

UNIVERSIDADE FEDERAL DO PAMPA

Wolgan Ens Quepfert

**Recomendação de Pesquisadores na
UNIPAMPA: Uma Proposta Baseada no
Conteúdo de Projetos de Pesquisa**

Alegrete
2018

Wolgan Ens Quepfert

**Recomendação de Pesquisadores na UNIPAMPA:
Uma Proposta Baseada no Conteúdo de Projetos de
Pesquisa**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Ciência da Computação da Universidade Federal do Pampa como requisito parcial para a obtenção do título de Bacharel em Ciência da Computação.

Orientador: Profa. Dra. Andréa Sabedra Bordin

Alegrete
2018

Wolgan Ens Quepfert

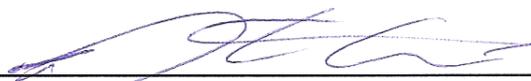
**Recomendação de Pesquisadores na UNIPAMPA:
Uma Proposta Baseada no Conteúdo de Projetos de
Pesquisa**

Trabalho de Conclusão de Curso apresentado
ao Curso de Graduação em Ciência da Com-
putação da Universidade Federal do Pampa
como requisito parcial para a obtenção do tí-
tulo de Bacharel em Ciência da Computação.

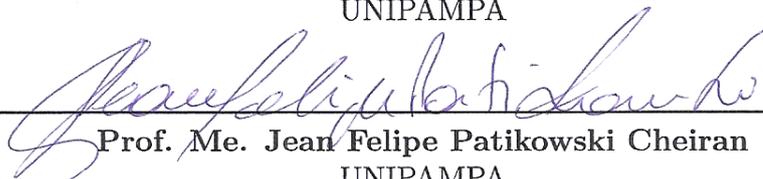
Trabalho de Conclusão de Curso defendido e aprovado em 06 de Dezembro de 2018.
Banca examinadora:



Prof. Dra. Andréa Sabedra Bordin
Orientador
UNIPAMPA



Prof. Dr. Alessandro Gonçalves Girardi
UNIPAMPA



Prof. Me. Jean Felipe Patikowski Cheiran
UNIPAMPA

Este trabalho é dedicado às crianças adultas que,
quando pequenas, sonharam em se tornar cientistas.

AGRADECIMENTOS

Agradeço a todos que de alguma forma participaram e contribuíram positivamente no meu período de graduação. Agradeço aos professores pelas oportunidades de participação em atividades que contribuíram com o meu desenvolvimento acadêmico e pelas orientações nos mais diversos trabalhos e aos meus familiares e entes queridos, por todos os tipos de suporte e apoio que me foram dados durante este período.

“Não vos amoldeis às estruturas deste mundo,
mas transformai-vos pela renovação da mente,
a fim de distinguir qual é a vontade de Deus:
o que é bom, o que Lhe é agradável, o que é perfeito.
(Bíblia Sagrada, Romanos 12:2)

RESUMO

Sistemas de recomendação (SR) são mecanismos populares no tratamento do problema de excesso de informação na Web 2.0. Grandes companhias, como Amazon, Netflix, Facebook e Youtube, são reconhecidas por utilizarem estes mecanismos. No meio acadêmico, sistemas de recomendação são mais populares na recomendação de artigos científicos para atividades de revisão de literatura (Google Scholar, Web of Science e Scopus). No entanto a recomendação de pesquisadores para colaboração científica em projetos de pesquisa é um domínio que precisa de maior atenção em SR. Neste contexto, a recomendação de pesquisadores pode fomentar a colaboração científica, a qual por sua vez, proporciona diversos benefícios, como compartilhamento de conhecimento, desenvolvimento de instituições e suas regiões, aumento na rede de contatos dos pesquisadores, resolução de problemas complexos, etc. Assim sendo, este trabalho tem como objetivo o desenvolvimento e avaliação de um sistema de recomendação de pesquisadores com potencial de colaboração científica na UNIPAMPA. O método de recomendação proposto consiste na avaliação da similaridade de perfis de pesquisadores construídos a partir das palavras-chave e referências bibliográficas dos projetos de pesquisa cadastrados no Sistema de Informação para Projetos de Pesquisa, Ensino e Extensão (SIPPEE). O sistema foi avaliado através de um instrumento de avaliação no formato de um questionário, com os pesquisadores que utilizam a SIPPEE. Como resultado, identificou-se que os dados da SIPPEE podem ser utilizados para identificar pesquisadores com potencial de colaboração.

Palavras-chave: Sistemas de Recomendação, Colaboração Científica, Acoplamento Bibliográfico, Palavras-chave.

ABSTRACT

Recommender Systems (RS) are important tools to overcome the Web 2.0 information overload problem. Big companies such as Amazon, Netflix, Facebook, YouTube, are well known for using these mechanisms. In the academic field, recommendation systems are more popular in recommending scientific articles for activities such literature review (Scholar, Web of Science and Scopus). However the recommendation of researchers for scientific collaboration in research projects is a domain that needs more attention in RS. In this context, the recommendation of researchers can foster scientific collaboration, which in turn provides several benefits, such as knowledge sharing, development of institutions and their regions, increase the researcher's network, resolution of complex problems, etc. Therefore, the goal of this work is to develop and evaluate a scientific collaborator recommender system for scientific collaboration at UNIPAMPA. The proposed recommendation method consists in evaluating the similarity of researcher profiles constructed from the keywords and bibliographic references of the research projects registered in the Information System for Research, Teaching and Extension Projects (SIPPEE). The system was evaluated through an interview with researchers using SIPPEE. As a result, it has been identified that SIPPEE data can be used to identify researchers with potential for collaboration.

Key-words: Recommender Systems, Collaboration, Bibliographic Coupling, Key-words

LISTA DE FIGURAS

Figura 1 – Rede de Colaboração nos 10 campi da UNIPAMPA.	26
Figura 2 – Cálculo da previsão de avaliação do item 3 para o usuário A	29
Figura 3 – Esquema simples de recomendação baseada em conteúdo.	30
Figura 4 – Similaridade do Cosseno entre dois pesquisadores V e W.	33
Figura 5 – Acoplamento Bibliográfico entre dois documentos A e B.	34
Figura 6 – Acoplamento Bibliográfico entre dois pesquisadores A e B.	35
Figura 7 – Fluxograma do processo de revisão sistemática da literatura	37
Figura 8 – Visão geral do sistema de recomendação.	43
Figura 9 – Processo de obtenção dos arquivos PDF dos projetos de pesquisa da SIPPEE.	44
Figura 10 – Formulário para pesquisa de projetos na plataforma SIPPEE.	45
Figura 11 – Tabela de projetos de pesquisa da SIPPEE.	45
Figura 12 – Processo de construção de perfil dos pesquisadores.	46
Figura 13 – Processo de cálculo de similaridade entre os perfis dos pesquisadores.	47
Figura 14 – Processo de construção da rede de colaboração.	48
Figura 15 – Exemplo de grafo de colaboração gerado pelo scriptLattes e armaze- nado no Neo4j.	48
Figura 16 – Grafo global de recomendações.	50
Figura 17 – Grafo individual de recomendações.	50
Figura 18 – Informações do perfil do pesquisador J.K.	51
Figura 19 – Exemplo de recomendação extraída do grafo global de recomendações.	53
Figura 20 – Informações dos perfis de dois pesquisadores com perfis similares.	54
Figura 21 – Repostas da questão: Os pesquisadores recomendados trabalham em linhas/áreas de pesquisa em que sou capaz de contribuir.	55
Figura 22 – Repostas da questão: Descubri pesquisadores com potencial de colabo- ração que eu não conhecia anteriormente.	56
Figura 23 – Repostas da questão: Foram sugeridos diversos pesquisadores.	56
Figura 24 – Repostas da questão: Consegui perceber porque os pesquisadores foram recomendados para mim.	57
Figura 25 – Repostas da questão: Fiquei entusiasmado com as novas possibilidades de colaboração.	58
Figura 26 – Repostas da questão: Não encontrei dificuldades na utilização do sistema.	58
Figura 27 – Repostas da questão: A tarefa de encontrar colegas para colaborar em projetos ficou mais fácil de ser executada.	59
Figura 28 – Repostas da questão: Caso o sistema fosse implantado, utilizaria o mesmo frequentemente.	59

LISTA DE TABELAS

Tabela 1 – Número de projetos de pesquisa por campus	27
Tabela 2 – Trabalhos relacionados levantados durante o processo de revisão sistemática da literatura.	39

SUMÁRIO

1	INTRODUÇÃO	21
1.1	Objetivos	23
1.2	Justificativa	23
1.3	Organização do trabalho	24
2	FUNDAMENTAÇÃO TEÓRICA	25
2.1	Colaboração Científica	25
2.1.1	Colaboração Científica na UNIPAMPA	26
2.2	Sistemas de Recomendação	27
2.2.1	Filtragem Colaborativa (FC)	27
2.2.2	Baseado em Conteúdo	29
2.3	Identificando Especialistas Acadêmicos com Potencial de Co- laboração	32
2.3.1	Palavras-chave e Similaridade do Cosseno	32
2.3.2	Acoplamento Bibliográfico	33
2.4	Análise de Redes Sociais	35
2.4.1	Ator (vértice)	36
2.4.2	Atores Adjacentes	36
2.4.3	Atributos	36
2.4.4	Relação (aresta)	36
3	TRABALHOS RELACIONADOS	37
3.1	Questões de Pesquisa	37
3.2	Processo de Pesquisa	38
3.3	Definição dos Critérios de Inclusão e Exclusão	38
3.4	Resultado da Revisão	38
3.4.1	Visão geral dos trabalhos	39
3.5	Lições do capítulo	42
4	SISTEMA DE RECOMENDAÇÃO DE PESQUISADORES	43
4.1	Descrição Geral	43
4.1.1	Obtenção dos Arquivos dos Projetos	44
4.1.2	Extração dos Dados e Construção dos Perfis	45
4.1.3	Cálculo da Similaridade entre Perfis	46
4.1.4	Construção da Rede de Coautoria dos Pesquisadores	47
4.1.5	Interface Gráfica de Usuário	49
4.1.6	Avaliação da Proposta	51
5	RESULTADOS	53

5.1	Resultados do Instrumento de Avaliação	54
6	CONSIDERAÇÕES FINAIS	61
	REFERÊNCIAS	63
	APÊNDICES	67
	APÊNDICE A – CONVITE PARA AVALIAÇÃO DO SISTEMA DE RECOMENDAÇÃO	69
	APÊNDICE B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO	71
	APÊNDICE C – AVALIAÇÃO DO SISTEMA DE RECOMENDAÇÃO	73
	ANEXOS	75

1 INTRODUÇÃO

Grande parte das informações disponíveis na internet atualmente deriva da evolução da *Web 2.0*, onde cada vez mais o usuário final deixa de ser apenas consumidor de informação e passa a participar da rede mundial de computadores como um produtor ativo de informações. Assim sendo, o excesso de informações presente na internet tornou necessária a criação de mecanismos que auxiliem o usuário final a encontrar informações *online* que sejam relevantes para ele, ou que satisfazem suas preferências. Estes mecanismos são conhecidos como Sistemas de Recomendação (SR).

Diversos tipos de aplicações incorporam sistemas de recomendação em seu funcionamento atualmente. Os itens recomendados variam entre: produtos (Amazon), filmes (Netflix), relações de amizade (Facebook), artigos científicos (Mendeley), etc. Estes sistemas e os tipos de itens por eles recomendados são bastante populares e atraem muitos esforços voltados para construção de sistemas nesses domínios. Por outro lado, existem domínios onde o desenvolvimento de sistemas recomendação ainda carece de maior atenção.

No meio acadêmico por exemplo, onde os pesquisadores exercem diversas atividades como a criação de projetos de pesquisa, extensão, ensino, publicação de artigos, a colaboração com outros pares que compartilham interesse em tópicos/linhas de pesquisa é desejável. Segundo Sonnenwald (2007) a colaboração científica é uma das bases para a solução de problemas complexos e para o avanço e desenvolvimento de uma região. Nesse contexto sistemas de recomendação podem ser construídos com o objetivo de identificar potenciais colaboradores em projetos científicos.

De fato, é sabido que a colaboração entre pesquisadores obtêm como resposta melhores resultados na solução de problemas e avanços no desenvolvimento da pesquisa. Zuckerman (1967) ao entrevistar 41 premiados com prêmio Nobel de ciência, identificou uma forte relação entre colaboração e produtividade. Wuchty, Jones e Uzzi (2007) e Sonnenwald (2007) afirmam haver uma tendência na mudança do modelo de avanço científico individual para modelos baseados em equipes de trabalho e que trabalhos em equipe atualmente dominam a distribuição de citações (medida de qualidade) em 4 áreas de pesquisa específicas. Melin (2000), ao questionar pesquisadores sobre os efeitos de suas recentes colaborações, obteve respostas como: aumento de visibilidade, novos tipos de conhecimento e expansão e aprofundamento de contatos, sendo assim, percebe-se também a existência de consequências positivas a nível individual resultantes da colaboração científica.

Hara et al. (2003a) afirmam que, apesar de várias pesquisas demonstrarem a importância da colaboração, poucos pesquisadores têm dado atenção à definição de colaboração e que em muitos casos, o termo “colaboração” é utilizado intuitivamente como “cooperação” ou “coordenação”. Na literatura, existem diversas definições para o termo “colaboração”, dentre elas, Mattessich e Monsey (1992) definem “colaboração” como uma relação mútua e bem definida, feita por duas ou mais organizações visando atingir obje-

tivos em comum.

A Universidade Federal do Pampa (UNIPAMPA) é uma instituição de ensino superior (IES) criada em 2006 com uma estrutura multicampi formada por 10 campi localizados em cidades ou áreas de fronteira com o Uruguai e Argentina. Dentre os seus objetivos, está o fomento ao desenvolvimento da metade sul do Rio Grande do Sul, objetivo este que pode ser atingido através de resultados obtidos pelo fomento à colaboração entre os pesquisadores da instituição.

Kelm e Bordin (2017) analisaram a rede de colaboração da UNIPAMPA, verificando especificamente as relações de coautoria existentes entre os campus de Bagé e Uruguaiana. Através deste estudo os autores concluíram que com a análise da rede de colaboração “é possível diagnosticar e tomar decisões tanto em nível individual, o do pesquisador, como em nível gerencial, como o das comissões locais e da pró-reitoria de pesquisa, no sentido de fomentar a criação de novos laços de colaboração e fortalecer os laços existentes”.

Projetos de diversas naturezas executados por membros de IESs são comumente gerenciados por sistemas de informação internos às instituições, dessa forma, é preciso tentar garantir que essas informações sejam utilizadas da melhor forma possível em prol dos objetivos específicos de cada instituição. Dentro de sua gama de sistemas internos de gerenciamento, a UNIPAMPA possui o Sistema de Informação de Projetos de Pesquisa, Ensino e Extensão (SIPPEE), responsável por armazenar e gerenciar todos os projetos vinculados à universidade. A plataforma SIPPEE mantém dados de projetos de pesquisa, ensino e extensão, criados por pesquisadores dos 10 campi da UNIPAMPA a partir do ano de 2010.

Sistemas de recomendação baseados em conteúdo utilizam descrições ou informações textuais, como por exemplo palavras-chave e referências bibliográficas de projetos de pesquisa para recomendar itens que coincidam com os interesses de determinados perfis de usuário. Neste contexto, um sistema de recomendação baseado em conteúdo que, através de técnicas de medição de similaridade como a similaridade do cosseno e acoplamento bibliográfico, indique potenciais colaboradores a partir do conteúdo de projetos cadastrados na plataforma SIPPEE, pode beneficiar a comunidade acadêmica ao possibilitar a aproximação e possivelmente colaboração entre os membros universidade.

Como consequência, pode-se inferir como resultado um maior desenvolvimento científico da instituição e uma ferramenta que pode auxiliar a instituição a cumprir seu objetivo no desenvolvimento da região. Estas são algumas das principais contribuições de um sistema de recomendação aplicado ao meio acadêmico para a descoberta de colaboradores.

1.1 Objetivos

Desenvolver um sistema de recomendação de pesquisadores acadêmicos baseado em conteúdo, utilizando os projetos de pesquisa da plataforma SIPPEE.

Os objetivos específicos do trabalho são:

- Modelar perfis dos pesquisadores através das palavras-chave e referências bibliográficas dos projetos de pesquisa da plataforma SIPPEE;
- Implementar os algoritmos de similaridade entre perfis de pesquisadores;
- Avaliar a adequabilidade dos algoritmos de similaridade;
- Implementar interface *web* de visualização das recomendações através de uma rede interativa de recomendações;
- Avaliar o sistema e as recomendações geradas com pesquisadores que utilizam a plataforma SIPPEE.

1.2 Justificativa

O desenvolvimento da pesquisa científica, depende fortemente da troca de conhecimento e experiências entre pesquisadores. Sem essas trocas, a pesquisa em si se tornaria inviável. O processo de interação entre cientistas promove a criação e o compartilhamento de conhecimento (HECK, 2016). Uma rede de pesquisadores é diretamente influenciada pelo seu ambiente de trabalho e muitas vezes esse ambiente pode dificultar o desenvolvimento da pesquisa.

Suponha uma instituição de ensino multicampi como a UNIPAMPA, caso um pesquisador envolvido em um determinado projeto esteja à procura de outro(s) pesquisador(es) da instituição para contribuir(em) com seu trabalho, ele teria que a princípio possuir ao menos o conhecimento dos trabalhos realizados pelos seus colegas, para através de uma análise manual, determinar quais teriam o perfil que melhor se enquadra com as características do projeto.

Neste contexto, um sistema de recomendação de pesquisadores com potencial de colaboração em projetos de pesquisa dentro da instituição, diminui o esforço necessário para encontrar colegas adequados para trabalhar em equipe, contribuindo para o fomento da colaboração científica.

Além disso, um sistema de recomendação de potenciais colaboradores que contribua com o aumento da colaboração científica de uma instituição, pode reduzir os custos de pesquisa e melhorar a comunicação dentro de organizações, garantindo que o conhecimento produzido em uma organização ou instituição seja mais facilmente acessado e utilizado (XU et al., 2012). Assim sendo decisões de gestão relacionadas à promoção da

pesquisa podem ser feitas em direção ao fomento da colaboração científica, promovendo um maior desenvolvimento e visibilidade da região onde a instituição está localizada.

1.3 Organização do trabalho

Este trabalho está dividido em 5 capítulos:

- Capítulo 2: são introduzidos os conceitos e fundamentações base para o desenvolvimento do trabalho;
- Capítulo 3: é feito um levantamento de trabalhos relacionados à construção do sistema de recomendação aqui proposto;
- Capítulo 4: neste capítulo toda a metodologia para o desenvolvimento do sistema é apresentada;
- Capítulo 5: são apresentados os resultados obtidos do desenvolvimento do sistema de recomendação e das avaliações do ponto de vista do usuário;
- Capítulo 6: neste capítulo são apresentadas as considerações finais do trabalho e aspectos relacionados a trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, são abordados conceitos e aspectos teóricos necessários para o desenvolvimento do trabalho. A seção 2.1 introduz conceitos relacionados à colaboração científica. A subseção 2.1.1 exhibe dados a respeito da colaboração entre os 10 campi da UNIPAMPA. A seção 2.2 introduz ao leitor o funcionamento de sistemas de recomendação. A subseção 2.2.1 explica o conceito de filtragem colaborativa e como ela é aplicada em sistemas de recomendação. A subseção 2.2.2 aborda o método de recomendação baseado em conteúdo. Na seção 2.3 são introduzidos aspectos relacionados à recomendação de especialistas acadêmicos com potencial de colaboração. Por fim, a seção 2.4 define alguns conceitos básicos sobre análise de redes sociais (ARS).

2.1 Colaboração Científica

O termo “colaboração científica”, com um certo nível de abstração, pode ser definido como um relacionamento estabelecido entre dois ou mais indivíduos com intuito de através do compartilhamento de conhecimento e divisão de tarefas atingir um objetivo em comum. Bozeman, Fay e Slade (2013) definem colaboração científica como um “processo social onde humanos dispõem de seu capital humano com o objetivo de produzir conhecimento”.

Em Gordon (1980), o autor demonstra uma relação relevante entre artigos com diversos níveis de autoria (número de autores) e suas respectivas submissões em um *journal* reconhecido na área de astronomia. Pela data da publicação, podemos verificar que a colaboração científica não é um conceito novo, mas sim uma tendência que vem sendo observada já há algumas décadas. Katz e Martin (1997) reportam que diversos estudos realmente indicam que alta produtividade (no que diz respeito a publicações de artigos) está de fato relacionada com altos níveis de colaboração, além disso, o fato da colaboração científica ser um fenômeno observado há décadas, também é reforçado pelos autores.

Por outro lado, Katz e Martin (1997) consideram que a maioria dos esforços realizados com intuito de observar o crescimento do fenômeno da colaboração científica são baseados nas relações de coautoria entre pesquisadores, e que nem sempre coautoria deve ser considerado um sinônimo de colaboração. Esta observação é feita devido a diversos aspectos sociais que muitas vezes são considerados para a definição dos papéis de autoria que cada autor irá exercer em publicações. De qualquer forma, os autores afirmam que a coautoria é uma métrica de muito mais fácil acesso do que a avaliação de aspectos sociais, e que devido a grande quantidade de informação disponível, esta se torna uma métrica com valor estatístico aceitável para mensurar a colaboração.

Para Hara et al. (2003b), a colaboração é um componente de pesquisa crítico no mundo da “*big science*”, o qual envolve projetos de grande porte, dominados por problemas complexos, rápidas mudanças de tecnologia, crescimento dinâmico do conhecimento e conhecimento altamente especializado, além disso, colaboração e comunicação fazem parte

da base do trabalho científico de um pesquisador.

2.1.1 Colaboração Científica na UNIPAMPA

Bordin e Kelm (2016) desenvolveram um estudo através do método de análise de redes de coautoria a fim de verificar a situação referente a colaboração científica entre pesquisadores dos campi da instituição. O estudo foi desenvolvido em três etapas: coleta dos currículos, processamento dos dados e análise dos resultados.

Na etapa de coleta, realizada no período entre Novembro de 2015 e Fevereiro de 2016, foram coletados 706 currículos de pesquisadores dos 10 campi da instituição. Na etapa de processamento, as relações de coautoria foram identificadas através da ferramenta scriptLattes de Mena-Chalco e Junior (2009). Na análise dos resultados, foi identificado o campus com mais atividade colaborativa, bem como os laços de colaboração intercampi. A Figura 1 ilustra a rede de colaboração de todos os dez campi da UNIPAMPA no período mencionado.

Figura 1 – Rede de Colaboração nos 10 campi da UNIPAMPA.



Fonte: o próprio autor

A análise das relações de coautoria intercampi demonstrou que o campus de Uruguaiana possui o maior número de relações de coautoria com outros campi.

Na análise individual de cada campus, foi verificado que o campus Uruguaiana também tem o maior grau médio de colaboração em relação a todos os outros campi. Adicionalmente, podemos também verificar na Tabela 1 que o campus de Uruguaiana possui também o maior número de projetos de pesquisa cadastrados na plataforma SIPPEE, além disso, também em Uruguaiana está situado o colaborador com maior grau de colaboração na análise individual dos pesquisadores.

Tabela 1 – Número de projetos de pesquisa por campus em 2018.

Campus	Número de Projetos
Alegrete	246
Bagé	332
Caçapava do Sul	100
Dom Pedrito	184
Itaqui	303
Jaguarão	106
Santana do Livramento	143
São Borja	222
São Gabriel	149
Uruguaiana	824

2.2 Sistemas de Recomendação

Sistemas de recomendação (SR) coletam informações sobre preferências de usuários a respeito de um conjunto de itens (filmes, músicas, livros, piadas, artigos científicos, destinos de viagens, etc) de forma explícita, perguntando ao usuário diretamente sobre suas preferências ou implícita, através da observação das ações do usuário, como livros lidos, músicas ouvidas, páginas visitadas, etc (BOBADILLA et al., 2013).

As informações sobre as preferências do usuário são utilizadas para a construção de seus perfis, e a partir destes, são aplicadas técnicas de recomendação para prever e recomendar novos itens provavelmente desejáveis aos usuários.

Os principais métodos de recomendação são filtragem colaborativa, fundamentada na ideia de que se usuários compartilham interesses no passado, eles também terão interesses em comum no futuro; baseada em conteúdo, na qual descrições sobre itens e perfis de usuários são utilizados para a recomendação e métodos híbridos, combinando os dois anteriores. A subseção 2.2.1 detalha o funcionamento da recomendação por filtragem colaborativa. A subseção 2.2.2 descreve o funcionamento de sistemas de recomendação baseados em conteúdo e a seção 2.3 introduz aspectos e abordagens relacionadas a identificação de especialistas acadêmicos para potencial colaboração.

2.2.1 Filtragem Colaborativa (FC)

Este método de recomendação baseia-se na ideia de que se usuários compartilham interesses no passado, eles também terão interesses em comum no futuro (JANNACH et

al., 2010). Um sistema de FC computa suas recomendações explorando os relacionamentos e similaridades entre itens e entre usuários. Estas relações podem ser extraídas através da interação dos usuários com os itens gerenciados pelo sistema (ELAHI; RICCI; RUBENS, 2016).

Por exemplo, suponha que dois usuários A e B de uma loja virtual têm um histórico de compras semelhante e que recentemente o usuário A adquiriu um item que o usuário B ainda não possui. Um sistema de recomendação que utiliza o método FC certamente recomendaria este item ao usuário A (JANNACH et al., 2010).

Uma das técnicas mais utilizadas neste tipo de sistema é a baseada na “melhor vizinhança (*Neighborhood based*)”. Esta técnica faz uma previsão do quanto um item ainda não adquirido por um usuário alvo de recomendações seria de interesse do mesmo com base nos itens avaliados ou adquiridos por usuários semelhantes (vizinhos).

Um conjunto de usuários semelhantes ao usuário alvo pode ser determinado de várias formas, uma técnica popularmente utilizada é o coeficiente de correlação de Pearson (*Pearson’s correlation coefficient*) (ELAHI; RICCI; RUBENS, 2016), na qual a similaridade $sim(u, v)$ entre dois usuários u e v , é calculada através da Equação 2.1:

$$sim(u, v) = \frac{\sum_{i \in I_{uv}} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I_{uv}} (r_{ui} - \bar{r}_u)^2 (r_{vi} - \bar{r}_v)^2}} \quad (2.1)$$

Na qual I_{uv} representa o conjunto de itens avaliados por usuários u e v , e \bar{r}_u a avaliação média do usuário u . Alguns usuários costumam avaliar itens de forma mais positiva (maior nota) do que outros usuários, e vice-versa, por isso, a avaliação média do usuário é subtraída para levar em conta o fato de que dois usuários podem ter padrões de avaliação semelhantes mesmo com valores distintos.

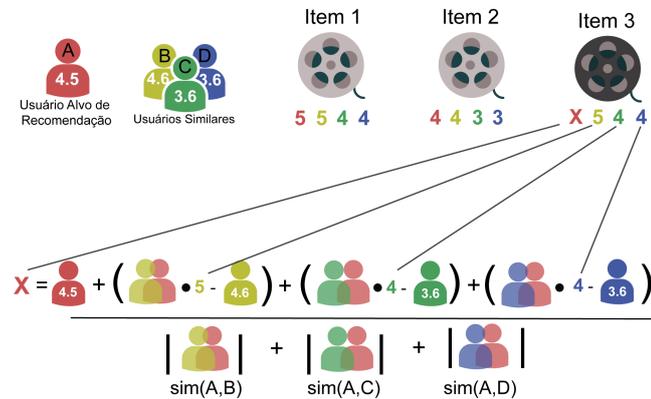
Com base nos vizinhos mais próximos (*nearest neighbors*), determinados pela Equação 2.1, pode-se calcular a avaliação estimada de um item para um usuário alvo de recomendações. Por exemplo, seja $r_{ui} \in \{1, \dots, 5\}$ a popular avaliação 5 estrelas de um usuário u sobre um item i , então a avaliação estimada de um usuário u sobre um item i pode ser calculada através da Equação 2.2:

$$\hat{r}_{ui} = \bar{r}_u + \frac{\sum_{v \in N_i(u)} sim(u, v)(r_{vi} - \bar{r}_v)}{\sum_{v \in N_i(u)} |sim(u, v)|} \quad (2.2)$$

A Figura 2 ilustra o cálculo da predição da avaliação feita pelo usuário A para o item 3 com base nas avaliações feitas por 3 usuários semelhantes ao usuário A (B, C e D), na qual a avaliação média dos usuários para todos os itens consta na imagem que o representa, e as avaliações de cada usuário para cada item estão localizadas abaixo dos itens.

A principal vantagem da recomendação baseada em filtragem colaborativa é que não é necessária uma especificação ou descrição detalhada dos itens. Por outro lado uma de suas principais desvantagens é o problema conhecido como *cold-start* (começo frio).

Figura 2 – Cálculo da previsão de avaliação do item 3 para o usuário A



Fonte: o próprio autor

Este problema se deve ao fato de que novos itens que ainda não foram avaliados podem ficar fora das recomendações, ou ainda, usuários sem um histórico significativo dificultam a geração de usuários semelhantes.

2.2.2 Baseado em Conteúdo

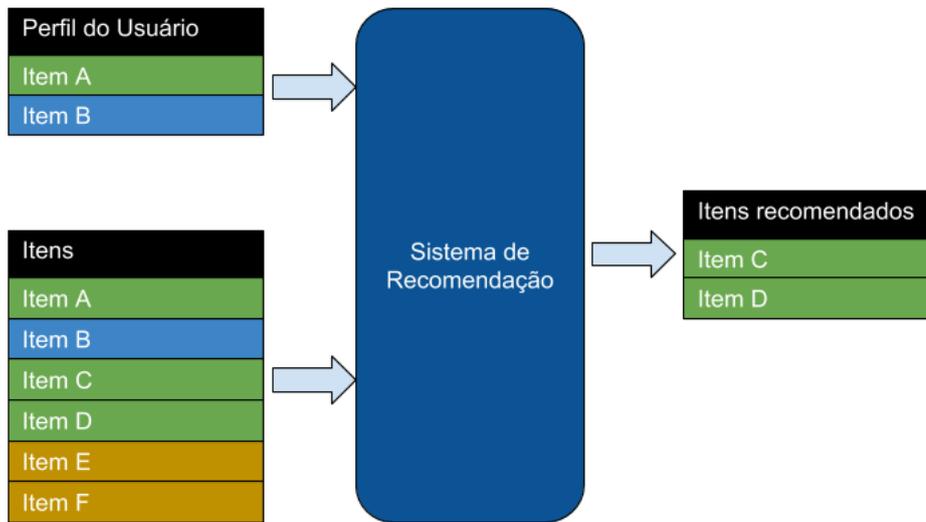
Este método de recomendação baseia-se na utilização da descrição de itens e na modelagem de perfis de usuário, construídos com o objetivo de descrever de algum modo os interesses do usuário em determinados tipos de itens do sistema. Com isso, a tarefa de recomendação basicamente se resume a determinar quais itens ainda não adquiridos pelo usuário coincidem com os seus interesses (histórico de itens adquiridos) (JANNACH et al., 2010).

Por exemplo: se um usuário de uma loja virtual adquiriu alguns filmes de ficção no passado, então um sistema de recomendação baseado em conteúdo (*content-based*) provavelmente irá recomendar novos filmes de ficção que o usuário ainda não adquiriu nessa loja. Outros exemplos típicos de sistemas de recomendação baseados em conteúdo são sistemas que recomendam artigos de notícias comparando as palavras-chave de um artigo de notícia em questão com as palavras-chave de outros artigos de notícia que um usuário já tenha acessado previamente. Neste tipo de sistema, os itens a serem recomendados são comumente chamados de “documentos” (BOBADILLA et al., 2013).

A Figura 3 ilustra um esquema simples de recomendação baseado em conteúdo, onde itens semelhantes são exibidos na mesma cor, e os itens previamente adquiridos pelo usuário compõem seu perfil.

Em sistemas baseados em conteúdo, os documentos são popularmente representados por modelos criados à partir de seu conteúdo, e os modelos de usuários são representados pelas características dos itens que o usuário já interagiu (comprou, leu, baixou,

Figura 3 – Esquema simples de recomendação baseada em conteúdo.



Fonte: o próprio autor

ouviu, etc). O modelo de representação de documentos e de perfis de usuários pode ser construído de várias maneiras, desde uma lista simples de termos de sua descrição, até uma representação no espaço vetorial (*Vector Space Model* ou VSM)(BEEL et al., 2016).

A medida de similaridade entre modelos de documentos baseados em listas de palavras-chave pode ser obtida através do coeficiente de Dice (*Dice Coefficient*). Considere $palavraschave(a)$ como sendo todas as palavras-chave de um documento a então a similaridade entre dois documentos a e b pode ser determinada através da Equação 2.3:

$$\frac{2 |palavraschave(a) \cap palavraschave(b)|}{|palavraschave(a)| + |palavraschave(b)|} \quad (2.3)$$

Os modelos de documentos e perfis de usuário no espaço vetorial podem ser representados pela frequência com que os termos (palavras-chave, *tokens*, etc) mais relevantes aparecem nos documentos, onde cada dimensão dos vetores consiste na medida de relevância ou peso de cada termo. Para calcular a relevância de um termo em uma coleção de documentos, utiliza-se a técnica TF-IDF (abreviação do inglês *term frequency–inverse document frequency*). Quando os dados são representados através de vetores, a medida de similaridade entre os mesmos pode ser obtida através da técnica similaridade do cosseno (do inglês *cosine similarity*).

TF-IDF é uma técnica derivada das áreas de mineração de texto e recuperação de informação, utilizada para mensurar a importância de uma palavra dentro de uma coleção ou corpus de documentos (ROBERTSON, 2004). Assumindo que a importância de uma palavra está relacionada com o número de vezes que a mesma ocorre, o valor de TF-IDF aumenta proporcionalmente ao número de vezes que uma palavra aparece

em um documento e tem seu valor ajustado de acordo com a frequência da palavra nos demais documentos do corpus, prevenindo com que palavras muito frequentes (não tão importantes) em todos os documentos interfiram na sua precisão.

A frequência de uma palavra em um documento (tf) é calculada pela Equação 2.4 (JANNACH et al., 2010). Seja d um documento qualquer e t um termo ou palavra qualquer pertencente ao documento d :

$$tf(t, d) = \frac{f_d(t)}{w \in d \max f_d(w)} \quad (2.4)$$

Onde $f_d(t)$ retorna a frequência da palavra t no documento d e $w \in d \max f_d(w)$ retorna a maior frequência entre as demais palavras w no documento d . Como documentos mais extensos tendem a conter uma maior frequência de palavras, esta divisão é feita a fim de normalizar a frequência das palavras com o tamanho do documento.

O cálculo da frequência inversa do documento (do inglês, *Inverse Document Frequency*, ou *IDF*) é realizado pela Equação 2.5 da seguinte forma: seja t um termo ou palavra qualquer e D uma coleção ou corpus de documentos.

$$idf(t, D) = \ln \left(\frac{|D|}{|d \in D : t \in d|} \right) \quad (2.5)$$

O número total de documentos é dividido pelo número de documentos em D que possuem o termo ou palavra t , e por fim, aplica-se a função logarítmica para alterar a escala do valor. O resultado final ($tfidf$, Equação 2.6) é obtido multiplicando o valor de tf (Equação 2.4) por idf (Equação 2.5).

$$tfidf(t, d, D) = tf(t, d) \cdot idf(t, D) \quad (2.6)$$

A medida de similaridade entre vetores conhecida como similaridade do cosseno, é dada através do cálculo do produto interno dos mesmos. Ao representarmos informações no espaço vetorial, podemos utilizar o cosseno do ângulo formado entre eles para comparar suas orientações. O cosseno de 0 é 1, ou seja, se o ângulo formado entre dois vetores é igual a 0 (paralelos) então eles são avaliados como “idênticos”, qualquer outro grau formado entre os vetores, resulta em um valor menor do que 1.

Dois vetores ortogonais tem similaridade 0 enquanto que vetores opostos tem similaridade igual a -1. Como apenas a orientação dos vetores é levada em consideração, fica claro que para esta medida, a magnitude dos vetores é irrelevante para o cálculo. Dado dois vetores de dimensão N , v e w , a similaridade entre eles pode ser calculada pela Equação 2.7.

$$\text{cosinSim}(\vec{v}, \vec{w}) = \frac{\vec{v} \bullet \vec{w}}{|\vec{v}| |\vec{w}|} = \frac{\sum_{i=1}^N v_i \times w_i}{\sqrt{\sum_{i=1}^N v_i^2} \sqrt{\sum_{i=1}^N w_i^2}} \quad (2.7)$$

2.3 Identificando Especialistas Acadêmicos com Potencial de Colaboração

Identificar especialistas acadêmicos com potencial para colaboração, consiste basicamente na tarefa de encontrar perfis de usuário que demonstrem de alguma forma, possuir interesses comuns em tópicos ou áreas de pesquisa. Uma das principais abordagens para a recomendação de potenciais colaboradores no meio acadêmico é baseada no conteúdo de suas produções (*expertise*).

Em Makarov, Bulanov e Zhukov (2016) os autores definem a similaridade de interesses entre dois autores com base no número de áreas comuns de publicação em revistas. No trabalho de Davoodi, Kianmehr e Afsharchi (2013), os autores modelam perfis de usuário com base em atributos como experiências de mercado de trabalho, histórico educacional, interesses e habilidades através da extração de diversas fontes *online*. Lopes et al. (2010) modelam perfis de pesquisadores através de vetores contendo a frequência em que pesquisadores publicam em determinadas áreas de pesquisa, onde a similaridade entre esses perfis é obtida através da similaridade do cosseno.

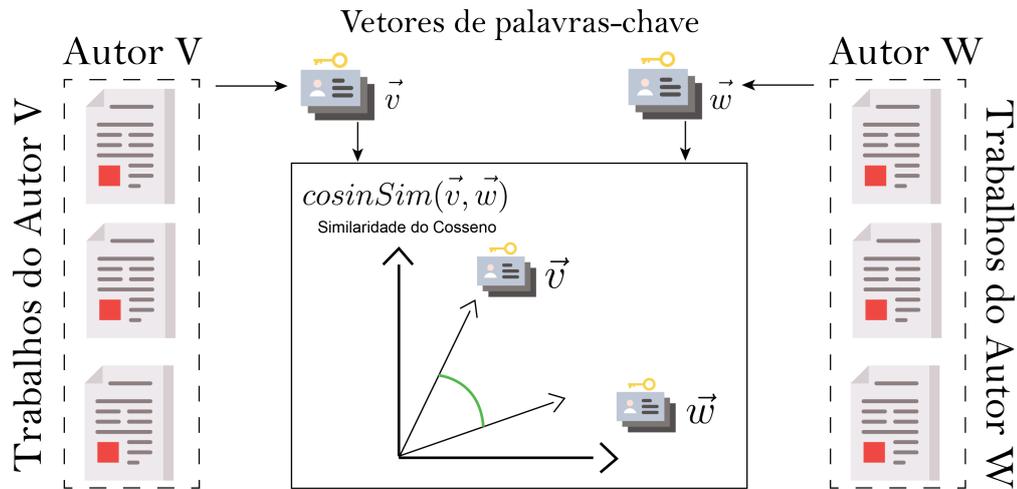
As próximas subseções (subseção 2.3.1, subseção 2.3.2), apresentam mecanismos para identificar pesquisadores com potencial de colaboração, com base nas palavras-chave e referências bibliográficas de suas obras.

2.3.1 Palavras-chave e Similaridade do Cosseno

Na subseção 2.2.2 foram apresentados métodos para identificar a similaridade entre documentos com base na relevância de seus termos ou palavras-chave. As palavras-chave de um trabalho de pesquisa tem como objetivo transmitir ao leitor, de uma forma sucinta, a temática do trabalho, assim como auxiliar ferramentas indexadoras e motores de busca a encontrar a publicação, colaborando em aspectos como alcance e citações do trabalho.

Palavras-chave são também utilizadas para identificar especialistas acadêmicos com potencial de colaboração. Ao estender o conceito de similaridade entre documentos criando vetores compostos pelas frequências das palavras-chave de todos os trabalhos dos pesquisadores, torna-se possível realizar o cálculo de similaridade entre pesquisadores através de técnicas como a similaridade do cosseno (Equação 2.7), como ilustrado na Figura 4.

Figura 4 – Similaridade do Cosseno entre dois pesquisadores V e W.



Fonte: o próprio autor.

2.3.2 Acoplamento Bibliográfico

Kessler (1962) foi o primeiro pesquisador a perceber que um número de artigos científicos tem uma relação significativa entre si quando estes possuem uma ou mais referências bibliográficas em comum. Dessa forma, a técnica Acoplamento Bibliográfico (AB) é utilizada para estabelecer uma relação de similaridade entre documentos utilizando a estrutura bibliográfica dos mesmos (BOYACK; KLAVANS, 2010). Quanto maior for o número de referências bibliográficas compartilhadas entre dois documentos, maior será a probabilidade dos mesmos tratarem de temas semelhantes. A Figura 5 ilustra o funcionamento da técnica.

Uma forma de se obter o valor AB entre dois documentos é através do coeficiente de Jaccard. Esta é uma técnica de medição de similaridade baseada na intersecção de elementos em conjuntos. A medição é feita utilizando a quantidade de elementos que os conjuntos compartilham e a quantidade total de elementos dos dois conjuntos.

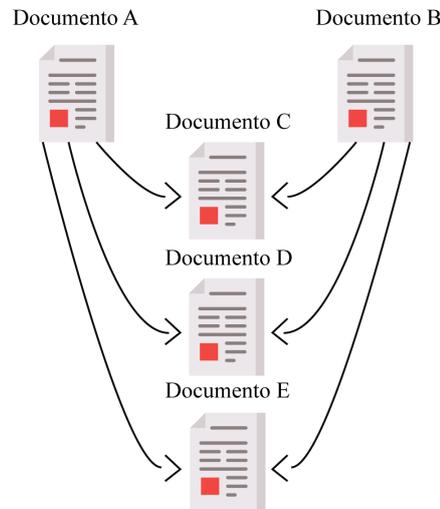
O leque de utilização desta técnica é muito amplo, podendo ser aplicada na medição de similaridade entre palavras-chave, remoção de ambiguidade de palavras, entre outras medições de similaridade onde as informações possam ser representadas como conjuntos. O coeficiente de Jaccard é definido pela Equação 2.8.

$$jaccard(A, B) = \frac{|A \cap B|}{|A| + |B| - |A \cap B|} \quad (2.8)$$

Através do coeficiente de Jaccard, o cálculo do AB entre dois documentos A e B pode ser obtido por meio da Equação 2.9:

$$bibCoup(A, B) = \frac{|ref(A) \cap ref(B)|}{|ref(A)| + |ref(B)| - |ref(A) \cap ref(B)|} \quad (2.9)$$

Figura 5 – Acoplamento Bibliográfico entre dois documentos A e B.



Fonte: o próprio autor.

Onde $ref(A)$ e $ref(B)$ são os conjuntos de referências bibliográficas dos documentos A e B , respectivamente.

Uma das principais características do método AB é a estaticidade temporal, ou seja, o valor de AB entre documentos não se altera no decorrer do tempo, uma vez que um trabalho, depois de publicado, não sofre alterações em suas referências bibliográficas.

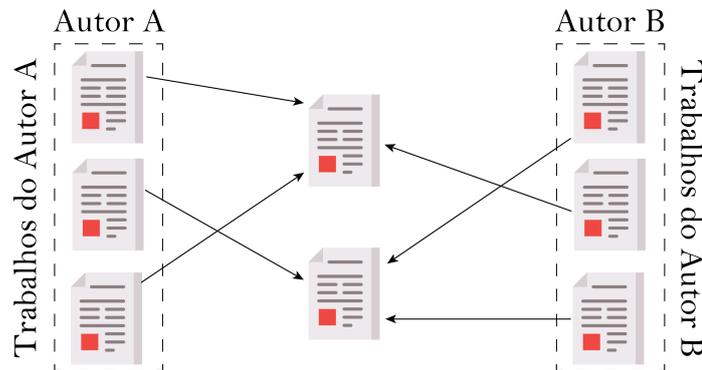
O Método Acoplamento Bibliográfico de Autor (ABA), é uma extensão do AB tradicional proposto inicialmente por Zhao e Strotmann (2008a), que visa analisar de maneira mais realista o estado corrente de atividades de pesquisa através do mapeamento de autores ativos em um determinado campo de pesquisa.

Diferentemente da técnica AB, a técnica ABA dá suporte ao estudo de mudanças nos campos de pesquisa ao longo do tempo, já que seu valor é obtido através do número de referências em comum entre todas as referências dos trabalhos de dois autores, podendo ser alterada com o tempo (desde que pelo menos um dos autores esteja ativo e publicando). Dessa forma, a técnica possui potencial para identificar interesses comuns em áreas/tópicos de pesquisa entre pesquisadores. A Figura 6 ilustra o funcionamento da técnica ABA.

Outra diferença importante entre as duas técnicas é a forma como o valor do acoplamento pode ser obtido, sendo que na técnica ABA o cálculo é diferente da AB tradicional. Ma (2012) apresenta três possíveis maneiras de se obter o valor (ou “força”) de ABA entre dois autores A e B :

- **Método simples.** Criam-se dois conjuntos $s1$ e $s2$, contendo todas as referências dos trabalhos dos autores A e B respectivamente, onde o valor ABA é o número de intersecções entre $s1$ e $s2$, ignorando reincidências de referências em $s1$ ou $s2$.

Figura 6 – Acoplamento Bibliográfico entre dois pesquisadores A e B.



Fonte: o próprio autor.

- **Método mínimo.** Criam-se dois conjuntos s_1 e s_2 , contendo todas as referências dos trabalhos dos autores A e B respectivamente, onde o valor ABA é a soma da menor frequência em que cada referência aparece nos dois conjuntos. Ex: se uma referência aparece 2 vezes em s_1 e 3 vezes em s_2 , o peso dessa referência na soma total será 2.
- **Método combinado.** Assim como no método mínimo, também considera referências repetidas, porém, ao invés de utilizar o menor valor de ocorrências de cada uma das referências, multiplica-se o número de vezes que uma referência aparece nos dois conjuntos. Utilizando o mesmo exemplo, o peso da referência na soma total seria 6 ao invés de 2.

Em seu experimento, o autor afirma que o método mínimo é o mais apropriado para obter o valor de ABA entre dois pesquisadores, Zhao e Strotmann (2008b) também utilizam um método similar ao método mínimo para o cálculo em seu trabalho.

2.4 Análise de Redes Sociais

Análise de redes sociais é o estudo de entidades sociais (pessoas em uma organização, empresas, pesquisadores, etc) e seus relacionamentos. As interações e relacionamentos podem ser representados através de grafos, onde cada vértice (ou nó) representa um ator, e cada aresta representa um relacionamento. A partir dessa rede, podemos estudar as propriedades de sua estrutura, avaliando aspectos de cada ator. Ainda sobre uma mesma rede, podem ser criados vários tipos de subgrafos, por exemplo: comunidades formadas por grupos de atores (LIU, 2007).

Nesta seção são introduzidos alguns conceitos básicos para o estudo de redes sociais. Como a teoria dos grafos está intimamente relacionada com ARS estes conceitos em sua grande maioria são conceitos pertencentes às estruturas de grafos.

2.4.1 Ator (vértice)

Atores são basicamente as entidades envolvidas em uma rede. Exemplos de atores são: indivíduos, empresas, pesquisadores em uma rede de coautoria, etc. Estas entidades são representadas como sendo os vértices em grafos que representam a rede social.

2.4.2 Atores Adjacentes

São aqueles atores que possuem algum tipo de relacionamento direto com outro ator qualquer, ou seja, a distância (número de arestas) entre os vértices que representam os atores é igual a 1.

2.4.3 Atributos

Os atributos de um ator são suas características individuais (idade, número de empregados, áreas de atuação, etc).

2.4.4 Relação (aresta)

É a ligação entre um par de atores pertencentes a uma rede, ligação esta que contém um conjunto de atributos que descrevem a relação entre os atores. Esta propriedade pode assumir dois valores, direcional ou não direcional. Uma relação direcional possui um ator como transmissor da relação e o outro ator como receptor, a relação não direcional ocorre quando a relação é recíproca.

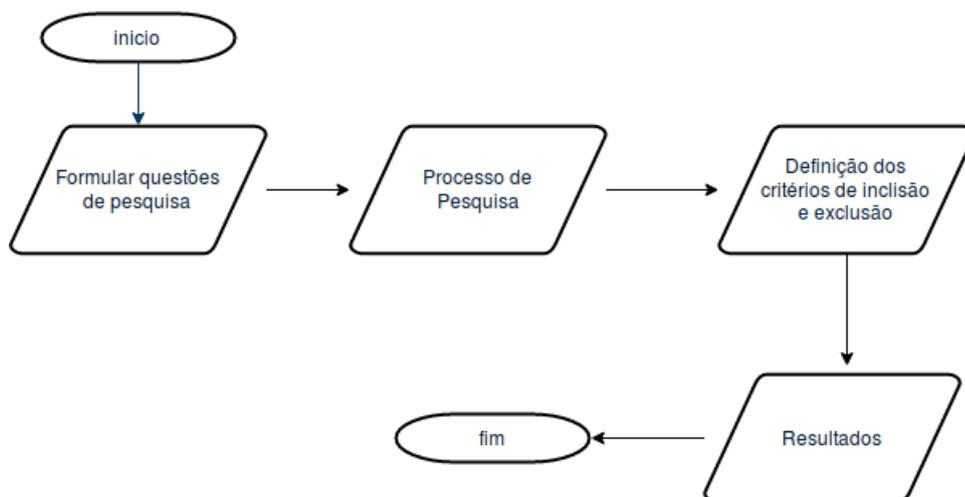
3 TRABALHOS RELACIONADOS

A pesquisa de trabalhos relacionados têm como propósito geral, verificar o estado atual da literatura de um determinado tópico ou área de pesquisa através da análise de documentos científicos relacionados.

Para o levantamento de trabalhos relacionados, foi realizada uma revisão sistemática da literatura (*Systematic Literature Review* ou SLR). Uma SLR é uma revisão metodologicamente rigorosa e tem como objetivo não só coletar todas as evidências existentes sobre uma questão de pesquisa, mas também apoiar o desenvolvimento de orientações baseadas em evidências (KITCHENHAM et al., 2009).

Durante a execução do processo de revisão, as seguintes etapas foram executadas: processo de pesquisa, definição dos critérios de inclusão e exclusão, extração de informações e resultados. O fluxograma da Figura 7 descreve o passo a passo do processo de revisão sistemática que foi adaptado de Kitchenham et al. (2009) e utilizado na revisão de literatura.

Figura 7 – Fluxograma do processo de revisão sistemática da literatura



Fonte: o próprio autor

3.1 Questões de Pesquisa

As questões de pesquisas (QP) são definidas objetivando obter-se uma visão geral dos aspectos que devem ser abordados pelos trabalhos relacionados. A questão de pesquisa geral deste trabalho é: Como é feita atualmente a recomendação de pesquisadores com potencial de colaboração? Com base nesta questão de pesquisa geral, as seguintes questões de pesquisa foram levantadas:

- QP1 - Qual a fonte de dados utilizada?

- QP2 - Como as técnicas foram utilizadas?
- QP3 - Como os dados são armazenados e exibidos?
- QP4 - Como o sistema foi avaliado?

3.2 Processo de Pesquisa

Como base de dados para a busca de trabalhos relacionados, foi utilizada a biblioteca digital *Google Scholar*. Com base nas questões de pesquisa, a *string* de busca definida para a pesquisa na biblioteca digital foi definida como: **expert or academic “recommender systems” “bibliographic coupling”**

Como resultado da busca na base de dados da biblioteca digital, 123 documentos foram retornados. Estes resultados passaram por um processo de filtragem definido através de critérios de inclusão e exclusão descritos na seção 3.3.

3.3 Definição dos Critérios de Inclusão e Exclusão

Os critérios para inclusão de trabalhos resultantes do processo de pesquisa são:

- Trabalhos onde a técnica *Bibliographic Coupling* foi utilizada;
- Trabalhos que relatam a implementação de sistemas de recomendação de especialistas no ambiente acadêmico;

Os critérios de exclusão são:

- Trabalhos que relatam a implementação de sistemas de recomendação de artigos científicos, ao invés de especialistas;
- Patentes;
- Citações;
- Revisões de literatura.

3.4 Resultado da Revisão

Dentre os trabalhos resultantes da pesquisa, 5 permaneceram após passar pelo processo de filtragem com base nos critérios de exclusão e inclusão definidos acima. As informações: título, autores e ano de publicação do trabalhos remanescentes são exibidos na Tabela 2.

Quando um trabalho atende, ou seja, responde a uma questão de pesquisa definida na seção 3.1, a coluna referente a questão de pesquisa (QP) é destacada com um “X”. O valor “-” na coluna de questões de pesquisa, significa que o trabalho respondeu

parcialmente à questão. Se a coluna estiver em branco, isto significa que o trabalho não atendeu ou respondeu a questão de pesquisa.

Tabela 2 – Trabalhos relacionados levantados durante o processo de revisão sistemática da literatura.

Título	Autores	Ano	QP1	QP2	QP3
Expert recommender systems: Establishing Communities of Practice based on social bookmarking systems	Tamara Heck; Isabella Peters	2010	X		-
Testing collaborative filtering against co-citation analysis and bibliographic coupling for academic author recommendation	Tamara Heck; Isabella Peters; Wolfgang G. Stock.	2011	X	X	-
Expert recommendation for knowledge management in academia	Tamara Heck; Isabella Peters; Wolfgang G. Stock.	2011	X	X	-
Recommendation for Social Networking in Academia	Tamara Heck.	2012	X	X	-
Expert Recommendation for the Academic Field	Tamara Heck.	2016	X	X	-

A subseção 3.4.1 apresenta uma visão geral sobre os trabalhos relacionados e também descreve como cada trabalho se relaciona com as questões de pesquisa levantadas.

3.4.1 Visão geral dos trabalhos

Heck e Peters (2010) propuseram um sistema de recomendação baseado em um sistema de *social bookmarking* chamado *CiteULike (CUL)*, para ajudar usuários a encontrar comunidades de práticas, onde pessoas compartilham o mesmo interesse e dão suporte umas às outras em seus trabalhos ou áreas de estudo. O sistema é baseado em filtragem colaborativa, feita em cima dos *bookmarks* dos usuários do CUL, e também sob as *tags* que os usuários usam para marcar os *bookmarks*. Os autores concluíram que para uma recomendação de especialistas mais adequada, mais usuários teriam que ser ativos no sistema de *social bookmarking*.

Heck, Peters e Stock (2011) através da extração de informações vindas de três bases de dados distintas: *Web of Science (WoS)*, *Scopus* e o serviço de *social bookmarking CiteULike*, geraram recomendações de autores que poderiam ser potenciais colaboradores de um cientista alvo de recomendações. A ideia principal do trabalho é avaliar a relevância de um grupo de autores recomendado através de métodos de filtragem colaborativa e como os resultados diferem em comparação às técnicas de *bibliographic coupling* e *co-citation*.

Um primeiro resultado é exibido e adicionalmente é feita uma avaliação explícita por cientistas parceiros do trabalho.

Heck, Hanraths e Stock (2011) desenvolveram um sistema de recomendação utilizando as mesmas técnicas de seu trabalho anterior Heck, Peters e Stock (2011), com o objetivo de encontrar potenciais colaboradores para um pesquisador à procura de colegas para trabalho em equipe. O diferencial deste trabalho para o anterior do mesmo ano, foi o objetivo das avaliações, onde Heck, Peters e Stock (2011) foram mais focado na comparação entre os resultados das diferentes técnicas aplicadas sob as bases de dados e neste o foco maior foi a avaliação da relevância das recomendações para cada entrevistado..

Heck (2012) é um capítulo do livro *Library and Information Science, Volume 5*, e trata-se de uma compilação de trabalhos anteriores (HECK; PETERS; STOCK, 2011; HECK; HANRATHS; STOCK, 2011). No capítulo, a autora argumenta sobre a utilização do método de filtragem colaborativa através de sistemas de *social bookmarking (CiteU-Like)* como sendo uma alternativa para a falta de informações de citação (*co-citation*) e referências em comum (*Bibliographic Coupling*) referentes a pesquisadores que estão iniciando em um uma área de pesquisa específica. Isto se deve ao fato de que pesquisadores novos em uma área, dificilmente são citados por outros trabalhos, ou por falta de publicações na área ou pelo tempo de resposta que uma publicação pode ter.

Heck (2016) sumariza todo seu estudo prévio na área de recomendação de pesquisadores para colaboração científica. Neste trabalho a autora sumariza o estado da pesquisa, introduz conceitos relacionados a colaboração científica. Apresenta as técnicas utilizadas nos demais trabalhos aqui citados, como por exemplo a combinação dos métodos de filtragem colaborativa com os métodos *co-citation* e *bibliographic coupling* aplicadas em 3 principais bases de dados utilizadas em seus trabalhos anteriores e no serviço de *social bookmarking CiteULike* e finaliza sumarizando resultados das entrevistas aplicadas em suas publicações. A seguir, as questões de pesquisa levantadas são abordadas para cada trabalho relacionado.

QP1: Qual a fonte de dados utilizada?

Heck e Peters (2010) criaram uma base de dados a partir da coleta de dados do serviço de *social bookmarking CiteULike (CUL)*. CUL é um serviço gratuito para gerenciamento e descoberta de novas referências, que podem ser compartilhadas com todos os usuários do serviço.

Heck, Peters e Stock (2011), Heck, Hanraths e Stock (2011), Heck (2012), utilizaram bases de dados de 3 serviços *web*, um deles é o mesmo utilizado em seu trabalho anterior citado acima (CUL). As outras bases de dados utilizadas pertencem aos serviços Scopus e WoS. A base de dados WoS foi escolhida em função da utilização da técnica *bibliographic coupling* pois ela oferece a busca de documentos relacionados a um determinado documento. A base Scopus foi utilizada para aplicação da técnica de Author Co-Citation (ACC), pois a base de dados WoS, possui uma limitação em sua estrutura

bibliográfica, onde apenas o nome do primeiro autor é exibido, não sendo adequada para a técnica ACC.

QP2: Como as técnicas foram utilizadas?

Heck e Peters (2010) utilizaram os *bookmarks* e *tags* em comum feitas pelos usuários no serviço CUL. A técnica *bibliographic coupling* foi utilizada para calcular a similaridade entre os interesses dos usuários através dos *bookmarks* que usuários mantêm no serviço.

Heck, Peters e Stock (2011), Heck, Hanraths e Stock (2011) a técnica *bibliographic coupling* foi utilizada sobre a base WoS para medir a similaridade através do algoritmo *cosine similarity*. Os autores alvo da recomendação posteriormente avaliaram as recomendações feitas através da análise da rede de potenciais colaboradores gerada pelo sistema de recomendação. A diferença encontrada entre os dois trabalhos de Tamara Heck publicados em 2011 é que o primeiro (HECK; PETERS; STOCK, 2011) realiza um comparativo entre as técnicas aplicadas nos trabalhos, e o segundo expõe as recomendações feitas para que usuários alvo (parceiros do trabalho) informem a relevância dos autores recomendados em cada uma das técnicas.

QP3: Como os dados são armazenados e exibidos?

Heck e Peters (2010) geraram gráficos da rede social de colaboração e apresentaram a entrevistados que tinham como tarefa avaliar o sistema. Heck, Peters e Stock (2011), Heck, Hanraths e Stock (2011) repetiram a abordagem do trabalho anterior exibindo os gráficos da rede para os avaliadores, a diferença é que foi gerado um gráfico da rede para cada uma das bases de dados utilizadas. Em ambos os trabalhos acima, os nós dos grafos apresentados, representam os autores e o peso das arestas é dado pelo *cosine similarity* entre os nós.

QP4: Como o sistema foi avaliado?

Heck e Peters (2010) realizaram uma pesquisa enviada para 363 empregados do instituto de pesquisa *Forschungszentrum Jülich*, questionando se os mesmos utilizavam serviços *desocial bookmarking*, porque utilizavam ou não esses serviços, se eles trabalhavam em comunidades de práticas e como elas foram formadas.

Heck, Peters e Stock (2011) colaboraram com 6 pesquisadores da mesma instituição do trabalho anterior citado acima, inicialmente *clusters* individuais foram construídos para cada um dos pesquisadores. O processo de avaliação foi dividido em 3 partes. A primeira parte se trata de uma entrevista semi-estruturada com questões sobre o comportamento dos pesquisadores. Na segunda parte da avaliação os pesquisadores deveriam ordenar os autores recomendados em ordem de relevância (1 = irrelevante e 10 = muito relevante) e com qual dos autores os pesquisadores poderiam colaborar. Para isto foram selecionados os 10 primeiros autores em ordem alfabética. Na terceira parte da avaliação os avaliadores deveriam dar notas de 1 a 10 para a relevância dos *clusters* gerados de acordo com a distribuição dos autores e os grupos gerados. Para isso, as seguintes questões foram feitas: 1. De acordo com a sua avaliação, a distribuição dos autores reflete a

perspectiva de realidade da comunidade de pesquisa e a colaboração entre eles? 2. Existem quaisquer outros autores importantes que você não se lembrava anteriormente? 3. Este grafo (*cluster*) ajudaria você a organizar um *workshop* ou encontrar parceiros para colaboração científica?

Heck, Hanraths e Stock (2011) aplicaram a mesma avaliação baseada em entrevistas do trabalho anterior (HECK; PETERS; STOCK, 2011), porém desta vez apenas as duas últimas etapas fizeram parte do processo de avaliação.

3.5 Lições do capítulo

A revisão da literatura através do processo de revisão sistemática proporcionou acesso aos documentos relevantes relacionados ao trabalho em desenvolvimento, auxiliando no processo de aquisição do conhecimento de domínio.

Pôde-se observar que a maioria dos sistemas de recomendação encontrados na revisão de literatura utilizam fontes de dados que não são internas a uma instituição específica, como por exemplo uma universidade ou centro de pesquisa, mas sim bases de dados de grandes serviços como bibliotecas digitais (WoS, Scopus, etc) e sistemas de *social bookmarking* (CUL). Estas grandes bases públicas, são propícias para a construção de sistemas de recomendação gerais, onde o histórico de publicações dos autores não pertencem a uma instituição específica, onde a definição das técnicas e algoritmos para gerar recomendações dependem fortemente da base de dados disponível.

Neste trabalho, as recomendações serão feitas dentro do ambiente de uma instituição de ensino, a partir de documentos de projetos internos à instituição. Sendo assim há uma limitação com relação às técnicas utilizadas nos trabalhos relacionados que podem ser aplicadas.

Dos trabalhos coletados através do processo de revisão da literatura, apenas os trabalhos da autora Heck tratam especificamente da tarefa de recomendação de pesquisadores para potencial colaboração. Sistemas de recomendações de produtos para *e-commerce*, filmes e livros parecem já ser um tema bastante abordado e com muitos trabalhos disponíveis na literatura, por outro lado, recomendação de pesquisadores para potencial colaboração parece ser um tema ainda amplamente aberto.

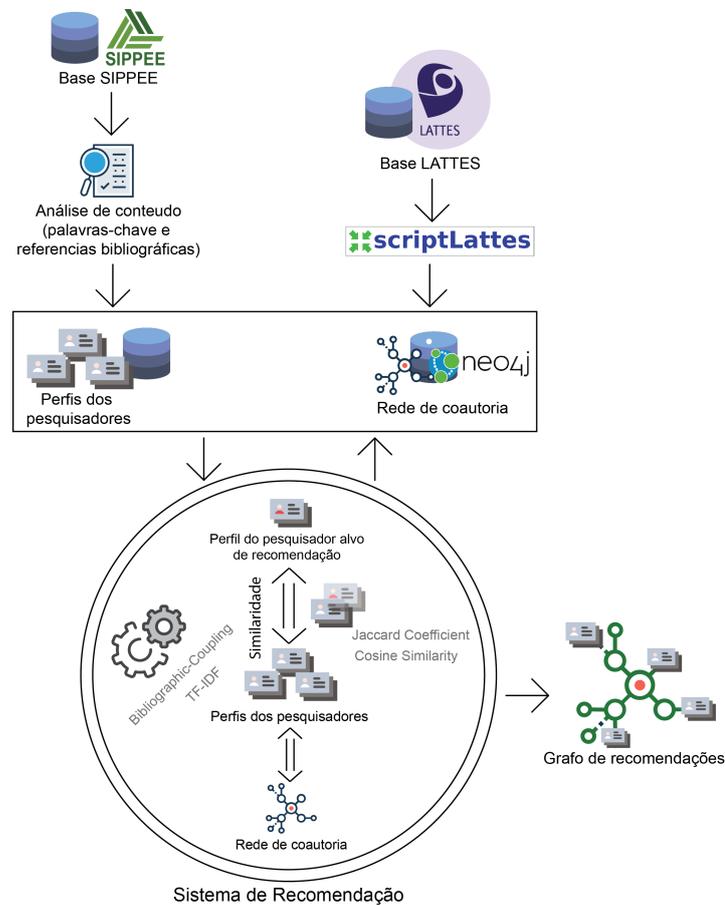
Nenhum dos trabalhos especifica como os dados utilizados nos sistemas são armazenados. Neste trabalho, com cunho prático de implementação e desenvolvimento, o processo de avaliação e escolha da tecnologia de armazenamento das informações utilizadas no processo de recomendação precisou ser então executado.

As técnicas *co-citation* e AB se tornaram medidas padrão na análise cientométrica para a verificação da similaridade de documentos científicos e conseqüentemente são técnicas fundamentais para a recomendação de potenciais colaboradores. Dentre algoritmos para o cálculo de similaridade, os mais aplicados foram: *cosine similarity*, *dice coefficient* e *jaccard index* ou *jaccard coefficient*.

4 SISTEMA DE RECOMENDAÇÃO DE PESQUISADORES

O objetivo geral deste trabalho é desenvolver um sistema de recomendação de pesquisadores com potencial de colaboração científica com base no conteúdo de seus projetos de pesquisa, mais especificamente das palavras-chave e referências bibliográficas, cadastrados na plataforma SIPPEE. Dessa forma, este capítulo tem o propósito de descrever a metodologia de desenvolvimento e avaliação deste sistema. A seção 4.1 aborda a descrição geral da solução proposta e a Figura 8 mostra uma visão geral da arquitetura do sistema proposto.

Figura 8 – Visão geral do sistema de recomendação.



Fonte: o próprio autor.

4.1 Descrição Geral

A base desta proposta é a recomendação de pesquisadores com perfis de pesquisa semelhantes. Entende-se que dois perfis de pesquisa são semelhantes quando ambos demonstram possuir interesses comuns em áreas ou tópicos de pesquisa. Nesta proposta, os perfis dos pesquisadores foram criados a partir das informações extraídas do conteúdo de

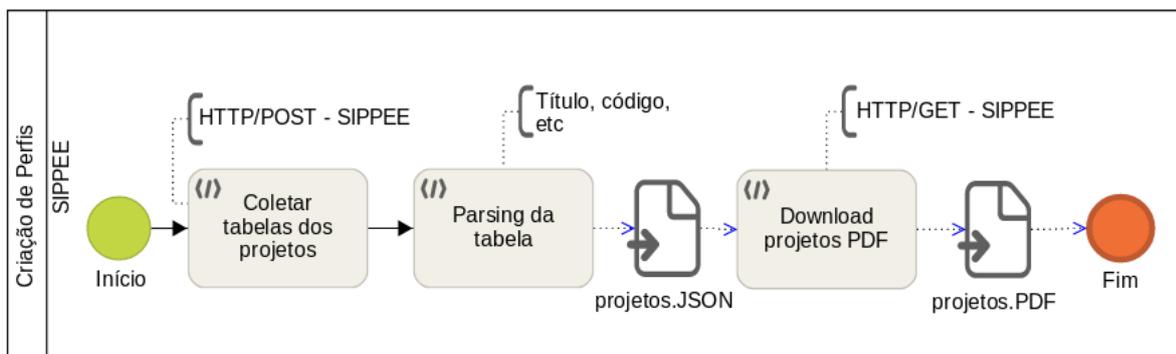
seus projetos na plataforma SIPPEE. A proposta da solução é dividida em quatro etapas principais:

1. Obtenção dos arquivos PDF dos projetos de pesquisa;
2. Construção dos perfis dos pesquisadores com base na extração de palavras-chave e referências bibliográficas dos arquivos PDF;
3. Cálculo de similaridade com base nas palavras-chave e referências bibliográficas;
4. Construção da rede de coautoria dos pesquisadores;
5. Construção da interface gráfica de usuário para exibição das recomendações.

4.1.1 Obtenção dos Arquivos dos Projetos

A obtenção dos arquivos PDF dos projetos de pesquisa da SIPPEE é realizada por um módulo do sistema de recomendação desenvolvido especificamente para extrair dados da plataforma. Este módulo tem como objetivo coletar as tabelas de projetos de pesquisa da página da plataforma SIPPEE, extrair dessas tabelas através de um parser HTML as informações dos projetos (coordenador, campus, código e título) e por fim baixar os arquivos PDF dos projetos. A Figura 9 ilustra o fluxo de execução do processo.

Figura 9 – Processo de obtenção dos arquivos PDF dos projetos de pesquisa da SIPPEE.



Fonte: o próprio autor.

Na etapa inicial do processo, o módulo do sistema de recomendação realiza uma requisição HTTP utilizando o método POST para solicitar a tabela de projetos de pesquisa. Esta requisição é a mesma realizada na página inicial da plataforma SIPPEE (Figura 10).

O retorno da requisição consiste no código fonte da página, que contém a tabela de projetos em formato HTML, dessa forma, através de um *parser* HTML pertencente ao módulo, o código fonte da tabela de projetos foi utilizado para extrair as informações contidas nos elementos que compõem a tabela (Figura 11).

Figura 10 – Formulário para pesquisa de projetos na plataforma SIPPEE.

Consulta aos projetos cadastrados

Plataforma	<input type="text" value="Pesquisa"/>	Título	<input type="text"/>
Unidade	<input type="text" value="Selecione um item"/>	Coordenador	<input type="text"/>
Edital	<input type="text" value="Selecione um item"/>	<input type="button" value=":: Pesquisar ::"/>	

Fonte: o próprio autor.

Figura 11 – Tabela de projetos de pesquisa da SIPPEE.

Pesquisa			
Plataforma	Unidade	Coordenador	Título
Pesquisa	Alegrete	ADRIANA GINDRI SALBEGO	Diagnóstico do sistema de transporte coletivo urbano: o caso do município de Alegrete/RS
Pesquisa	Alegrete	ADRIANA GINDRI SALBEGO	Projeto Piloto Cadastro dos Usuários da Sub-bacia do Ibicuí
Pesquisa	Alegrete	Aldoni Gabriel Wiedenhof	Montagem e instrumentação de uma máquina de ensaios de fadiga de cargas axial
Pesquisa	Alegrete	Aldoni Gabriel Wiedenhof	Avaliação e desempenho de juntas topo soldadas pelo processo "friction stir welding" em liga de alumínio.

Fonte: o próprio autor.

A consulta por projetos na base SIPPEE é feita de forma paginada, ou seja, um número determinado de projetos é listado por página e existem links que permitem a navegação entre as diversas páginas da consulta. Assim sendo, para garantir que todos os projetos resultantes sejam obtidos, o módulo do sistema extrai os links de navegação e realiza novas requisições para cada um deles, obtendo a tabela de projetos para cada uma das páginas.

Após obter as tabelas de todas as páginas, um arquivo contendo as informações dos projetos é gerado e armazenado. Este arquivo serve de entrada para o método que realiza o download dos arquivos PDF dos projetos.

Neste caso de estudo, 60 pesquisadores membros do corpo docente de Alegrete consentiram com uso de seus projetos. Assim sendo, antes de baixar os arquivos dos projetos, aqueles que não pertencem a esse grupo são excluídos do processo.

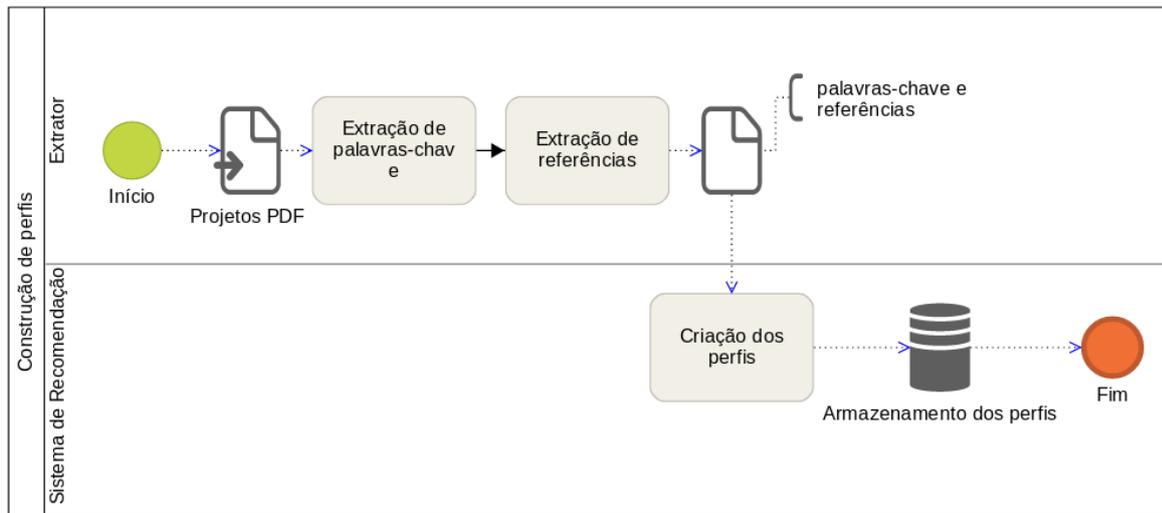
4.1.2 Extração dos Dados e Construção dos Perfis

Após coletar os arquivos PDF dos projetos, as palavras-chave e as referências bibliográficas de cada projeto são extraídas com a finalidade de construir os perfis dos pesquisadores. Esta etapa é executada através de dois módulos do sistema de recomendação.

O módulo Extrator é responsável por extrair dos arquivos PDF obtidos na etapa

1, as palavras-chave e referências bibliográficas. As informações extraídas são então utilizadas pelo módulo principal do sistema que após construir os perfis dos pesquisadores, armazena as informações em um banco de dados relacional utilizando o sistema de gerenciamento de banco de dados MySQL. A Figura 12 ilustra o processo da etapa 2.

Figura 12 – Processo de construção de perfil dos pesquisadores.



Fonte: o próprio autor.

4.1.3 Cálculo da Similaridade entre Perfis

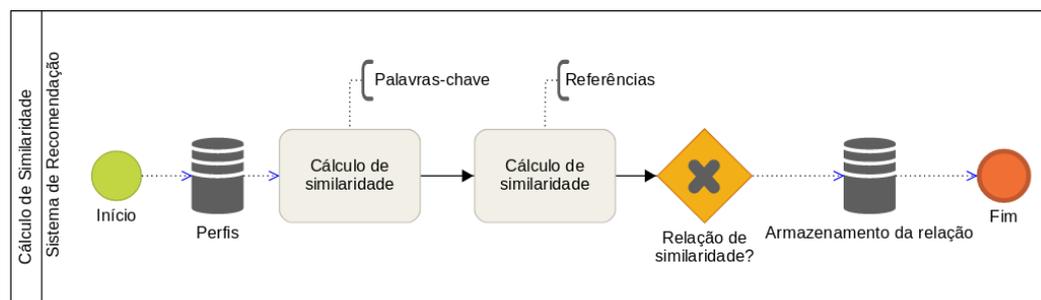
Com os perfis dos pesquisadores em mãos, na etapa 3 são calculados os valores de similaridade entre todos os perfis armazenados. O cálculo da similaridade entre palavras-chave é obtido através da similaridade do cosseno (subseção 2.3.1), onde as componentes dos vetores são compostas pela importância de cada palavra-chave, definida por sua frequência.

O cálculo da similaridade entre referências bibliográficas é obtido através do método mínimo do acoplamento bibliográfico de autor introduzido na subseção 2.3.2. A Figura 13 ilustra o processo da etapa 3.

Para o cálculo de similaridade entre referências bibliográficas através do método ABA, consideramos que duas cadeias de caracteres tratam da mesma referência bibliográfica caso possuam 80% (mesmo valor utilizado pela ferramenta scriptLattes para determinar títulos semelhante) ou mais de similaridade. Esta condição é válida tanto para identificar referências repetidas pertencentes ao perfil de um pesquisador, quanto para identificar referências em comum entre perfis de pesquisadores.

A medida de similaridade entre as referências é obtida através da Distância de Levenshtein, que calcula a similaridade entre duas cadeias de caracteres com base no número de operações de edição, inserção e remoção de caracteres que precisam ser feitas

Figura 13 – Processo de cálculo de similaridade entre os perfis dos pesquisadores.



Fonte: o próprio autor.

para transformar uma cadeia na outra. Isso é feito devido à liberdade que os pesquisadores tem para definir o formato das referências bibliográficas dos projetos.

O teste “Relação de Similaridade” verifica se alguma das similaridades entre dois perfis é maior que zero, dessa forma mantêm-se armazenadas as relações de similaridade entre os perfis cuja alguma das similaridades existe.

Como a rede de recomendações é exibida através de um grafo, diferentemente dos perfis de pesquisadores, as relações de similaridade são armazenadas através da plataforma de base de dados orientada a grafos Neo4j ¹.

4.1.4 Construção da Rede de Coautoria dos Pesquisadores

A fim de identificar recomendações feitas entre pesquisadores que já colaboraram no passado, o sistema de recomendação consulta a rede de colaboração construída através da ferramenta scriptLattes. Para construir a rede, a ferramenta recebe como entrada os arquivos HTML dos currículos oriundos da plataforma Lattes ² dos pesquisadores e gera um arquivo contendo uma matriz com as relações de coautoria existentes. O processo de construção da rede de colaboração é exibido na Figura 14.

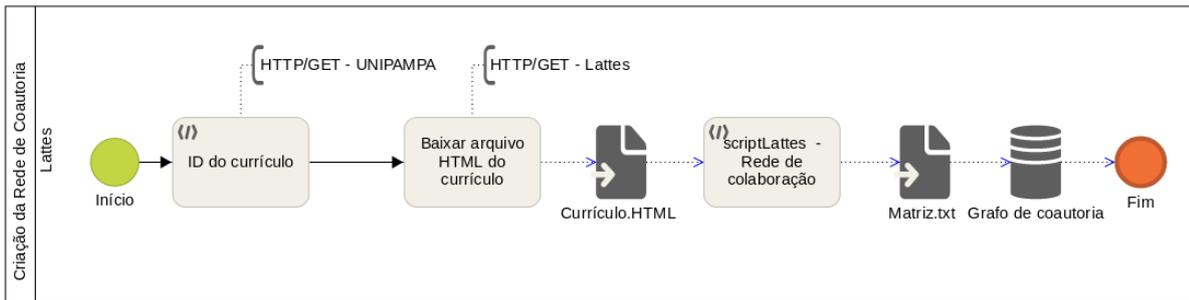
Na etapa inicial, o sistema “navega” entre os links das listas de docentes do site da UNIPAMPA e, caso conste o link para o currículo Lattes, o mesmo é extraído e armazenado. Após obter os links de todos os currículos Lattes do corpo docente, os arquivos HTML dos currículos são baixados diretamente da plataforma Lattes. Os arquivos HTML dos currículos são então utilizados pela ferramenta scriptLattes para gerar o arquivo da matriz de colaborações.

A matriz de colaboração gerada indica se dois pesquisadores quaisquer são coautores caso o valor dentro das posições dos pesquisadores na matriz seja diferente de 0. Dessa forma, a matriz é utilizada para armazenar o grafo de colaboração no banco de

¹ Neo4j Graph Platform – The Leader in Graph Databases <<https://neo4j.com/>>

² Plataforma Lattes <<http://lattes.cnpq.br/>>

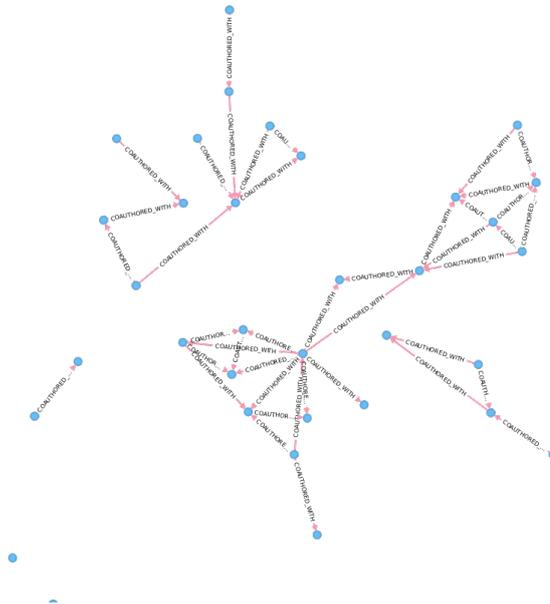
Figura 14 – Processo de construção da rede de colaboração.



Fonte: o próprio autor.

dados Neo4j. A Figura 15 mostra um exemplo do resultado do processo considerando pesquisadores do campus Alegrete.

Figura 15 – Exemplo de grafo de colaboração gerado pelo scriptLattes e armazenado no Neo4j.



Fonte: o próprio autor.

Com o grafo de colaboração armazenado no Neo4j, podemos verificar se dois pesquisadores $p1$ e $p2$ são coautores executando a seguinte consulta em linguagem Cypher (linguagem de consulta padrão do Neo4j):

```
MATCH p=({name: 'p1'}) -[r:COAUTHORED_WITH] -({name: 'p2'})
RETURN p
```

4.1.5 Interface Gráfica de Usuário

A interface gráfica de usuário (do inglês *Graphics User Interface* ou GUI) desenvolvida é uma aplicação *web* construída através de uma arquitetura cliente/servidor, onde o cliente realiza requisições para o servidor através de dispositivos como *notebooks*, *smartphones*, *tablets*, *PCs*, etc, utilizando o protocolo HTTP e o servidor por sua vez, atende essas requisições e retorna para o cliente as informações solicitadas (página, arquivo, JSON, etc).

A construção da aplicação no lado do cliente é feita com tecnologias convencionais para aplicações ricas para a *web*: HTML, JavaScript e CSS. Além disso, a biblioteca React³ foi utilizada para criar os componentes gráficos da interface. A ideia geral da utilização da biblioteca React é a criação de classes que representam os componentes da GUI, onde esses componentes são automaticamente atualizados quando atributos (estado) da classe são alterados. Dessa forma, podemos dizer que os elementos da página estão “amarrados” aos atributos da classe, fazendo desnecessária a reconstrução manual (convencionalmente realizada) de elementos da página.

O objetivo principal da GUI é permitir que o usuário visualize as recomendações de maneira dinâmica e interativa, além de ser utilizada para a realização da avaliação do ponto de vista do usuário, possuindo três componentes principais: grafo global de recomendações, grafo individual de recomendações e a página de informações do perfil de pesquisador.

No grafo global de recomendações, o usuário tem acesso a uma rede que permite visualizar todas as recomendações que foram feitas entre todos os pesquisadores participantes, onde os vértices representam os pesquisadores e as arestas representam as recomendações. O grafo global de recomendações pode ser visto na Figura 16.

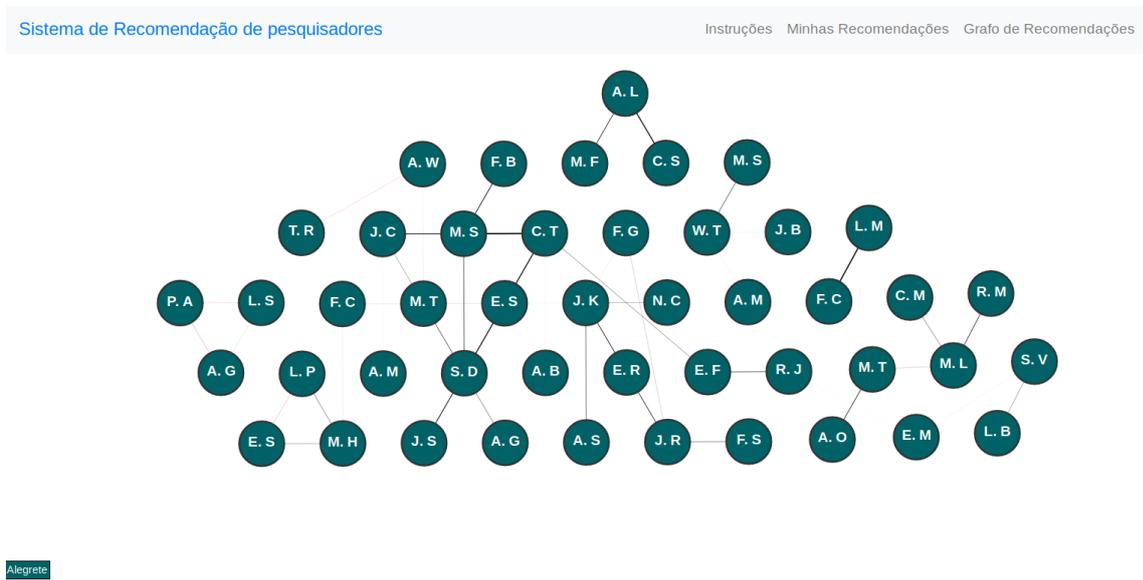
O grafo individual de recomendações permite ao usuário visualizar de maneira semelhante ao grafo global, as recomendações que foram feitas para um determinado usuário, podendo ser utilizado tanto para acessar as recomendações do usuário autenticado na aplicação, como para visualizar separadamente as recomendações de qualquer pesquisador presente no grafo global de recomendações. O componente é exibido na Figura 17.

Ao clicar em um dos vértices do grafo de recomendações individuais, o usuário é redirecionado para uma página contendo informações sobre o pesquisador selecionado, esta página tem como principal objetivo, auxiliar no entendimento das recomendações feitas e pode ser vista na Figura 18.

Nos grafos global e individual de recomendações, a espessura das arestas (linhas) que representam as recomendações indicam o quão forte é a similaridade entre os pesquisadores, ou seja, quanto mais espessa a linha, maior a similaridade. No grafo individual de recomendações, o valor em porcentagem abaixo do nó, indica o “quão recomendado” o pesquisador é com base em uma combinação linear simples das métricas utilizadas.

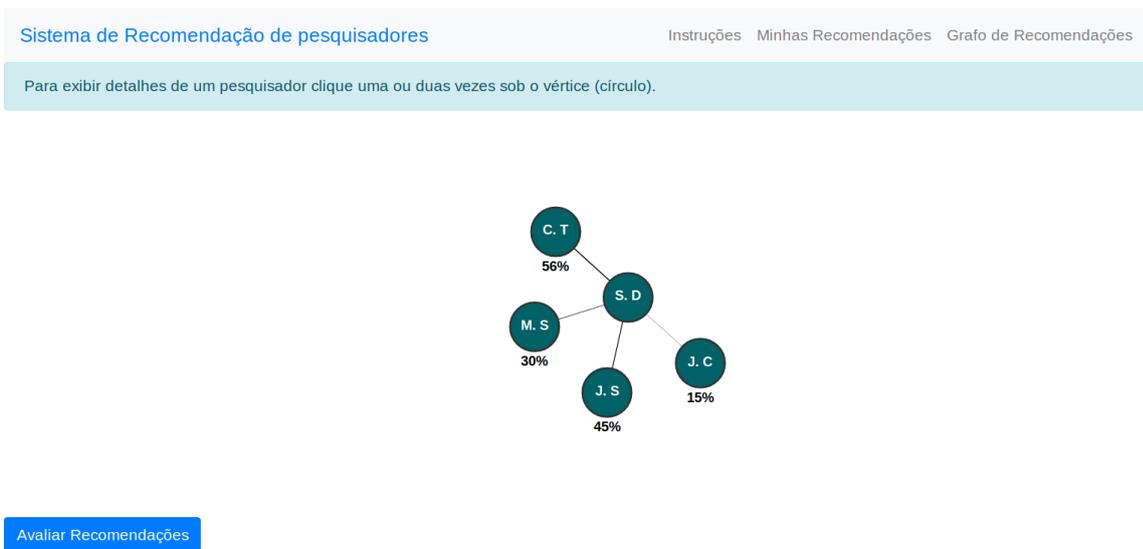
³ React – A JavaScript library for building user interfaces <<https://reactjs.org/>>

Figura 16 – Grafo global de recomendações.



Fonte: o próprio autor.

Figura 17 – Grafo individual de recomendações.



Fonte: o próprio autor.

Figura 18 – Informações do perfil do pesquisador J.K.

Sistema de Recomendação de pesquisadores

Instruções Minhas Recomendações Grafo de Recomendações

J.K.

Projeto SIPPEE 8753 Ver resumo do projeto

Palavras-chave

eficiência energética microgeração geração descentralizada recursos energéticos renováveis

Referências

- ABNT - ISO 50.001 - 2 - 3 - 4, Gestão de Energia, 2012
- ANEEL. Atlas de Energia Elétrica. <http://www.aneel.gov.br/>
- DUFFIE & BECKMAN, "Solar engineering of thermal processes"; Wiley y 2006
- ELEKTRO, Manuais Elektro de Eficiência Energética. Disponível em: http://elektro.com.br/clientescomerciais-eindustriais/manuais_eficiencia_a_energetica.html
- ELETROBRÁS - PROCEL. Bibliografia Específica sobre Eficiência Energética. Disponível em <http://www.procelinfo.com.br/main.asp>
- EVO - Efficiency Valuation Organization. "Protocolo Internacional de Medição e Verificação de Performance: Conceitos e Opções para a Determinação de Economias de Energia e de Água"; Abril 2007
- FARRET, F.A. Aproveitamento de pequenas fontes de energia elétrica. Ed. 02, p. 245, Editora da UFSM, 2010.
- FARRET, F.A.; SIMOES, M.G. Integration of alternative sources of energy. Ed. 01, p. 499, IEEE PRESS/Wiley & Sons, Inc., 2006.
- KAEHLER, J. W. M.; Eficiência Energética: da Avaliação Gerencial à Auditoria Energética; Alegrete, UNIPAMPA, Livro em Edição, 2016

Projeto SIPPEE 4298 Ver resumo do projeto

Palavras-chave

eficiência energética iluminação pública gestão pelo lado da demanda - gld

Referências

- ABNT, NBR 5101: Iluminação Pública, 2012;
- COMUNIDADE EUROPEIA, EN13201-4: Métodos de Medida das Performances Fotométricas, 2011;
- COPEL, Norma Técnica Projeto de Iluminação Pública, 2012
- CPFL, Norma Técnica Projeto de Iluminação Pública, 2012
- KE Lun, A Method of Light Reflectance Measurement, Dissertação de Mestrado, Universidade da Columbia Britânica, 1999
- PROCEL - RELUZ / ELETROBRÁS & IBAM; Iluminação Pública Eficiente, 2004;
- SANTOS, Eduardo Ribeiro, A Iluminação Pública como Elemento de Composição da Paisagem Urbana, Dissertação de Mestrado PROPAP /UFRGS, 2005

Projeto SIPPEE 4192 Ver resumo do projeto

Palavras-chave

eficiência energética gestão pelo lado da demanda - gld tarifa horo-sazonal tarifa branca

Referências

- CHOWDHURY & CROSSLEY, "Microgrids and Active Distribution Networks", IET Renewable Energy Series 6, 2009
- FOX-PENNER, Peter, "Smart Power: Climate Change, the Smart Grid and the Future of Electric Utilities", Island Press, 2010
- GELLINGS, Clark W. "The Smart Grid: Enabling Energy Efficiency and Demand Response", CRC - The Fairmont Press, 2009.
- GELLINGS, Clark W. "Saving Energy and Reducing CO2 Emissions with Electricity", CRC - The Fairmont Press, 2011.
- IEA/OECD, Strategic Plan for the demand-Side Management Programme: 2008 - 2012, [www.ieadsm.org/Files/ acesso em 15/07/2011](http://www.ieadsm.org/Files/acesso em 15/07/2011).

Fonte: o próprio autor.

4.1.6 Avaliação da Proposta

A fim de avaliar as recomendações geradas pelo grafo individual de recomendações, foi elaborado um instrumento de avaliação⁴ Apêndice C no formato de um questionário, onde os usuários do sistema (pesquisadores participantes da pesquisa) indicaram a relevância das recomendações feitas.

Os aspectos abordados neste instrumento de avaliação são adaptados de Pu, Chen e Hu (2011), onde os autores levantam aspectos que devem ser levados em consideração quando se avalia um sistema de recomendação do ponto de vista do usuário. Dentre os aspectos apresentados estão: precisão percebida, novidade (ou descoberta), atratividade, diversidade, além de questões como a usabilidade da interface e sua facilidade de utilização.

As questões propostas têm como objetivo tentar cobrir alguns dos principais aspectos do ponto de vista do usuário do sistema de recomendação. Os participantes foram convidados a responder as questões através de convites (Apêndice A) por e-mail enviados àqueles que tiveram recomendações feitas. As respostas das questões são valores numéricos em um intervalo de 1 ("discordo completamente") até 5 ("concordo completamente"), valores da escala Likert (LIKERT, 1932). Na lista abaixo constam as questões elaboradas, juntamente com o aspecto de avaliação do ponto de vista do usuário que ela visa cobrir.

- **Precisão percebida:** Os pesquisadores recomendados trabalham em linhas/áreas de pesquisa em que sou capaz de contribuir.

⁴ Instrumento de avaliação <<https://goo.gl/V5XnFa>>

- **Novidade (ou descoberta):** Descobri pesquisadores com potencial de colaboração que eu não conhecia anteriormente.
- **Diversidade:** Foram sugeridos diversos pesquisadores.
- **Transparência:** Consegui perceber porque os pesquisadores foram recomendados para mim.
- **Atratividade:** Fiquei entusiasmado com as novas possibilidades de colaboração.
- **Facilidade de uso:** Não encontrei dificuldades na utilização do sistema.
- **Utilidade percebida:** A tarefa de encontrar colegas para colaborar em projetos ficou mais fácil de ser executada.
- **Intenções do usuário:** Caso o sistema fosse implantado, utilizaria o mesmo frequentemente.

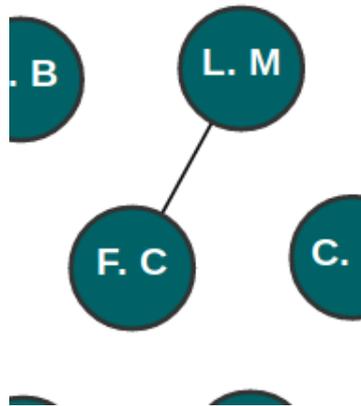
Os resultados das respostas poderão ser utilizados para o aperfeiçoamento contínuo do sistema.

5 RESULTADOS

Neste capítulo são apresentados os principais resultados do desenvolvimento do sistema de recomendação proposto. Inicialmente são apresentados dados relacionados ao sistema e às recomendações feitas e em seguida resultados extraídos do instrumento de avaliação realizada por usuários convidados.

Participaram neste caso de estudo, 60 pesquisadores do campus Alegrete. Dessa forma, 60 perfis dos pesquisadores foram modelados a partir da extração de palavras-chave e referências bibliográficas de 171 projetos, totalizando 517 palavras-chave e 2497 referências bibliográficas. A fim de demonstração, selecionou-se dois pesquisadores da rede global de recomendações, que indicam pela espessura da aresta, terem uma relação forte de similaridade, como ilustrado na Figura 19.

Figura 19 – Exemplo de recomendação extraída do grafo global de recomendações.



Fonte: O próprio autor.

Ao comparar os perfis dos dois pesquisadores selecionados (pesquisadores L. M. e F.C) através da página de informações de perfil (Figura 20), pode-se identificar evidências que confirmam a similaridade entre os perfis dos dois pesquisadores. A página de informação dos perfis dos pesquisadores mostra o compartilhamento de palavras-chave como "turbulência", assim como referências bibliográficas como os trabalhos dos autores Acevedo, O. C e Fitzjarrald, D.

Ainda observando os dados resultantes do grafo global de recomendações (Figura 16), pode-se identificar um grupo de pesquisadores (S.D, M.S, C.T, J.C) cujo sistema identificou que trabalham em áreas similares. Os pesquisadores membros desse grupo trabalham na área de Engenharia de Software e a grande maioria costuma inserir a palavra-chave “Engenharia de Software” em seus projetos, sendo este o principal fator da formação do mesmo.

Figura 20 – Informações dos perfis de dois pesquisadores com perfis similares.

The image displays two columns of research profiles. The left column is for 'F.C.' and the right for 'L.M.'. Each profile includes a project title (e.g., 'Projeto SIPPEE 10207'), a list of keywords (e.g., 'turbulência', 'camada limite estável'), and a list of references. The references for both profiles are highly similar, mentioning authors like ACEVEDO, O. C., COSTA, F. D., DEGRADIA, G. A., and FITZJARRALD, D. in the context of stable boundary layer research.

Fonte: O próprio autor.

O sistema gerou um total de 92 recomendações, onde cada pesquisador recebeu em média 2 recomendações. Dessa forma 46 pesquisadores tiveram pelo menos uma recomendação feita e 14 não tiveram candidatos identificados para uma recomendação, sendo 5 o maior número de recomendações feitas para um pesquisador. Ainda sobre as recomendações, 18 delas foram feitas entre pesquisadores que já haviam colaborado previamente, ou seja, 80% das recomendações ocorreram entre pares que ainda não colaboram. O número total de coautorias entre os pesquisadores participantes é 36.

Com relação as duas métricas, o valor médio de similaridade que a métrica similaridade do cosseno através das palavras-chave obteve é de 19,9%, enquanto que a média do acoplamento bibliográfico de autor é de 1%. A similaridade do cosseno mais alta é 68,3% enquanto a do acoplamento bibliográfico de autor é 5,2%. Pode-se verificar também que 69,5% das recomendações dependem exclusivamente da similaridade do cosseno, ou seja, 64 recomendações foram feitas sem nenhuma participação do acoplamento bibliográfico de autor. Além disso, em 70 recomendações, a similaridade por acoplamento bibliográfico de autor ficou abaixo de 1%, isto aparenta ser um indicativo de que, ao menos para a base de dados utilizada, o método não parece ser indicado para a tarefa de recomendação.

5.1 Resultados do Instrumento de Avaliação

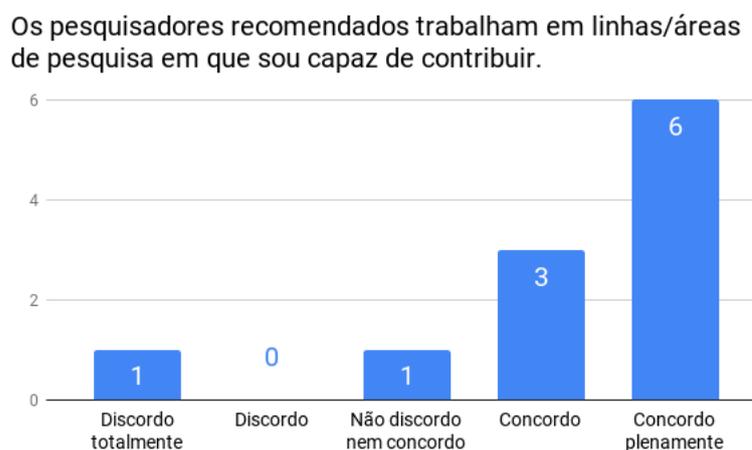
Nesta seção são apresentados dados resultantes da aplicação do instrumento de avaliação. Dos 60 pesquisadores participantes da pesquisa, 46 pesquisadores receberam o convite para avaliar o sistema e suas recomendações, sendo que desse número, 11 atenderam o convite e participaram do processo de avaliação.

Para cada questão referente a um dos aspectos de avaliação abordados, um gráfico com as respostas é exibido seguido de considerações a respeito dos dados.

A Figura 21 mostra as avaliações referentes a questão do instrumento que cobre o aspecto **precisão percebida**, onde podemos verificar que a maioria dos respondentes

concorda com o fato de que os pesquisadores recomendados trabalham em linhas/áreas nas quais podem contribuir.

Figura 21 – Repostas da questão: Os pesquisadores recomendados trabalham em linhas/áreas de pesquisa em que sou capaz de contribuir.



Fonte: O próprio autor.

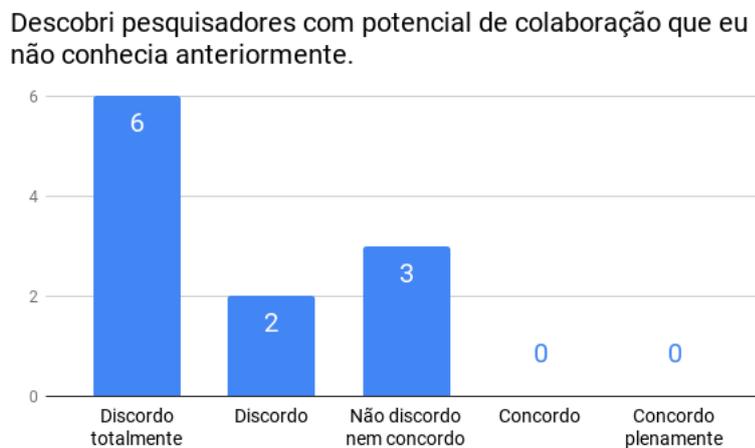
Em relação ao aspecto de avaliação **novidade (ou descoberta)**, já havia uma expectativa de que não seria satisfeito pelo sistema desenvolvido, uma vez que os pesquisadores participantes na pesquisa são docentes do mesmo campus. Ainda que não seja uma regra os pesquisadores conhecerem o trabalho dos seus colegas de campus, o fato da pesquisa abranger somente os pesquisadores de um mesmo campus tem um impacto no resultado. Adicionalmente, destaca-se que não se obteve uma participação completa de todos os pesquisadores do campus. Figura 22 apresenta os dados da questão que cobre este aspecto.

Uma observação a ser feita sobre a questão formulada, é que ela está fortemente calcada no desconhecimento da existência do pesquisador recomendado. Dessa forma, ela não abrange os casos em que o pesquisador alvo de recomendação descobre que uma recomendação de um pesquisador já conhecido trabalha em uma linha/área de pesquisa em que eles podem colaborar e que ele não tinha ciência que fazia parte das linhas do pesquisador recomendado.

O gráfico da Figura 23 mostra as respostas da questão que visa cobrir o aspecto **diversidade**. Observando os dados, pode-se perceber que a maior parte dos respondentes não soube afirmar quanto a diversidade de recomendações. Isto provavelmente se deve ao fato dos participantes não terem ciência do número médio de recomendações ou qual o número de recomendações que deve ser considerado alto ou baixo.

Acredita-se que uma fatia considerável dos pesquisadores do campus tem uma ideia de quantos e quais pesquisadores trabalham em linhas/áreas de possível colaboração e que

Figura 22 – Repostas da questão: Descobri pesquisadores com potencial de colaboração que eu não conhecia anteriormente.

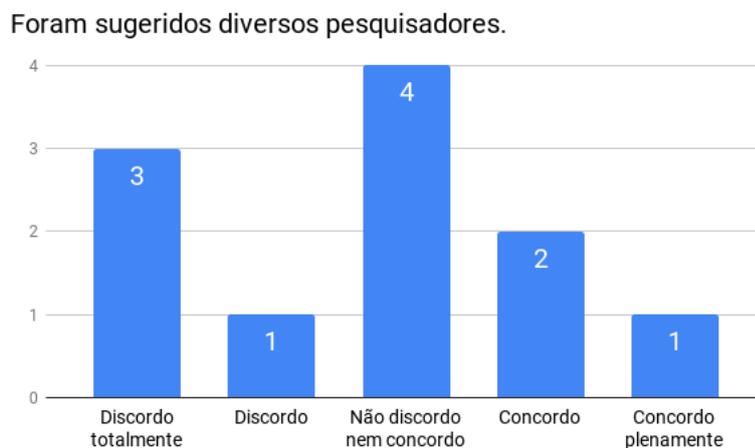


Fonte: O próprio autor.

a ciência desse fato pode também ter impedido que mais respostas estivessem concentradas na opção “discordo totalmente”.

Dois dos pesquisadores que responderam “Discordo totalmente” tiveram apenas 1 recomendação feita, o que facilita a escolha pela opção, diminuindo as possíveis implicações do fato exposto no parágrafo anterior.

Figura 23 – Repostas da questão: Foram sugeridos diversos pesquisadores.

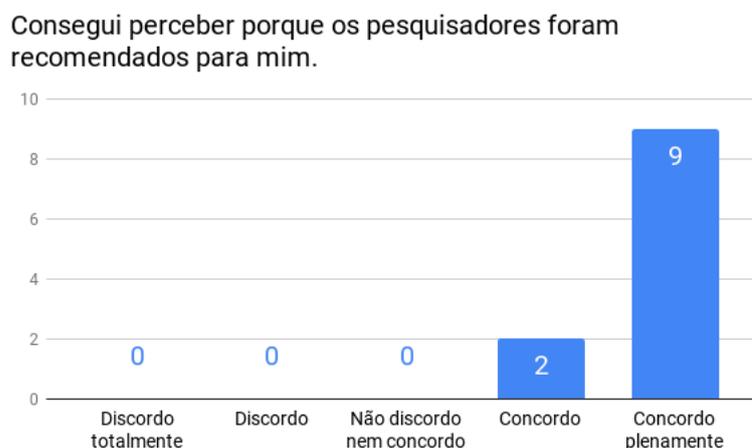


Fonte: O próprio autor.

A Figura 24 indica que o sistema de recomendação se mostrou transparente com relação às recomendações feitas, ou seja, os usuários conseguem perceber a origem das

recomendações. Este aspecto é fundamental para que o sistema não provoque o sentimento de que está de posse de fontes de informações sobre o usuário que o mesmo desconhece.

Figura 24 – Respostas da questão: Consegui perceber porque os pesquisadores foram recomendados para mim.



Fonte: O próprio autor.

O fato dos pesquisadores não obterem diversas recomendações, aliado a recomendação de colaboradores já conhecidos implica que não houve muitas oportunidades para novas colaborações. Dessa forma, podemos observar na Figura 25 que a maior parte dos respondentes não se sentiu entusiasmo por novas possíveis colaborações oriundas das recomendações.

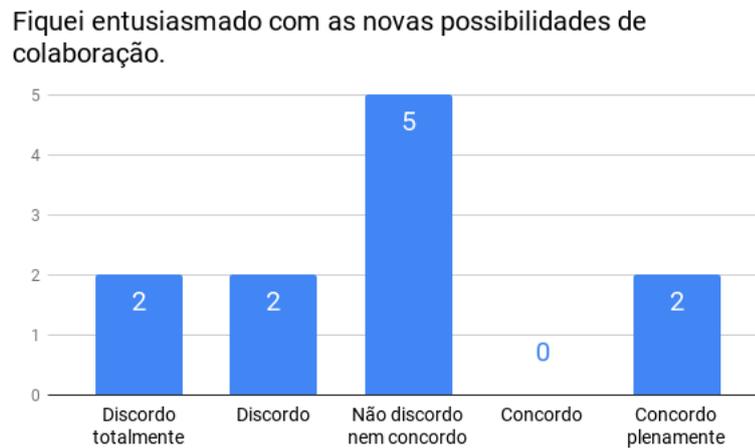
Quanto ao aspecto **facilidade de uso**, não houve relatos nas respostas, que indiquem qualquer dificuldade na utilização do sistema de recomendação. As respostas indicam que os usuários participantes da avaliação conseguiram visualizar e avaliar as recomendações sem obstáculos de usabilidade.

Na Figura 27 relacionada ao aspecto de **utilidade percebida**, observa-se que os pesquisadores reconhecem a utilidade que o sistema de recomendação tem no processo de busca por potenciais colaboradores.

Alinhado aos resultados obtidos na questão anterior, percebe-se que caso o sistema de recomendação fosse implantado (e.g: um sistema interno da instituição), a maioria dos pesquisadores utilizaria com frequência o sistema de recomendação, como pode ser observado na Figura 28.

Além das questões relacionadas aos aspectos de avaliação elencados para esta pesquisa, os participantes foram encorajados a fazer quaisquer tipos de observações através de um campo de texto livre inserido no final do instrumento. Dois pesquisadores utilizaram este campo para sugerir que referências bibliográficas em diferentes idiomas (ex: edições de livros) sejam consideradas, a utilização de outras fontes de dados para construção de

Figura 25 – Repostas da questão: Fiquei entusiasmado com as novas possibilidades de colaboração.



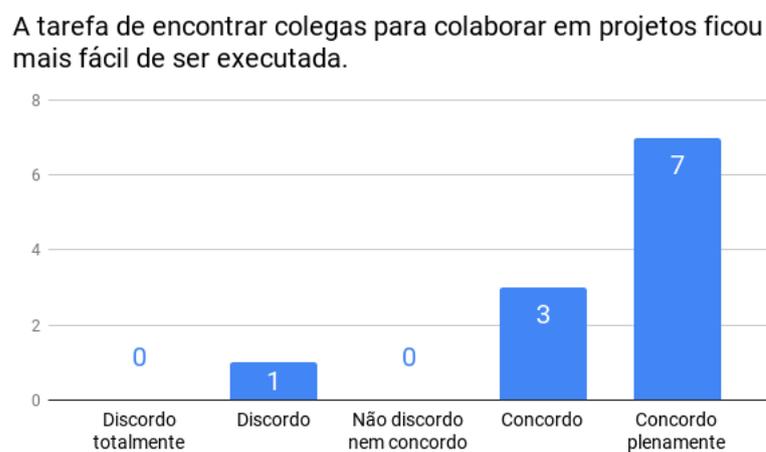
Fonte: O próprio autor.

Figura 26 – Repostas da questão: Não encontrei dificuldades na utilização do sistema.



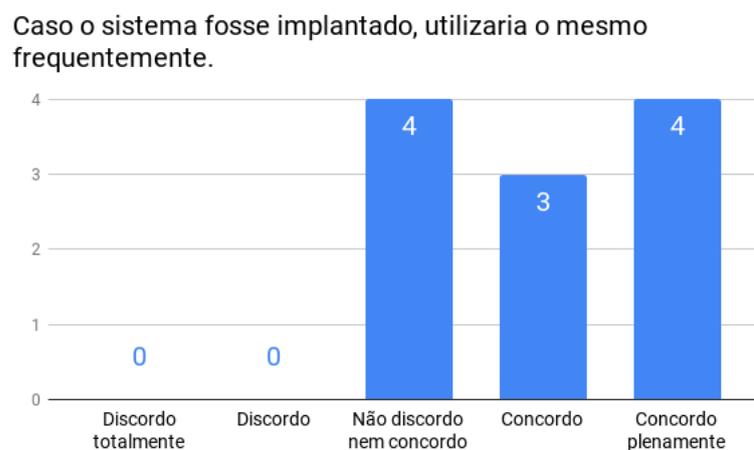
Fonte: O próprio autor.

Figura 27 – Repostas da questão: A tarefa de encontrar colegas para colaborar em projetos ficou mais fácil de ser executada.



Fonte: O próprio autor.

Figura 28 – Repostas da questão: Caso o sistema fosse implantado, utilizaria o mesmo frequentemente.



Fonte: O próprio autor.

perfis e para ressaltar a importância deste tipo de mecanismo através da sentença: “..os resultados do estudo são relevantes, já que permite conhecer linhas de pesquisas similares desenvolvidas por outros pesquisadores”.

6 CONSIDERAÇÕES FINAIS

Este trabalho apresentou uma motivação que aborda a importância e os impactos positivos da colaboração científica diante de uma tendência evidente do crescimento de produção de conhecimento feita de forma colaborativa, que deixa pra trás antigos modelos onde se supervalorizava a criação individual de conhecimento.

Nesse contexto observou-se uma oportunidade para a proposição de um sistema computacional que fomentasse um cenário mais colaborativo de pesquisa. A revisão de literatura revelou que sistemas de recomendação que recomendem pesquisadores com perfis de pesquisa semelhantes é um domínio com oportunidades de pesquisa ainda bastante abertas.

Os benefícios que a colaboração científica proporciona aliado a essa oportunidade de recomendar pesquisadores conjuntamente com a existência de uma base de projetos de pesquisa como a SIPPEE culminou na proposição de um sistema que visa estimular a colaboração científica entre os pesquisadores dos 10 campi da UNIPAMPA, fazendo com que a plataforma deixe de ser apenas um sistema interno de gerenciamento e passe a contribuir para o objetivo maior da universidade, que é o desenvolvimento regional.

Sobre sistemas de recomendação, um aspecto que aparenta não estar bem estabelecido é a forma como eles são avaliados, onde diversas métricas de avaliação das recomendações tem um foco muito maior na “precisão” dos algoritmos do que com a real satisfação dos usuários. Por outro lado os diversos esforços encontrados que tentam avaliar sistemas de recomendação do ponto de vista do usuário dependem de métricas bastante subjetivas e abstratas que também são de difícil mensuração.

Neste trabalho foram selecionados alguns aspectos de avaliação do ponto de vista do usuário cuja aplicação aos itens recomendados se mostrou mais adequada. Todavia, em trabalhos futuros pretende-se aprofundar a avaliação, explorando outros aspectos não utilizados, inclusive oriundos de outros trabalhos além do utilizado.

Além de todos os benefícios relacionadas ao fomento da colaboração científica expostos até aqui, o desenvolvimento do sistema de recomendação proposto neste trabalho servirá como base para futuros estudos na área de sistemas de recomendação visando melhorar seu método, tendo em vista que este é um trabalho pioneiro nesta área dentro da instituição.

Através dos resultados extraídos do instrumento de avaliação do ponto de vista do usuário, pode-se perceber que foi possível identificar pesquisadores com potencial colaboração utilizando informações extraídas dos projetos de pesquisa da SIPPEE. Ainda, os pesquisadores se mostraram bastante pró-ativos com relação as intenções de utilização do sistema de recomendação caso implantado.

A métrica que acabou identificando quase que totalmente as relações de similaridade, foi a similaridade do cosseno com base nos vetores de frequências das palavras-chave. Assim sendo, ao menos para esta base de dados, percebemos que o método do acopla-

mento bibliográfico de autor não é uma boa alternativa como métrica, cabendo verificar se isto se aplica a recomendação de pesquisadores utilizando diferentes bases de dados.

Uma limitação presente no desenvolvimento do trabalho foi a inviabilidade de acesso aos dados dos projetos da plataforma SIPPEE, que tornou necessária a utilização de um termo de autorização, disponível no Apêndice B, assinado pelos participantes neste caso de estudo. Este fato resultou em uma diminuição significativa de escopo. Assim sendo, considera-se de extrema importância para as instituições, que os dados das produções internas ou, ao menos, parte do conteúdo das produções sejam de livre acesso, para que as próprias instituições possam fazer um melhor aproveitamento do capital intelectual produzido e armazenado em seus sistemas internos.

Ainda relacionado aos dados da plataforma SIPPEE, uma dificuldade encontrada foi a falta de padrão nas informações presentes nos arquivos PDF dos projetos. Apesar de ser um campo obrigatório, muitos projetos não continham palavras-chave especificadas, além disso, o formulário padrão da instituição para cadastro de projetos nem sempre era utilizado e as referências bibliográficas podem ser inseridas em qualquer formato. Solucionar essas questões pode facilitar muito o processo de extração de informações, permitindo que mais soluções automatizadas possam ser implantadas.

Em trabalhos futuros, deseja-se incorporar no sistema de recomendação aspectos oriundos da análise da rede de colaboração, utilizada neste caso de estudo, apenas para identificar aqueles pesquisadores que já são colaboradores, para integrar fatores sociais como métricas adicionais de similaridade, construindo assim um sistema de recomendação híbrido.

Além dos aspectos da Análise de Redes Sociais, deseja-se identificar outras métricas de similaridade e outras fontes de informação sobre os pesquisadores para enriquecer a recomendação baseada em conteúdo. Isto será feito através da extração de dados em bibliotecas digitais, redes sociais de pesquisa, etc.

REFERÊNCIAS

- BEEL, J. et al. Research-paper recommender systems: a literature survey. **International Journal on Digital Libraries**, Springer, v. 17, n. 4, p. 305–338, 2016. Citado na página 30.
- BOBADILLA, J. et al. Recommender systems survey. **Knowledge-based systems**, Elsevier, v. 46, p. 109–132, 2013. Citado 2 vezes nas páginas 27 e 29.
- BORDIN, A. S.; KELM, W. Análise de redes de colaboração científica em uma universidade multicampi. In: UNIPAMPA. **Relatório técnico do projeto de pesquisa Análise e Recomendação da Rede de Colaboração Científica da Unipampa**. [S.l.], 2016. p. 1–12. Citado na página 26.
- BOYACK, K. W.; KLAUVANS, R. Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? **Journal of the American Society for Information Science and Technology**, Wiley Online Library, v. 61, n. 12, p. 2389–2404, 2010. Citado na página 33.
- BOZEMAN, B.; FAY, D.; SLADE, C. P. Research collaboration in universities and academic entrepreneurship: the-state-of-the-art. **The Journal of Technology Transfer**, Springer, v. 38, n. 1, p. 1–67, 2013. Citado na página 25.
- DAVOODI, E.; KIANMEHR, K.; AFSHARCHI, M. A semantic social network-based expert recommender system. **Applied intelligence**, Springer, v. 39, n. 1, p. 1–13, 2013. Citado na página 32.
- ELAHI, M.; RICCI, F.; RUBENS, N. A survey of active learning in collaborative filtering recommender systems. **Computer Science Review**, Elsevier, v. 20, p. 29–50, 2016. Citado na página 28.
- GORDON, M. A critical reassessment of inferred relations between multiple authorship, scientific collaboration, the production of papers and their acceptance for publication. **Scientometrics**, Akadémiai Kiadó, co-published with Springer Science+ Business Media BV, Formerly Kluwer Academic Publishers BV, v. 2, n. 3, p. 193–201, 1980. Citado na página 25.
- HARA, N. et al. An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration. **Journal of the American Society for Information Science and Technology**, Wiley Subscription Services, Inc., A Wiley Company, v. 54, n. 10, p. 952–965, 2003. ISSN 1532-2890. Disponível em: <<http://dx.doi.org/10.1002/asi.10291>>. Citado na página 21.
- HARA, N. et al. An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration. **Journal of the American Society for Information science and Technology**, Wiley Online Library, v. 54, n. 10, p. 952–965, 2003. Citado na página 25.
- HECK, T. Chapter 11 recommendation for social networking in academia. In: **Social Information Research**. [S.l.]: Emerald Group Publishing Limited, 2012. p. 239–265. Citado na página 40.

- HECK, T. **Expert Recommendation for the Academic Field**. Tese (Doutorado) — Dissertation, Düsseldorf, Heinrich-Heine-Universität, 2015, 2016. Citado 2 vezes nas páginas 23 e 40.
- HECK, T.; HANRATHS, O.; STOCK, W. G. Expert recommendation for knowledge management in academia. **Proceedings of the Association for Information Science and Technology**, Wiley Online Library, v. 48, n. 1, p. 1–4, 2011. Citado 3 vezes nas páginas 40, 41 e 42.
- HECK, T.; PETERS, I. Expert recommender systems: Establishing communities of practice based on social bookmarking systems. In: **Proceedings of I-Know**. [S.l.: s.n.], 2010. p. 458–464. Citado 3 vezes nas páginas 39, 40 e 41.
- HECK, T.; PETERS, I.; STOCK, W. G. Testing collaborative filtering against co-citation analysis and bibliographic coupling for academic author recommendation. In: **Proceedings of the 3rd ACM RecSys' 11 Workshop on Recommender Systems and the Social Web**. [S.l.: s.n.], 2011. p. 16–23. Citado 4 vezes nas páginas 39, 40, 41 e 42.
- JANNACH, D. et al. **Recommender systems: an introduction**. Shaftesbury Rd, Cambridge CB2 8BS, UK: Cambridge University Press, 2010. Citado 3 vezes nas páginas 28, 29 e 31.
- KATZ, J. S.; MARTIN, B. R. What is research collaboration? **Research policy**, Elsevier, v. 26, n. 1, p. 1–18, 1997. Citado na página 25.
- KELM, W. K.; BORDIN, A. S. Uma análise das relações de coautoria mais frequentes em uma universidade multicampi. **Anais do Salão Internacional de Ensino, Pesquisa e Extensão**, v. 8, n. 2, 2017. Citado na página 22.
- KESSLER, M. M. **An experimental study of bibliographic coupling between technical papers**. [S.l.], 1962. Citado na página 33.
- KITCHENHAM, B. et al. Systematic literature reviews in software engineering—a systematic literature review. **Information and software technology**, Elsevier, v. 51, n. 1, p. 7–15, 2009. Citado na página 37.
- LIKERT, R. A technique for the measurement of attitudes. **Archives of psychology**, 1932. Citado na página 51.
- LIU, B. **Web data mining: exploring hyperlinks, contents, and usage data**. [S.l.]: Springer Science & Business Media, 2007. Citado na página 35.
- LOPES, G. R. et al. Collaboration recommendation on academic social networks. In: SPRINGER. **International Conference on Conceptual Modeling**. [S.l.], 2010. p. 190–199. Citado na página 32.
- MA, R. Author bibliographic coupling analysis: A test based on a chinese academic database. **Journal of Informetrics**, Elsevier, v. 6, n. 4, p. 532–542, 2012. Citado na página 34.
- MAKAROV, I.; BULANOV, O.; ZHUKOV, L. E. Co-author recommender system. In: SPRINGER. **International Conference on Network Analysis**. [S.l.], 2016. p. 251–257. Citado na página 32.

- MATTESSICH, P. W.; MONSEY, B. R. **Collaboration: what makes it work. A review of research literature on factors influencing successful collaboration.** 451 Lexington Pkwy N, St Paul, MN 55104, EU: ERIC, 1992. Citado na página 21.
- MELIN, G. Pragmatism and self-organization: Research collaboration on the individual level. **Research policy**, Elsevier, v. 29, n. 1, p. 31–40, 2000. Citado na página 21.
- MENA-CHALCO, J. P.; JUNIOR, R. M. C. Scriptlattes: an open-source knowledge extraction system from the lattes platform. **Journal of the Brazilian Computer Society**, Springer, v. 15, n. 4, p. 31–39, 2009. Citado na página 26.
- PU, P.; CHEN, L.; HU, R. A user-centric evaluation framework for recommender systems. In: ACM. **Proceedings of the fifth ACM conference on Recommender systems.** [S.l.], 2011. p. 157–164. Citado na página 51.
- ROBERTSON, S. Understanding inverse document frequency: on theoretical arguments for idf. **Journal of documentation**, Emerald Group Publishing Limited, v. 60, n. 5, p. 503–520, 2004. Citado na página 30.
- SONNENWALD, D. H. Scientific collaboration. **Annual review of information science and technology**, Wiley Online Library, v. 41, n. 1, p. 643–681, 2007. Citado na página 21.
- WUCHTY, S.; JONES, B. F.; UZZI, B. The increasing dominance of teams in production of knowledge. **Science**, American Association for the Advancement of Science, v. 316, n. 5827, p. 1036–1039, 2007. Citado na página 21.
- XU, Y. et al. Combining social network and semantic concept analysis for personalized academic researcher recommendation. **Decision Support Systems**, Elsevier, v. 54, n. 1, p. 564–573, 2012. Citado na página 23.
- ZHAO, D.; STROTMANN, A. Author bibliographic coupling: Another approach to citation-based author knowledge network analysis. **Proceedings of the American Society for Information Science and Technology**, Wiley Online Library, v. 45, n. 1, p. 1–10, 2008. Citado na página 34.
- ZHAO, D.; STROTMANN, A. Evolution of research activities and intellectual influences in information science 1996–2005: Introducing author bibliographic-coupling analysis. **Journal of the American Society for Information Science and Technology**, Wiley Online Library, v. 59, n. 13, p. 2070–2086, 2008. Citado na página 35.
- ZUCKERMAN, H. Nobel laureates in science: Patterns of productivity, collaboration, and authorship. **American Sociological Review**, [American Sociological Association, Sage Publications, Inc.], v. 32, n. 3, p. 391–403, 1967. ISSN 00031224. Disponível em: <<http://www.jstor.org/stable/2091086>>. Citado na página 21.

Apêndices

APÊNDICE A – CONVITE PARA AVALIAÇÃO DO SISTEMA DE RECOMENDAÇÃO

Prezado(a) pesquisador(a), primeiramente agradecemos pela sua colaboração com a autorização do uso dos formulários de projetos de pesquisa da SIPPEE.

Nesta última etapa, gostaríamos de convidá-lo para realizar uma (Apêndice A) pelo sistema .

O sistema recomenda pesquisadores com potencial para colaboração científica, através da identificação de similaridade das (Apêndice A) e **referências bibliográficas** de projetos de pesquisa.

Para realizar a avaliação, desenvolvemos um sistema *web*, onde você terá acesso às recomendações de pesquisadores que lhe foram feitas.

Depois de verificar as recomendações, solicitamos que você responda 10 questões em um formulário que aborda aspectos que devem ser considerados para a avaliação das recomendações do sistema do ponto de vista do usuário final.

Para acessar a aplicação é importante que você esteja **em um dos computadores da instituição**, pois a mesma está hospedada dentro de um serviço interno na rede da UNIPAMPA.

Observações importantes: Além das suas recomendações, você terá acesso a uma visão geral de todas as recomendações feitas para todos os pesquisadores, porém, no momento da avaliação, pedimos que **considere apenas as recomendações que foram feitas para você** através do link “Minhas recomendações” no menu superior.

Para acessar a aplicação e obter instruções mais detalhadas, basta acessar a página: <<http://200.132.146.13:8080/>>

APÊNDICE B – TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Prezado(a) pesquisador(a), convidamos você a colaborar em uma pesquisa de Trabalho de Conclusão de Curso (TCC) intitulada Recomendação de Pesquisadores na UNIPAMPA: Uma Proposta Baseada no Conteúdo de Projetos de Pesquisa, do acadêmico de Ciência da Computação Wolgan Ens, orientado pela profa. Andréa Sabedra Bordin, que tem por objetivo recomendar pesquisadores com interesses de pesquisa semelhantes para outros pesquisadores da Unipampa.

A metodologia da pesquisa baseia-se no método de acoplamento bibliográfico e do perfil das palavras-chave utilizadas nos projetos de pesquisa submetidos ao SIPPEE. Assim, serão extraídos exclusivamente dois conjuntos de dados dos projetos de pesquisa: palavras-chave e referências bibliográficas.

A colaboração aqui solicitada se caracteriza pelo consentimento de utilização dos formulários de projeto de pesquisa submetidos ao SIPPEE desde o seu ingresso na Unipampa.

Como benefícios dessa colaboração destaca-se a oportunidade de verificar como os dados produzidos e armazenados na instituição podem ser utilizados para auxiliar pesquisadores a encontrar pesquisadores com interesses semelhantes, fomentando assim a colaboração acadêmica e, conseqüentemente, a qualificação das pesquisas.

A qualquer tempo você poderá solicitar esclarecimentos adicionais sobre o estudo em qualquer aspecto que desejar. Também poderá retirar seu consentimento a qualquer momento, sem sofrer qualquer tipo de penalidade ou prejuízo.

Com exceção das palavras-chave e referências bibliográficas, os demais dados (seções) do projeto não serão utilizados na pesquisa. Após a extração dos dados requeridos, os arquivos dos projetos não serão mantidos pelos pesquisadores.

Nome do pesquisador:

Assinatura

APÊNDICE C – AVALIAÇÃO DO SISTEMA DE RECOMENDAÇÃO

A avaliação do usuário visa cobrir diversos aspectos que devem ser considerados para a avaliação das recomendações feitas pelo sistema. Para isso, aspectos do ponto de vista do usuário final são abordados através das questões presentes abaixo. Os aspectos aqui mencionados são os propostos no trabalho de Pu, Chen e Hu (2011): precisão percebida, novidade (ou descoberta), diversidade, transparência, atratividade, facilidade de uso, utilidade percebida, intenções do usuário

Nome:

-
1. Os pesquisadores recomendados trabalham em linhas/áreas de pesquisa em que sou capaz de contribuir. *
 2. Descubri pesquisadores com potencial de colaboração que eu não conhecia anteriormente. *
 3. Foram sugeridos diversos pesquisadores. *
 4. Consegui perceber porque os pesquisadores foram recomendados para mim. *
 5. Fiquei entusiasmado com as novas possibilidades de colaboração. *
 6. Não encontrei dificuldades na utilização do sistema. *
 7. A tarefa de encontrar colegas para colaborar em projetos ficou mais fácil de ser executada. *
 8. Caso o sistema fosse implantado, utilizaria o mesmo frequentemente. *
 9. Caso deseje inserir informações adicionais ou escrever sobre suas escolhas nas opções anteriores, utilize este campo.
-

Anexos

