



UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO

CENTRO DE CIÊNCIAS EXATAS E TECNOLOGIA

PROGRAMA DE PÓS-GRADUAÇÃO EM INFORMÁTICA

RECOMENDAÇÕES DE SESSÕES PARA APOIAR A PARTICIPAÇÃO EM EVENTOS  
CIENTÍFICOS

Henrique Gurgacz de Almeida

**Orientadores**

Sean Wolfgang Matsui Siqueira

Leila Cristina Vasconcellos de Andrade

RIO DE JANEIRO – RJ – BRASIL

SETEMBRO – 2018

## **RECOMENDAÇÕES DE SESSÕES PARA APOIAR A PARTICIPAÇÃO EM EVENTOS CIENTÍFICOS**

Dissertação apresentada ao Programa de Pós-Graduação em Informática da Universidade Federal do Estado do Rio de Janeiro, como pré-requisito para a obtenção do grau de Mestre em Curso de Informática.

Orientação: Sean Wolfgang Matsui Siqueira  
e Leila Cristina Vasconcelos de Andrade

RIO DE JANEIRO – RJ – BRASIL

SETEMBRO – 2018

A447 Almeida, Henrique Gurgacz de  
Recomendação de sessões para apoiar a participação  
em eventos científicos / Henrique Gurgacz de  
Almeida. -- Rio de Janeiro, 2018.  
69

Orientador: Sean Wolfgang Matsui Siqueira.  
Coorientadora: Leila Cristina Vasconcelos de  
Andrade.

Dissertação (Mestrado) - Universidade Federal do  
Estado do Rio de Janeiro, Programa de Pós-Graduação  
em Informática, 2018.

1. Sistemas de Recomendação. 2. Eventos  
Científicos. I. Siqueira, Sean Wolfgang Matsui,  
orient. II. Andrade, Leila Cristina Vasconcelos de,  
coorient. III. Título.

RECOMENDAÇÃO DE SESSÕES PARA APOIAR A PARTICIPAÇÃO EM EVENTOS CIENTÍFICOS

Henrique Gurgacz de Almeida

DISSERTAÇÃO APRESENTADA COMO REQUISITO PARCIAL PARA OBTENÇÃO DO TÍTULO DE MESTRE PELO PROGRAMA DE PÓSGRADUAÇÃO EM INFORMÁTICA DA UNIVERSIDADE FEDERAL DO ESTADO DO RIO DE JANEIRO (UNIRIO). APROVADA PELA COMISSÃO EXAMINADORA ABAIXO ASSINADA.

Aprovada por:

---

Sean Wolfgang Matsui Siqueira, D.Sc (Orientador) – UNIRIO

---

Leila Cristina Vasconcelos de Andrade, D.Sc (Orientador) – UNIRIO

---

Bernardo Pereira Nunes, D.Sc – UNIRIO

---

Jairo Francisco de Souza, D.Sc – UFJF

RIO DE JANEIRO, RJ - BRASIL

SETEMBRO - 2018

“A vida é construída nos sonhos e concretizada no amor”  
Chico Xavier

## **AGRADECIMENTOS**

Agradeço a DEUS.

A toda a minha família pelo apoio, incentivo, paciência e revisões de trabalhos ao longo de todo curso.

Aos meus orientadores Sean e Leila pela oportunidade, paciência, sabedoria e apoio nesta jornada.

Aos membros da comissão organizadora do Simpósio Brasileiro de Sistemas de Informação 2018 pelo apoio no processo de avaliação da presente pesquisa.

Aos professores do PPGI da UNIRIO por apresentar o meio científico e as pesquisas acadêmicas, pela contribuição na formação de uma nova forma de pensar e por abrir novos horizontes.

Aos colegas do grupo de pesquisa que contribuíram nas discussões, trocas de experiências e deram garra para seguir em frente.

ALMEIDA, Henrique Gurgacz de. **Recomendação de sessões para apoiar a participação em eventos científicos**. UNIRIO, 2018. 69 páginas. Dissertação de Mestrado. Departamento de Informática Aplicada, UNIRIO.

## RESUMO

Os eventos científicos são um importante meio de publicação e de troca de experiências que auxilia no desenvolvimento das pesquisas. Além disso, a maior parte dos eventos científicos possui muitas sessões, temas e salas com conteúdo simultâneos. Apesar da importância para o pesquisador em obter o melhor aproveitamento dos eventos, as principais ferramentas de *software* utilizadas nos eventos científicos não disponibilizam recursos que auxiliam na participação. Neste contexto, os Sistemas de Recomendação mostram-se como uma possível solução para esses problemas. Apesar dos Sistemas de Recomendação serem aplicados em várias áreas, quando trata-se de eventos científicos, as particularidades deste meio trazem a necessidade de fazer uma análise específica e de definir um novo método de recomendação. Temas como temporalidade, tratamento do espaço físico nas recomendações e regras específicas dos eventos precisam ser consideradas antes de qualquer recomendação. Há também a necessidade de mapeamento do perfil dos usuários, para que a recomendação possa ser feita de forma automática. Neste caso, faz-se necessário a identificação das melhores fontes de dados de assuntos relacionados a pesquisas e temas de interesse. A solução proposta neste trabalho tratou de todos os aspectos acima descritos, sugerindo assim um método de recomendação específico para eventos científicos. A partir do método proposto foi construído um *software* chamado Conference Recommender que foi utilizado em um evento científico real como estudo de caso. As recomendações foram avaliadas pelos participantes do evento e como resultado foi obtida a maioria de avaliações positivas. Apesar de ainda necessitar de melhorias em trabalhos

futuros, percebeu-se que o método criado pode auxiliar e trazer um melhor aproveitamento na participação de eventos científicos.

**Palavras-chave:** Sistemas de Recomendação, Eventos Científicos



## ABSTRACT

The scientific conferences are an important way of publication and an opportunity to exchange experiences, helping in the researches developments. However, most of the scientific conferences have many sessions, topics and rooms with simultaneous presentations. Even though of the importance to researchers to obtain the best of the scientific conferences, the main tools used in these events do not provide features to help them over the event. In this context the Recommendation Systems are shown as a possible solution to these problems. Despite the recommendation systems are applied in many areas, the scientific conferences characteristics demands a specific analysis and recommendation method. Moreover, topics like temporality, treatment of physical spaces in the recommendations, and specific rules for scientific conferences needs to be considered for all recommendations. Furthermore, the user profile mapping and construction need to be considered to allow automatic recommendations. In this case, the best data sources regarding researches and scientific interest needs to be identified. The proposal solution in this master thesis deals with all the aspects mentioned, thus suggesting a recommendation method specifically to scientific conferences. From the proposed method, a software was designed, known as Conference Recommender, which was applied in a real scientific event as a case study. The recommendations were rated by the participants of the event, thus obtaining many positive rates. Although it still needs future improvements, it was noticed that the designed method can help and bring better use in the participation of scientific events.

**Keywords:** Recommendation Systems, Scientific Conference.

# SUMÁRIO

<b>1. INTRODUÇÃO .....</b>	<b>12</b>
<b>1.1. Motivação .....</b>	<b>18</b>
<b>1.2. Problema de Pesquisa .....</b>	<b>19</b>
<b>1.3. Objetivo de Pesquisa.....</b>	<b>20</b>
<b>1.4. Metodologia de Pesquisa .....</b>	<b>20</b>
<b>1.5. Estrutura da Dissertação.....</b>	<b>21</b>
<b>2. CONCEITOS FUNDAMENTAIS .....</b>	<b>22</b>
<b>2.1. Sistemas de Recomendação em Eventos Científicos .....</b>	<b>22</b>
<b>2.1.1. Evento Científico .....</b>	<b>23</b>
<b>2.1.2. Ferramentas Tecnológicas para Eventos Científicos .....</b>	<b>24</b>
<b>3. EXPERIMENTO CBIE 2017.....</b>	<b>26</b>
<b>3.1. Fonte de dados de Pesquisadores e Sessões de eventos científicos .....</b>	<b>27</b>
<b>3.2. Ferramentas de Análise de Texto .....</b>	<b>29</b>
<b>3.3. Método Utilizado no Experimento .....</b>	<b>32</b>
<b>3.4. Limpeza dos Dados e Análises .....</b>	<b>33</b>
<b>3.5. Conclusões e Resultados do Experimento .....</b>	<b>35</b>
<b>4. PROPOSTA DO SISTEMA DE RECOMENDAÇÃO.....</b>	<b>37</b>
<b>4.1. Método de Recomendação.....</b>	<b>37</b>
<b>4.1.1. Extração de Dados das Sessões .....</b>	<b>38</b>
<b>4.1.2. Extração de dados dos Inscritos.....</b>	<b>39</b>
<b>4.1.3. Análise e Recomendação.....</b>	<b>40</b>
<b>4.2. Arquitetura e Artefatos .....</b>	<b>42</b>
<b>4.3. Validação da Solução.....</b>	<b>47</b>
<b>5. ANÁLISE DOS RESULTADOS DO ESTUDO DE CASO.....</b>	<b>52</b>
<b>6. CONSIDERAÇÕES FINAIS .....</b>	<b>59</b>

<b>6.1. Contribuições .....</b>	<b>59</b>
<b>6.2. Conclusão.....</b>	<b>60</b>
<b>6.3. Trabalhos Futuros .....</b>	<b>61</b>
<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>63</b>

## ÍNDICE DE FIGURAS

<b>Figura 1: Tela inicial da aplicação Web utilizada no experimento no CBIE 2017.....</b>	<b>27</b>
<b>Figura 2: Exemplo de Resumo de um trabalho analisado no CBIE 2017 .....</b>	<b>31</b>
<b>Figura 3: Método utilizado no experimento.....</b>	<b>33</b>
<b>Figura 4: Diagrama do método de recomendação.....</b>	<b>38</b>
<b>Figura 5: Módulos da solução proposta .....</b>	<b>42</b>
<b>Figura 6: Módulos da solução e atividades de cada módulo .....</b>	<b>43</b>
<b>Figura 7: Fluxo dos dados entre os módulos da solução.....</b>	<b>45</b>
<b>Figura 8: Tela inicial da solução .....</b>	<b>46</b>
<b>Figura 9: Tela interna da solução após acesso do usuário.....</b>	<b>46</b>
<b>Figura 10: Pagina de programação do SBSI 2018 com o link para o Conference Recommender .....</b>	<b>50</b>
<b>Figura 11: E-mail enviado para os inscritos pelo Conference Recommender.....</b>	<b>51</b>
<b>Figura 12: Quantidade de recomendações por grupo de usuários .....</b>	<b>54</b>
<b>Figura 13: Quantidade de recomendações por categoria do usuário .....</b>	<b>55</b>
<b>Figura 14: Quantidade de avaliações ruins por média de peso.....</b>	<b>56</b>
<b>Figura 15: Quantidade de avaliações eliminando baixo peso.....</b>	<b>57</b>

## ÍNDICE DE TABELAS

<b>Tabela 1: Comparação dos recursos das principais ferramentas utilizadas em eventos científicos .....</b>	<b>24</b>
<b>Tabela 2: Relação de Informações disponibilizadas na Plataforma Lattes e no ResearchGate .....</b>	<b>29</b>
<b>Tabela 3: Comparação das ferramentas de análise de texto .....</b>	<b>31</b>
<b>Tabela 4: Palavras-Chave não congruentes retornadas pelas ferramentas de análise de texto.....</b>	<b>32</b>
<b>Tabela 5: Usuários participantes do experimento.....</b>	<b>34</b>
<b>Tabela 6: Similaridade de usuários com sessões.....</b>	<b>34</b>
<b>Tabela 7: Números do módulo tradutor e analisador de texto .....</b>	<b>52</b>
<b>Tabela 8: Dados Gerais do uso do Conference Recommender no SBSI 2018 .....</b>	<b>53</b>
<b>Tabela 9: Avaliação das recomendações de acordo com o perfil dos usuários .....</b>	<b>55</b>
<b>Tabela 10: Avaliação dos usuários baseados na quantidade de palavras-chave similares .....</b>	<b>56</b>
<b>Tabela 11: Relação entre o tamanho do perfil do usuário e a avaliação dada.....</b>	<b>57</b>

## LISTA DE NOMENCLATURAS

AAAI - *Association for the Advancement of Artificial Intelligence*

ACM – *Association for Computing Machinery*

CBIE - Congresso Brasileiro de Informática na Educação

CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico

DSR - *Design Science Research*

DSRM - *Design Science Research Methodology*

IEEE - *Institute of Electrical and Electronics Engineers*

SBSI – Simposio Brasileiro de Sistemas de Informação

UNIRIO – Universidade Federal do Estado do Rio de Janeiro

## 1. INTRODUÇÃO

O Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), agência do Ministério da Ciência, Tecnologia, Inovação e Comunicação, criado em 1951, tem como principais atribuições fomentar a pesquisa científica e tecnológica e incentivar a formação de pesquisadores brasileiros (CNPQ, 2018). Os indicadores do CNPq, disponibilizados de forma aberta no seu site, demonstram o crescimento da produção científica brasileira em todos os aspectos.

Diretamente associada a produção científica, estão os eventos, as conferências e os simpósios. Anualmente novos eventos são criados, contando com cada vez mais participantes (ARBOIT e BUFREM, 2011). De acordo com Lacerda *et al.* (2008), os eventos científicos assumem um papel de grande importância no processo da comunicação científica na medida em que a transmissão de ideias e fatos novos chegam ao conhecimento da comunidade científica de maneira mais rápida, quando comparados com os veiculados pelos meios formais de comunicação.

Ainda de acordo com Lacerda *et al.* (2008) em uma de suas pesquisas, mais de 97% dos entrevistados informaram que a participação em eventos científicos é importante ou muito importante, especialmente para o desenvolvimento do conhecimento científico.

Apesar da importância dos eventos científicos, a participação traz alguns desafios. Para Corrêa *et al.* (2009), em grandes eventos ocorrem simultaneamente diversas atividades, o que pode dificultar a escolha ou a participação por parte dos participantes nos melhores temas.

Além da dificuldade de escolha das atividades dos eventos, as atuais ferramentas tecnológicas usadas nos eventos científicos são sites e aplicativos para dispositivos móveis que, em sua grande maioria, permitem apenas a inscrição, o acompanhamento e a visualização da programação do evento, das informações de contato e do endereço do evento. A ausência de

inteligência e recursos de recomendações nestas ferramentas fazem com que o usuário não tenha um bom auxílio tecnológico na participação dos eventos científicos.

Neste cenário, uma importante ferramenta para auxiliar nas decisões do dia-a-dia são os sistemas de recomendação. Tais sistemas tem por objetivo melhorar a experiência dos usuários com os sistemas de informação e trazer maior produtividade, uma vez que eles conseguem combinar técnicas computacionais para selecionar itens personalizados com base nos interesses dos usuários e conforme o contexto (RICCI, ROKACH e SHAPIRA, 2015).

Sistemas de recomendação surgiram como uma área de pesquisa independente em meados dos anos 90 e até hoje apresenta grande interesse de estudo e aumento no número de pesquisas (RICCI, ROKACH e SHAPIRA, 2015). O uso constante destes sistemas em importantes soluções de mercado como Youtube<sup>1</sup>, Netflix<sup>2</sup>, Spotify<sup>3</sup> e Facebook<sup>4</sup>, demonstram um pouco da importância e da relevância dos sistemas de recomendação na atualidade.

A recomendação é parte natural no processo social de tomada de decisão, em que pessoas costumam usar recomendações de outros indivíduos ou autoridades em seus processos de escolha (RESNICK e VARIAN, 1997). Com esta observação e também em resposta às necessidades dos usuários de receberem recomendações dos sistemas de informação, Tapestry trouxe os primeiros sistemas de recomendação, nesta época denominado de filtragem colaborativa. Apenas em 1997 Paul Resnick e Hal Varian criaram o referido termo (RESNICK e VARIAN, 1997).

Para Ricci *et al.* (2015) e para Motta *et al.* (2012), a abundância de informações disponíveis na Web e nas bibliotecas digitais, em combinação com sua natureza dinâmica e heterogênea, determinou uma dificuldade crescente em encontrar o que se procura, de forma precisa e de uma maneira que melhor atenda aos requisitos desejados. Sendo este o grande impulsionador dos sistemas de recomendação.

Ricci *et al.* (2015) afirmam também que os sistemas de recomendação são ferramentas de *softwares* e técnicas que provêm sugestões para itens a serem utilizados por usuários. Para eles, as sugestões estão relacionadas a vários processos de tomada de decisão, como por exemplo a compra de um livro, a escolha de uma música para ouvir ou uma notícia para ler.

---

<sup>1</sup> Youtube: [www.youtube.com](http://www.youtube.com)

<sup>2</sup> Netflix: [www.netflix.com](http://www.netflix.com)

<sup>3</sup> Spotify: [www.spotify.com](http://www.spotify.com)

<sup>4</sup> Facebook: [www.facebook.com](http://www.facebook.com)



Desta forma, com o avanço ao longo dos anos das tecnologias de informação e comunicação, bem como com o incremento do uso de sistemas e o aumento da massa de dados, é cada vez mais importante que os usuários recebam recomendações dos sistemas, uma vez que a variedade de opções e escolhas que os usuários podem tomar, passam a ser, muitas vezes, extremamente ampla, inviabilizando a tomada de decisão. Por este motivo os sistemas de recomendação, mais especificamente a filtragem dos dados, passam a ser uma ferramenta muito importante para os usuários, uma vez que os dados a serem escolhidos são filtrados, permitindo assim uma escolha mais rápida e adequada a seu desejo (MOTTA, GARCIA, *et al.*, 2012).

Para Ricci, Rokach e Shapira (2015) e Adomavicius e Tuzhilin (2005), os métodos clássicos de sistemas de recomendação podem ser classificados em três diferentes categorias:

- filtragem baseada em conteúdo: analisa documentos, dados ou descrições de itens previamente classificados por um usuário e constroem, com estes dados, o perfil de interesses do usuário. Uma vez que o perfil esteja devidamente construído, as recomendações podem ser feitas através da associação de características do perfil do usuário com características do conteúdo de outros objetos ou usuários (RICCI, ROKACH e SHAPIRA, 2015).
- filtragem colaborativa: filtra itens para serem recomendados baseados em comportamento, avaliação e preferência de outros usuários (SILVA, 2017). Há duas possíveis abordagens, na abordagem baseada em vizinhança o foco está no relacionamento entre itens ou entre usuários. Ou seja, usuários diferentes que fizeram boa avaliação em produtos semelhantes são chamados de vizinhos e estes por sua vez possivelmente possuem interesses semelhantes, facilitando assim a recomendação. Já na abordagem de fatoração de matrizes é realizada a associação da nota dos usuários aos itens e, com modelos matemáticos, são extraídos os fatores latentes com o grau de interesse dos usuários para cada fator, permitindo fazer as devidas recomendações (JÚNIOR, 2017) e (RICCI, ROKACH e SHAPIRA, 2015).
- filtragem híbrida: criado com o intuito de mitigar limitações existentes em cada um dos métodos de recomendação, muitos sistemas têm utilizado um modelo híbrido, em que as abordagens do sistema por filtragem colaborativa são combinadas com as abordagens dos sistemas com filtragem baseada em contexto (RICCI, ROKACH e SHAPIRA, 2015).

Entre as principais vantagens dos sistemas de recomendação com filtragem baseada em conteúdo podem ser citadas (RICCI, ROKACH e SHAPIRA, 2015):

- independência de outros usuários: para recomendar itens a um usuário, o sistema necessita de dados apenas deste mesmo usuário;
- transparência: o sistema é capaz de fornecer detalhes sobre cada item recomendado e suas características, uma vez que é feita uma comparação de similaridade entre o item e o usuário;
- item novo: para que um novo item seja recomendado basta que suas características sejam similares às do perfil de algum usuário. Não é necessário avaliar, comentar ou obter dados do usuário sobre o item novo.

Por outro lado, esta abordagem de recomendação também possui alguns problemas (RICCI, ROKACH e SHAPIRA, 2015):

- análise limitada de conteúdo: a extração das características dos itens para serem recomendados é uma tarefa muito complexa e que possui limitações. Como exemplo, um item bem avaliado sobre *rally*, do qual foi extraída a característica “corrida de carros”, pode levar o sistema a inferir que o usuário gosta de corridas de carros, quando na verdade ele gosta de *rally*, mas odeia Fórmula Indy;
- super especialização: uma vez que esta abordagem de recomendação recomenda apenas itens com alto grau de similaridade, itens com certo nível de novidade podem acabar não sendo recomendados;
- problema do novo usuário: o sistema apenas consegue prover recomendações confiáveis para um usuário quando conseguir construir o perfil do mesmo.

Além dos sistemas de recomendação com filtragem baseada em contexto, colaborativa e híbrida, há também outros modelos mais recentes, foco de trabalhos científicos atuais, são eles (JÚNIOR, 2017) e (COSTA e MACEDO, 2012):

- recomendação com filtragem contextual: os sistemas de recomendação com filtragem contextual exploram o contexto no qual o usuário está inserido para fazer recomendações adequadas ao momento, local e dispositivo utilizado;
- recomendação baseada em conhecimento: sistemas baseados em conhecimento realizam suas recomendações com base em um domínio de conhecimento específico, buscando itens conforme a necessidade e a preferência do usuário;

- recomendação comunitárias: a recomendação comunitária é uma das abordagens mais populares e atuais, principalmente devido ao aumento do uso de redes sociais. Neste modelo a recomendação toma como base toda a rede ou a comunidade de pessoas, suas interações, colaborações e classificações, para dar recomendações a partir de tais dados;
- recomendação demográficas: os sistemas de recomendação demográficos consideram que os interesses dos usuários variam de acordo com o tempo e o espaço. Por este motivo neste modelo é necessário analisar o posicionamento e o horário das ações dos usuários e utilizar tais dados para realizar as recomendações.

Os sistemas de recomendação com filtragem baseada em contexto têm como base a necessidade de capturar a maior quantidade possível de informações para realizar a recomendação. Porém muitas destas informações estão expostas no formato textual, como por exemplo em documentos, Web sites, mensagens de texto, redes sociais, entre outras formas não estruturadas (ADOMAVICIUS e TUZHILIN, 2005) e (POIRIER, TELLIER, *et al.*, 2010).

Tendo em vista esta necessidade, a análise de texto, a mineração de texto, o processamento de linguagem natural e a análise de conteúdo são técnicas importantes para converter informações não estruturadas em palavras-chave ou conceitos estruturados, permitindo armazenar, consultar e executar os métodos de recomendação nos dados extraídos (RICCI, ROKACH e SHAPIRA, 2015) e (PEREIRA E SILVA, 2015).

De acordo com Brito (2017), uma das derivações da mineração de texto é a classificação de texto, técnica que atribui a um ou mais documentos palavras-chave ou classes que ajudem a categorizar os documentos.

A classificação de texto passa na maior parte dos casos pelas seguintes etapas (BASONI, 2015):

- análise léxica: elimina do texto imagens, dígitos e caracteres especiais e deixa o texto inteiro em maiúsculo ou minúsculo;
- eliminação de palavras não discriminativas: elimina do texto as palavras que não contribuem para identificar o assunto do texto. Alguns exemplos de palavras não discriminativas são: sua, seu, esse, meu, tinha, foram, será, qual, quando, muito, há, mesmo, entre outros;

- lematização: identifica as variantes morfológica das palavras e transforma as mesmas no seu radical. Um exemplo são as palavras representação, representado e representando, que podem ser classificados apenas como “represent”;
- frequência do termo: termos com baixa frequência na análise de texto podem não ser relevantes e por este motivo podem ser eliminados;
- ponderação de termos: em casos onde há uma grande lista de termos que classificam um documento, pode ser relevante para a técnica de classificação retornar o peso de um determinado termo frente a todos os demais. Para isso algumas técnicas matemáticas de análise de frequência do termo podem ser aplicadas para calcular o peso.

Neste sentido, por exemplo, a execução de algoritmos de análise de texto no currículo de uma pessoa pode retornar palavras-chave que caracterizam experiência, formação, áreas de interesse e demais informações relevantes de uma pessoa, sendo tais informações muito úteis para os sistemas de recomendação. Por este motivo a análise de texto é uma ferramenta muito importante no processo de recomendação.

Sempre que há a necessidade de extração de palavras-chave ou termos de textos através do processo de análise de texto, podem ser retornados termos não esperados ou pouco relevantes. Neste sentido um processo de normalização de termos se faz necessário.

Para Moraes e Ambrósio (2007), há duas importantes etapas para normalizar os termos. São eles:

- Remoção de *Stopwords*: Nesta etapa as palavras consideradas não relevantes na análise de textos devem ser eliminadas. Estas palavras caracterizam-se justamente por não traduzirem sua essência, por exemplo: sua, seu, onde, contudo e isso;
- Normalização morfológica: nesta etapa procura-se identificar dois ou mais termos que se diferem apenas por questões morfológicas, mas que na essência são termos sinônimos. Como por exemplo, dois termos diferentes simplesmente por um estar no singular e outro no plural.

A remoção de *Stopwords* pode utilizar listas padronizadas, chamadas de *Stoplists*, que são disponibilizadas livremente na internet, porém com pouca ou nenhuma atualização e que

no geral possuem termos genéricos, aplicável para qualquer área de conhecimento no idioma em questão (SAIF, FERNANDEZ, *et al.*, 2014).

Ainda de acordo com Saif *et al.* (2014), é possível que áreas e temas específicos possam utilizar listas especiais de *Stopwords*, visto que tais áreas podem conter termos não relevantes ou então termos muito comuns, que não devem ser considerados em um processo de análise de texto e de extração de palavras-chave.

Para a normalização morfológica, procura-se na maioria das vezes reduzir os termos ao seu radical ou a sua raiz, com isso eliminam-se características de gênero, número ou grau dos termos gerados. Esta técnica de normalização morfológica, que é a mais utilizada, é conhecida por método de lematização. Há outras técnicas como de Porter e Lovins, porém são menos aplicadas e limitadas ao idioma inglês (MORAIS e AMBRÓSIO, 2007).

Desta forma, o uso dos sistemas de recomendação em conjunto com as técnicas comum de análise e normalização de texto, apresenta-se como uma possível solução para apoiar nas dificuldades supracitadas durante a participação de pesquisadores em eventos científicos.

## **1.1. Motivação**

Os eventos científicos são de grande importância para os pesquisadores, uma vez que permitem expor seus próprios trabalhos, conhecer novas pesquisas e pesquisadores, trocar experiências e entender as pesquisas de outras instituições. Naturalmente os eventos científicos ajudam ainda a dar visibilidade e relevância aos trabalhos apresentados, principalmente por serem ambientes respeitados e com grande parte do público altamente capacitado.

Apesar do aumento no número de eventos científicos e da importância dos eventos para o desenvolvimento científico, há uma série de desafios por parte dos pesquisadores para participarem no evento, pode-se citar como exemplo:

- A necessidade de deslocamento até o local do evento;
- A necessidade de ausentar-se do local de trabalho e poder impactar as disciplinas lecionadas ou as pesquisas em andamento;

- O grande número de sessões e artigos apresentados de forma simultânea, exigindo do participante do evento a escolha das melhores sessões;
- A dificuldade de obter o melhor aproveitamento e aprendizado.

Além da dificuldade supracitadas, as poucas pesquisas e divulgação de dados referente a participação e satisfação dos participantes em eventos científicos traz ainda mais relevância para esse tema, uma vez que tais dados podem apresentar novas descobertas.

Outro importante motivador para a realização da presente pesquisa refere-se aos sistemas que são utilizados para apoio nos eventos científicos, visto que grande parte deles não trazem nenhum recurso de análise de perfil, recomendação ou inteligência para de fato apoiar a participação dos pesquisadores no evento.

## **1.2. Problema de Pesquisa**

Apesar do uso dos sistemas de recomendação ser algo usual em diversas áreas, quando tratamos de eventos científicos o mesmo ainda é pouco explorado. Um dos principais problemas de pesquisa abordado nesta dissertação trata-se das particularidades dos eventos científicos para a aplicabilidade do sistema de recomendação. Podem ser citadas como particularidades:

- O fato do evento científico ser temporal, possuindo sessões com horários definidos, faz com que as recomendações precisem tratar tal temporalidade e nunca duas recomendações podem ser feitas para um mesmo usuário no mesmo horário, uma vez que o usuário apenas pode estar em um local por horário;
- O fato do evento científico envolver espaços físicos, de forma que uma sessão é na maioria das vezes realizadas em um único local físico, logo um usuário não pode receber recomendações em locais muito distantes ou então locais diferentes em um mesmo horário;
- As regras particulares dos eventos científicos precisam ser tratadas, uma vez que cada evento científico se difere dos demais quanto ao tema ou público-alvo e um método de recomendação precisa ser amplo e viável para qualquer tipo de evento científico.

Além das particularidades dos eventos científicos, outro importante problema de pesquisa tratado é a identificação da fonte dos dados para o mapeamento do perfil, permitindo

assim a recomendação. Uma vez que se espera que o usuário não precise descrever seu perfil no sistema de recomendação e sim que o sistema identifique automaticamente os interesses do usuário, faz-se necessário encontrar repositórios de dados de pesquisadores e das sessões dos eventos para permitir o mapeamento do perfil.

### **1.3. Objetivo de Pesquisa**

Esta dissertação visa demonstrar que um sistema de recomendação pode auxiliar pesquisadores e demais participantes nos eventos científicos, permitindo assim recomendar aos usuários sessões dos eventos que estejam relacionadas aos seus objetivos de pesquisa, viabilizando indiretamente melhor aproveitamento e aprendizado.

De forma específica, esta dissertação trata os seguintes objetivos:

- Identificar fontes de dados que permitam o mapeamento automático do perfil dos pesquisadores, sem a necessidade de questionar o perfil para cada usuário;
- Propor, construir e avaliar um método e um sistema de recomendação que una ferramentas e recursos existentes ou novos para permitir alcançar o objetivo principal de recomendar sessões de eventos científicos para pesquisadores;
- Explicitar e tratar no método e no sistema de recomendação as especificidades dos eventos científicos para um sistema de recomendação.

### **1.4. Metodologia de Pesquisa**

Esta dissertação seguiu inicialmente uma abordagem de pesquisa exploratória, que para Raul Sidnei Wazlawick define-se como:

“... aquela em que o autor não tem necessariamente uma hipótese ou objetivo definido em mente. Ela pode ser considerada, muitas vezes, como o primeiro estágio de um processo de pesquisa mais longo. Na pesquisa exploratória, o autor vai examinar um conjunto de fenômenos, buscando anomalias que não sejam ainda conhecidas e que possam ser,

então, a base para uma pesquisa mais elaborada. ” (WAZLAWICK, 2014)

Assim como citado por Wazlawick, neste trabalho buscou-se o entendimento dos eventos científicos, as particularidades e características deste meio. Para isso uma revisão da literatura foi realizada, permitindo assim identificar problemas e oportunidades para serem tratadas.

Na sequência, a partir de um conjunto de dados obtidos em um evento científico, foi realizado um experimento em laboratório para auxiliar na formação do método de recomendação proposto nesta dissertação.

Uma vez que o experimento realizado permitiu a construção do método de recomendação, o próximo passo da pesquisa envolveu a construção de uma solução de *software* utilizando o método em questão. Por fim para avaliação e confirmação do atingimento dos objetos da pesquisa, o *software* criado foi avaliado em um estudo de caso durante um evento científico real. Foi aplicado um estudo de caso explanatório (RECKER, 2013), no qual foi possível obter os resultados e conclusões deste trabalho.

## **1.5. Estrutura da Dissertação**

Esta dissertação está organizada em seis capítulos, sendo o primeiro deles o presente capítulo, em que foi apresentada uma breve introdução e contextualização quanto a problemas, motivações, metodologias e objetivos da pesquisa.

O segundo capítulo visa a contextualizar o leitor quanto aos conceitos fundamentais referentes a pesquisa, o terceiro capítulo apresenta detalhes e resultados de um experimento realizado durante a pesquisa e que foi fundamental para a elaboração da proposta do método de recomendação e da arquitetura final da solução, esta que por sua vez é apresentada no capítulo quatro juntamente com a apresentação do método de avaliação da solução.

Por fim, no capítulo cinco são apresentados os resultados do estudo de caso da solução proposta e no capítulo seis são apresentadas conclusões, considerações finais, contribuições e trabalhos futuros.



## **2. CONCEITOS FUNDAMENTAIS**

Neste capítulo são apresentados os principais conceitos e referências para permitir o entendimento de termos, técnicas e métodos utilizados na construção da presente dissertação de mestrado.

### **2.1.Sistemas de Recomendação em Eventos Científicos**

Os indicadores do CNPq, disponibilizados de forma aberta no seu site, demonstram o crescimento da produção científica brasileira em todos os aspectos. Diretamente associadas a produção científica estão os eventos, as conferências e os simpósios. Anualmente novos eventos são criados e estão ganhando cada vez mais participantes (ARBOIT e BUFREM, 2011).

Apesar da importância dos eventos científicos, a participação traz alguns desafios. Para Correa *et al.* (2009), em grandes eventos ocorrem simultaneamente diversas atividades e sessões, o que pode dificultar a escolha ou a participação por parte dos participantes nos melhores temas.

A dificuldade na escolha das sessões em eventos científicos por parte dos pesquisadores e demais participantes dos eventos caracteriza-se como um problema que costuma ter o apoio dos sistemas de recomendação, uma vez que sessões podem ser recomendadas aos participantes do evento por meio destes sistemas. Desta forma, neste capítulo são apresentadas informações e características relevantes referentes aos eventos científicos e ferramentas frequentemente utilizadas.

### 2.1.1. Evento Científico

Evento é uma concentração ou reunião formal e solene de pessoas e/ou entidades realizada em data e local especial, com objetivo de celebrar acontecimentos importantes e significativos, além de estabelecer contatos de natureza comercial, cultural, esportiva, social, familiar, religiosa, científica, etc. (ZANELLA, 2003).

Para a comunidade científica a geração de novos conhecimentos está diretamente relacionada ao desenvolvimento de pesquisas. O suporte para as pesquisas são principalmente as revistas científicas, os capítulos de livros, os trabalhos apresentados em congressos e os demais canais públicos que permitem deixar público as pesquisas e os resultados alcançados. Os eventos científicos fornecem publicação rápida de novas informações e submetem essas informações a críticas imediatas (SANTOS, 2013); (RAJGOLI, 2011).

Além disso, vários estudos, principalmente na área da ciência da computação, demonstram que os eventos científicos possuem um fator de impacto maior do que a publicação em periódicos (RAHM, 2008); (BAR-ILAN, 2010). O relatório de avaliação quadrienal da CAPES para a área de ciência da computação, publicado em 2017, reforça a importância dos eventos científicos para a área e demonstra também importantes números referentes a qualidade, aumento da publicação e participação em tais eventos (NAVAUX, 2017).

Os eventos científicos são uma forma de divulgação de trabalhos presentes em quase todos os campos da ciência. Muitos destes eventos são locais de publicação altamente respeitados, muitos de seus comitês de programa científico incluem líderes e pesquisadores em ascensão de todo o mundo, com membros que se sobrepõem a conselhos editoriais dos principais periódicos da área de pesquisa (CAIRES, 2015).

Ainda de acordo com Caires (2015) os eventos científicos na área de ciência da computação são organizados periodicamente em uma temporada bem definida, por muitas décadas, com os mais antigos fundados no final dos anos 1960 e endossados por associações científicas internacionais estabelecidas, como ACM, IEEE, AAAI.

Entre os trabalhos submetidos em eventos e conferências científicas na área de ciência da computação, algumas bancas avaliadoras chegam a aceitar apenas de 10% a 25% dos trabalhos, demonstrando alta competitividade nestes eventos (CAIRES, 2015).

Apesar da alta concorrência para submissão de trabalhos, os eventos científicos são importantes também para o público em geral, principalmente estudantes e pesquisadores, uma vez que é a oportunidade para atualizarem-se quanto às pesquisas mais recentes nas diversas áreas.

Em uma pesquisa na Universidade Federal de Santa Catarina, Lacerda *et al.* (2008) demonstraram que mais de 97% dos entrevistados acreditam que a participação em eventos científicos é importante ou muito importante, principalmente para o desenvolvimento do conhecimento científico.

### 2.1.2. Ferramentas Tecnológicas para Eventos Científicos

As principais ferramentas tecnológicas usualmente aplicadas nos eventos científicos, tais como (Connfa, 2018), (Galoá, 2018), (Conference Navigator, 2018), (Eventool, 2018), (mConference, 2018) e (Eventmobi, 2018) limitam-se a sites e aplicativos para dispositivos móveis. Na grande maioria, estas ferramentas permitem o acompanhamento e a visualização da programação do evento, informações de contato e endereço do evento.

Na Tabela 1 é possível verificar a lista das ferramentas e as funcionalidades presentes em cada uma delas. As funcionalidades marcadas com o símbolo ✓ indicam que há recursos na ferramenta, já os casos marcados com o símbolo ✕ indicam a ausência da funcionalidade.

**Tabela 1:** Comparação dos recursos das principais ferramentas utilizadas em eventos científicos

Funcionalidade	Connfa	Galoá	Conference Navigator	Eventool	mConference	Eventmobi
Realizar inscrição	✕	✓	✕	✕	✕	✕
Visualizar programação	✓	✓	✓	✓	✓	✓
Visualizar detalhes das sessões do evento	✓	✓	✓	✓	✓	✓
Visualizar informações de contato e endereço do evento	✓	✓	✕	✓	✕	✓
Marcar sessões como favoritas	✓	✓	✕	✓	✓	✓
Integração com redes sociais	✓	✓	✓	✕	✕	✓
Planta do local do evento	✓	✕	✕	✕	✓	✕
Disponível para dispositivo móvel	✓	✓	✓	✓	✓	✓
Notificar usuário no horário das sessões favoritas	✕	✓	✕	✕	✓	✕

Criar notificações e alertas para todos os usuários	×	✓	×	✓	✓	×
Realizar enquetes ou pesquisas	×	✓	×	×	×	✓
Permitir interação entre participantes do evento	×	✓	✓	×	×	×
Recomendar sessões para assistir	×	×	×	×	×	×

A ausência de inteligência e recursos de recomendações é notada nas ferramentas apresentadas na Figura 1. Tais recursos fazem com que o participante não tenha auxílio tecnológico nas tomadas de decisão, como por exemplo de sessões para assistir, que são realizadas durante a participação em eventos científicos.

### 3. EXPERIMENTO CBIE 2017

Com o intuito de avaliar a participação dos pesquisadores em eventos científicos, foi desenvolvida uma aplicação Web que permitisse mapear a participação de pesquisadores no Congresso Brasileiro de Informática na Educação (CBIE) 2017, que ocorreu na Universidade Federal de Pernambuco em Recife entre os dias 30 de outubro e 02 de novembro de 2017.

Os principais objetivos deste experimento foram:

- gerar uma base de dados de usuários e sessões assistidas;
- permitir analisar a coerência da escolha das sessões pelos usuários;
- utilizar a base de dados gerada e as descobertas para auxiliar na criação do algoritmo e método propostos.

Para atingir tais objetivos o experimento envolveu as seguintes etapas:

- identificação de participantes: nesta fase foi selecionado um grupo limitado de pesquisadores inscritos no CBIE 2017. A seleção foi feita com o auxílio de pesquisadores da UNIRIO e com a divulgação em redes sociais. Cada participante interessado no experimento forneceu alguns dados pessoais para o mapeamento de seus perfis;
- mapeamento da participação: com o apoio de uma solução Web, demonstrada na Figura 1, foi possível que cada participante indicasse em tempo real as salas que estavam e com isso, as sessões que estavam assistindo no CBIE 2017;
- classificação das sessões e dos participantes: nesta etapa as sessões e os participantes do experimento tiveram os dados manualmente classificados com o uso de ferramentas de análise de texto;

- consolidação dos resultados: nesta etapa foi possível comparar o resultado da análise de texto de todas as ferramentas avaliadas, bem como identificação da coesão entre os usuários e as sessões assistidas.



Figura 1: Tela inicial da aplicação Web utilizada no experimento no CBIE 2017

O detalhamento das etapas do experimento, especialmente a identificação das fontes de dados, a análise de texto para gerar o perfil para as sessões e o perfil para os usuários são apresentados nas seções a seguir.

### 3.1. Fonte de dados de Pesquisadores e Sessões de eventos científicos

A base dos sistemas de recomendação com filtragem baseada em conteúdo está na captura e análise dos textos, dados e documentos relacionados aos itens a serem recomendados ou aos usuários que receberão as recomendações. O resultado da análise de texto é, na maior parte das vezes, uma lista de palavras-chave para cada item e cada usuário, permitindo realizar a recomendação.

Para efetuar a recomendação é necessário gerar a lista de palavras-chave que classifique adequadamente cada perfil e assim a geração das palavras-chave é feita baseada em textos, documentos e dados. Fez-se necessário identificar algum repositório como fonte dos dados tanto de usuários pesquisadores como de sessões de eventos científicos.

Ao analisar possíveis fontes de dados de pesquisadores, foram mapeados dois sistemas que permitem gerar palavras-chave para gerar o perfil do pesquisador.

Lançado em 2008, o ResearchGate<sup>5</sup> é uma rede social destinada a pesquisadores e acadêmicos de qualquer área do conhecimento com atuação em diversos países. Esta caracteriza-se por ser uma plataforma gratuita que permite a membros interagirem e colaborarem com colegas de trabalho e de campos de estudo, oferecendo diversas ferramentas exclusivas. Diferente do Facebook e Twitter, os usuários das redes sociais para pesquisadores não interagem uns com os outros por postagem de fotos pessoais ou compartilhamento de acontecimentos de sua vida particular, mas mantém o foco nas pesquisas realizadas por cada membro (DAMASCENO, 2016).

Segundo Sanchez, Granado e Antunes (2014)

O grande objetivo do ResearchGate é o de promover a ligação entre cientistas de todo o mundo, para que consigam colaborar mais entre si e dessa forma acelerar o processo de aumento do conhecimento.

No Brasil, além do ResearchGate, outra importante ferramenta com dados de pesquisadores é a Plataforma Lattes<sup>6</sup>, uma base de dados de currículos, grupos de pesquisa e instituições mantida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (CNPQ, 2018).

O Currículo Lattes se tornou um padrão nacional no registro da vida pregressa e atual dos estudantes e pesquisadores do país, e é hoje adotado pela maioria das instituições de fomento, universidades e institutos de pesquisa do País. Por sua riqueza de informações e sua crescente confiabilidade e abrangência, se tornou elemento indispensável e compulsório à análise de mérito e competência dos pleitos de financiamentos na área de ciência e tecnologia (CNPQ, 2018).

Para Fábio e Johanna (2009) a Plataforma Lattes representa a experiência do CNPQ na integração de bases de dados de currículos e de instituições na ciência e tecnologia. Sua base de dados já ultrapassou 1 milhão de registros (currículos) e as informações contidas na plataforma servem tanto para as atividades de gestão, como no apoio a formulação de políticas para a área de ciência e tecnologia.

---

<sup>5</sup> ResearchGate: Rede profissional para cientistas e pesquisadores que ajuda os usuários a se conectarem e facilita o compartilhamento e o acesso à produção científicos. Disponível em <https://www.researchgate.net/> (Researchgate, 2018)

<sup>6</sup> Plataforma Lattes: base de dados de currículos, de grupos de pesquisa e instituições mantida pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico. Disponível em <http://lattes.cnpq.br/> (CNPQ, 2018)

A Tabela 2 apresenta a relação de informações disponibilizadas na Plataforma Lattes e no ResearchGate relacionadas aos pesquisadores. Os dados mais relevantes para o mapeamento do perfil, como por exemplo texto de apresentação, tópicos de interesses e habilidades, projetos de pesquisa e produção científica, podem ser encontrados em ambas as soluções.

**Tabela 2:** Relação de Informações disponibilizadas na Plataforma Lattes e no ResearchGate

<b>Informações Disponibilizadas</b>	<b>Plataforma Lattes</b>	<b>ResearchGate</b>
Texto de Apresentação	✓	✓
Relação de Tópicos de Interesses e Habilidades	✓	✓
Formação Acadêmica e Titulação	✓	✓
Atuação Profissional	✓	✓
Projetos de Pesquisa (Título e Resumo)	✓	✓
Prêmio e título	✓	✓
Produção Científica (Título e Resumo)	✓	✓
Participação em Eventos, Congressos e Feiras	✓	×
Orientações em Andamento	✓	×
Rede Social, contatos, seguidores, autores	×	✓
Acesso Publicações e Artigos Completo	×	✓

Além do mapeamento de dados referente aos pesquisadores, faz-se necessária a identificação de conteúdo que permita o levantamento de perfil das sessões de eventos científicos. Muitas das sessões de eventos científicos estão associadas a artigos ou produções científicas, que por sua vez possuem título, resumo e conteúdo de pesquisa, podendo ser utilizados como base para a análise de texto para mapeamento do perfil.

### **3.2. Ferramentas de Análise de Texto**

Uma vez que as fontes dos dados dos pesquisadores e sessões do evento científico foram identificadas e seus conteúdos foram extraídos, foi feita a análise do conteúdo e com isso armazenado as palavras-chave de cada pesquisador e sessão.

Desta forma, para cumprimento do escopo do experimento, optou-se por mapear as características de algumas ferramentas de análise de texto disponíveis para o amplo uso. Como resultado desta pesquisa as seguintes ferramentas foram analisadas:



- Google Cloud Natural Language: ferramenta disponibilizada pela Google na Web como um serviço. A ferramenta permite identificar o idioma, estrutura sintática, significado, classificação e sentimentos de um texto, sendo ainda a única ferramenta que apresenta peso das análises nos resultados (Google Cloud, 2018). Apesar da vasta documentação técnica para uso da ferramenta, a Google não disponibiliza de forma oficial mais informações sobre o funcionamento, dos algoritmos desta solução;
- Aylien: empresa irlandesa de tecnologia e pesquisas que possui uma ferramenta de análise de conteúdo e tratamento de linguagem natural (Aylien, 2018). Assim como na ferramenta da Google a Aylien também permite identificar sentimentos e entidades dos textos processados, porém há limitações de recursos para alguns idiomas;
- uClassify: ferramenta disponibilizada na Web para fazer classificação de textos, sentimentos, tom da escrita (formal ou informal), identificação de idioma e identificação de classificações para temas específicos (esportes, saúde, artes) (uClassify, 2018). O idioma português não é reconhecido por esta ferramenta;
- ParallelDots: empresa indiana de tecnologia e pesquisas com foco em inteligência artificial. Entre as ferramentas disponibilizadas na Web, a principal delas permite a análise de textos sob a perspectiva dos sentimentos, da geração de palavras-chave, da identificação de entidades, das análises semântica e sintática, bem como da identificação de idioma (ParallelDots, 2018). O idioma português não é reconhecido por esta ferramenta;
- Tag The Web: solução de análise de textos baseado na estrutura taxonômica e do grafo de categorias da Wikipédia (MEDEIROS, NUNES, *et al.*, 2018). Apesar desta solução permitir múltiplos idiomas, apenas o idioma inglês está disponibilizado, uma vez que foi desenvolvido com a extração da base da Wikipédia com este idioma;

Para validar as ferramentas de análise de texto, foram selecionados resumos de 4 trabalhos apresentados no CBIE 2017. Como algumas ferramentas apenas analisam textos em inglês optou-se por selecionar trabalhos que possuíam resumo em inglês (*abstract*). Para cada resumo selecionado, solicitou-se que 3 alunos de pós-graduação da UNIRIO marcassem as palavras-chave que caracterizavam o texto. Para exemplificar este processo, na Figura 2 pode

ser visualizado o resumo em inglês de um dos trabalhos (PAZ e CAZELLA, 2017), que passou por este procedimento. Nesta figura as palavras-chave indicadas pelos estudantes de pós-graduação da UNIRIO foram destacadas em vermelho e sublinhadas.

“This article presents a case study that sought to identify profiles of students with potential for evasion in a community university through the application of the Knowledge Discovery in Database (DCBD) process. The data were collected and refer to undergraduate students enrolled in the second semester of 2016. Data were mined applying the data classification task, using the J48 algorithm and the decision tree technique. The results obtained showed an accuracy of 90%, and it is possible to identify profiles of students evading the Higher Education Institution (HEI) in their campuses and courses.”

Figura 2: Exemplo de Resumo de um trabalho analisado no CBIE 2017

Para finalizar a análise das ferramentas, foram comparadas as palavras-chave previamente selecionadas com o resultado de cada ferramenta de análise de texto. A comparação para o artigo apresentado na Figura 2 pode ser visualizada na Tabela 3, em que foi marcado ✓ os casos de similaridade e ✗ os casos de ausência da palavras-chave no retorno da ferramenta de análise de texto. O mesmo procedimento foi executado com os demais 3 trabalhos selecionados.

Tabela 3: Comparação das ferramentas de análise de texto

Palavra-Chave	Tag the Web	ParallelDots	Uclassify	Aylien	Google
Case study	✓	✓	✗	✗	✓
Profiles of students	✗	✓	✗	✓	✓
Evasion	✗	✓	✗	✓	✓
Community university	✗	✓	✗	✓	✓
Knowledge Discovery in database	✓	✓	✗	✓	✓
Undergraduate students	✗	✓	✗	✓	✓
Data classification task	✓	✓	✗	✗	✗
J48 algorithm	✗	✓	✗	✗	✓
Decision tree technique	✗	✓	✗	✗	✓
Accuracy of 90%	✓	✓	✗	✗	✗

O resultado apresentado na Tabela 3 foi semelhante aos demais 3 trabalhos, em que algumas ferramentas conseguiram mapear a maior parte das palavras-chave previamente

selecionadas, enquanto outras ferramentas não tiveram um resultado satisfatório. Para complementar a análise, foi avaliada a quantidade de palavras-chave retornada pelas ferramentas com pouca coerência com o texto e as palavras-chave inicialmente identificadas. A Tabela 4 apresenta a tabela comparativa entre as ferramentas.

**Tabela 4:** Palavras-Chave não congruentes retornadas pelas ferramentas de análise de texto

<b>Critério</b>	<b>Tag the Web</b>	<b>ParallelDots</b>	<b>Uclassify</b>	<b>Aylien</b>	<b>Google</b>
Quantidade de palavras-chave retornadas	33	31	10	20	23
Quantidade de palavras-chave não congruentes com o texto	17	0	7	0	0
Percentual de palavras-chave congruentes com o texto	48%	100%	30%	100%	100%

Comparando os resultados apresentados nas Tabelas 3 e 4, pode ser concluído que as ferramentas Uclassify e Tag The Web retornam poucas palavras coesas com o texto validado e não retornaram a maior parte das palavras-chave previamente mapeadas. O principal motivo deste resultado para os objetivos deste experimento foi pelo fato destas ferramentas retornarem principalmente palavras-chave de alto nível, como por exemplo: Ciência da Computação, Algoritmos e Sistemas. Apesar destas palavras serem coerentes com o texto, elas estão em um nível muito alto de abstração o que não permite classificar adequadamente o teor do texto avaliado.

Com isso, apesar das ferramentas Google e ParallelDots serem as melhores avaliadas nos testes realizados, optou-se por seguir com a ferramenta Google por ter como diferencial a possibilidade de análise de textos no idioma português e peso nas palavras-chave.

### **3.3. Método Utilizado no Experimento**

Considerando a existência da base de dados da participação de pesquisadores no CBIE 2017, a identificação das melhores fontes de dados de sessões e de pesquisadores e a ferramenta mais adequada para análise de textos, a próxima etapa do experimento refere-se à criação do método responsável por identificar as similaridades e associações possíveis entre as sessões e os pesquisadores do CBIE 2017, permitindo analisar a coerência da escolha das sessões.

O método completo utilizado no experimento foi trabalhado em duas etapas. Na primeira foram identificadas as palavras-chave dos artigos científicos relacionadas as sessões assistidas pelos participantes do experimento no CBIE 2017. Para isso, foi extraído manualmente o resumo e o título dos artigos e executada a ferramenta de análise de texto do Google que gerou a lista de palavras-chave. Na segunda etapa, manualmente buscou-se os perfis dos participantes do experimento no ResearchGate e no Currículo Lattes. Os dados de trabalhos recentes dos pesquisadores foram extraídos e aplicados na ferramenta de análise de texto do Google, que por sua vez gerou a lista de palavras-chave. O método completo pode ser visualizado na Figura 3.

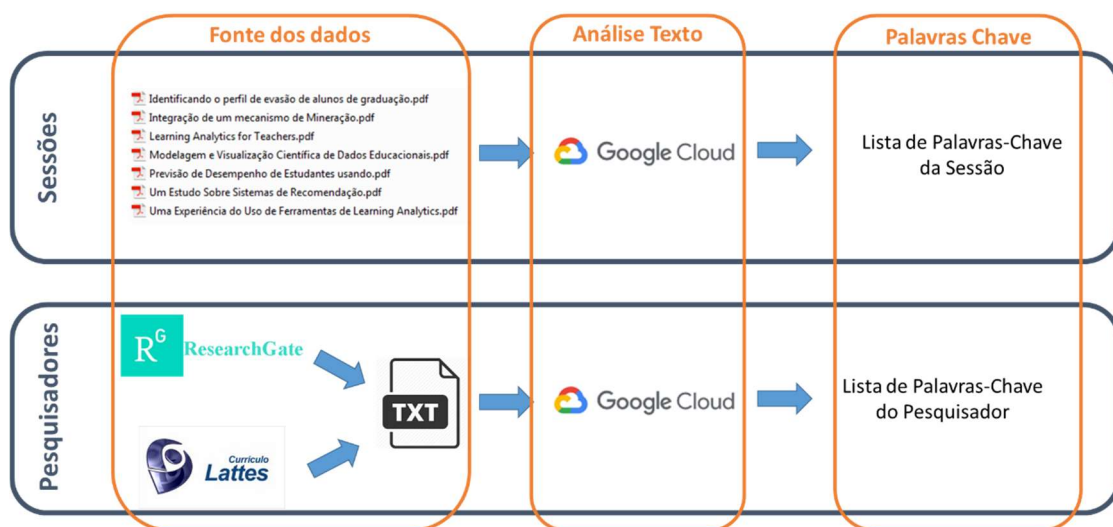


Figura 3: Método utilizado no experimento

A partir das palavras-chave identificadas entre as sessões e os participantes, foi possível fazer a comparação por similaridade entre as palavras-chave, permitindo assim analisar a coerência da escolha dos pesquisadores no CBIE 2017.

### 3.4.Limpeza dos Dados e Análises

Ao final do uso da aplicação Web criada para este experimento, uma pequena base de dados pode ser formada com dados dos pesquisadores e das sessões assistidas, o que permitiu o cumprimento de um dos objetivos propostos para o experimento.

Os dados extraídos com a aplicação Web precisaram passar por um tratamento, uma vez que foram detectados usuários duplicados, indicação de salas e sessões diferentes em horários similares para o mesmo usuário e indicação de sala duplicada para o mesmo usuário. Desta forma para não comprometer o experimento, foram selecionados apenas 3 usuários que tiveram boa qualidade nos dados, uma vez que não houve duplicidade, os horários de entrada nas salas e a permanência coincidiram com os horários das sessões. A partir dos 3 usuários selecionados foi executado o método proposto no experimento, em que foi feita a avaliação do perfil dos usuários e das sessões assistidas. A Tabela 5 apresenta cada um dos usuários avaliados, a fonte de dados do usuário, a quantidade de palavras-chave identificada para o usuário e a quantidade de sessões assistidas.

**Tabela 5:** Usuários participantes do experimento

<b>Usuário</b>	<b>Fonte dos Dados</b>	<b>Palavras-Chave do usuário</b>	<b>Sessões Assistidas</b>
Usuário 1	ResearchGate	21	7
Usuário 2	ResearchGate	18	7
Usuário 3	Currículo Lattes	17	3

O próximo passo envolveu comparar manualmente a similaridade de cada sessão com os usuários. Nesta etapa, ao analisar os resultados e os dados entre os usuários e sessões foram identificados problemas, os quais são tratados na seção de resultados. Tais problemas geraram um número baixo de similaridade, conforme apresentado na Tabela 6, em que há a relação de usuários e a quantidade de palavras-chave comuns entre o usuário e as sessões.

**Tabela 6:** Similaridade de usuários com sessões

	<b>Sessões</b>						
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
Usuário 1	5	3	4	2	4	4	6
Usuário 2	2	3	0	5	0	4	2
Usuário 3	6	8	7				

De acordo com o resultado da Tabela 6, se for considerado que as palavras-chave do usuário são termos fortes e de grande relevância, uma vez que foram extraídos do seu perfil, pode ser considerado que 1 ou 2 palavras fortes podem ser utilizadas por um sistema de recomendação para direcionar a melhor sessão a partir de uma lista de sessões. Como exemplo, para o usuário 1 da Tabela 6, as palavras Mineração de Dados, Aprendizado e Ontologias foram

importantes no seu perfil e apareceram na maioria das sessões que ele assistiu, demonstrando assim alguma coesão nas escolhas.

### **3.5. Conclusões e Resultados do Experimento**

Apesar de ser um experimento preliminar, para identificar possíveis correlações entre os participantes de um evento e a escolha das sessões para assistir, o experimento envolveu certa complexidade, pesquisas e análises, permitindo assim a formação de uma estratégia concisa para a continuidade do estudo, uma vez que o experimento não objetivou a criação do algoritmo de recomendação, mas sim os estudos associados para a formação do algoritmo proposto.

Ao longo do experimento foram selecionadas, a partir de análises pontuais já apresentadas, as fontes dos dados dos pesquisadores (ResearchGate e Currículo Lattes) e também a ferramenta do Google para a análise dos textos. Porém ao analisar os dados do experimento, percebeu-se que algumas recomendações de sessões não ocorreriam por tratamento especiais não terem sido considerados no método proposto, tais como:

- tratamento de regras específicas do evento: alguns eventos possuem regras específicas que precisam ser consideradas, como por exemplo público-alvo específico para determinadas sessões ou então sessões que exigem inscrição prévia e neste caso só faz sentido recomendar para aqueles que se inscreveram. Este tratamento não foi feito para o experimento do CBIE e mostra-se relevante uma vez que a maior parte dos eventos científicos possuem regras específicas para algumas sessões;
- recomendação obrigatória: em alguns casos a recomendação de sessões deve ser obrigatória independente das palavras-chave, isso ocorre para autores de trabalhos de determinadas sessões, palestrantes ou então em sessões com inscrição prévia;
- uso de palavras comuns sem relevância: frequentemente algumas palavras que não caracterizam interesses dos usuários ou assuntos das sessões foram destacadas como palavras-chave, como por exemplo: Estudo de Caso, Artigo, Experimento e Revisão Sistemática. Tais palavras são muito comuns nos

resumos e títulos dos trabalhos e isso faz com que ocorram similaridades apenas pela existência de estudos de caso de um usuário com uma sessão, o que não necessariamente caracteriza um interesse do usuário ou o assunto de uma sessão;

- ausência de pesos nas palavras-chave: apesar do algoritmo de análise de texto utilizado no método deste experimento retornar o peso de cada palavra-chave, esta informação não foi utilizada no método. Ao analisar os resultados percebeu-se que o peso tem grande relevância, uma vez que ele é gerado com base na frequência da palavra no texto analisado. Logo, por exemplo um pesquisador que publicou 15 trabalhos que possuem as palavras-chave educação e ontologia, possivelmente possui um interesse maior nestes temas se comparado com a palavra-chave “Aprendizado de máquina” identificado apenas em um trabalho que ele publicou.

Apesar da massa de dados reduzida utilizada no experimento, todos os problemas citados acima apontam para melhorias no método para encontrar a similaridade entre sessões e usuários, de forma que não foram identificados problemas com as fontes de dados ou com o algoritmo de análise de texto selecionado.

Por fim, ao analisar os 3 usuários do experimento e as sessões assistidas, o método proposto, mesmo com os problemas citados, teria recomendado 15 das 17 sessões assistidas, demonstrando assim que há coesão em termos de palavras-chave entre o perfil destes usuários com as sessões assistidas.

Ao comparar os objetivos iniciais do experimento com os resultados encontrados, pode-se afirmar que foi possível gerar a base de dados de usuários e sessões de um evento real, foi possível realizar comparações das sessões com o perfil de cada usuário identificando a coesão da escolha das sessões e por fim propor um método, ferramenta e fontes de dados para identificar similaridade entre as sessões e os usuários

## **4. PROPOSTA DO SISTEMA DE RECOMENDAÇÃO**

As descobertas identificadas no experimento realizado no CBIE 2017 foram fundamentais para a formação do método de recomendação proposto nesta dissertação. Este método tomou como base a solução para dificuldades e problemas ocorridos no experimento.

Assim, neste capítulo serão tratados o método de recomendação, a arquitetura da solução proposta, os artefatos gerados e o método de avaliação da solução.

### **4.1. Método de Recomendação**

Para atender à necessidade de recomendação de sessões para pesquisadores e demais participantes de eventos científicos, a base da recomendação utiliza dados expostos textualmente. Tais dados apresentam-se textualmente nos artigos referentes às sessões e também textualmente em páginas eletrônicas, redes sociais e currículos dos participantes. Desta forma, pela necessidade de análise de conteúdo, a técnica de recomendação aplicada neste trabalho é a recomendação com filtragem baseada em conteúdo.

Para realizar a análise do conteúdo e extrair palavras-chave tanto das sessões como também dos pesquisadores, foram replicadas as mesmas técnicas aplicadas no experimento realizado no CBIE. Tais técnicas tiveram bons resultados, porém precisaram de melhorias, como por exemplo o uso de uma ferramenta eletrônica de tradução, para garantir palavras-chave em um idioma padronizado, tratamento de regras específicas do evento científico, melhorias na assertividade das recomendações com o uso de pesos nas palavras-chave e não analisar dados antigos no perfil dos usuários. Os detalhes destas modificações no método serão apresentações nas seções a seguir.



Considerando a experiência do experimento no CBIE 2017 e agregando as melhorias supracitadas, foi proposto um método de recomendação para sessões em eventos científicos conforme diagrama na Figura 4. O diagrama foi separado em três grandes grupos destacados na figura e detalhados nas seções a seguir.

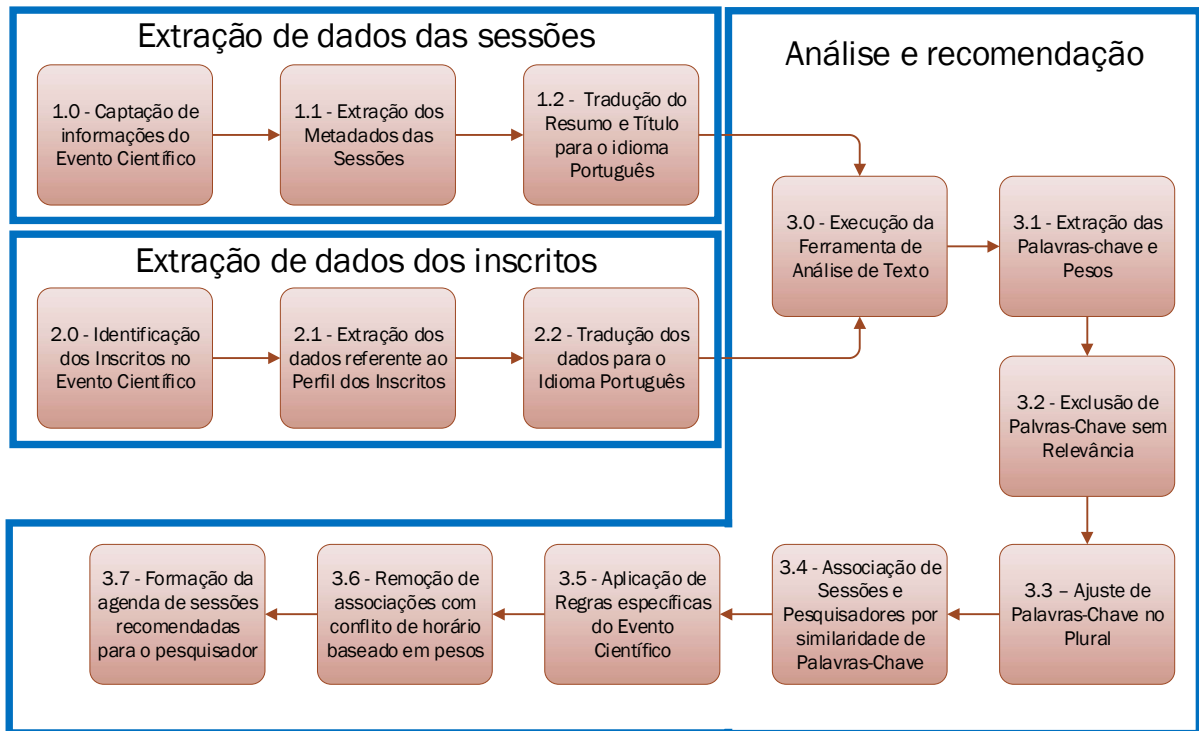


Figura 4: Diagrama do método de recomendação

#### 4.1.1. Extração de Dados das Sessões

Nesta etapa três atividades são realizadas, conforme Figura 4, são elas:

1.0 – Captação de informações do evento científico: a primeira etapa envolveu captar as informações sobre o evento científico, como por exemplo: agenda do evento, salas, sessões, característica de cada sessão (público alvo, tema, título), datas e horários. Em alguns eventos tais dados são disponibilizados no próprio site, já em outros é necessário solicitar para a comissão organizadora.

1.1 – Extração dos metadados<sup>7</sup> das sessões: obtidas as informações do evento, a próxima etapa visa a capturar em cada sessão do evento quais artigos serão apresentados, de forma a conseguir os metadados de cada artigo. Os principais dados a serem capturados são: título, resumo, autores e sessão associada.

1.2 – Tradução do resumo e do título para o idioma português: com o objetivo de padronização do idioma entre as sessões e os pesquisadores, optou-se em utilizar uma ferramenta do Google para tradução do título e do resumo para o idioma português.

#### **4.1.2. Extração de dados dos Inscritos**

Nesta etapa, conforme a Figura 4, três atividades são realizadas, a saber:

2.0 – Identificação dos inscritos no evento científico: para o funcionamento da solução de recomendação, faz-se necessário que seja fornecida a lista de inscritos no evento ou então que seja feita uma integração com o sistema de inscrições. Os principais dados necessários referentes aos inscritos são o nome, a categoria do inscrito, como por exemplo: professor, pesquisador, aluno de pós-graduação e aluno de graduação, e-mail de inscrição e se possível algum link da Web para o perfil do inscrito (Facebook, ResearchGate, Currículo lattes, etc.). Algumas informações complementares também são relevantes, como por exemplo as inscrições em sessões especiais que exigem inscrição prévia.

2.1 – Extração dos dados referentes ao perfil dos inscritos: nesta etapa é necessário identificar o perfil dos inscritos com base no conteúdo da Web. Conforme já identificado durante o experimento realizado no CBIE, as principais bases de informações para este fim são o ResearchGate e o Currículo Lattes. Desta forma, baseado nos dados fornecidos pelos organizadores do evento sobre cada inscrito, é necessário buscar publicações recentes, temas de interesse e demais dados que são disponibilizados nas bases supracitadas. Diferente do trabalho realizado no experimento, nesta etapa, como proposta final, optou-se em buscar apenas dados recentes limitados a 5 anos da data atual. Nos casos de inscritos com pouco (3 ou menos) publicações no período de 5 anos, optou-se em usar todo o histórico de publicação. Esta regra

---

<sup>7</sup> Metadados: habitualmente definidos simplesmente como dados descrevendo outros dados, ou então dado que descreve atributos de um recurso, caracteriza suas relações, apoia sua descoberta e uso efetivo, e existe em um ambiente eletrônico (CAMPOS, 2007).

de limite de tempo foi adotada para evitar a geração de palavras-chave referente a pesquisas muito antigas do pesquisador e com isso foram privilegiadas pesquisas recentes e temas atuais trabalhos pelo usuário.

2.2 - Tradução dos dados para o idioma português: assim como no item 1.2 da Figura 4 deste método, faz-se necessário utilizar a ferramenta do Google para tradução dos dados capturados nos sites ResearchGate e Currículo Lattes referente aos inscritos. Isso é necessário, visto que os usuários possuem publicações com título e resumo em inglês e tais dados são publicados muitas vezes em inglês.

### **4.1.3. Análise e Recomendação**

A etapa final do método compreende oito atividades, são elas:

3.0 – Execução da ferramenta de análise de texto: uma vez que os dados dos artigos de cada sessão e também dos inscritos foram obtidos e estão padronizados de forma textual e no idioma português, é possível executar a ferramenta de análise de texto, neste caso Google Natural Language, para com isso receber a lista de palavras-chave de cada artigo de cada sessão e também de cada inscrito.

3.1 – Extração das palavras-chave e pesos: o resultado do processamento do Google Natural Language é um vetor de palavras-chave e do peso de cada palavra-chave. O vetor precisa ser devidamente armazenado e associado à sessão e artigo ou então ao inscrito.

3.2 – Exclusão de palavras-chave sem relevância: nesta etapa, palavras-chave com pouca relevância precisam ser eliminadas de forma a evitar associações indevidas de termos comuns tanto para pesquisadores como para sessões, mas que não caracterizam a pesquisa ou interesse do pesquisador. As palavras-chave a serem eliminadas nesta etapa foram as mesmas identificadas no experimento do CBIE. As principais palavras-chave eliminadas no CBIE foram: Análise, Abordagem, Uso, Proposta, Caso, Estudo, Resultado, Literatura, Revisão, Tipo, Avaliação, Abordagem, Contexto, Artigo e Metodologia.

3.3 – Ajuste de palavras-chave no plural: uma vez eliminadas as palavras-chave sem relevância é necessário ajustar termos sem similaridade por estarem, por exemplo, flexionados de acordo com o gênero ou o número. Para isso uma das melhores técnicas que pode ser aplicada

é o algoritmo de lematização, previamente contextualizado nos capítulos anteriores. A execução deste algoritmo em todas as palavras-chave cria uma normalização e padronização permitindo seguir para a próxima etapa.

3.4 – Associação de sessões e pesquisadores por similaridade de termos: nesta etapa ocorrem duas importantes ações. A primeira envolve associar todos os casos os quais há alguma palavra-chave similar entre o perfil do usuário inscrito no evento com os artigos das sessões, de forma que um único termo similar seja suficiente para tornar o artigo e sua sessão elegível a ser recomendado. A segunda ação agrupa os artigos em que houve alguma similaridade em sessões do evento e neste momento os pesos das palavras-chave são somados. Por exemplo, em uma sessão com 3 artigos com um total de 15 palavras-chave, provavelmente o peso da somatória dessas palavras será maior que uma sessão com apenas 1 artigo de apenas 3 palavras-chave. Ainda assim, se nesta segunda sessão as 3 palavras-chave tiverem um peso muito alto no perfil do usuário, o peso final da sessão poderá ser alto.

3.5 – Aplicação de regras específicas do evento científico: alguns eventos científicos possuem sessões especiais ou específicas para um determinado público. Além disso algumas regras se aplicam a qualquer evento científico e precisam ser tratadas nesta etapa do método. Alguns exemplos de regras que precisam ser aplicadas são: (i) sempre associar usuários inscritos com possíveis sessões as quais seus artigos serão expostos, considerando assim que os autores sempre estarão presentes no seu próprio artigo e ao longo de toda a sessão; (ii) algumas sessões são pagas separadamente do restante do evento ou então precisam ser agendadas antecipadamente, logo entende-se que sempre devem ser recomendadas essas sessões para estes inscritos e (iii) algumas sessões podem possuir um público alvo especial (estudantes, graduandos, professores, etc.). Neste caso, se houve a informação de público alvo das sessões e a categoria dos inscritos um peso adicional deve ser considerado quando a similaridade ocorrer e o público alvo também for alcançado.

3.6 – Remoção de associações com conflito de horário baseado em pesos: uma vez que todas as associações de sessão com usuários foram feitas e possíveis regras especiais foram aplicadas, faz-se necessário remover possíveis conflitos de horário. Essa etapa é necessária pois se considera que possivelmente haverá no perfil do usuário ao menos uma palavra-chave comum em cada artigo de cada sessão, logo todas as sessões serão recomendadas para os usuários. Como várias sessões ocorrem simultaneamente é importante então recomendar para cada horário a sessão com o maior peso e maior similaridade de palavras-chave entre as sessões

e o usuário. Neste sentido são excluídas de cada usuário sessões com menor peso para cada dia e horário do evento.

3.7 – Formação da agenda de sessões recomendadas para o inscrito: nesta etapa a agenda final de sessões a serem recomendadas é formada e enviada para cada inscrito. Neste momento não poderão haver sessões sobrepostas e nenhum usuário poderá ficar sem recomendação.

## 4.2. Arquitetura e Artefatos

Para atender ao método de recomendação supracitado, criou-se a aplicação chamada de Conference Recommender. Dada as diversas atividades do método e o fato de cada etapa ser executada em momentos diferentes, a aplicação foi criada de forma modularizada, conforme apresentado na Figura 5.



Figura 5: Módulos da solução proposta

Para facilitar o entendimento e o escopo de cada módulo é apresentada a Figura 6, uma vez que cada módulo apresentado na Figura 5 refere-se a um conjunto de atividades representadas no diagrama do método de recomendação na Figura 4.

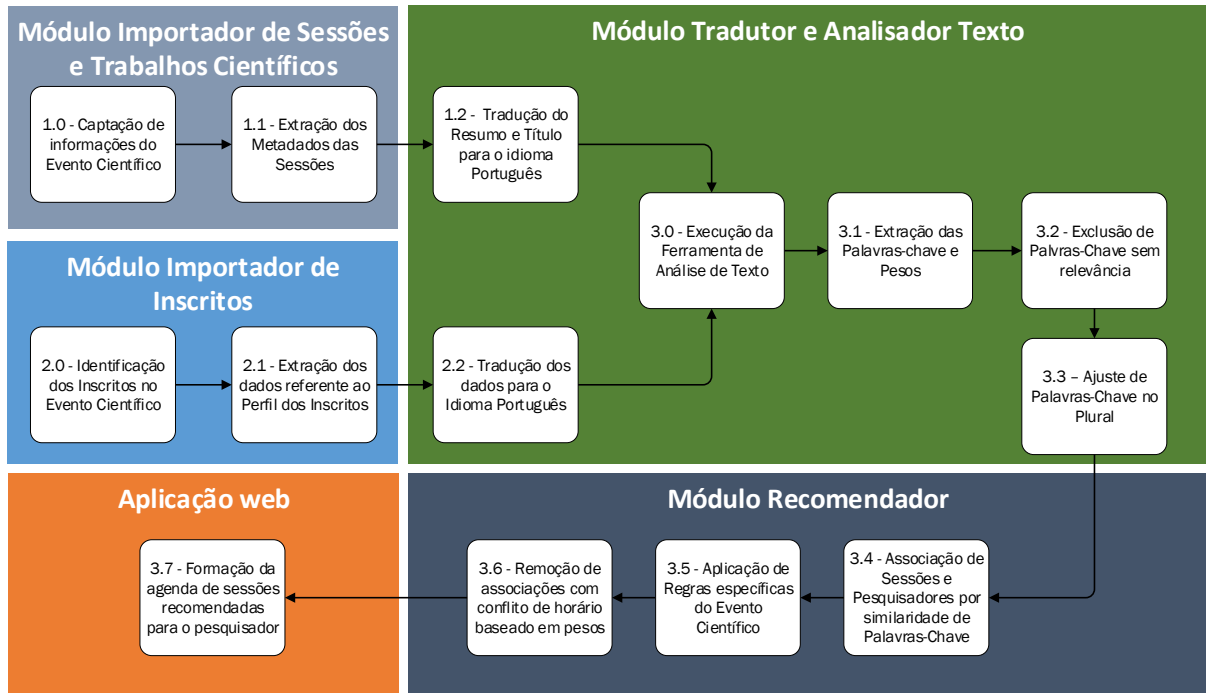


Figura 6: Módulos da solução e atividades de cada módulo

O módulo importador de sessões e trabalhos científicos é executado quando os dados do evento científico são disponibilizados. A função do módulo é importar os dados do formato Microsoft Excel para a base de dados da solução.

O módulo importador de inscritos é executado quando é disponibilizada a lista dos inscritos no evento e a lista de inscritos em sessões especiais. A função do módulo é importar os inscritos no formato Microsoft Excel para a base de dados do Conference Recommender e também importar os dados de forma textual do ResearchGate e do Currículo Lattes associando tais dados a cada inscrito previamente importado.

O módulo tradutor e analisador de texto é responsável por traduzir os textos dos trabalhos científicos e das sessões do evento científico, bem como o conteúdo extraído do ResearchGate e do Currículo Lattes de cada inscrito e transformar o texto traduzido em vetores de palavras-chave. Também é responsabilidade deste módulo efetuar as limpezas e normalizações dos dados. Este módulo é executado sempre que novas importações de sessões ou de inscritos são efetuadas no sistema.

O módulo recomendador é executado sempre que ocorrem mudanças na importação de dados de inscritos ou dados do evento. Este módulo é responsável por interpretar os dados mais

recentes disponibilizados pelo módulo tradutor e analisador de texto e efetuar a devida associação dos inscritos com as sessões do evento, além também de realizar a priorização das sessões a serem recomendadas.

Por fim, a aplicação Web é o módulo responsável por apresentar o resultado final da recomendação no formato de uma agenda do evento. Este módulo é executado sempre que um usuário acessa a aplicação Web para visualizar a agenda de eventos.

Na aplicação Web, houve uma grande preocupação em permitir o amplo uso por qualquer inscrito no evento científico, sem a necessidade de cadastro prévio e sem a necessidade do usuário informar seu perfil ao acessar a solução. Desta forma, bastando informar o e-mail, a aplicação Web carrega os dados mais atualizados referentes ao evento e apresenta a recomendação das sessões para o usuário.

Visto a possibilidade de alguns usuários inscreverem-se no evento após o envio da lista de inscritos pela comissão organizadora, foi complementado na aplicação Web a possibilidade de qualquer usuário informar seus dados e inscrever-se para utilizar a ferramenta. Neste caso, após a inscrição, o usuário precisará aguardar o processamento manual dos dados para identificar o perfil no ResearchGate ou no Currículo Lattes.

Devido ao ResearchGate e ao Currículo Lattes não disponibilizarem recursos nativos de extração de dados através de outros sistemas, a captura de tais dados foi considerada no escopo desta solução como uma etapa manual, ou seja, uma vez importados os inscritos, é necessário também importar informações textuais que permita mapear seus perfis. Na Figura 7 é apresentado o fluxo dos dados na solução criada, nela é possível identificar a importação dos dados manualmente coletados do ResearchGate e do Currículo Lattes.

Ainda na Figura 7 é possível perceber a camada de banco de dados, que é representada pelo sistema gerenciador de banco de dados relacional da Microsoft, o SQL Server versão 2018. Essa camada é usada como apoio para todos os demais módulos, de forma que os módulos importadores incluem os dados para o banco de dados, o módulo tradutor e analisador de texto, faz a leitura, as atualizações e as manipulações nos dados no banco de dados e a aplicação Web lê os dados e formata da melhor forma para entregar ao usuário final da solução.

Desta forma, todos os dados transitados no Conference Recommender passam necessariamente por um banco de dados e muitas operações envolvem rotinas de seleção e

manipulação de dados na linguagem SQL<sup>8</sup>, que no caso do Microsoft SQL Server 2018 é a linguagem T-SQL<sup>9</sup>.

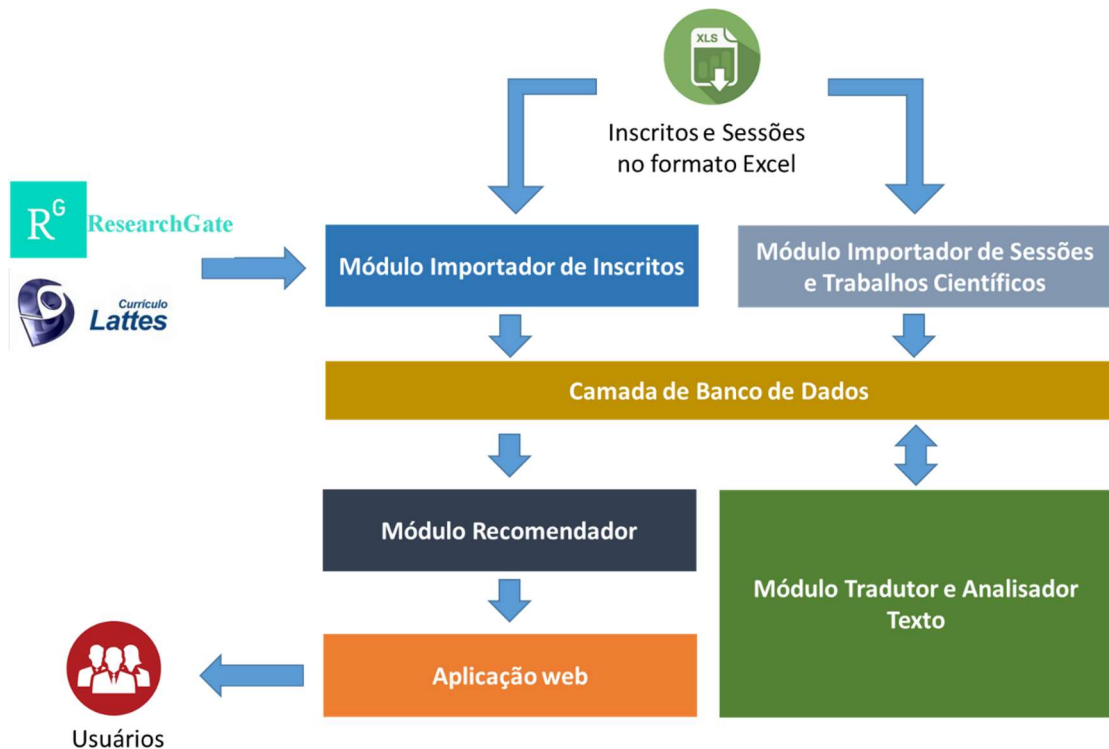


Figura 7: Fluxo dos dados entre os módulos da solução

Todos os módulos foram desenvolvidos na linguagem C# utilizando a versão 4.6 do Microsoft Framework .net. Na Figura 8 é possível visualizar a tela inicial do Conference Recommender e na Figura 9 a tela com a lista de recomendação de sessões para um usuário. Tais telas possuem a logomarca SBSI 2018, referenciando ao evento científico utilizado para validação da solução, assunto que será tratado na seção seguinte.

<sup>8</sup> SQL: *Structured Query Language* (SQL) é uma linguagem para desenvolvimento e manipulação de objetos e dados em sistemas gerenciadores de banco de dados (LIMA, 2005).

<sup>9</sup> T-SQL: A linguagem T-SQL (*Transact -SQL*) é uma extensão ao padrão SQL-92, sendo a linguagem utilizada por desenvolvedores na construção de aplicações que manipulam dados mantidos no Sistema Gerenciador de Banco de dados Microsoft SQL Server (SILVA, 2017).





Figura 8: Tela inicial da solução



Figura 9: Tela interna da solução após acesso do usuário

### 4.3. Validação da Solução

Uma vez que o método de recomendação foi concebido e desenvolvido baseado nas experiências obtidas no experimento realizado no CBIE 2017, a próxima etapa envolveu a validação da solução a partir de um estudo de caso em um evento real. Para tal, optou-se por escolher um evento científico que atendesse aos seguintes requisitos:

- Possuir sessões simultâneas que justificasse o uso de uma solução de recomendação para escolha das melhores sessões;
- Evento científico de ampla divulgação, reconhecido pela sua área de pesquisa e com grande número de participantes;
- Evento científico realizado em datas condizentes com o cronograma de construção desta pesquisa e dissertação;
- Comissão organizadora do evento aceitar o uso do Conference Recommender.

Desta forma, um dos eventos que atendeu todos os requisitos foi o Simpósio Brasileiro de Sistemas de Informação (SBSI), realizado no ano de 2018 na Universidade de Caxias do Sul no Rio Grande do Sul no período de 4 a 8 de junho.

O SBSI é voltado à apresentação de trabalhos científicos e à discussão de temas relevantes na área de Sistemas de Informação. A programação do evento é bastante diversificada, incluindo sessões técnicas, palestras, minicursos e workshops. Atualmente o SBSI é reconhecido como o evento mais importante na área de Sistemas de Informação no cenário nacional (SBSI 2017, 2017) e (Sociedade Brasileira de Computação, 2018).

O SBSI 2018 teve o período de inscrição aberto em fevereiro de 2018, contando com as seguintes categorias de inscritos:

- Estudantes de Ensino médio;
- Estudantes de Graduação;
- Estudantes de Pós-graduação;
- Professores de Ensino médio;
- Profissionais.

Além disso, o evento contou com uma série de minicursos que foram ofertados aos inscritos, mas que exigiu o pagamento de inscrição adicional de acordo com a categoria do inscrito.

O SBSI 2018 contou com sessões técnicas, palestras de convidados, minicursos, painéis e *workshops*. Na sua trilha principal uma ampla lista de tópicos de interesse foi disponibilizada para submissão de artigos, tais como: educação em sistemas de informação, inovação em sistemas de informação, integração de informações e de aplicações, gestão de pessoas em sistemas de informações e qualidade em sistemas de informações.

Além da trilha principal ocorreu também no mesmo evento o Fórum de Educação em Sistemas de Informação (FESI), o Workshop de Iniciação Científica em Sistemas de Informação (WICSI), o Workshop de Teses e Dissertações em Sistemas de Informação (WTDSI) e o Encontro de Inovação em SI (EISI).

Desta forma o primeiro passo para a realização do estudo de caso no SBSI 2018 foi a importação da lista de sessões e artigo do evento. Os seguintes dados foram disponibilizados no formato Microsoft Excel:

- Título: título do trabalho aceito.
- Tema: tema principal do trabalho aceito. Esta informação não foi importada para o Conference Recommender, visto que muitos trabalhos no arquivo em questão estavam sem este campo preenchido;
- Autores: nome completo dos autores do trabalho;
- E-mail: lista dos e-mails dos autores do trabalho;
- Resumo: texto do resumo do trabalho. Informação de grande importância para o Conference Recommender conseguir fazer a categorização da sessão baseado na análise deste texto.

As informações de sala, nome da sessão associada, dia e horário da apresentação de cada artigo não foi disponibilizado no arquivo. Com isso, foi necessário cruzar cada artigo de cada sessão com a programação do SBSI 2018 disponibilizado no site do evento, permitindo mapear os dados faltantes.

Em relação à lista dos potenciais inscritos no SBSI, foi disponibilizado um arquivo em formato Excel com 1 semana de antecedência ao início do evento. Os seguintes dados foram disponibilizados no arquivo:

- Nome: nome completo do inscrito;
- Categoria: categoria do inscrito que foi agrupado nas seguintes opções:
  - Estudante de Graduação;
  - Profissional;
  - Estudante de Pós-Graduação;
  - Estudando do Ensino médio;
- Afiliação: nome da universidade de vínculo principal com o inscrito;
- Status: informação sobre a situação da inscrição. Os inscritos com a situação cancelada não tiveram seus dados importados para o Conference Recommender;
- E-mail: e-mail do inscrito;

Após a recepção e a importação dos dados para o Conference Recommender os seguintes números foram obtidos: 238 inscritos, 105 artigos, 280 autores de artigos, 48 sessões, palestras ou minicursos e 6 salas onde ocorreram as sessões.

Para cada inscrito importado para o Conference Recommender foi feita a captura dos dados do Perfil no ResearchGate e no Currículo Lattes, nesta etapa da preparação do estudo de caso para avaliação da ferramenta notou-se duas dificuldades:

- Dos 238 inscritos, 116 eram estudantes de graduação ou ensino médio, para estes usuários não foi identificado perfil no ResearchGate e nem no Currículo Lattes, logo não foi possível levantar um texto do perfil do usuário para permitir a análise do texto e identificação de palavras-chave;
- A maior parte dos usuários que possuem perfil no ResearchGate possuem cadastro também no Currículo Lattes, logo ao extrair dados de ambas as fontes muitos dados se repetiram e geraram palavras-chave com peso maior pelo fato da duplicidade de dados. Por este motivo usuários que possuíam cadastro em ambas as fontes optou-se por utilizar o site com o perfil atualizado mais recentemente.

Para os minicursos do SBSI 2018 foi disponibilizada uma lista de inscritos por minicurso, este caso foi inserido manualmente na base de dados do Conference Recommender,

uma vez que não foi criado um módulo especial para importação deste tipo de conteúdo. Com isso, os usuários inscritos em minicurso tiveram a recomendação automática para tais sessões.

Os inscritos no SBSI 2018 puderam, alguns dias antes e durante o evento, acessar o Conference Recommender utilizando o mesmo e-mail da inscrição do evento. Após o acesso os usuários foram direcionados para a tela com as recomendações, conforme já apresentado na seção anterior na Figura 9, em que os usuários tiveram a possibilidade de avaliar cada recomendação em três categorias: ruim, regular ou boa.

O apoio da comissão organizadora do evento foi fundamental para a maior utilização da solução antes e durante o evento. Na Figura 10 pode ser vista a página oficial do SBSI 2018 com o atalho criado para o Conference Recommender, orientando os usuários interessados a utilizarem a solução de recomendação, além do atalho na página do evento, os inscritos receberam também o e-mail apresentado na Figura 11 convidando-os para utilizar a solução.

The image shows a screenshot of the SBSI 2018 website. At the top, there are logos for SBC (Secretaria Brasileira de Computação) and SBSI 2018, along with the event title 'XIV SIMPÓSIO BRASILEIRO DE SISTEMAS DE INFORMAÇÃO' and dates '4 A 8 DE JUNHO - CAMPUS-SEDE - CAXIAS DO SUL/RS'. Below the header is a navigation menu with links: INICIAL, SBSI, INSCRIÇÕES, CHAMADA DE TRABALHOS, PALESTRANTES, HACKATHON, PROGRAMAÇÃO (highlighted), LOCAL, and HOSPEDAGEM/PASSAGEM. The main content area is titled 'Programação' and features a section for 'Conference Recommender' with a brief description. Below that, it lists activities for 'Dia 4 de junho - Segunda-feira', including 'Das 9h às 10h - Credenciamento - UCS Teatro' and 'Das 10h às 11h30min'. A dropdown menu shows 'Encontro de Inovação em Sistemas de Informação (EISI) - Sala 1'.

Figura 10: Pagina de programação do SBSI 2018 com o link para o Conference Recommender



Figura 11: E-mail enviado para os inscritos pelo Conference Recommender

Para apoiar na avaliação da solução, um grupo fechado de doze pesquisadores do programa de pós-graduação em Informática da UNIRIO foi convidado a avaliar as recomendações do Conference Recommender, mesmo considerando que nenhum membro deste grupo tenha participado do SBSI 2018.

## 5. ANÁLISE DOS RESULTADOS DO ESTUDO DE CASO

Após a importação dos dados de inscritos, sessões e artigos aceitos no SBSI 2018, o próximo passo envolveu a execução do módulo tradutor e analisador de texto. Essa execução permitiu a extração das palavras-chave referentes ao perfil dos inscritos e das sessões do evento conforme números apresentados na Tabela 7.

**Tabela 7:** Números do módulo tradutor e analisador de texto

	<b>Artigos</b>	<b>Inscritos</b>
Total de Palavras-Chave	3852	2995
Total de Palavras-Chave distintas	1423	1351
Palavras-Chave por Artigo ou por Inscrito	36	13

Conforme exposto na apresentação do estudo de caso (seção 4), o estudo de validação do Conference Recommender ocorreu baseado em dois grupos de usuários:

- Inscritos: usuários que participaram do SBSI 2018 e tiveram seus dados fornecidos pela organização do evento ou então por livre cadastro dos usuários no Conference Recommender;
- Não inscritos: grupo de pesquisadores do Programa de Pós-Graduação em Informática da UNIRIO e que não estavam no SBSI 2018.

A primeira análise de resultados que pode ser observada do estudo de caso é referente ao tamanho da amostragem da avaliação. Na Tabela 8 podem ser analisados dados gerais e iniciais relacionados ao uso do Conference Recommender durante o SBSI 2018. Nesta tabela pode ser verificado que de 238 inscritos importados na solução, 97 tiveram o perfil analisado, isso devido à dificuldade supracitada de análise do perfil de usuários categorizados como “Graduação”. A lista de 97 usuários com perfil analisado gerou, conforme Tabela 8, 1702

recomendações de sessões, totalizando aproximadamente 18 sessões recomendadas para cada usuário para todos os dias do evento.

Ainda referente a Tabela 8, um total de 64 usuários acessaram e visualizaram as recomendações no Conference Recommender, representando 66% dos usuários com perfil analisado. Dos 64 usuários, 28 não apenas acessaram e visualizaram as recomendações, como também avaliaram cada uma das sessões recomendadas, totalizando assim 329 recomendações que foram avaliadas nos critérios bom, regular ou ruim.

**Tabela 8:** Dados Gerais do uso do Conference Recommender no SBSI 2018

	<b>Inscritos</b>	<b>Não Inscritos</b>	<b>Total</b>
Usuários que tiveram o perfil analisado	85	12	97
Total de Sessões Recomendadas	1494	208	1702
Usuários que acessaram o Conference Recommender	55	9	64
Usuários que avaliaram as recomendações	19	9	28
Total de sessões avaliadas	204	125	329

Desta forma na Figura 12 é possível visualizar o resultado da avaliação dos usuários referente as sessões recomendadas divididas em dois grupos de usuários, os inscritos e os não inscritos. É importante ressaltar que os números apresentados na Figura 12 se referem ao número de recomendações avaliadas e não ao número de usuários, uma vez que cada usuário pode avaliar individualmente cada sessão recomendada. É possível perceber na Figura 12 que os usuários não inscritos avaliaram relativamente melhor as recomendações do que os inscritos. Os principais motivos que podem justificar esse cenário são:

- Os usuários não inscritos no SBSI 2018 possivelmente não conheciam as restrições ou regras específicas do evento, que possibilitariam uma avaliação ruim, como por exemplo a recomendação de um minicurso (que exige pagamento adicional para assistir) ou então a recomendação de uma sessão com público alvo direcionado para professores, mas que o tema é de interesse mesmo para aqueles que não são professores. Ou seja, pode ter ocorrido uma avaliação positiva pelo desconhecimento de regras e critérios do evento que são normalmente conhecidos por aqueles que já se inscreveram e tem a participação confirmada;



- Os avaliadores possivelmente não conheciam a agenda completa do evento e não tinham a informação de outras sessões. Para alguns usuários o Conference Recommender pode ter sido o primeiro contato com as sessões do SBSI 2018.

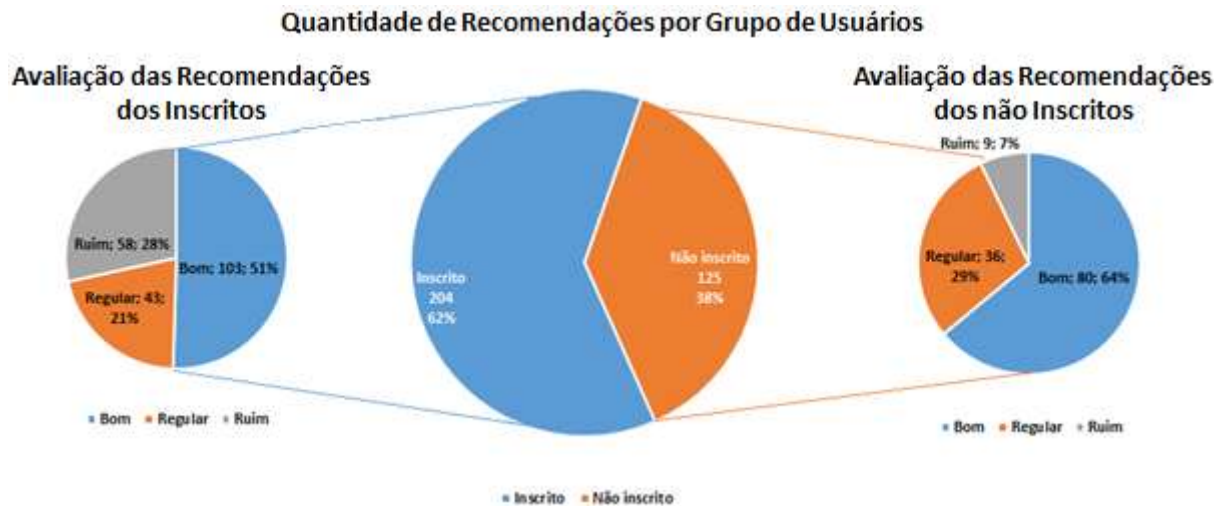


Figura 12: Quantidade de recomendações por grupo de usuários

Ainda visando a identificar as diferentes avaliações entre diferentes grupos de usuários, na Figura 13 é apresentado o grupo dos usuários categorizados como pós-graduação com 74% das avaliações e o grupo dos profissionais, com apenas 26% das avaliações identificadas. Ao comparar a resposta das avaliações entre os diferentes grupos, percebeu-se a maioria de avaliação positiva para ambos os grupos, porém a avaliação das recomendações com o critério “regular” foi maior para o grupo de pós-graduação, o que pode ser categorizado como alguma incerteza por parte dos usuários quanto ao conteúdo da sessão recomendada e por isso pode ter sido avaliado como regular.

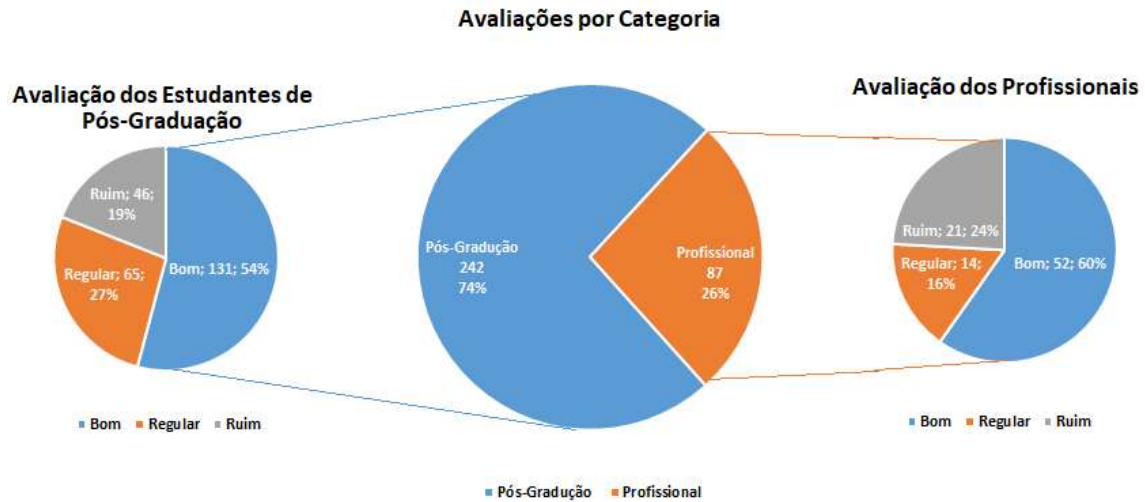


Figura 13: Quantidade de recomendações por categoria do usuário

Outra importante análise efetuada com os resultados do estudo de caso foi baseada nas palavras-chave encontradas no perfil de cada usuário. Na Tabela 9 é possível identificar que o maior percentual de avaliações positivas foi dos usuários em que o perfil continha mais de 41 palavras-chave.

Tabela 9: Avaliação das recomendações de acordo com o perfil dos usuários

Quantidade de palavras-chave no Perfil dos Usuários	Qtd. de usuários	Média de Similaridade (Sessão x Perfil)	Avaliação Dada		
			Bom	Regular	Ruim
19 ou menos	8	3,38	51%	27%	22%
De 20 a 40	16	4,54	54%	24%	22%
41 ou mais	4	6,76	72%	12%	16%

Ainda na Tabela 9 é apresentada a média de palavras-chave similares entre as sessões e o perfil do usuário (sem considerar o peso das palavras-chave). É possível visualizar que quanto maior a quantidade de palavras-chave no perfil do usuário, maior é o número de palavras-chave similares. O desvio padrão encontrado nas similaridades entre sessão e perfil foi de 2,89.

O peso das palavras-chave, tanto para os usuários como para as sessões, tiveram um papel importante para a recomendação, uma vez que a recomendação que totalizou o maior peso foi recomendada para o usuário final. Desta forma, na Figura 14 é apresentado o comparativo das avaliações ruins em comparação com o peso das palavras-chave do perfil do usuário e das sessões. Nesta figura é possível perceber que apenas 24% das avaliações ruins

estavam na faixa de maior peso (0,70 ou maior) enquanto 46% das avaliações ruins estavam nas recomendações com peso menor que 0,39. Apenas

Quantidade de Avaliações Ruins Por Média de Peso

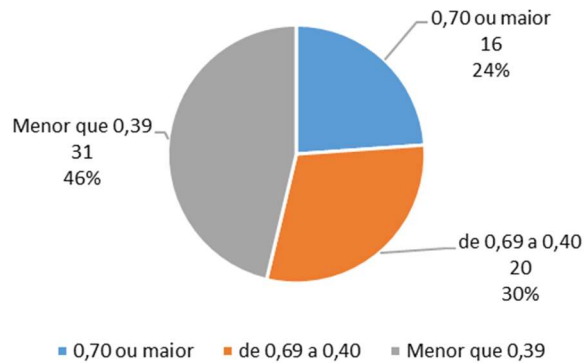


Figura 14: Quantidade de avaliações ruins por média de peso

Outra análise efetuada referente a avaliação dos usuários refere-se à quantidade de palavras-chave similares entre sessões e usuários. Na Tabela 10 é possível visualizar que quanto maior a quantidade de palavras-chave similares entre as sessões e os usuários, melhor foi a avaliação dos mesmos.

**Tabela 10:** Avaliação dos usuários baseados na quantidade de palavras-chave similares

Quantidade de palavras-chave similar entre as Sessões e os Usuários	Avaliação Dada		
	Bom	Regular	Ruim
4 ou menos	53%	22%	25%
De 5 a 9	60%	29%	11%
10 ou Mais	67%	13%	20%

Ainda para complementar a análise, foi comparado também se os usuários com perfis mais completos e detalhados, em termos de quantidade de caracteres no perfil, tiveram alguma diferença na avaliação das recomendações. A Tabela 11 apresenta esse resultado. Pode ser identificado que tanto os perfis pequenos como os grandes tiveram um percentual de avaliação similar em todos os critérios de avaliação.

**Tabela 11:** Relação entre o tamanho do perfil do usuário e a avaliação dada

Tamanho do Perfil dos usuários (em quantidade de caracteres)	Avaliação		
	Bom	Regular	Ruim
Mais de 1000	65,15%	18,18%	16,67%
De 500 a 999	46,67%	30,67%	22,67%
Até 499	61,95%	18,58%	19,47%

Por fim, o gráfico apresentado na Figura 15 demonstra a avaliação de todos os usuários para as recomendações, porém eliminando recomendações com pesos muito baixos. Neste caso é possível perceber que se assumir que pesos baixos geram avaliações ruins, conforme apresentado na Figura 14, então ao eliminar essas recomendações e suas avaliações tem-se um panorama sugerido para uma futura melhoria no método, a qual recomendações de baixo peso não são geradas. Com isso é possível perceber uma melhora no resultado das avaliações, se comparado com o número inicialmente apresentado, passando assim para 60% de avaliações boas referentes as recomendações apresentadas pelo método.

#### Quantidade de Avaliações Eliminando Baixo Peso

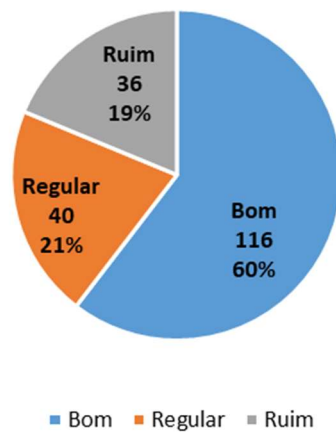


Figura 15: Quantidade de avaliações eliminando baixo peso

Foram identificados também autores de artigos de algumas sessões que avaliaram a recomendação da sua própria sessão como regular ou ruim. No total houve 57 recomendações de sessões para seus próprios autores, 15 foram avaliadas, as quais 3 foram avaliadas como regular ou ruim. Para estes casos duas possíveis conclusões podem ser consideradas: (i) o

usuário não percebeu que a sessão que estava avaliando no Conference Recommender era a sessão onde estaria seu próprio trabalho, ou (ii) apesar do usuário ser um dos autores do artigo, ele não possuía interesse em assistir seu próprio trabalho e por isso avaliou a recomendação como regular ou ruim.

Por fim, cabe destacar os minicursos. Apesar de ter sido importado para o Conference Recommender a lista dos inscritos nos minicursos, nenhum dos inscritos acessaram a solução e fizeram avaliações. Ainda assim, mesmo sabendo que os minicursos eram sessões pagas, optou-se por recomendar os minicursos, mesmo para usuários não inscritos, uma vez que poderia interessar aos usuários. Tal situação de recomendação ocorreu para os casos onde o minicurso teve maior similaridade de perfil com o usuário se comparado com todas as demais sessões no mesmo horário. No total foram 75 recomendações neste cenário e destas, 16 tiveram avaliação. O resultado foi de aproximadamente 56% das avaliações apontadas como ruim ou regular. Isso possivelmente pode ter ocorrido por usuários que perceberam que os minicursos eram pagos e por não terem se inscrito anteriormente não poderiam assistir.

## 6. CONSIDERAÇÕES FINAIS

Nesta seção são apresentadas as contribuições desta dissertação, bem como as conclusões e as recomendações de trabalhos ou pesquisas futuras que possam dar continuidade ou complementar o presente trabalho.

### 6.1. Contribuições

O número reduzido de trabalhos de pesquisas que aplicam as técnicas de sistemas de recomendação para eventos científicos, demonstra o potencial do tema, uma vez que novas descobertas e aprendizados podem ser extraídos em trabalhos nestas linhas de pesquisa.

Apesar do método proposto nesta dissertação utilizar técnicas simples e comuns em muitos trabalhos que propõem sistemas de recomendação, neste trabalho foi possível efetuar o mapeamento e validação de regras e características específicas de recomendação aplicáveis apenas no cenário dos eventos científicos. Neste cenário identificou-se, entre outras regras, a necessidade de tratar a temporalidade e os aspectos físicos dos eventos, uma vez que apenas uma recomendação por data, horário e sala é permitida, considerando que um usuário não pode estar em dois locais ao mesmo tempo.

Além disso, o estudo de redes sociais com foco em trabalhos e interesses de pesquisa científicos também se apresentou como um diferencial e importante contribuição, uma vez que fez-se necessário a identificação de dados que permitissem a criação de perfil e interesse de pesquisas de cada usuário da solução.

A identificação automática dos perfis, a partir da leitura de dados de redes sociais, sem a necessidade de intervenção ou colaboração dos usuários da solução, pode ser considerada uma

inovação na aplicação para eventos científicos, principalmente se comparado com os poucos trabalhos de pesquisa neste tema.

Por fim, a comparação dos recursos funcionais das ferramentas que costumam ser utilizadas no evento científico demonstrou a ausência de recursos de recomendação em todos os casos comparados, mostrando assim maior importância em pesquisas e estudos neste tema.

## 6.2. Conclusão

Esta dissertação visou demonstrar que um sistema de recomendação pode auxiliar pesquisadores e demais participantes de eventos científicos, a partir de recomendações que estejam relacionadas aos objetivos de pesquisa do usuário.

Para chegar a essa demonstração foi proposta nesta dissertação a criação de uma nova técnica a ser aplicada em um sistema de recomendação. Esta técnica considerou o tratamento das particularidades dos eventos científicos, bem como a extração de dados de redes sociais e sites que contém informações referentes a pesquisas científicas, permitindo assim gerar o perfil dos usuários.

Para validação da técnica de recomendação foi criada a solução de *software* chamada Conference Recommender, que teve como objetivo recomendar sessões de um evento científico real para usuários inscritos no evento. Ao final do uso da solução os usuários foram convidados a avaliar as recomendações, permitindo assim identificar a aceitação e coerência da recomendação por parte dos usuários. Para essa validação e avaliação foi aplicada a solução no SBSI realizado em 2018 na cidade de Caxias do Sul no estado do Rio Grande do Sul.

Com isso, após a aplicação do Conference Recommender no evento, foi possível extrair uma série de aprendizados com os resultados gerados. Inicialmente percebeu-se a avaliação positiva na maior parte das avaliações das recomendações feitas pela solução. Demonstrando assim assertividade da recomendação com o perfil dos usuários. Além disso percebeu-se a importância da riqueza de dados no perfil do usuário que recebe a recomendação, visto que as melhores avaliações foram de usuários em que o perfil continha uma maior quantidade de palavras-chave. Ainda nos resultados, foi possível visualizar que a técnica criada conseguiu bons resultados mesmo com perfis de usuário com pouco conteúdo, ou seja, a técnica conseguiu fazer a extração correta de palavras-chave, o cálculo do peso e a recomendação acertada mesmo

para usuários com pouco conteúdo nos sites ou redes sociais utilizadas para extração do perfil. Por fim, o peso das palavras-chave foi apresentado como um fator relevante nas avaliações, uma vez que as avaliações ruins, foram na maior parte das vezes, para recomendações com baixo peso.

De forma indireta foi possível perceber também que o Conference Recommender, não apenas conseguiu recomendar na maior parte das vezes sessões coerentes aos objetivos de pesquisa e interesses dos usuários, como também apresentou-se como uma ferramenta para auxiliar e orientar os usuários quanto a participação no evento, especialmente porque o SBSI 2018 contou com diversas sessões simultâneas e a escolha da melhor sessão por parte dos usuários poderia ser uma tarefa onerosa.

De forma específica, esta dissertação tratou dos seguintes objetivos descritos inicialmente:

- Identificação e uso de fontes de dados para o mapeamento automático do perfil dos pesquisadores, sem a necessidade de questionar o perfil para cada usuário;
- Criação e avaliação do método de recomendação e desenvolvimento do Conference Recommender para permitir recomendar sessões de eventos científicos para pesquisadores e demais inscritos nos eventos.

### **6.3. Trabalhos Futuros**

As poucas ou limitadas pesquisas que tratam de recomendações em eventos científicos ajudam a trazer relevância para este tema e com isso favorece traçar trabalhos futuros que complementam o presente trabalho e podem trazer novas descobertas. Alguns trabalhos recomendados são:

- Conversão do modelo de recomendação apresentado nesta dissertação para um modelo híbrido, utilizando não apenas conteúdo do perfil dos usuários, mas também dados da rede de amizades, contribuições e demais colaboração em redes sociais;
- Melhorar a usabilidade do *software* criado, apresentando para o usuário mais informações sobre a sessão, os trabalhos e o resumo dos trabalhos, permitindo



assim avaliar a recomendação com um conteúdo mais rico. Além disso a inclusão de uma avaliação geral referente a todas as recomendações e também a inclusão de justificativa para a avaliação dada contribuiria para uma avaliação qualitativa das recomendações.

- Integração ao modelo de recomendação deste trabalho o uso de dispositivos da Internet das Coisas, de forma que seja possível mapear o posicionamento dos usuários durante o evento, as salas na qual cada usuário encontra-se e também o tempo do pesquisador em cada sala ou sessão. A captura de tais dados permitiria mapear com precisão a participação dos usuários no evento e ajudaria na avaliação das sessões assistidas e nas desistências durante determinadas sessões. Ainda neste cenário seria possível efetuar recomendações de sessões baseadas na participação de outros usuários usando dados do perfil e também do posicionamento de outros usuários.
- Automatizar a extração dos dados das redes sociais dos usuários, facilitando assim a construção do perfil e eliminando os procedimentos manuais citados no método deste trabalho.

## REFERÊNCIAS BIBLIOGRÁFICAS

ADOMAVICIUS, G.; TUZHILIN, A. Toward the Next Generation of Recommender Systems: A Survey of the State-of-Art and Possible Extensions. **IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING**, v. 17, n. 6, 2005.

ARBOIT, A. E.; BUFREM, L. S. Produção de Trabalhos Científicos em Eventos Nacionais da Área de Ciência da Informação. **Transinformação**, v. 23, p. 207-217, Dezembro 2011.

AYLIEN. **Artificial Intelligence Applied to Textual Content**, 2018. Disponível em: <<https://aylien.com/>>. Acesso em: 23 Jun 2018.

BAR-ILAN, J. Web of Science with the Conference Proceedings Citation Indexes: the case of computer science. **Scientometrics**, Australia, v. 83, p. 809–824., 2010.

BASONI, H. G. **Modelo assistente para classificação de dados provenientes de redes sociais: um estudo de caso com dados do twitter**. M.Sc. Dissertation. Universidade Federal do Espírito Santo. [S.l.], p. 81. 2015.

BRITO, E. M. N. **Mineração de textos: detecção automática de sentimentos em comentários nas mídias sociais**. M.Sc dissertation. Fundação Mineira de Educação e Cultura. Belo Horizonte. 2017.

CAIRES, L. Again, the Role of Conference Papers in Computer Science and Informatics. **NOVA Laboratory for Computer Science and Informatics - Universidade Nova de Lisboa**, Lisboa - Portugal, p. 0, 2015.

CAMPOS, L. F. D. B. METADADOS DIGITAIS: revisão bibliográfica da evolução e tendências por meio de categorias funcionais. **Revista Eletrônica Bibliotecon**, Florianópolis, v. 23, p. 16 - 46, 2007.

CNPQ. Sobre a plataforma Lattes. **Plataforma lattes - CNPQ**, 2018. Disponível em: <<http://lattes.cnpq.br/>>. Acesso em: 16 Junho 2018.

CONFERENCE Navigator. **Conference Navigator**, 2018. Disponível em: <<http://halley.exp.sis.pitt.edu/cn3/portalindex.php>>. Acesso em: 30 Jun 2018.

CONNFA. **Open Source iOS & Android App for Conferences and Events**, 2018. Disponível em: <<http://connfa.com/>>. Acesso em: 30 Jun 2018.

CORRÊA, E. J.; VASCONCELOS, M.; SOUZA, M. S. D. L. **Iniciação à metodologia científica: participação em eventos e elaboração de textos científicos**. Belo Horizonte - Brasil: Coopmed, 2009.

- COSTA, H.; MACEDO, L. **Sistema de recomendação baseado em conhecimento**. Universidade de Coimbra. Coimbra - Portugal, p. 124. 2012.
- COUTINHO, R. M. **Organização de eventos**. Manaus - Brasil: e-Tec/MEC, 2010.
- DAMASCENO, M. D. M. Tecnologias Digitais e Comunicação: A colaboração científica na rede social virtual ResearchGate. **M.Sc. Dissertation. Universidade Federal do Amazonas**, Manaus, p. 94, 2016.
- EVENTMOBI. **Eventmobi**, 2018. Disponível em: <<https://www.eventmobi.com/>>. Acesso em: 30 Jun 2018.
- EVENTOOL. **Eventool**, 2018. Disponível em: <<https://www.eventool.com/pt>>. Acesso em: 30 Jun 2018.
- GALOÁ. **Galoá**, 2018. Disponível em: <<https://galoa.com.br/>>. Acesso em: 30 Jun 2018.
- GOOGLE Cloud. **Cloud Natural Language**, 2018. Disponível em: <<https://cloud.google.com/natural-language/>>. Acesso em: 23 jun 2018.
- GRAVA, A. P. **Sistema de recomendação de artigos científicos utilizando dados sociais**. USP. [S.l.], p. 94. 2016.
- JÚNIOR, S. M. D. S. **Recomendação de conteúdo baseada em informações semânticas extraídas de bases de conhecimento**. M.Sc. dissertation, USP. São Carlos. 2017.
- LACERDA, A. L. et al. A Importância dos Eventos Científicos na Formação Acadêmica: Estudantes de Biblioteconomia. **Revista ABC: Biblioteconomia em Santa Catarina**, Santa Catarina - Brasil, v. 13, n. 1, p. 130 - 144, Jan - Jun 2008.
- LIMA, A. G. D. A. Padrão SQL e sua Evolução. **CCUEC - Unicamp**, Campinas, 2005. 1-14.
- MCONFERENCE. **mConference**, 2018. Disponível em: <<https://www.mconference.com.br/>>. Acesso em: 30 Jun 2018.
- MEDEIROS, J. F. et al. **TagTheWeb**: Using Wikipedia Categories to Automatically Categorize Resources on the Web. European Semantic Web Conference 2018. Heraklion - Grecia: [s.n.]. 2018. p. 5.
- MORAIS, E. A. M.; AMBRÓSIO, A. P. L. **Mineração de Textos**. Universidade Federal de Goiás. [S.l.], p. 30. 2007.
- MOTTA, C. L. R. D. et al. Sistemas de recomendação. In: PIMENTEL, M.; FUKS, H. **Sistemas Colaborativos**. : Elsevier, 2012. Cap. 15, p. .
- NAVAUX, P. O. A. **Relatório de Avaliação - Ciência da Computação**. CAPES. [S.l.], p. 39. 2017.
- PARALLELDOTS , 2018. Disponível em: <<https://www.paralleldots.com/text-analysis-apis>>. Acesso em: 23 Jun 2018.
- PAZ, F. J.; CAZELLA, S. C. **Identificando o perigo de evasão de alunos de graduação através de mineração de dados educacionais**: um estudo de caso de uma universidade comunitária. VI Congresso Brasileiro de Informática na Educação. Recife - Brasil: [s.n.]. 2017.
- PEREIRA E SILVA, C. K. **Extração de características de perfil e de contexto em redes sociais para recomendação de recursos educacionais**. M.Sc dissertation. Universidade Federal de Juiz de Fora. Juiz de Fora, p. 140. 2015.

- POIRIER, D. et al. Towards text-based recommendations. **Adaptivity, Personalization and Fusion of Heterogeneous Information**, Paris, v. 30, p. 136-137, Abril 2010.
- RAHM, E. Comparing the scientific impact of conference and journal publications in computer science. **Information Services & Use**, Leipzig - Alemanha, v. 28, p. 127 - 128, 2008.
- RAJGOLI, I. U. Conference proceedings as a source of information in LIS research in India: A study based on citations. **Annals of Library and Information Studies**, Bangalore - India, v. 58, p. 346-354, Dezembro 2011.
- RECKER, J. **Scientific Research in Information System - A Beginner's Guide**. Brisbane - Australia: Springer, 2013.
- RESEARCHGATE. **Researchgate - Share and discovery research**, 2018. Disponível em: <<https://www.researchgate.net/>>. Acesso em: 11 Ago 2018.
- RESNICK, P.; VARIAN, H. R. Recommender systems. **Communications of the ACM**, v. 40, n. 3, p. 56, Março 1997.
- RICCI, F.; ROKACH, L.; SHAPIRA, B. **Recommender Systems Handbook**. 2nd. ed. New York: Springer, 2015.
- SAIF, H. et al. **On Stopwords, Filtering and Data Sparsity for Sentiment Analysis of Twitter**. Ninth International Conference on Language Resources and Evaluation. Reykjavik - Irlanda: [s.n.]. 2014. p. 810–817.
- SANCHEZ, A.; GRANADO, A.; ANTUNES, J. L. **Redes Sociais para Cientistas**. 1st. ed. Lisboa - Portugal: Nova Escola Doutoral, 2014.
- SANTOS, T. D. C. S. **Gestão de projetos: evolução do conhecimento nos eventos científicos na área de administração e engenharia da produção no Brasil e sua aplicação prática**. Universidade Nove de Julho - UNINOVE. São Paulo, p. 114. 2013.
- SBSI 2017. **Histórico de Edições do SBSI**, 2017. Disponível em: <<http://sbsi2017.dcc.ufla.br/sbsi/historico.html>>. Acesso em: 07 jul. 2018.
- SILVA, A. R. D. Gamificação e inteligência coletiva para promover a participação em sistemas de bate-papo para educação. **Dissertação de Mestrado - Universidade Federal do Estado do Rio de Janeiro**, Rio de Janeiro, 2016. 119f.
- SILVA, F. M.; SMIT, J. W. organização da informação em sistemas eletrônicos abertos de informação científica & tecnológica: Análise da Plataforma Lattes. **perspectivas em ciência da informação**, v. 14, n. 1, p. 22, Jan-Abr 2009.
- SILVA, J. V. B. C. D. Trabalhando com a linguagem T-SQL. **Devmedia**, 2017. Disponível em: <<https://www.devmedia.com.br/trabalhando-com-a-linguagem-t-sql/38126>>. Acesso em: 07 jul. 2018.
- SILVA, L. S. D. **MMRecommender: Arquitetura Aberta para Sistemas de Recomendação**. M.Sc. dissertation, Universidade Federal de Juiz de Fora., p. 90. 2017.
- SOCIEDADE Brasileira de Computação. **SBSI - Simpósio Brasileiro de Sistemas de Informação**, 2018. Disponível em: <<http://www.sbc.org.br/2-uncategorised/1817-sbsi-simpósio-brasileiro-de-sistemas-de-informacao>>. Acesso em: 07 jul. 2018.

UCLASSIFY. **About uClassify**, 2018. Disponível em: <<https://uclassify.com/about>>. Acesso em: 23 Jun 2018.

WAZLAWICK, R. S. **Metodologia de Pesquisa para Ciência da Computação**. 2nd. ed. Rio de Janeiro: Elsevier, 2014.

ZANELLA, L. C. **Manual de Organização de Eventos: Planejamento e Operacionalização**. São Paulo: Atlas, 2003.