

## COMPONENTE DE RECOMENDAÇÃO PARA SUPORTE A TOMADA DE DECISÃO EM PROGRAMAS DE PÓS-GRADUAÇÃO

THIAGO HERON ALBANO DE ÁVILA<sup>1</sup>;  
TIAGO THOMPSEN PRIMO<sup>2</sup>

<sup>1</sup> Universidade Federal de Pelotas – thiagoheronavila@gmail.com

<sup>2</sup> Universidade Federal de Pelotas – tiagoprimo@gmail.com

### 1. INTRODUÇÃO

Os sistemas de recomendação, vêm de sua definição, onde autores como RESNICK; VARIAN (1997), BURKE (2002), MAHMOOD; RICCI (2009) os definem como sendo ferramentas de software que propõem sugestões de itens que podem ser do interesse de um usuário ou grupo de usuários.

O presente trabalho está diretamente relacionada às necessidades do produto Cientum, da empresa parceira Indeorum, que permite a coordenadores de pós-graduação e gestores de instituições de ensino a gerenciar a produção científica de grupos de pesquisa e programas em geral. Essa ferramenta é utilizada de forma a tornar seus programas de pós estratégicos através da análise temporal, comparativa dos grupos e até mesmo individual dos pesquisadores.

Através do contato da ferramenta Cientum com os clientes foi percebida a necessidade da implementação de um mecanismo de recomendação de locais de publicação e recomendação de pesquisadores com interesses em comum para que gestores e coordenadores de pós-graduação possam sugerir aos pesquisadores da instituição, os periódicos, conferências ou parcerias que tragam benefícios tangíveis as suas instituições.

Com base nisso, o presente trabalho apresenta um sistema de recomendação para suprir as necessidades do produto. Para a realização deste, foi necessário desenvolver modos de recuperação de dados em questão, como dados de pesquisadores quanto locais de publicação, modelagem e estruturação destes para criar o sistema de recomendação explorando diversas técnicas de recomendação e utilizando uma poderosa ferramenta denominada *Elasticsearch*.

### 2. METODOLOGIA

Para o desenvolvimento do trabalho, foi de necessidade a criação de uma base e dados de currículos de pesquisadores e locais de publicação, como conferências e periódicos, entre outros. Esses dados foram obtidos através das Plataformas Lattes, *WikiCFP: A Wiki For Calls For Papers* e a *DBLP:Digital Bibliography & Library Project*.

Em relação à base de dados de pesquisadores, a empresa parceira do trabalho, Indeorum, disponibilizou cinquenta mil currículos de pesquisadores brasileiros. Além disso, para a fase inicial do desenvolvimento, foi necessário a

inclusão de pesquisadores mais próximos da Universidade Federal de Pelotas, com o objetivo de efetuar os primeiros testes do algoritmo para verificar se o escopo do resultado era compatível com a realidade de atuação desses pesquisadores mais próximos.

Logo, foi escolhido os docentes do Centro de Desenvolvimento Tecnológico(CDTEC), pela aproximação e conhecimento do grupo e a área em que atuam. Para a recuperação de dados destes, foi realizado um crawler na Plataforma Lattes através da API Scrapy, respeitando as normas impostas pela plataforma, para conseguir os currículos dos pesquisadores no formato XML.

Enquanto a recuperação de dados de locais de publicação, da mesma forma que foi necessário trabalhar com um escopo menor na parte dos currículos, foi preciso delimitar um escopo de locais de publicação para a fase inicial do sistema. Portanto, através de uma técnica de *crawling*, para a obtenção de dados da *WikiCFP* e *DBLP* que contém os dados de conferências internacionais, workshops, seminários, eventos, periódicos principalmente da área da ciência e outros campos da tecnologia.

Por fim, para a realização da recomendação em si, foi utilizado o *Elasticsearch*, uma poderosa ferramenta que permite a realização de buscas e análise de dados de grande volume de dados, permitindo indexar documentos e realizar buscas quase em tempo real. Devido a quantidade de documentos recuperados, essa ferramenta possibilita o processamento de uma análise ou busca na base inteira, e retornar um resultado em um tempo muito curto, praticamente em tempo real.

Logo, essa ferramenta foi utilizada para a indexação de todos os dados recuperados, tanto dos currículos de pesquisadores quanto os locais de publicação. Porém, a escolha do *Elasticsearch* para a implementação da recomendação, vai além da busca de documentos em tempo real, mas principalmente pela suas formas de realização de buscas de documentos.

Há ferramenta disponibiliza diversas formas de realizar buscas na base de dados, o que permite uma diversificação na forma como recuperar dados para o melhor aproveitamento do sistema de recomendação. A principal forma utilizada na pesquisa, é denominada pela documentação *More Like This*, que dado um documento de entrada é retornado documentos similares a este. Aliado a isso, há diversos parâmetros que definem o quão preciso podem ser esses documentos.

### 3. RESULTADOS E DISCUSSÃO

Como o trabalho inicial foi realizado para desenvolver um sistema de recomendação de pesquisadores com interesses em comum e locais de publicação para que os gestores e coordenadores de pós-graduação possam sugerir aos pesquisador da instituição, meios para publicação ou parcerias que tragam benefícios tanto para o pesquisador quanto para a instituição. Para a avaliação disso, foi tomado como base principalmente os pesquisadores mais

próximos da instituição e seus principais meios de publicação para a identificação de uma sugestão dentro do escopo destes.

Posteriormente, foi selecionado alguns professores do CDTEC e submetido aos resultados, para que estes dessem seus *feedbacks* para o aperfeiçoamento do sistema de recomendação.

Logo abaixo, na Tabela 1, pode-se analisar os resultados obtidos dado o perfil, ou seja, o currículo lattes de um docente da área da ciência da computação, Marcelo Schiavon Porto. Com seus resultados pode-se perceber a similaridade de áreas que foram obtidas.

**Tabela 1. Resultados de similaridades do docente da área da ciência da computação, Marcelo Schiavon Porto, que possui pontuação máxima de 4565.79 na busca da *More Like This*.**

Pontuação	Nome	Área de Atuação
2851.91	Luciano Volcan Agostini	Pesquisa em projeto de hardware para compressão de imagens e vídeos, como: AV1, VVC, HEVC e H.264/AVC
2501.5708	Bruno Zatt	Processamento de sinais, arquiteturas computacionais e sistemas embarcados desenvolvendo algoritmos, arquiteturas e circuitos integrados para processamento multimídia, especialmente compressão de vídeo 2D
1921.8013	Daniel Munari Palomino	Área de codificação de imagens e vídeos digitais explorando soluções nos diferentes níveis de hardware e software (como desenvolvimento de arquiteturas de hardware dedicadas, desenvolvimento de soluções no nível do sistema e no nível da aplicação). Padrões de codificação como MPEG, H.264/AVC e HEVC

#### 4. CONCLUSÕES

Conclui-se indica-se com as última versão desenvolvida, apresentado o resultado a Tabela 1, é possível acertar o escopo da área de atuação dos pesquisadores. Sendo assim, sugerindo pesquisadores colaboradores e locais de publicação para esses tanto por Filtragem Baseada em Conteúdo quanto por Filtragem Colaborativa.



Posteriormente, pretende-se agregar novos mecanismos para a sugestão de locais de publicação, consequentemente ampliando a base de dados para outras áreas, o mesmo para sugestões de pesquisadores. A partir disso, pode-se explorar novos métodos de recomendação, como o Método Híbrido entre as duas abordagens citadas, além de outros tipos de algoritmos, principalmente modelos de redes neurais para identificar e aproximar pesos melhores, *boosting*, para as áreas estudadas.

## 5. REFERÊNCIAS BIBLIOGRÁFICAS

RESNICK, P.; VARIAN, H. R. Recommender systems. *Commun. ACM*, [S.I.], v.40, n.3, p.56–58, 1997. BURKE, R. Hybrid Recommender Systems: survey and experiments. In: **USER MODELING AND USER-ADAPTED INTERACTION**, Hingham, MA, USA. Anais. . . Kluwer Academic Publishers, 2002. v.12, n.4, p.331–370.

MAHMOOD, T.; RICCI, F. Improving recommender systems with adaptive conversational strategies. In: **ACM CONFERENCE ON HYPERTEXT AND HYPERMEDIA**, 20., New York, NY, USA. Proceedings. . . ACM, 2009. p.73–82. (HT '09)

PRIMO, Tiago Thompsen. **Método de representação de conhecimento baseado em Ontologias para apoiar Sistemas de Recomendação Educacionais**. 2013. Tese parcial de Doutorado em Ciência da Computação. Programa de Pós-Graduação em Computação. Universidade Federal de Pelotas