da Universidade Federal de Ouro Preto, o objetivo deste trabalho consiste em propor, desenvolver e validar um serviço de recomendação para tal biblioteca, com o intuito de indicar produtos educacionais que sejam do interesse de usuários por meio de abordagens baseadas em conteúdo, colaborativa e híbrida. O serviço proposto e desenvolvido envolve três estratégias distintas de recomendação baseadas no XML do currículo Lattes de um usuário, em um produto educacional do acervo da BDMPEC, e em um texto livre fornecido pelo usuário. Os resultados obtidos foram satisfatórios e validaram o serviço de recomendação proposto e desenvolvido, chegando a alcançar 90% de eficácia.

Palavras-chave: Serviço de recomendação. Biblioteca digital. Modelo vetorial. Filtragem baseada em conteúdo. Filtragem colaborativa. Filtragem híbrida.

Abstract

Recommendation services provide items suggestions to be recommended to a user. With the emergence of the digital libraries, and lots of digital objects stored in its collections, to a user find the objects more applicable is very hard. Trying to improve the user satisfaction on the usage of a digital library, its necessary the development of assist services that assist on the recuperation of digital objects that are of the interest of the user. So, considering the Digital Library of Professional Master's Degree in Science Teach of Federal University of Ouro Preto, the objective of this work is propose, develop and validate a recommendation service for such library, in order to indicate educational products that are of the interest of the users through content based, collaborative and hybrid approaches. The proposed and developed service involve three distinct strategies of recommendation based on XML of a user's Lattes resume, on a educational product of the BDMPEC's collection, and on a free text provided by the user. The results obtained were satisfactory and validated the proposed and developed recommendation service, reaching 90% effectiveness.

Keywords: Recommendation service. Digital library. Vector model. Content based filtration. Collaborative filtration. Hybrid filtration.

Lista de Ilustrações

| | |
|---|----|
| Figura 2.1 – Arquitetura do serviço de auto-arquivamento proposto para a BDMPEC. | 7 |
| Figura 2.2 – Arquitetura básica de um serviço de recomendação. | 8 |
| Figura 2.3 – Funcionamento da FC. | 9 |
| Figura 2.4 – Funcionamento da FBC. | 10 |
| Figura 2.5 – Função do cosseno. | 11 |
| Figura 2.6 – Exemplo do modelo vetorial. | 12 |
| Figura 2.7 – Texto de exemplo para cálculo de palavras-chave. | 13 |
| Figura 2.8 – Palavras-chave candidatas analisadas a partir do texto de entrada. | 14 |
| Figura 2.9 – Matriz de co-ocorrências dos termos das palavras-chave candidatas. | 15 |
| Figura 2.10 – Termos das palavras-chave candidatas e suas pontuações. | 15 |
| Figura 2.11 – Comparação das palavras-chave extraídas pelo algoritmo RAKE com as palavras-chave manualmente atribuídas ao texto. | 16 |
| Figura 3.1 – Arquitetura geral do serviço de recomendação proposto. | 21 |
| Figura 3.2 – Arquitetura de leitura de informações para a estratégia baseada no XML do currículo Lattes. | 22 |
| Figura 3.3 – Arquitetura de leitura de informações para a estratégia baseada na seleção de um produto educacional. | 22 |
| Figura 3.4 – Arquitetura de leitura de informações para a estratégia baseada no texto livre. | 22 |
| Figura 3.5 – Página inicial do serviço de recomendação. | 25 |
| Figura 3.6 – Página com os produtos educacionais recomendados. | 26 |
| Figura 3.7 – Página de apresentação dos produtos educacionais consultados. | 26 |
| Figura 3.8 – Página de informações a respeito do produto educacional selecionado. | 27 |

Lista de Tabelas

| | |
|--|----|
| Tabela 2.1 – Mapeamento entre um Produto MPEC e o <i>Dublin Core</i> simples. | 5 |
| Tabela 2.2 – Mapeamento entre um Produto MPEC e o <i>Dublin Core</i> qualificado. | 6 |
| Tabela 3.1 – Similaridade entre os produtos educacionais e os dados fornecidos pelo usuário (vide Equação 2.1). | 24 |
| Tabela 3.2 – Similaridade entre os produtos educacionais e os dados fornecidos pelo usuário considerando o número de acessos (vide Equação 2.1). | 24 |
| Tabela 4.1 – Casos de teste: currículo Lattes http://lattes.cnpq.br/3216449144279833 . . . | 31 |
| Tabela 4.2 – Casos de teste: currículo Lattes http://lattes.cnpq.br/3537386569474804 . . . | 32 |
| Tabela 4.3 – Caso de teste: produto educacional 1. | 33 |
| Tabela 4.4 – Caso de teste: produto educacional 2. | 33 |
| Tabela 4.5 – Caso de teste: texto livre - resumo do artigo Pott e Estrela (2017). | 34 |
| Tabela 4.6 – Caso de teste: texto livre - resumo do artigo Bacci e Pataca (2008). | 34 |
| Tabela 4.7 – Casos de teste utilizando o número de acesso dos produtos. | 34 |

Lista de Abreviaturas e Siglas

| | |
|---------|---|
| BDMPEC | Biblioteca Digital de Mestrado Profissional em Ensino de Ciências |
| UFOP | Universidade Federal de Ouro Preto |
| MPEC | Mestrado Profissional em Ensino de Ciências |
| MP | Mestrado Profissional |
| FC | Filtragem Colaborativa |
| FBC | Filtragem Baseada em Conteúdo |
| RI | Recuperação de Informação |
| BDDECOM | Biblioteca Digital do Departamento de Computação |

Sumário

| | | |
|----------|---|-----------|
| 1 | Introdução | 1 |
| 1.1 | Justificativa | 2 |
| 1.2 | Objetivos Geral e Específicos | 2 |
| 1.3 | Organização do Trabalho | 3 |
| 2 | Revisão Bibliográfica | 4 |
| 2.1 | Fundamentação Teórica | 4 |
| 2.1.1 | Biblioteca Digital de Mestrado Profissional em Ensino de Ciências | 4 |
| 2.1.2 | Serviço de recomendação | 7 |
| 2.1.2.1 | Filtragem colaborativa | 8 |
| 2.1.2.2 | Filtragem baseada em conteúdo | 9 |
| 2.1.2.3 | Filtragem híbrida | 10 |
| 2.1.3 | Modelo Vetorial | 11 |
| 2.1.4 | Algoritmo RAKE | 13 |
| 2.2 | Trabalhos Relacionados | 16 |
| 3 | Serviço de Recomendação Proposto | 20 |
| 3.1 | Arquitetura de funcionamento | 20 |
| 3.1.1 | Leitura de informações | 21 |
| 3.1.2 | Extração de dados | 22 |
| 3.1.3 | Recomendação | 23 |
| 3.2 | Interface da BDMPEC | 25 |
| 4 | Experimentação Prática | 28 |
| 4.1 | Descrição dos experimentos | 28 |
| 4.2 | Análise dos resultados obtidos | 29 |
| 5 | Considerações Finais | 36 |
| 5.1 | Conclusões | 36 |
| 5.2 | Perspectivas de trabalho futuro | 36 |
| | Referências | 38 |

1 Introdução

Devido à consolidação da *Web* e à facilidade de acesso a informação nela contidas, pesquisas científicas são divulgadas rapidamente a nível mundial. Porém, como grande parte dessas pesquisas encontram-se espalhadas em diversos repositórios da *Web*, tornam-se necessários mecanismos que promovam, de uma forma melhor e mais eficiente, a organização e a facilidade de acesso. Para isso, como possíveis mecanismos, surgiram as bibliotecas digitais que, segundo [Andrade et al. \(2020\)](#), são sistemas eletrônicos de armazenamento de informações com capacidade para selecionar, estruturar, oferecer acesso intelectual, interpretar, distribuir, preservar a integridade e garantir a persistência, ao longo do tempo, de coleções de objetos digitais, para que estejam legíveis e disponíveis para uso por um conjunto de comunidades específicas. Ademais, bibliotecas digitais oferecem serviços que permitem aos seus usuários gerenciar e disseminar objetos digitais.

No Brasil, existem bibliotecas digitais que suprem a carência de bibliotecas em municípios e escolas brasileiras. É o caso do Portal Domínio Público¹, do Ministério da Educação, que disponibiliza gratuitamente quase 200 mil arquivos distribuídos em diversos tipos de mídias: sons, imagens, vídeos e textos. Entre eles, está a obra completa de Machado de Assis.

Em 2017, foi desenvolvida por [Andrade et al. \(2020\)](#) a BDMPEC da UFOP que se encontra no site <http://www.bdmpec.ufop.br/>. Tal biblioteca está voltada, em um primeiro momento, para a catalogação, a disponibilização e a pesquisa de produtos gerados por egressos do programa de Mestrado Profissional em Ensino de Ciências (MPEC) da UFOP, com potencial para ser estendida e utilizada por outros programas de Mestrado Profissional (MP) relativos ao ensino devido à similaridade das características dos produtos educacionais confeccionados por esses programas. Neste caso, o objeto digital tratado pela BDMPEC consiste no produto educacional final gerado por egressos do MPEC/UFOP. Por meio de uma experimentação realizada com distintos participantes, foi constatado que seus serviços são eficazes e de fácil utilização e aprendizagem: avaliação positiva da BDMPEC, quanto ao seu uso.

[Mathias \(2016\)](#) desenvolveu outros serviços para a BDMPEC, sendo eles: serviço de avaliação, por meio do qual usuários podem avaliar um produto cadastrado na biblioteca; serviço de auto-arquivamento, por meio do qual usuários podem cadastrar novos produtos na biblioteca, fornecendo os dados relativos e requisitando posteriormente o cadastro; e serviço de busca na *Web*, por meio do qual usuários podem utilizar dados de um produto cadastrado para realizar uma busca na *Web*. Com isso, complementando os serviços adicionais oferecidos pela BDMPEC, a fim de melhorar a experiência dos usuários ao utilizarem a biblioteca, oferecendo uma forma eficiente de encontrar produtos educacionais que sejam do interesse dos mesmos, este trabalho

¹ É uma biblioteca digital mantida pelo Ministério da Educação do Brasil, lançada em 2004. Em 2014, ultrapassou os 198 mil títulos em seu acervo: 182 mil em arquivos de texto e 15 mil em outras mídias.

consiste no desenvolvimento e na validação de um serviço de recomendação para a BDMPEC.

1.1 Justificativa

Atualmente, para um usuário da BDMPEC encontrar um produto que seja de seu interesse, pode ser necessário a realização de várias buscas utilizando vários termos de pesquisa diferentes, já que a única forma de acesso aos produtos educacionais é o serviço de acesso ao acervo.

De acordo com Ricci, Rokach e Shapira (2010), um serviço de recomendação combina várias técnicas computacionais para selecionar itens personalizados com base nos interesses dos usuários e conforme o contexto no qual estão inseridos. Esses itens podem assumir formas variadas como, por exemplo, livros, filmes, notícias, músicas, vídeos e anúncios. No caso da BDMPEC, esses itens são os produtos educacionais arquivados na biblioteca.

Para breve conhecimento, um dos mais personalizados serviços de recomendação existente é utilizado pela empresa Netflix. Ela adaptou seus algoritmos de recomendação de uma maneira que, segundo Andrade (2017), 75% do conteúdo assistido no ano de 2012 foi proveniente de algum tipo de recomendação. O serviço de recomendação proposto neste trabalho busca melhorar a forma como um usuário pode encontrar produtos MPEC desejados na BDMPEC.

1.2 Objetivos Geral e Específicos

Esse trabalho possui, como objetivo geral, propor, desenvolver e validar um serviço de recomendação para a BDMPEC, com o intuito de indicar produtos educacionais cadastrados no repositório da biblioteca que sejam de interesse de seus usuários, filtrando aqueles relativos às necessidades dos mesmos.

As estratégias de recomendação propostas no serviço podem ocorrer por meio: (a) da seleção de um produto educacional já cadastrado na biblioteca, no intuito de serem recomendados outros produtos similares ao selecionado; (b) da indicação do currículo Lattes de um determinado usuário por meio do qual informações são usadas para realizar a recomendação; e (c) do fornecimento de um texto livre, ou seja, termos a serem usados em uma recomendação. Os objetos digitais da BDMPEC, que serão retornados pelo serviço de recomendação, correspondem a produtos educacionais gerados por egressos do MPEC da UFOP que se encontram no repositório da biblioteca. Desta forma, para se processar uma recomendação, quanto ao lado da BDMPEC, cada produto educacional armazenado será representado por um conjunto de termos importantes, obtidos a partir de seus dados, que o caracterizem apropriadamente.

Alguns objetivos específicos, alcançados neste trabalho, são:

- melhorar a satisfação dos usuários ao utilizarem a BDMPEC;

- reduzir o tempo que um usuário gasta para encontrar um produto de seu interesse;
- consolidar as distintas estratégias de recomendação propostas neste trabalho, como parte do serviço proposto, por meio da análise e comparação de resultados experimentais.

1.3 Organização do Trabalho

Os próximos capítulos deste trabalho organizam-se da seguinte forma: no Capítulo 2, é realizada a revisão bibliográfica, que consiste no referencial teórico, para dar base ao entendimento do trabalho, e nos trabalhos relacionados que apresentam soluções já desenvolvidas semelhantes à solução proposta neste trabalho; no Capítulo 3, é apresentada a metodologia deste trabalho, envolvendo a arquitetura de funcionamento e a interface do serviço de recomendação proposto; no Capítulo 4, são apresentados os resultados experimentais, quanto à utilização do serviço de recomendação desenvolvido, e as discussões relativas aos mesmos; por fim, no Capítulo 5, são descritas as conclusões e as perspectivas de trabalho futuro.

2 Revisão Bibliográfica

Conforme já mencionado, este trabalho possui, como objetivo geral, a proposta, o desenvolvimento e a validação de um serviço de recomendação para a BDMPEC. Sendo assim, este Capítulo apresenta a revisão bibliográfica, para dar suporte ao alcance desse objetivo. Na Seção 2.1, é apresentada a fundamentação teórica necessária à realização deste trabalho e, na Seção 2.2, são apresentados alguns trabalhos relacionados.

2.1 Fundamentação Teórica

Nesta seção, é apresentada a fundamentação teórica de suporte ao desenvolvimento adequado deste trabalho. Encontra-se estruturada da seguinte forma: na Subseção 2.1.1, é apresentada a BDMPEC, contexto da aplicação do serviço de recomendação proposto; na Subseção 2.1.2, é descrito, de uma forma genérica, serviço de recomendação, contemplando algumas abordagens que podem ser utilizadas; na Subseção 2.1.3, é apresentado o modelo vetorial, modelo da área de Recuperação de Informação utilizado pelo serviço de recomendação proposto; e, por fim, na Subseção 2.1.4, é apresentado o algoritmo RAKE, utilizado na extração de palavras-chave de textos, também utilizado pelo serviço de recomendação proposto.

2.1.1 Biblioteca Digital de Mestrado Profissional em Ensino de Ciências

De acordo com [Andrade et al. \(2020\)](#), programas de Mestrado Profissional em Ensino devem atuar na capacitação de profissionais docentes no intuito de que seus produtos gerados sejam desenvolvidos e empregados da melhor forma possível; além disso, devem também propor soluções para as demandas sociais, políticas e econômicas associadas ao cenário educacional brasileiro. Neste contexto, o egresso de programas de Mestrado Profissional em Ensino deve elaborar produtos focados nessas soluções, mais especificamente, ao Ensino de Ciências na Educação Básica ou Superior.

Desenvolvida em 2017 por [Andrade et al. \(2020\)](#), a BDMPEC/UFOP apresenta o objetivo de catalogar, disponibilizar e permitir avaliação de produtos gerados por egressos do programa de MPEC da UFOP. Além disso, a BDMPEC oferece alguns serviços adicionais, desenvolvidos por [Mathias \(2016\)](#), sendo eles: serviço de auto-arquivamento, que possibilita que usuários arquivem novos produtos educacionais na BDMPEC; e serviço de busca na *Web*, que permite a localização na *Web* de documentos e páginas relacionados aos produtos educacionais catalogados na BDMPEC.

A fim de permitir a interoperabilidade com outras bibliotecas digitais, ou seja, garantir a capacidade das bibliotecas digitais trocarem e compartilharem produtos e serviços, a BDMPEC

segue o padrão *Dublin Core* que tem, como objetivo, descrever, de forma eficaz, os metadados de um produto catalogado (objeto digital). Para se obter os metadados relativos a um determinado produto catalogado em seu repositório, o banco de dados da BDMPEC armazena:

- os temas abordados pelo produto MPEC;
- a área de concentração a qual o produto MPEC pertence;
- o nível de ensino associado ao produto MPEC;
- as palavras-chave relacionadas ao produto MPEC;
- informações como título, data da defesa, resumo e *link* de acesso, relacionadas a dissertação do produto MPEC;
- o autor, o orientador e o co-orientador do produto MPEC;
- a instituição de ensino;
- a categoria associada ao produto MPEC, podendo ser apostila, apresentação digital, cartilha, curso, hipertexto, jogo, modelo educacional, livro paradidático, hipermídia, objeto de ensino, oficina, página *web*, palestra, proposta de disciplina, protótipo, sequência didática.

Com base em tais informações, foram definidos os metadados simples e qualificados apresentados, respectivamente, nas Tabelas 2.1 e 2.2.

Tabela 2.1 – Mapeamento entre um Produto MPEC e o *Dublin Core* simples.

| <i>Dublin Core</i> Simples | Produto do MPEC |
|-----------------------------------|----------------------------------|
| Title | Título do Produto |
| Creator | Autor do Produto |
| Subject | Palavra chave |
| Description | Descrição |
| Date | Data da defesa |
| Contributor | Orientador |
| Identifier | Identificador único do produto |
| Relation | Áreas de concentração do produto |
| Source | URL de acesso ao produto |

Fonte: Andrade et al. (2020)

Tabela 2.2 – Mapeamento entre um Produto MPEC e o *Dublin Core* qualificado.

| Dublin Core Qualificado | Produto do MPEC |
|--------------------------------|--------------------------------|
| coContributor | Co-orientador |
| Série_de_Destino | Categoria do produto elaborado |
| Dissertação | Título da dissertação |
| Nível_de_Ensino | Nível de ensino |

Fonte: [Andrade et al. \(2020\)](#)

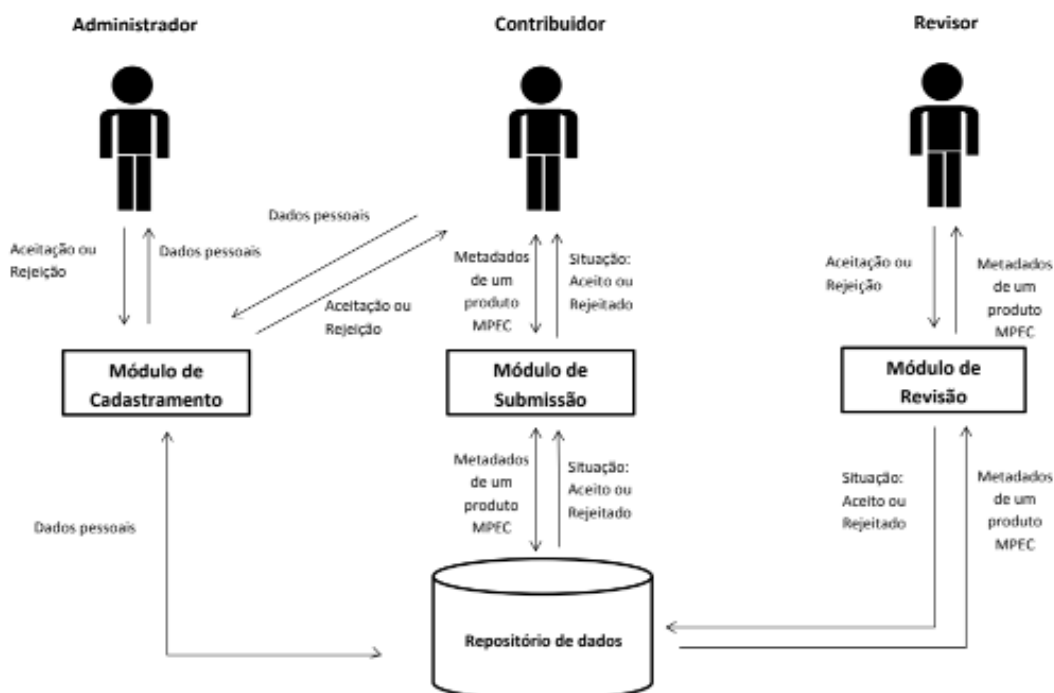
Todos os metadados que descrevem os produtos catalogados no banco de dados da BDMPEC são extremamente importantes para a proposta e a utilização do serviço de recomendação a ser desenvolvido neste trabalho.

Além disso, a BDMPEC possui quatro tipos de usuários, sendo eles:

- administrador: que administra a BDMPEC, tendo acesso a todos os serviços da mesma, além de ser o responsável em permitir o cadastramento dos demais tipos de usuário;
- revisor: que pode fazer a revisão de produtos submetidos na BDMPEC por meio do serviço de revisão;
- avaliador: que pode fazer a avaliação de produtos armazenados na BDMPEC por meio do serviço de avaliação;
- contribuidor: que submete produtos a serem adicionados no repositório da BDMPEC por meio do serviço de auto-arquivamento.

No intuito de melhor compreender as principais ações dos tipos de usuário administrador, revisor e contribuidor, a Figura 2.1 apresenta a arquitetura do serviço de auto-arquivamento proposto para a BDMPEC. Tal serviço, como já mencionado, permite usuários contribuidores arquivarem novos produtos educacionais na BDMPEC. Esse auto-arquivamento é importante pois gerará os valores descritivos de um determinado novo produto, que serão utilizados pelo serviço de recomendação proposto neste trabalho.

Figura 2.1 – Arquitetura do serviço de auto-arquivamento proposto para a BDMPEC.



Fonte: Mathias (2016)

De acordo com a Figura 2.1, para um usuário se tornar usuário contribuidor ele deve submeter seus dados pessoais, que serão salvos no repositório de dados e disponibilizados para o administrador aceitá-lo ou não como contribuidor; todo o controle dessa parte é feita pelo módulo de cadastramento. Caso seja aceito como usuário contribuidor, será possível submeter produtos MPEC; para isso, o usuário colaborador deve submeter os metadados do produto MPEC a ser catalogado, que serão analisados pelo usuário revisor; que, por sua vez, pode aceitar ou rejeitar o produto MPEC cadastrado. Toda a parte de submissão do produto é gerenciada pelo módulo de submissão, e toda a parte de revisão do produto é gerenciada pelo módulo de revisão. O repositório de dados corresponde ao repositório que armazena todas as informações da BDMPEC; no caso do serviço de auto-arquivamento, esse repositório armazena os dados dos usuários (administrador, contribuidor e revisor), e os produtos cadastrados pelo usuário contribuidor.

2.1.2 Serviço de recomendação

Segundo ONU News (2017), o uso da internet continua a crescer globalmente. Atualmente, mais de 4,1 bilhões de pessoas utilizam a rede mundial. Diante disso, cada vez mais dados são gerados na internet, o que leva aos usuários a uma exposição maior de dados cada vez mais variados.

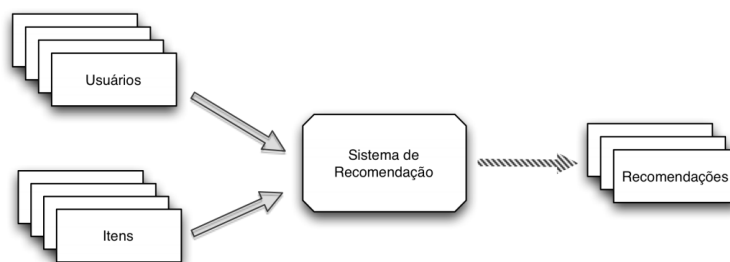
Serviços de recomendação estão diretamente ligados ao grande volume de dados disponíveis na internet, pois um dos seus principais objetivos é filtrar os dados disponíveis para entregar

aqueles que sejam de interesse do usuário, facilitando, assim, o acesso aos dados desejados.

Atualmente, de acordo com Novoseltseva (2018), 35% das vendas realizadas pela Amazon são feitas devido ao serviço de recomendação usado pela empresa, o que indica a grande importância desse serviço para o mercado de *e-commerce*.

A Figura 2.2 exemplifica a arquitetura básica de um serviço de recomendação.

Figura 2.2 – Arquitetura básica de um serviço de recomendação.



Fonte: Retirada da Web¹

De acordo com a Figura 2.2, o serviço de recomendação ocorre da seguinte forma: o componente Serviço de Recomendação coleta dados dos Usuário e dos Itens a serem recomendados; para exemplificação, no caso da BMPEC, os itens correspondem a produtos MPEC catalogados na biblioteca e, no caso da Amazon, são os produtos vendidos por ela: cada serviço lida com um tipo de item. Após a coleta dos dados dos Usuários e dos Itens pelo Serviço de Recomendação, são aplicadas técnicas de recomendação para que as Recomendações geradas, ao final, sejam as mais próximas possíveis do que os Usuários desejam.

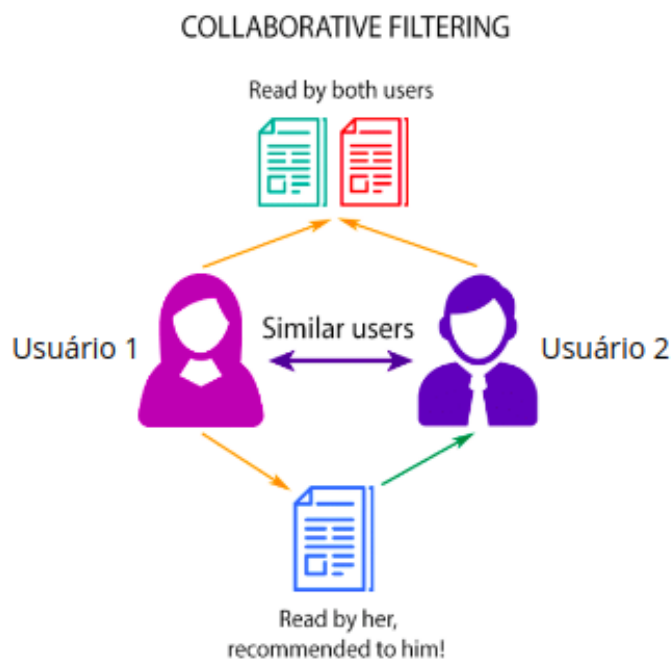
As principais técnicas de recomendação em que os serviços de recomendação baseiam-se são: filtragem colaborativa, filtragem baseada em conteúdo e filtragem híbrida. Tais técnicas encontram-se descritas nas subseções seguintes. As técnicas de recomendação representam formas de obter os dados do usuário-alvo e dos documentos que serão recomendados. Para gerar os valores de similaridade entre os dados, é necessário utilizar algum modelo de cálculo de similaridade como, por exemplo, o modelo vetorial.

2.1.2.1 Filtragem colaborativa

Segundo ADOMAVICIUS e TUZHILIN (2005), a técnica de Filtragem Colaborativa (FC) desempenha um papel muito significativo no processo de recomendação e, por isso, é a técnica mais utilizada. Nesta técnica, a recomendação é feita de acordo com as preferências de usuários que sejam similares ao usuário-alvo Ortega et al. (2013). Para melhor compreensão, a Figura 2.3 exemplifica o funcionamento dessa técnica.

¹ Disponível em: <https://miro.medium.com/max/1400/1*kgBD8OuP7SMsL63gcsQ90Q.png>. Acesso em: 02 ago. 2021.

Figura 2.3 – Funcionamento da FC.



Fonte: Retirada da Web ²

De acordo com a Figura 2.3, considerando os usuários como leitores de artigos e os documentos a ser recomendados a eles sendo artigos, a técnica de FC encontrou uma similaridade entre dois usuários devido a artigos igualmente lidos por eles; com isso, a técnica recomendou ao usuário 1 um artigo lido pelo usuário 2, já que os dois usuários foram considerados similares e tal artigo não foi lido pelo usuário 1.

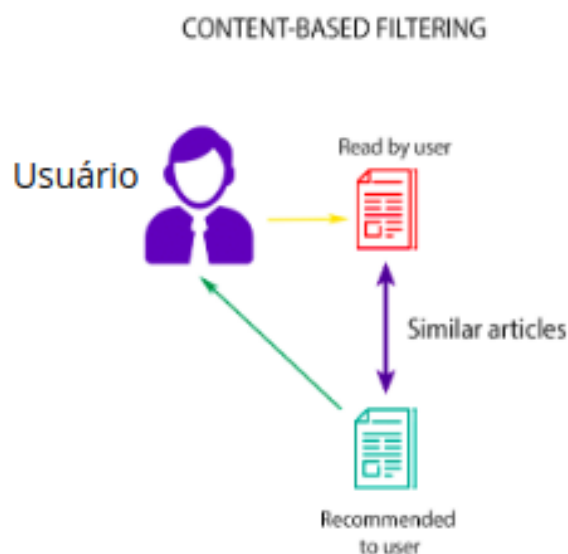
Neste trabalho, a FC é utilizada considerando o número de acessos a um produto, como pode ser visto na Subseção 3.1.3, na composição do valor de recomendação do produto em questão.

2.1.2.2 Filtragem baseada em conteúdo

Segundo LU et al. (2012), a técnica de Filtragem Baseada em Conteúdo (FBC) é dependente das escolhas dos usuários; desta forma, nessa técnica, a recomendação é feita de acordo com escolhas anteriormente utilizadas pelo usuário-alvo. FBC permite que o serviço de recomendação obtenha as preferências do usuário-alvo com maior precisão, retornando, assim, recomendações mais autênticas. A Figura 2.4 representa o funcionamento da FBC por meio de um exemplo.

² Disponível em: <<https://medium.com/data-hackers/deep-learning-para-sistemas-de-recomenda%C3%A7%C3%A3o-parte-1-introdu%C3%A7%C3%A3o-b19a896c471e>>. Acesso em: 02 ago. 2021.

Figura 2.4 – Funcionamento da FBC.



Fonte: Retirada da Web ³

De acordo com a Figura 2.4, considerando os usuários como leitores de artigos e os documentos a serem recomendados a eles sendo artigos, a partir de documentos lidos pelo usuário-alvo, a técnica FBC permite encontrar outros similares a eles, e recomendá-los ao usuário-alvo.

Neste trabalho, a FBC é utilizada em todas as estratégias propostas de recomendação, (vide Capítulo 3), a saber: recomendação utilizando os dados do currículo Lattes de um usuário-alvo, recomendação utilizando os dados de um produto MPEC selecionado pelo usuário-alvo, e recomendação baseada no texto livre fornecido pelo usuário-alvo.

2.1.2.3 Filtragem híbrida

Segundo NJ et al. (2012), a técnica de filtragem híbrida consiste na combinação das técnicas de FC e FBC, unindo os resultados obtidos pelas duas. Seu principal objetivo é gerar um resultado final melhor do que os resultados obtidos pelas duas filtragens separadamente.

Para realizar a filtragem híbrida, deve-se definir uma forma de junção para os resultados obtidos separadamente por cada filtragem; neste trabalho, a filtragem híbrida consiste, para cada produto recomendado em um processo de recomendação, no cálculo da média ponderada entre o número de acessos de um produto e o seu próprio valor de similaridade de recomendação.

³ Disponível em: <<https://medium.com/data-hackers/deep-learning-para-sistemas-de-recomenda%C3%A7%C3%A3o-parte-1-introdu%C3%A7%C3%A3o-b19a896c471e>>. Acesso em: 02 ago. 2021.

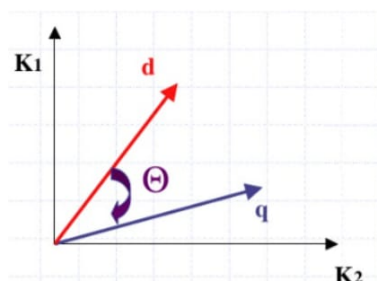
2.1.3 Modelo Vetorial

Segundo Baeza-Yates, Ribeiro-Neto et al. (1999), a área de Recuperação de Informação (RI) lida com a representação, o armazenamento, a organização e o acesso a documentos, que geralmente são textos ou parte de textos, com o objetivo de recuperar informações que possam ser úteis ou relevantes para o usuário. A representação e a organização dos documentos devem fornecer ao usuário fácil acesso à informação que é de seu interesse.

Atualmente, é possível observar a RI em mecanismos de pesquisa, como a máquina de busca do *Google* já que, nesse caso, o usuário não precisa de muitas informações específicas a respeito dos documentos que ele gostaria de acessar, bastando fazer uma pesquisa com a especificação de alguns termos. Feita a pesquisa, o mecanismo de RI atua para selecionar, dentre os documentos presentes em um repositório, aqueles que representam, qualitativamente, a pesquisa do usuário.

Existem vários modelos que tratam a recuperação de documentos por meio de consultas; dentre eles, está o modelo vetorial. De acordo com Baeza-Yates, Ribeiro-Neto et al. (1999), o Modelo Vetorial corresponde a um mecanismo no qual é possível recuperar documentos de um determinado repositório que respondem, similar e parcialmente, a uma expressão de busca. Para isso, o modelo vetorial calcula a similaridade dos documentos do repositório com a expressão de busca fornecida. O cálculo da similaridade é realizado utilizando a função do cosseno, representada pela Figura 2.5, onde, considerando o contexto de um serviço de recomendação, d corresponde aos termos pertencentes aos documentos candidatos à recomendação e q representa os termos pertencentes ao usuário-alvo cuja recomendação será feita. No exemplo apresentado, o espaço é bidimensional, apresentando apenas os termos K_1 e K_2 . A representação computacional dos termos é realizada por meio de vetores de termos. Com isso, d é um vetor de termos que melhor representa os documentos candidatos à recomendação, e q é um vetor de termos que melhor representa o usuário-alvo.

Figura 2.5 – Função do cosseno.



Fonte: Baeza-Yates, Ribeiro-Neto et al. (1999)

De acordo com a Figura 2.5, o cosseno do ângulo formado entre os vetores d e q determina a proximidade da ocorrência. O cálculo da similaridade, Equação 2.1, é baseado no ângulo entre os vetores d , que representa os dados do documento a ser recomendado, e q , que representa os

dados do usuário-alvo. O valor obtido no cálculo se encontra entre 0 e 1; quanto mais similar, mais próximo de 1.

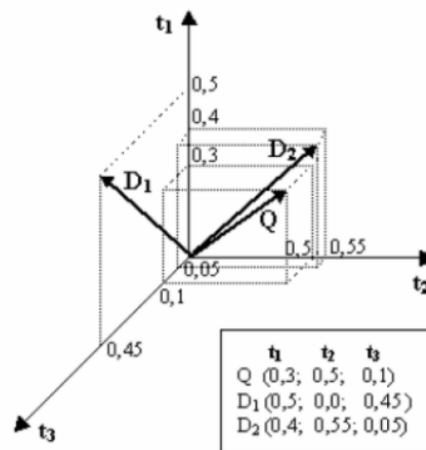
$$\text{Similaridade}(Q, D) = \frac{\sum_{k=1}^t W_{qk} \times W_{dk}}{\sqrt{\sum_{k=1}^t (W_{qd})^2 \times \sum_{k=1}^t (W_{jk, q})^2}} \quad (2.1)$$

Onde,

- Q representa os termos de consulta;
- D representa os termos do documento;
- W_{qk} representa o peso de relevância do termo k na consulta Q ; e
- W_{dk} representa o peso de relevância do termo k no documento D .

Para ilustrar a aplicação do modelo vetorial, a Figura 2.6 ilustra um espaço vetorial de três dimensões, representado pelos eixos t_1 , t_2 e t_3 . Os vetores de termos de dois documentos são representados por D_1 e D_2 e o vetor de termos dos dados do usuário-alvo por Q . Os valores numéricos, que representam os pesos relativos aos termos nos documentos e na consulta, também são mostrados.

Figura 2.6 – Exemplo do modelo vetorial.



Fonte: Santos (2016)

Considerando a Figura 2.6, aplicando o cálculo da similaridade definido pela Equação 2.1, tem-se os seguintes valores:

- Similaridade (Q, D_1) = 0,4899

- Similaridade (Q, D2) = 0,9915

Com isso, pode-se afirmar que a similaridade de $D2$ em relação à Q foi maior que a similaridade de $D1$ em relação à Q . Assim, considerando o contexto desse trabalho, com o resultado do modelo vetorial, é possível montar uma lista ordenada decendentemente pela similaridade dos documentos candidatos à recomendação e, assim, recomendar os que mais possuem similaridade com o vetor que representa o que se deseja relativo ao usuário-alvo para o qual a recomendação é apresentada.

Neste trabalho, todas as estratégias distintas do serviço proposto de recomendação utilizam o modelo vetorial para o cálculo da similaridade entre os dados do usuário-alvo e dos produtos MPEC a serem recomendados. Entende-se dados do usuário-alvo como as informações fornecidas por ele, dependente da estratégia utilizada, para que a recomendação seja realizada.

2.1.4 Algoritmo RAKE

Atualmente, existem vários algoritmos para extração de palavras-chave de textos; porém, a maioria aborda apenas um idioma, pois normalmente esses algoritmos utilizam conceitos para identificar e extrair as palavras-chave dos textos. O algoritmo RAKE é um algoritmo de extração de palavras-chave que suporta vários idiomas, dentre eles, o português.

De acordo com [Rose et al. \(2010\)](#), o algoritmo RAKE é baseado na observação de que as palavras-chave frequentemente possuem múltiplas palavras e raramente possuem pontuação padrão ou palavras de parada, como as palavras funcionais *e*, *o*, e *de*, entre outras com significado lexical mínimo. Os parâmetros de entrada do algoritmo RAKE incluem um texto que será analisado, uma lista de palavras de parada, um conjunto de delimitadores de frases e um conjunto de delimitadores de palavras. Para melhor explicação do algoritmo, será utilizado o texto apresentado na Figura 2.7. A Figura 2.7 apresenta também uma lista de palavras-chave manualmente selecionadas para comparação com os resultados gerados pelo algoritmo.

Figura 2.7 – Texto de exemplo para cálculo de palavras-chave.

Compatibility of systems of linear constraints over the set of natural numbers

Criteria of compatibility of a system of linear Diophantine equations, strict inequations, and nonstrict inequations are considered. Upper bounds for components of a minimal set of solutions and algorithms of construction of minimal generating sets of solutions for all types of systems are given. These criteria and the corresponding algorithms for constructing a minimal supporting set of solutions can be used in solving all the considered types of systems and systems of mixed types.

Manually assigned keywords:
linear constraints, set of natural numbers, linear Diophantine equations, strict inequations, nonstrict inequations, upper bounds, minimal generating sets

Fonte: [Rose et al. \(2010\)](#)

O primeiro passo do algoritmo RAKE consiste em separar o texto em um conjunto de palavras-chave candidatas, que são sequências de palavras de conteúdo (cujo significado léxico não seja mínimo) da forma que elas ocorrem no texto, vide Figura 2.8. Para isso, são usadas as palavras de paradas e os delimitadores de frases fornecidos na entrada de dados. Para o texto de exemplo apresentado na Figura 2.7, as palavras de parada foram: *of; a; and; are; for; all; these; the; can; be; e in*, e os delimitadores de frase foram as pontuações ponto e vírgula.

Figura 2.8 – Palavras-chave candidatas analisadas a partir do texto de entrada.

Compatibility – systems – linear constraints – set – natural numbers – Criteria – compatibility – system – linear Diophantine equations – strict inequations – nonstrict inequations – Upper bounds – components – minimal set – solutions – algorithms – minimal generating sets – solutions – systems – criteria – corresponding algorithms – constructing – minimal supporting set – solving – systems – systems

Fonte: Rose et al. (2010)

O segundo passo do algoritmo RAKE consiste em gerar a matriz de co-ocorrências (Figura 2.9) que associa cada termo de uma palavra-chave candidata com outros termos de uma palavra-chave candidata (vide Figura 2.10). Quando a matriz de co-ocorrência está completa, é calculado uma pontuação para cada termo de uma palavra-chave candidata. Para calcular a pontuação de cada termo, são utilizadas as seguintes métricas: (1) frequência do termo ($freq(w)^4$), (2) grau do termo ($deg(w)^5$), e (3) proporção de grau por frequência ($deg(w)/freq(w)$).

⁴ $freq(w)$ consiste no número de vezes que o termo w aparece no texto.

⁵ $deg(w)$ consiste na soma do número de vezes que o termo w aparece no texto junto com outro termo qualquer.

Figura 2.9 – Matriz de co-ocorrências dos termos das palavras-chave candidatas.

| | algorithms | bounds | compatibility | components | constraints | constructing | corresponding | criteria | diophantine | equations | generating | inequations | linear | minimal | natural | nonstrict | numbers | set | sets | solving | strict | supporting | system | systems | upper |
|---------------|------------|--------|---------------|------------|-------------|--------------|---------------|----------|-------------|-----------|------------|-------------|--------|---------|---------|-----------|---------|-----|------|---------|--------|------------|--------|---------|-------|
| algorithms | 2 | | | | | | 1 | | | | | | | | | | | | | | | | | | |
| bounds | | 1 | | | | | | | | | | | | | | | | | | | | | | | 1 |
| compatibility | | | 2 | | | | | | | | | | | | | | | | | | | | | | |
| components | | | | 1 | | | | | | | | | | | | | | | | | | | | | |
| constraints | | | | | 1 | | | | | | | 1 | | | | | | | | | | | | | |
| constructing | | | | | | 1 | | | | | | | | | | | | | | | | | | | |
| corresponding | 1 | | | | | | 1 | | | | | | | | | | | | | | | | | | |
| criteria | | | | | | | | 2 | | | | | | | | | | | | | | | | | |
| diophantine | | | | | | | | | 1 | 1 | | | 1 | | | | | | | | | | | | |
| equations | | | | | | | | | 1 | 1 | | | 1 | | | | | | | | | | | | |
| generating | | | | | | | | | | | 1 | | 1 | | | | | | 1 | | | | | | |
| inequations | | | | | | | | | | | | 2 | | | | 1 | | | | | 1 | | | | |
| linear | | | | | 1 | | | | 1 | 1 | | 2 | | | | | | | | | | | | | |
| minimal | | | | | | | | | | | 1 | | 3 | | | | | 2 | 1 | | | 1 | | | |
| natural | | | | | | | | | | | | | | | 1 | | 1 | | | | | | | | |
| nonstrict | | | | | | | | | | | | 1 | | | | 1 | | | | | | | | | |
| numbers | | | | | | | | | | | | | | | | 1 | 1 | | | | | | | | |
| set | | | | | | | | | | | | | 2 | | | | | 3 | | | | | 1 | | |
| sets | | | | | | | | | | 1 | | | 1 | | | | | | 1 | | | | | | |
| solving | | | | | | | | | | | | | | | | | | | | 1 | | | | | |
| strict | | | | | | | | | | | | 1 | | | | | | | | | 1 | | | | |
| supporting | | | | | | | | | | | | | | 1 | | | | | | | | 1 | | | |
| system | | | | | | | | | | | | | | | | | | | | | | | 1 | | |
| systems | | | | | | | | | | | | | | | | | | | | | | | | 4 | |
| upper | 1 | | | | | | | | | | | | | | | | | | | | | | | | 1 |

Fonte: Rose et al. (2010)

Figura 2.10 – Termos das palavras-chave candidatas e suas pontuações.

| | algorithms | bounds | compatibility | components | constraints | constructing | corresponding | criteria | diophantine | equations | generating | inequations | linear | minimal | natural | nonstrict | numbers | set | sets | solving | strict | supporting | system | systems | upper |
|------------------|------------|--------|---------------|------------|-------------|--------------|---------------|----------|-------------|-----------|------------|-------------|--------|---------|---------|-----------|---------|-----|------|---------|--------|------------|--------|---------|-------|
| deg(w) | 3 | 2 | 2 | 1 | 2 | 1 | 2 | 2 | 3 | 3 | 3 | 4 | 5 | 8 | 2 | 2 | 2 | 6 | 3 | 1 | 2 | 3 | 1 | 4 | 2 |
| freq(w) | 2 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 2 | 3 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 1 | 1 | 4 | 1 |
| deg(w) / freq(w) | 1.5 | 2 | 1 | 1 | 2 | 1 | 2 | 1 | 3 | 3 | 3 | 2 | 2.5 | 2.7 | 2 | 2 | 2 | 2 | 3 | 1 | 2 | 3 | 1 | 1 | 2 |

Fonte: Rose et al. (2010)

O terceiro passo do algoritmo RAKE consiste em procurar termos de palavras-chave adjacentes. Já que o RAKE é muito interessante por sua habilidade em escolher terminologia altamente específica, faz-se necessário também a identificação de palavras-chave que possuem palavras de parada como "serviço de recomendação". Para encontrar essas palavras-chave, o RAKE encontra pares de termos que se juntam ao menos duas vezes no mesmo texto e na mesma ordem. Com isso, uma nova palavra-chave candidata é criada com a combinação desses termos e a palavra de parada interior.

Para calcular a pontuação das palavras-chave, basta somar os valores de $deg(w)/freq(w)$ de cada termo w pertencente à palavra-chave; por exemplo, a palavra-chave *minimal generating*

sets tem a seguinte soma $2.7 + 3.0 + 3.0 = 8.7$. Após calculada a pontuação das palavras-chave candidatas, as T palavras-chave com maior pontuação são selecionadas como palavras-chave do texto. T consiste em um terço do número de palavras na matriz. Os resultados do exemplo e a comparação entre as palavras-chave manualmente atribuídas com as palavras chaves encontradas pelo algoritmo são apresentados na Figura 2.11.

Figura 2.11 – Comparação das palavras-chave extraídas pelo algoritmo RAKE com as palavras-chave manualmente atribuídas ao texto.

| Extracted by RAKE | Manually assigned |
|------------------------------|------------------------------|
| minimal generating sets | minimal generating sets |
| linear diophantine equations | linear Diophantine equations |
| minimal supporting set | |
| minimal set | |
| linear constraints | linear constraints |
| natural numbers | |
| strict inequations | strict inequations |
| nonstrict inequations | nonstrict inequations |
| upper bounds | upper bounds |
| | set of natural numbers |

Fonte: Rose et al. (2010)

A partir da Figura 2.11, é possível constatar que existem, portanto, três falsos positivos no conjunto de palavras-chave extraídas, resultando em uma precisão de 67%. Comparar os seis verdadeiros positivos dentro do conjunto de palavras-chave extraídas com o total de sete palavras-chave atribuídas manualmente resulta em uma revocação de 86%. Ponderando igualmente a precisão e a revocação, gera-se uma media harmônica de 75%.

Neste trabalho, a fim de obter uma melhor precisão dos dados utilizados no serviço de recomendação proposto, tanto do usuário-alvo quanto do produto MPEC, o algoritmo RAKE é utilizado para encontrar as palavras-chave em dados que utilizam textos muito grandes e que, possivelmente, possuem muitas informações desnecessárias para a recomendação. Exemplos destes dados são: resumo da dissertação de um produto MPEC e texto livre fornecido por um usuário a ser utilizado na recomendação.

2.2 Trabalhos Relacionados

Nesta seção, são apresentados alguns trabalhos relacionados ao desenvolvimento do serviço de recomendação para a BDMPEC, discutindo semelhanças e diferenças.

No trabalho de [NJ et al. \(2012\)](#), foi desenvolvido um serviço de recomendação híbrido para bibliotecas digitais que tem, como objetivo principal, a recomendação de artigos científicos, considerando informações obtidas dos currículos Lattes de usuários e avaliações dos usuários sobre um conjunto de artigos. Para isso, um determinado usuário deve se cadastrar no serviço e submeter um arquivo XML correspondente ao seu currículo Lattes; após isso, são enviadas requisições OAI-PMH a um ou vários provedores de dados cadastrados no serviço, tentando colher metadados dos documentos digitais específicos de cada biblioteca digital considerada. Há um módulo de avaliação que fornece avaliações na escala de 1-5 aos artigos das áreas de interesse do usuário, cadastradas no Lattes. O serviço também possui uma base de dados, denominada dados históricos, a qual armazena informações sobre itens consultados ou adquiridos pelos usuários, evitando-se assim o envio de artigos que o mesmo já tenha visto ou avaliado. Por fim, já com todas informações obtidas, a recomendação é acionada e ocorrerá.

O serviço de recomendação desenvolvido por [NJ et al. \(2012\)](#) possui a função de recomendar artigos baseado no perfil de cada um deles. Para isso, ele possui um algoritmo misto que implementa os métodos de filtragem colaborativa e baseada em conteúdo separadamente e combina as previsões geradas por ambos. A lista final de recomendação é gerada da seguinte forma: cada item presente em ambas as listas é adicionado à lista final com uma pontuação. Essa pontuação é a soma das posições desses itens nas listas de recomendações inicialmente geradas. Com isso, a lista é ordenada de forma crescente, onde quanto menor a pontuação de um artigo, mais próximo do início ele está. Além do algoritmo misto, foi desenvolvido também um algoritmo ponderado, onde o peso de um determinado item recomendado é computado dos resultados de todas as técnicas disponíveis de recomendação presentes no serviço; para isso, é feita a média ponderada de todos os resultados obtidos em cada técnica. Os resultados obtidos no trabalho de [NJ et al. \(2012\)](#) mostraram que a abordagem híbrida melhorou tanto os resultados obtidos pela abordagem colaborativa quanto pela abordagem baseada em conteúdo.

A Biblioteca Digital do Departamento de Computação da Universidade Federal de Ouro Preto (BDDECOM), de acordo com [Santos \(2016\)](#), possui o mesmo objetivo que a BDMPEC, porém focada em armazenar monografias apresentadas por alunos do curso de Ciência da Computação da UFOP. [Santos \(2016\)](#) desenvolveu, em sua monografia, um serviço de recomendação para a mesma; nesse serviço, é utilizado o currículo Lattes de um usuário da BDDECOM para coletar os dados do mesmo e, por meio do modelo vetorial, é calculada a similaridade entre cada monografia armazenada no repositório da BDDECOM e os dados coletados do usuário, para que o serviço recomende as mais bem ranqueadas. O serviço de recomendação desenvolvido fornece a possibilidade de o usuário selecionar quais dados ele deseja que sejam usados na recomendação, sendo eles: para as monografias da BDDECOM, título, descrição e palavras-chave e, para o currículo Lattes do usuário, linhas de pesquisa, áreas de atuação e palavras-chave dos trabalhos de formação acadêmica. Por meio dos experimentos realizados, os melhores resultados foram encontrados ao se combinar todos os dados do currículo Lattes com todos os dados das monografias em um processo de recomendação, alcançando uma média de 0,8942 de F1: métrica

que consiste na média harmônica entre as métricas precisão e revocação.

Furlan et al. (2018) desenvolveram um serviço para recomendar artigos acadêmicos disponíveis no Google Acadêmico para refinar uma pesquisa sobre artigos. Para o desenvolvimento desse sistema, foram necessárias três etapas: (1) coleta de dados do currículo Lattes do usuário, (2) uso do Google Acadêmico como fonte de artigos acadêmicos para o usuário, e (3) aplicação de técnicas de recomendação colaborativa e baseada em conteúdo para exibir os resultados. Em um determinado processo de recomendação, na primeira etapa, são extraídos dados do currículo Lattes do usuário, que são usados, na segunda etapa, para coletar artigos já existentes no Google Acadêmico, por meio de *Web Crawlers*⁶; na terceira etapa, a fim de garantir melhores resultados na recomendação utilizando processos já realizados pelo Google Acadêmico, são utilizadas as avaliações dos usuários sobre os artigos acadêmicos e a classificação dos artigos no Google Acadêmico, pois o mesmo classifica seus resultados pesando o texto de cada documento, o local de publicação, por quem foi escrito, e o número de citações que esse documento possui em demais publicações *About. . .* (). Com isso, na terceira etapa, ocorre a recomendação em si, por meio de técnicas de recomendação colaborativa, agregando avaliações e reconhecendo as semelhanças entre usuários cadastrados no sistema com base em suas classificações, ou por meio de técnicas de recomendação baseada em conteúdo, aplicando-se heurísticas para pesar o tanto que um determinado artigo acadêmico é útil para o usuário em questão, ao considerar o conhecimento adquirido sobre o mesmo Ricci, Rokach e Shapira (2010). Desta forma, ao se cadastrar no serviço de recomendação, o usuário possui duas opções de recomendação citadas anteriormente, tendo que escolher entre uma delas. A avaliação do sistema foi feita de forma qualitativa, por meio de testes realizados por um grupo de seis pessoas de forma local, que constataram o satisfatório funcionamento do serviço.

Tian et al. (2019) desenvolveram um serviço para recomendar livros aos estudantes da *College of Data Science and Application* da *Inner Mongolia University of Technology* na China, devido ao grande volume de livros existentes em sua biblioteca. Tal sistema utiliza recomendação híbrida. Para o desenvolvimento desse trabalho, foram utilizadas filtragem colaborativa e filtragem baseada em conteúdo. Na filtragem colaborativa, a recomendação foi baseada em usuário similares ao usuário-alvo. Já na técnica baseada em conteúdo, a recomendação foi baseada na descrição dos livros e no perfil do usuário em questão: a técnica recomenda livros que são similares aos que o usuário já tenha utilizado e aprovado; para isso, são utilizados pesos de importância dados aos aprovados pelo usuário. Por fim, para recomendar os livros, as duas técnicas foram aplicadas separadamente; após isso, os resultados retornados são combinados, e, assim, a lista de recomendação final foi gerada. Os resultados experimentais obtidos mostraram que a precisão melhora na proporção que a quantidade de livros para treinamento do serviço de recomendação aumenta, e que a recomendação híbrida, ou seja, a combinação das duas técnicas de recomendação, melhora ainda mais a precisão.

⁶ *Web Crawlers* são algoritmos que analisam códigos de sites disponíveis na Web e coleta informações deles.

O serviço de recomendação, desenvolvido neste trabalho, diferencia-se dos demais apresentados ao oferecer ao usuário três possibilidades de utilização do serviço na BDMPEC, sendo baseadas em: arquivo XML do currículo Lattes de um determinado usuário-alvo, produto MPEC selecionado por um usuário-alvo, e texto livre fornecido por um usuário-alvo; além disso, ao utilizar o algoritmo RAKE para extrair palavras-chave importantes dos dados textuais grandes considerados em uma recomendação, a fim de trazer mais eficácia às estratégias propostas do serviço de recomendação. E é similar aos demais trabalhos relacionados nas abordagens utilizadas pois, entre eles, foram utilizadas as abordagens colaborativa, baseada em conteúdo e híbrida; neste trabalho, todas essas são também utilizadas.

3 Serviço de Recomendação Proposto

Neste capítulo, é abordado o desenvolvimento do serviço de recomendação para a BDMPEC; para tanto, a Seção 3.1 descreve a sua arquitetura de funcionamento e a Seção 3.2 apresenta a interface da BDMPEC, relacionada ao mesmo.

3.1 Arquitetura de funcionamento

O serviço de recomendação proposto e desenvolvido neste trabalho é composto por diferentes estratégias de recomendação, sendo elas: (a) uma estratégia que utiliza o arquivo XML, referente ao currículo Lattes de um usuário, para obter informações a serem utilizadas na recomendação; (b) uma estratégia baseada na seleção, pelo usuário, de um produto educacional do acervo da BDMPEC; e (c) uma estratégia baseada em um texto livre, fornecido pelo usuário.

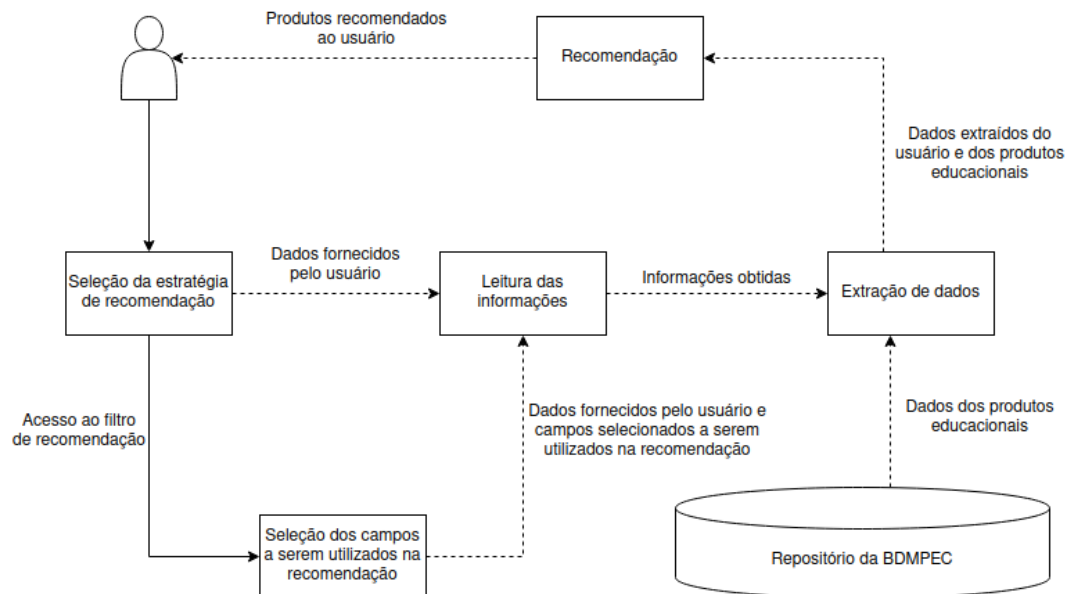
De uma forma geral, a arquitetura do serviço de recomendação, independente da estratégia utilizada, consiste em basicamente três módulos: (a) leitura das informações fornecidas pelo usuário; (b) extração dos dados do usuário e dos produtos educacionais presentes no repositório da BDMPEC; e (c) recomendação propriamente dita de produtos educacionais ao usuário. A visão geral da arquitetura de funcionamento do serviço de recomendação encontra-se na Figura 3.1.

Conforme apresentado na Figura 3.1, o serviço de recomendação, de modo geral, comporta-se da seguinte maneira: o usuário tem a opção de selecionar os campos a serem utilizados na recomendação; caso ele opte por não selecionar, são utilizados os campos-padrão definidos como sendo os melhores. A partir daí, é realizada a leitura das informações estabelecidas (arquivo do currículo Lattes do usuário ou produto educacional selecionado pelo usuário ou texto livre fornecido pelo usuário), que são passadas para o módulo de extração de dados, responsável em extrair os dados relativos às informações fornecidas pelo usuário e aos produtos educacionais presentes no repositório da BDMPEC, necessários para realização da estratégia de recomendação. Após extraídos tais dados, eles são enviados para o módulo que recomenda os produtos educacionais que são mais relevantes para o usuário. Como já mencionado, além de selecionar a estratégia de recomendação, o usuário pode selecionar os campos que devem ser considerados no momento da recomendação. Quanto aos produtos presentes no repositório da BDMPEC, os campos disponíveis para seleção são: título, descrição e palavras-chave; quanto ao currículo Lattes, utilizado em apenas uma das estratégias de recomendação, os campos disponíveis para seleção são: linhas de pesquisa, áreas de atuação e palavras-chave.

O restante desta seção encontra-se organizada como se segue. A Subseção 3.1.1 descreve o primeiro módulo da arquitetura geral de funcionamento, onde ocorre a leitura das informações,

fornecidas pelo usuário, baseada na estratégia selecionada; a Subseção 3.1.2 descreve o segundo módulo, onde são extraídos os dados relevantes à recomendação; e a Subseção 3.1.3 descreve o terceiro módulo, onde é feita a recomendação propriamente dita para o usuário.

Figura 3.1 – Arquitetura geral do serviço de recomendação proposto.



Fonte: Elaborada pelo autor.

3.1.1 Leitura de informações

O módulo de leitura de informações possui as funções de ler e validar os dados fornecidos pelo usuário, a partir da estratégia de recomendação selecionada e os campos para recomendação estabelecidos se for o caso, e enviar essas informações para o módulo de extração de dados. Para validar os dados fornecidos pelo usuário, o módulo reconhece os parâmetros de entrada do usuário e valida os mesmos. A validação depende da estratégia de recomendação a ser seguida. Para tanto, as Figuras 3.2, 3.3 e 3.4 definem as arquiteturas do módulo de leitura de informações quanto as estratégias baseada no arquivo XML do currículo Lattes do usuário, baseada na seleção de um produto educacional no acervo da BDMPEC, e baseada no texto livre, respectivamente.

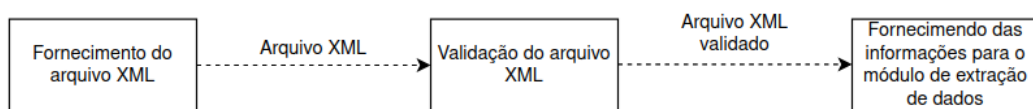
Conforme apresentado na Figura 3.2, para a estratégia de recomendação baseada no fornecimento do arquivo XML correspondente ao currículo Lattes do usuário em questão, o módulo de leitura de informações consiste em receber e validar o arquivo XML; para isso, é verificado se o arquivo realmente é do tipo XML e se ele contém informações de um currículo Lattes. Após isso, o arquivo XML é repassado para o módulo de extração de forma que ele fique salvo apenas localmente para a recomendação, sem ser necessário salvá-lo no banco de dados.

Conforme apresentado na Figura 3.3, para a estratégia de recomendação baseada em um produto educacional selecionado do acervo da biblioteca por um determinado usuário, o módulo de leitura de informações lê os campos estabelecidos do produto educacional selecionado e extrai

as palavras-chave do resumo, se for um dos campos estabelecidos, utilizando o algoritmo RAKE; após isso, repassa os campos e as palavras-chave extraídas do resumo, relativos ao produto educacional selecionado ao módulo de extração de dados.

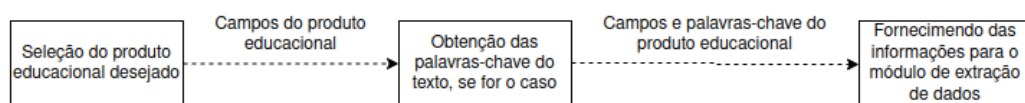
Conforme apresentado na Figura 3.4, para a estratégia de recomendação baseada no texto livre fornecido pelo usuário, o módulo de leitura das informações extrai as palavras-chave do texto fornecido, utilizando o algoritmo RAKE, e repassa as palavras-chave extraídas para o módulo de extração de dados.

Figura 3.2 – Arquitetura de leitura de informações para a estratégia baseada no XML do currículo Lattes.



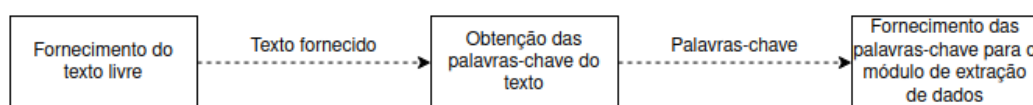
Fonte: Elaborada pelo autor.

Figura 3.3 – Arquitetura de leitura de informações para a estratégia baseada na seleção de um produto educacional.



Fonte: Elaborada pelo autor.

Figura 3.4 – Arquitetura de leitura de informações para a estratégia baseada no texto livre.



Fonte: Elaborada pelo autor.

3.1.2 Extração de dados

A principal função deste módulo é extrair os dados que representam cada produto educacional catalogado na BDMPEC e os dados que representam o usuário. Visando os dados que representam os produtos educacionais, ocorre o acesso ao repositório da BDMPEC para extrair, em princípio (campos-padrão), o título, o resumo e as palavras-chave de todos os produtos educacionais presentes no repositório. Visando os dados que representam o usuário, para cada estratégia de recomendação, dados diferentes são extraídos.

Para a estratégia de recomendação que utiliza o XML do currículo Lattes de um usuário, é feita a extração no currículo, em princípio (campos-padrão), dos campos linhas de pesquisa, áreas de atuação e palavras-chave (de todos os trabalhos cadastrados no currículo Lattes do usuário em questão). Para a estratégia baseada em um produto educacional selecionado no acervo, é feita a extração, em princípio (campos-padrão), dos campos título, resumo e palavras-chave do produto escolhido pelo usuário. Para a estratégia baseada no texto livre, é feito um corte relativo ao número de palavras-chave, definidas pelo RAKE, a serem consideradas no processo de recomendação; o corte consiste nas 5 palavras-chave de maior relevância, já que fazer um corte utilizando uma porcentagem das palavras-chave retornadas ou utilizando uma relevância mínima não é viável, devido à imprevisibilidade do número de palavras-chave a serem consideradas podendo prejudicar o processo de recomendação. Desta forma, esses são os dados que representam o usuário no momento da recomendação.

Após definidos os dados que representam o usuário e os dados que representam cada produto educacional, o módulo de extração repassa essas informações para o módulo de recomendação, que é o último passo no processo de recomendação proposto neste trabalho.

3.1.3 Recomendação

O módulo de recomendação possui a finalidade de recomendar os produtos educacionais que mais se assemelham aos dados fornecidos pelo usuário. Para isso, inicialmente, é realizado o cálculo de similaridade, utilizando o modelo vetorial (vide Subseção 2.1.3), entre os dados que representam o usuário com os dados que representam cada produto educacional. A escala de similaridade vai de 0 a 1: quanto mais perto de 1, maior será a similaridade entre os dados fornecidos.

Para o melhor entendimento, independente da estratégia de recomendação utilizada, considerando todos os produtos educacionais presentes no repositório da BDMPEC, o algoritmo proposto de recomendação possui a função de calcular a similaridade entre cada um deles e os dados fornecidos pelo usuário, de acordo com a estratégia de recomendação em questão, e realizar a ordenação desses produtos de forma decrescente considerando as similaridades dos mesmos. Uma vez realizada a ordenação, o módulo de recomendação retorna os produtos educacionais recomendados ao usuário: os produtos mais similares estarão no topo da lista (vide Tabela 3.1).

Tabela 3.1 – Similaridade entre os produtos educacionais e os dados fornecidos pelo usuário (vide Equação 2.1).

| Produto educacional | Similaridade | Acessos |
|----------------------------|---------------------|----------------|
| Produto 8 | 0.98 | 15 |
| Produto 3 | 0.91 | 18 |
| Produto 9 | 0.84 | 14 |
| Produto 1 | 0.75 | 25 |
| Produto 6 | 0.60 | 12 |
| Produto 7 | 0.52 | 11 |
| Produto 10 | 0.34 | 15 |
| Produto 2 | 0.0 | 10 |
| Produto 5 | 0.0 | 17 |
| Produto 4 | 0.0 | 16 |

Fonte: Elaborada pelo autor.

Após estabelecer a similaridade de cada produto, por meio do modelo vetorial, pode-se ou não utilizar a informação, presente na BDMPEC, referente ao número de acessos aos produtos no intuito de tentar aprimorar as similaridades dos mesmos. Para tanto, para cada produto com similaridade não-nula, a mesma é atualizada pela média aritmética entre a calculada até então e o nível de acesso dos produtos, o qual é calculado pela normalização do número de acessos de todos os produtos retornados. A Tabela 3.2 apresenta as novas similaridades de recomendação dos produtos, considerando a informação referente ao número de acessos aos mesmos, estando os mais similares no topo da lista.

Tabela 3.2 – Similaridade entre os produtos educacionais e os dados fornecidos pelo usuário considerando o número de acessos (vide Equação 2.1).

| Produto educacional | Similaridade Modelo Vetorial | Número de acessos ao produto | Nível de acesso ao produto | Similaridade final de recomendação |
|----------------------------|-------------------------------------|-------------------------------------|-----------------------------------|---|
| Produto 1 | 0.75 | 25 | 1 | 0.875 |
| Produto 3 | 0.91 | 18 | 0.72 | 0.815 |
| Produto 8 | 0.98 | 15 | 0.60 | 0.79 |
| Produto 9 | 0.84 | 14 | 0.56 | 0.70 |
| Produto 6 | 0.60 | 12 | 0.48 | 0.54 |
| Produto 7 | 0.52 | 11 | 0.44 | 0.48 |
| Produto 10 | 0.34 | 15 | 0.60 | 0.47 |
| Produto 5 | 0.0 | 17 | 0.68 | 0.0 |
| Produto 4 | 0.0 | 16 | 0.64 | 0.0 |
| Produto 2 | 0.0 | 10 | 0.40 | 0.0 |

Fonte: Elaborada pelo autor.

3.2 Interface da BDMPEC

Nesta Subseção, é apresentada a interface da BDMPEC, quanto ao serviço de recomendação proposto neste trabalho. De acordo com a Figura 3.5, é possível observar a página inicial do Serviço de Recomendação. Nesta página, o usuário pode fazer o *upload* de um arquivo XML de um currículo Lattes ou inserir um texto livre, e clicar no botão com texto **Recomendar**; desta forma, ele pode utilizar as estratégias baseada no arquivo XML do currículo Lattes e a baseada no texto livre, respectivamente. Por padrão, os campos selecionados na recomendação já estão marcados; porém, o usuário pode alterar e realizar a configuração que ele preferir. Vale ressaltar que, uma vez selecionando o campo **Número de acessos** para os produtos da BDMPEC, os números de acessos aos produtos serão considerados na composição dos valores de similaridade de recomendação dos mesmos, conforme exemplificado na Tabela 3.2. Uma vez estabelecendo os campos desejados, ao clicar no botão **Recomendar**, a recomendação propriamente dita é realizada (vide Figura 3.6).

Para utilizar a estratégia de recomendação baseada em um produto educacional existente no acervo da BDMPEC, o usuário deve, inicialmente, acessar o acervo e buscar por um produto; o resultado de uma busca é representado pela Figura 3.7. Após selecionar o produto educacional desejado, é exibida a página representada na Figura 3.8, que informa os dados do produto em questão e apresenta o botão **Recomendações**; ao selecionar tal botão, é realizada a recomendação propriamente dita e os produtos são exibidos como na Figura 3.6.

Figura 3.5 – Página inicial do serviço de recomendação.

Serviço de Recomendação

Arquivo Lattes (XML): No file selected.

Texto livre:

Definições de critérios para recomendação, OPCIONAL:

| | |
|--|--|
| Lattes | BDMPEC |
| <input checked="" type="checkbox"/> Linhas de Pesquisa | <input checked="" type="checkbox"/> Título |
| <input checked="" type="checkbox"/> Areas de Atuação | <input checked="" type="checkbox"/> Resumo |
| <input checked="" type="checkbox"/> Palavras-Chave | <input checked="" type="checkbox"/> Palavras-Chave |
| | <input type="checkbox"/> Número de acessos |

Fonte: <https://bdmpec.ufop.br>

Figura 3.6 – Página com os produtos educacionais recomendados.

Serviço de Recomendação

Arquivo Lattes (XML): No file selected.

Texto livre:

Definições de critérios para recomendação, OPCIONAL:

| | |
|--|--|
| Lattes | BDMPEC |
| <input checked="" type="checkbox"/> Linhas de Pesquisa | <input checked="" type="checkbox"/> Título |
| <input checked="" type="checkbox"/> Áreas de Atuação | <input checked="" type="checkbox"/> Resumo |
| <input checked="" type="checkbox"/> Palavras-Chave | <input checked="" type="checkbox"/> Palavras-Chave |
| | <input type="checkbox"/> Número de acessos |

Adaptação dos seres vivos

Autor: Namyna Fagna de Souza

Descrição Geral: Paradidático em formato de histórias em quadrinhos sobre o tema "Adaptação dos seres vivos" para o trabalho em aulas de ciências no ensino fundamental.

0 Acessos ao Produto

0 Downloads da Dissertação

Racismo ambiental: Caderno de Oficina

Autor: Ingridy Nathaly Santos Moreira

Descrição Geral: Guia para desenvolvimento de ações de formação inicial de professores sobre questões relacionados ao racismo ambiental.

0 Acessos ao Produto

0 Downloads da Dissertação

Fonte: <https://bdmpec.ufop.br>

Figura 3.7 – Página de apresentação dos produtos educacionais consultados.

Acesso ao Acervo

Digite os termos de sua pesquisa... [Pesquisa Avançada](#)

9 resultados encontrados para "mestrado"

Biblioteca Digital do Mestrado Profissional em Ensino de Ciências (BDMPEC)

Autor: Douglas Andrade

Descrição Geral: A BDMPEC permite a catalogação e a disponibilização de produtos gerados por egressos do programa de MPEC da UFOP com potencial para ser estendida e utilizada por outros programas de MP relativos ao ensino, devido à similaridade das características dos produtos educacionais confeccionados por esses programas.

8 Acessos ao Produto

8 Downloads da Dissertação

Química

Caderno de oficina para utilização na Educação Básica

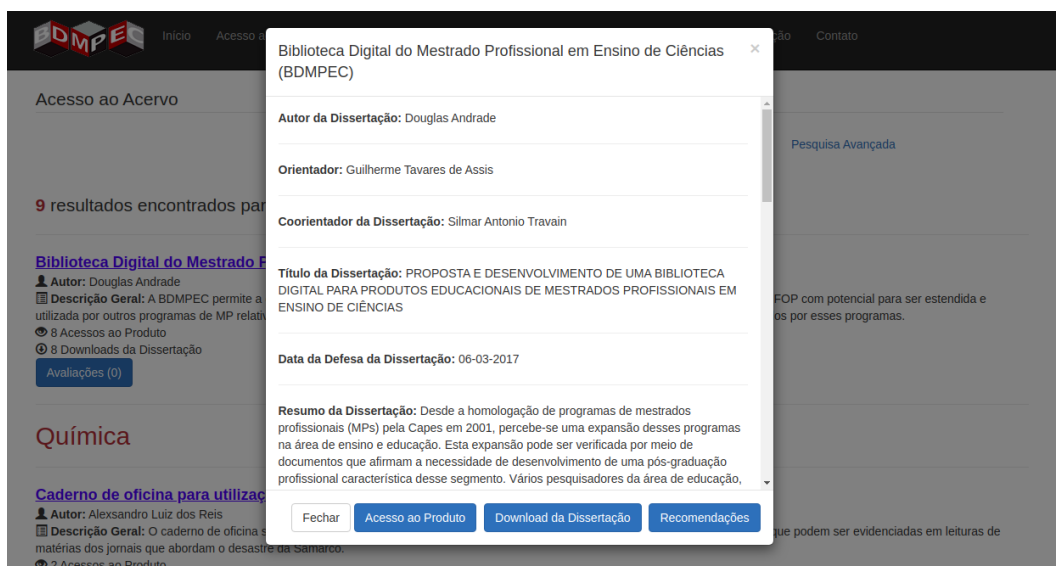
Autor: Alexandro Luiz dos Reis

Descrição Geral: O caderno de oficina se constitui em um instrumento para a reflexão, discussão e debate em torno das controvérsias, que podem ser evidenciadas em leituras de matérias dos jornais que abordam o desastre de Samarco.

0 Acessos ao Produto

Fonte: <https://bdmpec.ufop.br>

Figura 3.8 – Página de informações a respeito do produto educacional selecionado.



Fonte: <https://bdmpec.ufop.br>

4 Experimentação Prática

Neste capítulo, são apresentados e analisados os experimentos práticos realizados, considerando as estratégias de recomendação desenvolvidas quanto ao serviço proposto neste trabalho. A Seção 4.1 descreve os experimentos realizados e a Seção 4.2 apresenta e avalia os resultados obtidos por meio de tais experimentos.

4.1 Descrição dos experimentos

Com o intuito de validar as estratégias de recomendação desenvolvidas, foram realizados experimentos de validação do serviço de recomendação proposto. Para a estratégia de recomendação baseada no XML de um currículo Lattes, foram considerados 2 currículos, obtidos a partir do corpo docente do MPEC da UFOP¹; para a estratégia de recomendação baseada na seleção de um produto educacional presente no acervo, foram considerados 2 produtos educacionais selecionados; e para a estratégia baseada no texto livre, foram considerados 2 textos livres retirados de Pott e Estrela (2017) e Bacci e Pataca (2008), respectivamente. Os experimentos consideraram também o repositório completo da BDMPEC que, no atual momento, contém apenas 22 produtos educacionais catalogados.

Para cada currículo Lattes, produto educacional e texto livre, foi gerado um gabarito contendo os produtos educacionais relevantes dentre os produtos educacionais presentes no repositório da BDMPEC no momento. Para cada processo de recomendação realizado, a análise dos resultados obtidos foi feita considerando os 1 , $n/2$ e n primeiros produtos recomendados pela estratégia de recomendação ao usuário, onde n representa o número de produtos educacionais presentes no gabarito.

Para avaliar a qualidade de um *ranking* gerado pelo módulo de recomendação da Figura 3.1, foram utilizadas as seguintes métricas:

- **Precisão:** razão entre o número de produtos educacionais relevantes presentes no *ranking* gerado, de acordo com o gabarito correspondente estabelecido, e o número total de produtos educacionais do próprio *ranking*;
- **Revocação:** razão entre o número de produtos educacionais relevantes presentes no *ranking* gerado, de acordo com o gabarito correspondente estabelecido, e o número total de produtos educacionais presentes no próprio gabarito;
- **F1:** média harmônica (vide Equação 4.1) entre a precisão (Prec) e a revocação (Rev).

¹ <https://mpec.ufop.br/docentes>

$$F1 = \frac{2 \times Prec \times Rev}{Prec + Rev} \quad (4.1)$$

4.2 Análise dos resultados obtidos

Nesta seção, são apresentados e analisados os resultados obtidos por meio dos experimentos práticos realizados, envolvendo as distintas opções para os campos utilizados, tanto nas informações fornecidas pelo usuário (currículo Lattes, produto educacional selecionado ou texto livre) quanto nos produtos educacionais catalogados na BDMPEC. Como já mencionado, os experimentos consideraram 2 currículos Lattes; no caso, para cada currículo Lattes, foram executados 49 casos de teste referentes às diferentes combinações dos campos. Além disso, foram executados 7 casos de teste para cada produto educacional selecionado e cada texto livre. Por último, foram executados 6 casos de teste adicionais considerando o número de acessos dos produtos educacionais no cálculo da similaridade final de recomendação. Para facilitar a visualização dos resultados obtidos e a análise dos mesmos, todos os casos de teste foram tabelados.

Quanto às Tabelas 4.1 e 4.2, que apresentam os casos de teste associados à estratégia de recomendação baseada no currículo Lattes, as colunas utilizadas são:

- Dados extraídos do currículo Lattes: campos do currículo Lattes que foram utilizados no caso de teste;
- Metadados da BDMPEC: campos dos produtos educacionais catalogados na BDMPEC que foram utilizados no caso de teste;
- Resultados: resultados em precisão, revocação e F1 considerando, como já dito anteriormente, 1, n e $n/2$ produtos retornados, onde n é o número de produtos educacionais existentes no gabarito.

De uma forma geral, muitos casos de teste possuem os mesmos resultados já que, como já dito, a BDMPEC possui apenas 22 produtos educacionais no momento. Particularmente, quanto aos resultados obtidos nos casos de teste apresentados na Tabela 4.1, observa-se que:

- todos tiveram precisão 100% ao considerar o primeiro produto recomendado;
- a precisão destacou-se das outras métricas, o que indica que a assertividade das recomendações em relação ao gabarito foi muito boa;
- o valor da revocação e F1 ficaram mais significativos à medida que o número de produtos considerados aumentou, já que tal número foi ficando mais próximo do tamanho do gabarito;

- os melhores valores atingidos para todas as métricas foram obtidos ao considerar apenas o campo palavras-chave do currículo Lattes; ademais, quanto melhor preenchido está um currículo Lattes, mais eficaz é a recomendação;
- considerando os campos dos produtos da BDMPEC, os melhores resultados foram obtidos por todas as métricas quando todos os campos são utilizados na recomendação.

Quanto aos resultados obtidos nos casos de teste apresentados na Tabela 4.2, observa-se que:

- considerando 10 produtos retornados, as métricas ficaram todas acima de 70%;
- nos melhores casos, todas as métricas atingiram 90%;
- considerando apenas 1 produto, a precisão ficou nula em alguns casos de teste por retornar um produto não presente no gabarito em questão; porém, ao aumentar o número de produtos considerados, os níveis de precisão aumentaram também;
- os melhores valores atingidos para todas as métricas foram obtidos ao considerar ao menos 2 campos do currículo Lattes;
- considerando os campos dos produtos da BDMPEC, os melhores resultados foram obtidos por todas as métricas ao considerar ao menos 2 campos na recomendação.

Tabela 4.1 – Casos de teste: currículo Lattes <http://lattes.cnpq.br/3216449144279833>.

| Caso | Dados extraídos do currículo Lattes | | | | Metadados da BDMPEEC | | | | Resultados | | | | | | | |
|------|-------------------------------------|------------------|----------------|--------|----------------------|----------------|-----------|-----------|------------|-----------|------------|-----------|-----------|-------|----------|-------|
| | Linhas de Pesquisa | Áreas de Atuação | Palavras-chave | Título | Resumo | Palavras-chave | 1 produto | | 4 produtos | | 8 produtos | | Revocação | F1 | Precisão | F1 |
| | | | | | | | Precisão | Revocação | Precisão | Revocação | Precisão | Revocação | | | | |
| 1 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 2 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 3 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 4 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 5 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.5 | 0.25 | 0.3333 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 6 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 7 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.5 | 0.25 | 0.3333 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 8 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 9 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 10 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 11 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 12 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.5 | 0.25 | 0.3333 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 13 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 14 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.5 | 0.25 | 0.3333 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 15 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 16 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 17 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 18 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 19 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.5 | 0.25 | 0.3333 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 20 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 21 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.5 | 0.25 | 0.3333 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 22 | x | x | x | x | x | x | 0.125 | 0.2222 | 1 | 0.5 | 0.6666 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 23 | x | x | x | x | x | x | 0.125 | 0.2222 | 1 | 0.5 | 0.6666 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 24 | x | x | x | x | x | x | 0.125 | 0.2222 | 1 | 0.5 | 0.6666 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 25 | x | x | x | x | x | x | 0.125 | 0.2222 | 1 | 0.5 | 0.6666 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 26 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.5 | 0.25 | 0.3333 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 27 | x | x | x | x | x | x | 0.125 | 0.2222 | 1 | 0.5 | 0.6666 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 28 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.5 | 0.25 | 0.3333 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 29 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 30 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 31 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 32 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 33 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.5 | 0.25 | 0.3333 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 34 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 35 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.5 | 0.25 | 0.3333 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 36 | x | x | x | x | x | x | 0.125 | 0.2222 | 1 | 0.5 | 0.6666 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 37 | x | x | x | x | x | x | 0.125 | 0.2222 | 1 | 0.5 | 0.6666 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 38 | x | x | x | x | x | x | 0.125 | 0.2222 | 1 | 0.5 | 0.6666 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 39 | x | x | x | x | x | x | 0.125 | 0.2222 | 1 | 0.5 | 0.6666 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 40 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.5 | 0.25 | 0.3333 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 41 | x | x | x | x | x | x | 0.125 | 0.2222 | 1 | 0.5 | 0.6666 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 42 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.5 | 0.25 | 0.3333 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 43 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 44 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 45 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 46 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 47 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.5 | 0.25 | 0.3333 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |
| 48 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.75 | 0.75 | 0.75 | 0.75 | 0.75 |
| 49 | x | x | x | x | x | x | 0.125 | 0.2222 | 0.5 | 0.25 | 0.3333 | 0.625 | 0.625 | 0.625 | 0.625 | 0.625 |

Tabela 4.2 – Casos de teste: currículo Lattes <http://lattes.cnpq.br/3537386569474804>.

| Caso | Dados extraídos do currículo Lattes | | | | Metadados da BDMPec | | | | Resultados | | | | | | | | | |
|------|-------------------------------------|------------------|----------------|--------|---------------------|----------------|-----------|----------|------------|------------|----------|--------|-------------|----------|-----|--|--|--|
| | Linhas de Pesquisa | Áreas de Atuação | Palavras-chave | Título | Resumo | Palavras-chave | 1 produto | | | 5 produtos | | | 10 produtos | | | | | |
| | | | | | | | Previsão | Revoação | FI | Previsão | Revoação | FI | Previsão | Revoação | FI | | | |
| 1 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 1 | 0.5 | 0.6666 | 0.9 | 0.9 | 0.9 | | | |
| 2 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.8 | 0.4 | 0.5333 | 0.9 | 0.9 | 0.9 | | | |
| 3 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.8 | 0.4 | 0.5333 | 0.9 | 0.9 | 0.9 | | | |
| 4 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 1 | 0.5 | 0.6666 | 0.9 | 0.9 | 0.9 | | | |
| 5 | x | x | x | x | x | x | 0 | 0 | 0 | 0.6 | 0.3 | 0.4 | 0.8 | 0.8 | 0.8 | | | |
| 6 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 1 | 0.5 | 0.6666 | 0.9 | 0.9 | 0.9 | | | |
| 7 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.8 | 0.4 | 0.5333 | 0.9 | 0.9 | 0.9 | | | |
| 8 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 1 | 0.5 | 0.6666 | 0.9 | 0.9 | 0.9 | | | |
| 9 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.8 | 0.4 | 0.5333 | 0.9 | 0.9 | 0.9 | | | |
| 10 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.8 | 0.4 | 0.5333 | 0.9 | 0.9 | 0.9 | | | |
| 11 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 1 | 0.5 | 0.6666 | 0.9 | 0.9 | 0.9 | | | |
| 12 | x | x | x | x | x | x | 0 | 0 | 0 | 0.6 | 0.3 | 0.4 | 0.8 | 0.8 | 0.8 | | | |
| 13 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 1 | 0.5 | 0.6666 | 0.9 | 0.9 | 0.9 | | | |
| 14 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.8 | 0.4 | 0.5333 | 0.9 | 0.9 | 0.9 | | | |
| 15 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 1 | 0.5 | 0.6666 | 0.9 | 0.9 | 0.9 | | | |
| 16 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.8 | 0.4 | 0.5333 | 0.9 | 0.9 | 0.9 | | | |
| 17 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.8 | 0.4 | 0.5333 | 0.9 | 0.9 | 0.9 | | | |
| 18 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 1 | 0.5 | 0.6666 | 0.9 | 0.9 | 0.9 | | | |
| 19 | x | x | x | x | x | x | 0 | 0 | 0 | 0.8 | 0.4 | 0.5333 | 0.8 | 0.8 | 0.8 | | | |
| 20 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 1 | 0.5 | 0.6666 | 0.9 | 0.9 | 0.9 | | | |
| 21 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.8 | 0.4 | 0.5333 | 0.9 | 0.9 | 0.9 | | | |
| 22 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 1 | 0.5 | 0.6666 | 0.9 | 0.9 | 0.9 | | | |
| 23 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.8 | 0.4 | 0.5333 | 0.9 | 0.9 | 0.9 | | | |
| 24 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.8 | 0.4 | 0.5333 | 0.9 | 0.9 | 0.9 | | | |
| 25 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 1 | 0.5 | 0.6666 | 0.9 | 0.9 | 0.9 | | | |
| 26 | x | x | x | x | x | x | 0 | 0 | 0 | 0.6 | 0.3 | 0.4 | 0.8 | 0.8 | 0.8 | | | |
| 27 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 1 | 0.5 | 0.6666 | 0.9 | 0.9 | 0.9 | | | |
| 28 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.8 | 0.4 | 0.5333 | 0.9 | 0.9 | 0.9 | | | |
| 29 | x | x | x | x | x | x | 0 | 0 | 0 | 0.6 | 0.3 | 0.4 | 0.8 | 0.8 | 0.8 | | | |
| 30 | x | x | x | x | x | x | 0 | 0 | 0 | 0.6 | 0.3 | 0.4 | 0.8 | 0.8 | 0.8 | | | |
| 31 | x | x | x | x | x | x | 0 | 0 | 0 | 0.6 | 0.3 | 0.4 | 0.8 | 0.8 | 0.8 | | | |
| 32 | x | x | x | x | x | x | 0 | 0 | 0 | 0.6 | 0.3 | 0.4 | 0.8 | 0.8 | 0.8 | | | |
| 33 | x | x | x | x | x | x | 0 | 0 | 0 | 0.6 | 0.3 | 0.4 | 0.8 | 0.8 | 0.8 | | | |
| 34 | x | x | x | x | x | x | 0 | 0 | 0 | 0.6 | 0.3 | 0.4 | 0.8 | 0.8 | 0.8 | | | |
| 35 | x | x | x | x | x | x | 0 | 0 | 0 | 0.6 | 0.3 | 0.4 | 0.8 | 0.8 | 0.8 | | | |
| 36 | x | x | x | x | x | x | 0 | 0 | 0 | 0.6 | 0.3 | 0.4 | 0.8 | 0.8 | 0.8 | | | |
| 37 | x | x | x | x | x | x | 0 | 0 | 0 | 0.6 | 0.3 | 0.4 | 0.8 | 0.8 | 0.8 | | | |
| 38 | x | x | x | x | x | x | 0 | 0 | 0 | 0.6 | 0.3 | 0.4 | 0.8 | 0.8 | 0.8 | | | |
| 39 | x | x | x | x | x | x | 0 | 0 | 0 | 0.6 | 0.3 | 0.4 | 0.8 | 0.8 | 0.8 | | | |
| 40 | x | x | x | x | x | x | 0 | 0 | 0 | 0.6 | 0.3 | 0.4 | 0.8 | 0.8 | 0.8 | | | |
| 41 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.4 | 0.2 | 0.2666 | 0.7 | 0.7 | 0.7 | | | |
| 42 | x | x | x | x | x | x | 0 | 0 | 0 | 0.6 | 0.3 | 0.4 | 0.8 | 0.8 | 0.8 | | | |
| 43 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.4 | 0.2 | 0.2666 | 0.7 | 0.7 | 0.7 | | | |
| 44 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.4 | 0.2 | 0.2666 | 0.7 | 0.7 | 0.7 | | | |
| 45 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.4 | 0.2 | 0.2666 | 0.7 | 0.7 | 0.7 | | | |
| 46 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.4 | 0.2 | 0.2666 | 0.7 | 0.7 | 0.7 | | | |
| 47 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.4 | 0.2 | 0.2666 | 0.7 | 0.7 | 0.7 | | | |
| 48 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.4 | 0.2 | 0.2666 | 0.7 | 0.7 | 0.7 | | | |
| 49 | x | x | x | x | x | x | 1 | 0.1 | 0.1818 | 0.4 | 0.2 | 0.2666 | 0.7 | 0.7 | 0.7 | | | |

Quanto às Tabelas 4.3 e 4.4, que apresentam os casos de teste associados à estratégia de recomendação baseada na seleção de um produto educacional, considerando que a coluna “Metadados da BDMPEC” corresponde aos campos dos produtos educacionais considerados em ambos os lados de recomendação, observa-se que:

- ao considerar apenas 1 produto, a precisão foi de 100% em todos os casos;
- considerando o mesmo número de produtos que existe no gabarito, todas as métricas ficaram acima de 80%;
- as métricas foram estáveis em todos os casos de teste;
- os melhores casos de testes atingiram métricas de 85.71% em todas as métricas.

Tabela 4.3 – Caso de teste: produto educacional 1.

| Caso | Metadados da BDMPEC | | | Resultados | | | | | | | | |
|------|---------------------|--------|----------------|------------|-----------|--------|------------|-----------|------|------------|-----------|-----|
| | Título | Resumo | Palavras-chave | 1 produto | | | 3 produtos | | | 5 produtos | | |
| | | | | Precisão | Revocação | F1 | Precisão | Revocação | F1 | Precisão | Revocação | F1 |
| 1 | x | x | x | 1 | 0.2 | 0.3333 | 1 | 0.6 | 0.75 | 0.8 | 0.8 | 0.8 |
| 2 | x | x | | 1 | 0.2 | 0.3333 | 1 | 0.6 | 0.75 | 0.8 | 0.8 | 0.8 |
| 3 | x | | x | 1 | 0.2 | 0.3333 | 1 | 0.6 | 0.75 | 0.8 | 0.8 | 0.8 |
| 4 | | x | x | 1 | 0.2 | 0.3333 | 0.6666 | 0.4 | 0.5 | 0.8 | 0.8 | 0.8 |
| 5 | x | | | 1 | 0.2 | 0.3333 | 0.6666 | 0.4 | 0.5 | 0.8 | 0.8 | 0.8 |
| 6 | | x | | 1 | 0.2 | 0.3333 | 1 | 0.6 | 0.75 | 0.8 | 0.8 | 0.8 |
| 7 | | | x | 1 | 0.2 | 0.3333 | 1 | 0.6 | 0.75 | 0.8 | 0.8 | 0.8 |

Tabela 4.4 – Caso de teste: produto educacional 2.

| Caso | Metadados da BDMPEC | | | Resultados | | | | | | | | |
|------|---------------------|--------|----------------|------------|-----------|------|------------|-----------|--------|------------|-----------|--------|
| | Título | Resumo | Palavras-chave | 1 produto | | | 4 produtos | | | 7 produtos | | |
| | | | | Precisão | Revocação | F1 | Precisão | Revocação | F1 | Precisão | Revocação | F1 |
| 1 | x | x | x | 1 | 0.1428 | 0.25 | 1 | 0.5714 | 0.7272 | 0.8571 | 0.8571 | 0.8571 |
| 2 | x | x | | 1 | 0.1428 | 0.25 | 1 | 0.5714 | 0.7272 | 0.8571 | 0.8571 | 0.8571 |
| 3 | x | | x | 1 | 0.1428 | 0.25 | 1 | 0.5714 | 0.7272 | 0.8571 | 0.8571 | 0.8571 |
| 4 | | x | x | 1 | 0.1428 | 0.25 | 1 | 0.5714 | 0.7272 | 0.8571 | 0.8571 | 0.8571 |
| 5 | x | | | 1 | 0.1428 | 0.25 | 1 | 0.5714 | 0.7272 | 0.8571 | 0.8571 | 0.8571 |
| 6 | | x | | 1 | 0.1428 | 0.25 | 1 | 0.5714 | 0.7272 | 0.8571 | 0.8571 | 0.8571 |
| 7 | | | x | 1 | 0.1428 | 0.25 | 1 | 0.5714 | 0.7272 | 0.8571 | 0.8571 | 0.8571 |

Quanto às Tabelas 4.5 e 4.6, que apresentam os casos de teste associados à estratégia de recomendação baseada no texto livre, considerando que a coluna “Metadados da BDMPEC” corresponde aos campos dos produtos educacionais catalogados na BDMPEC, observa-se que:

- considerando 1 produto educacional, apenas o caso 7 da Tabela 4.5 não recomendou um produto existente no gabarito;
- os melhores resultados na Tabela 4.5 foram ao considerar 2 produtos, chegando a 80% de eficácia;
- os melhores resultados na Tabela 4.6 foram ao considerar o resumo e as palavras-chave, chegando a 75% de eficácia.

Tabela 4.5 – Caso de teste: texto livre - resumo do artigo Pott e Estrela (2017).

| Caso | Metadados da BDMPEC | | | Resultados | | | | | | | | |
|------|---------------------|--------|----------------|------------|-----------|-----|------------|-----------|-----|------------|-----------|--------|
| | Título | Resumo | Palavras-chave | 1 produto | | | 2 produtos | | | 3 produtos | | |
| | | | | Precisão | Revocação | F1 | Precisão | Revocação | F1 | Precisão | Revocação | F1 |
| 1 | x | x | x | 1 | 0.3333 | 0.5 | 1 | 0.6666 | 0.8 | 0.6666 | 0.6666 | 0.6666 |
| 2 | x | x | | 1 | 0.3333 | 0.5 | 1 | 0.6666 | 0.8 | 0.6666 | 0.6666 | 0.6666 |
| 3 | x | | x | 1 | 0.3333 | 0.5 | 1 | 0.6666 | 0.8 | 0.6666 | 0.6666 | 0.6666 |
| 4 | | x | x | 1 | 0.3333 | 0.5 | 1 | 0.6666 | 0.8 | 0.6666 | 0.6666 | 0.6666 |
| 5 | x | | | 1 | 0.3333 | 0.5 | 1 | 0.6666 | 0.8 | 0.6666 | 0.6666 | 0.6666 |
| 6 | | x | | 1 | 0.3333 | 0.5 | 1 | 0.6666 | 0.8 | 0.6666 | 0.6666 | 0.6666 |
| 7 | | | x | 0 | 0 | 0 | 0.5 | 0.3333 | 0.4 | 0.3333 | 0.3333 | 0.3333 |

Tabela 4.6 – Caso de teste: texto livre - resumo do artigo Bacci e Pataca (2008).

| Caso | Metadados da BDMPEC | | | Resultados | | | | | | | | |
|------|---------------------|--------|----------------|------------|-----------|-----|------------|-----------|--------|------------|-----------|------|
| | Título | Resumo | Palavras-chave | 1 produto | | | 2 produtos | | | 4 produtos | | |
| | | | | Precisão | Revocação | F1 | Precisão | Revocação | F1 | Precisão | Revocação | F1 |
| 1 | x | x | x | 1 | 0.25 | 0.4 | 0.5 | 0.25 | 0.3333 | 0.25 | 0.25 | 0.25 |
| 2 | x | x | | 1 | 0.25 | 0.4 | 0.5 | 0.25 | 0.3333 | 0.25 | 0.25 | 0.25 |
| 3 | x | | x | 1 | 0.25 | 0.4 | 1 | 0.5 | 0.6666 | 0.5 | 0.5 | 0.5 |
| 4 | | x | x | 1 | 0.25 | 0.4 | 0.5 | 0.25 | 0.3333 | 0.75 | 0.75 | 0.75 |
| 5 | x | | | 1 | 0.25 | 0.4 | 0.5 | 0.25 | 0.33 | 0.25 | 0.25 | 0.25 |
| 6 | | x | | 1 | 0.25 | 0.4 | 0.5 | 0.25 | 0.3333 | 0.75 | 0.75 | 0.75 |
| 7 | | | x | 1 | 0.25 | 0.4 | 0.5 | 0.25 | 0.3333 | 0.75 | 0.75 | 0.75 |

A Tabela 4.7 apresenta alguns casos de testes presentes nas Tabelas 4.1, 4.2, 4.3, 4.4, 4.5 e 4.6, considerando o número de acessos, aos produtos educacionais presentes na BDMPEC, na composição de seus valores de similaridade de recomendação. A partir dos dados apresentados, observa-se que:

- os casos 1 da Tabela 4.1, 29 da Tabela 4.2, 1 da Tabela 4.4 e 7 da Tabela 4.5 obtiveram melhores resultados;
- o caso 4 da Tabela 4.3 e o caso 5 da Tabela 4.6 obtiveram o mesmo resultado;
- para a amostra considerada, o número de acessos dos produtos mostrou-se eficaz, já que melhorou a maioria dos casos considerados.

Tabela 4.7 – Casos de teste utilizando o número de acesso dos produtos.

| Caso | Tabela | Resultados | | | | | | | | |
|------|--------|------------|-----------|--------|--------------|-----------|--------|------------|-----------|--------|
| | | 1 produto | | | n/2 produtos | | | n produtos | | |
| | | Precisão | Revocação | F1 | Precisão | Revocação | F1 | Precisão | Revocação | F1 |
| 1 | 4.1 | 1 | 0.125 | 0.2222 | 0.75 | 0.375 | 0.5 | 0.875 | 0.875 | 0.875 |
| 29 | 4.2 | 1 | 0.1 | 0.1818 | 1 | 0.5 | 0.6666 | 0.9 | 0.9 | 0.9 |
| 4 | 4.3 | 1 | 0.2 | 0.3333 | 1 | 0.6 | 0.75 | 0.8 | 0.8 | 0.8 |
| 1 | 4.4 | 1 | 0.1428 | 0.25 | 1 | 0.5714 | 0.7272 | 0.8571 | 0.8571 | 0.8571 |
| 7 | 4.5 | 1 | 0.3333 | 0.5 | 0 | 0.5 | 0.3333 | 0.6666 | 0.6666 | 0.6666 |
| 5 | 4.6 | 1 | 0.25 | 0.4 | 0.5 | 0.25 | 0.3333 | 0.25 | 0.25 | 0.25 |

Além dos casos de testes mencionados, foi realizada uma experimentação adicional para se verificar se o número de acessos aos produtos educacionais, conforme mencionados na Tabela 3.2, podem contribuir ou não positivamente nas estratégias de recomendação proposta.

Essa experimentação é demonstrada na Tabela 4.7, onde observa-se que o número de acessos aos produtos catalogados na BDMPEC pode contribuir positivamente para as estratégias de recomendação propostas neste trabalho. Cada um dos casos considerados na amostra utilizada na Tabela 4.7 foi retirado de cada uma das tabelas de caso de testes executados anteriormente; com isso, a amostra utilizou casos de todas as estratégias desenvolvidas.

Para a experimentação realizada, tanto nos casos que envolvem a estratégia baseada no texto livre quanto nos casos que envolvem o resumo de um produto catalogado na BDMPEC, que são os casos onde o algoritmo RAKE aplica-se, foram utilizadas as 5 palavras-chave mais relevantes retornadas pelo RAKE, já que fazer um corte utilizando uma porcentagem das palavras-chave retornadas ou utilizando uma relevância mínima não é viável, devido à imprevisibilidade do número de palavras-chave a serem consideradas podendo prejudicar o processo de recomendação.

Após realizados os experimentos de validação das estratégias de recomendação propostas para a BDMPEC, com base nos melhores valores obtidos de precisão e F1, foi estabelecida a melhor configuração para o serviço de recomendação desenvolvido, a saber:

- combinação de todos os campos estabelecidos para o critério "Dados extraídos do currículo Lattes";
- combinação de todos os campos estabelecidos para o critério "Metadados da BDMPEC".

Ademais, por meio dos experimentos realizados, foi possível constatar que a estratégia de recomendação baseada em um produto educacional catalogado na BDMPEC destacou-se entre as demais estratégias, considerando os valores obtidos pelas métricas de precisão, revocação e F1, já que seus valores foram, de uma forma geral, os maiores e mais constantes entre todos os casos executados.

5 Considerações Finais

Neste capítulo, são apresentadas as considerações finais do presente trabalho. A Seção 5.1 apresenta as conclusões, relativas aos resultados experimentais obtidos, e a Seção 5.2 apresenta as perspectivas de trabalho futuro.

5.1 Conclusões

Conforme apresentado, o objetivo geral desse trabalho consistiu em propor, desenvolver e validar um serviço de recomendação para a BDMPEC. Para tanto, foram desenvolvidas três estratégias distintas de recomendação, baseadas no arquivo XML de um currículo Lattes, em um produto educacional selecionado no acervo e em um texto livre fornecido.

Por meio dos experimentos realizados, no intuito de validar o serviço de recomendação desenvolvido, pode-se concluir que: (a) os resultados obtidos foram satisfatórios; (b) a precisão foi a métrica que se destacou considerando todos os casos de teste, o que indica a assertividade das recomendações; (c) as métricas de revocação e F1 destacaram-se mais nos casos de teste que consideraram os mesmos número de produtos educacionais presentes no gabarito; (d) a estratégia baseada no arquivo XML do currículo Lattes, considerando vários produtos retornados, chegou a atingir 90% de precisão, revocação e F1; (e) a estratégia baseada na seleção de um produto educacional catalogado na BDMPEC, considerando vários produtos retornados, chegou a atingir 87.5% de precisão, revocação e F1; (f) a estratégia baseada no texto livre fornecido pelo usuário, considerando vários produtos retornados, chegou a atingir 66.6% de precisão, revocação e F1, sendo a estratégia, de forma geral, com menores valores; (g) utilizar o número de acesso dos produtos foi positivo no processo de recomendação; e (h) a melhor estratégia de recomendação foi baseada na seleção de um produto educacional catalogado na BDMPEC por apresentar, de uma forma geral, os melhores níveis de precisão, revocação e F1 nos casos de teste considerados.

5.2 Perspectivas de trabalho futuro

Como perspectivas de trabalho futuro para aprimorar o serviço de recomendação proposto neste trabalho, tem-se: (1) realizar experimentos de validação do algoritmo RAKE (vide Subseção 2.1.4), para determinar, de uma forma mais eficiente, quantas palavras-chave devem ser utilizadas para representar um campo de valor textual longo em um processo de recomendação; (2) criar e validar uma nova estratégia de recomendação, baseada no cadastro de um usuário no serviço de recomendação, podendo assim utilizar o histórico de produtos acessados pelo mesmo; (3) realizar mais experimentos quando a BDMPEC possuir um número maior de produtos educacionais em seu acervo, a fim de validar a utilização ou não do número de acesso aos produtos como *default*;

(4) utilizar dados vindos de outras bibliotecas digitais para realizar a recomendação; e (5) realizar testes de usabilidade no serviço.

Referências

- ABOUT google scholar. Google. Disponível em: <<https://scholar.google.com.br/intl/pt-BR/scholar/about.html>>.
- 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.
- ANDRADE, D.; MATHIAS, W. da S. F.; TRAVAIN, S. A.; ASSIS, G. T. de. Uma biblioteca digital para produtos educacionais de mestrados profissionais em ensino de ciências. In: 5TH CONGRESSO NACIONAL DE PESQUISA E ENSINO EM CIENCIAS. Campina Grande, Brasil, 2020. p. 1–2.
- ANDRADE, M. H. S. *Entenda como funcionam os Sistemas de Recomendação*. 2017. Disponível em: <<https://www.igti.com.br/blog/como-funcionam-os-sistemas-de-recomendacao/>>. Acesso em: 01 de março de 2021.
- BACCI, D. d. L. C.; PATACA, E. M. Educação para a água. *Estudos avançados*, SciELO Brasil, v. 22, p. 211–226, 2008.
- BAEZA-YATES, R.; RIBEIRO-NETO, B. et al. *Modern information retrieval*. [S.l.]: ACM press New York, 1999. v. 463.
- FURLAN, L. A. da R.; ZAMBERLAN, A. de O.; VIEIRA, S. A. G.; CANAL, A. P. Desenvolvimento de um sistema de recomendação para bibliotecas digitais. v. 19, p. 87 – 103, 2018. ISSN 2176-462X. 11th CIRP Conference on Industrial Product-Service Systems.
- LU, L.; MEDO, M.; YEUNG, C. H.; ZHANG, Y.-C.; ZHANG, Z.-K.; ZHOU, T. Recommender systems. *Physics Reports*, Elsevier BV, v. 519, n. 1, p. 1–49, Oct 2012. ISSN 0370-1573. Disponível em: <<http://dx.doi.org/10.1016/j.physrep.2012.02.006>>.
- MATHIAS, W. da S. F. Proposta, desenvolvimento e validação dos serviços de autoarquivamento e busca na web para a biblioteca digital de produtos gerados por programas de mestrado profissional em ensino de ciências. Ouro Preto, p. 53, 08 2016.
- NJ, H. M.; COSTA, E. B.; OLIVEIRA, T. T.; SILVA, A. P.; BITTENCOURT, I. I. Sistema de recomendação híbrido para bibliotecas digitais que suportam o protocolo oai-pmh. In: *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*. [S.l.: s.n.], 2012. v. 1, n. 1.
- NOVOSELTSEVA, E. *ARTIFICIAL INTELLIGENCE IN ECOMMERCE: BENEFITS, STATISTICS, FACTS, USE CASES CASE STUDIES*. [S.l.], 2018. <<https://news.un.org/pt/story/2019/11/1693711>>.
- ONU News. *Estudo da ONU revela que mundo tem abismo digital de gênero*. 2017. Disponível em: <<https://news.un.org/pt/story/2019/11/1693711>>. Acesso em: 01 de março de 2021.
- ORTEGA, F.; SÁNCHEZ, J.-L.; BOBADILLA, J.; GUTIÉRREZ, A. Improving collaborative filtering-based recommender systems results using pareto dominance.

Information Sciences, v. 239, p. 50 – 61, 2013. ISSN 0020-0255. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0020025513002004>>.

POTT, C. M.; ESTRELA, C. C. Histórico ambiental: desastres ambientais e o despertar de um novo pensamento. *Estudos avançados*, SciELO Brasil, v. 31, p. 271–283, 2017.

RICCI, F.; ROKACH, L.; SHAPIRA, B. Recommender systems handbook. In: _____. [S.l.: s.n.], 2010. v. 1-35, p. 1–35.

ROSE, S.; ENGEL, D.; CRAMER, N.; COWLEY, W. Automatic keyword extraction from individual documents. *Text mining: applications and theory*, Citeseer, v. 1, p. 1–20, 2010.

SANTOS, G. R. do. Proposta e desenvolvimento de um serviço de recomendação para a biblioteca digital do curso de ciência da computação da universidade federal de ouro preto. Ouro Preto, p. 42, 03 2016.

TIAN, Y.; ZHENG, B.; WANG, Y.; ZHANG, Y.; WU, Q. College library personalized recommendation system based on hybrid recommendation algorithm. *Procedia CIRP*, v. 83, p. 490 – 494, 2019. ISSN 2212-8271. 11th CIRP Conference on Industrial Product-Service Systems. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S2212827119307401>>.