

Tenebris: Um Sistema de Recomendação de Obras Acadêmicas baseado em Filtragem Híbrida

Tenebris: A Recommendation System for Academic Papers based in Hybrid Filtering

Fábio S. Silva¹, Leticia C. Passos¹, Lucas Gabriel C. Evangelista¹ e Thiago M. Rocha²

¹Escola Superior de Tecnologia (EST) – Universidade do Estado do Amazonas(UEA)
Av. Djalma Batista – 3578, Flores – Manaus, AM – Brasil.
Caixa postal: 69055-035

²Samsung Ocean Center – Samsung
Av. Djalma Batista – 1200, Flores – Manaus, AM – Brasil.
Caixa postal: 69055-035

{fssilva, lgce.eng, lcps.eng}@uea.edu.br, tm.uea.2014@gmail.com

Abstract. *With the diversity of information, content, products and services currently available, the search for items that are of interest to the user became a chore. This article presents a proposal for a Recommender System of academic items, based on a hybrid recommendation approach, which aims to facilitate access to academic items and support the research of the teachers and students.*

Resumo. *Com a diversidade de informações, conteúdos, produtos e serviços disponíveis atualmente, a busca por itens que sejam de interesse do usuário tornou-se uma tarefa árdua. Este artigo apresenta a proposta de um Sistema de Recomendação de obras acadêmicas baseado em uma abordagem de recomendação híbrida que visa facilitar o acesso a itens acadêmicos e apoiar a pesquisa de professores e alunos.*

1. Introdução

A evolução tecnológica tem contribuído para que cada vez mais os usuários tenham acesso a um grande volume de informações, conteúdos, produtos, e serviços. Consequentemente, torna-se uma tarefa árdua para um usuário escolher uma opção entre as várias disponíveis. De acordo com [Baeza-Yates and Ribeiro-Neto 1999], este problema é conhecido na literatura como Sobrecarga de Informação e tem motivado o desenvolvimento de ferramentas para facilitar o acesso pelos usuários.

No ambiente acadêmico também ocorre o problema da Sobrecarga de Informação, pois uma diversidade de obras acadêmicas é produzida por instituições de ensino e pesquisa. Neste contexto, na Universidade do Estado do Amazonas (UEA), é produzido um grande volume de obras acadêmicas. No entanto, seus professores, alunos e pesquisadores ainda não dispõem de uma ferramenta que possa facilitar o acesso as obras acadêmicas produzidas pela instituição. Vale ressaltar que ferramentas de apoio à pesquisa, normalmente, são disponibilizadas em bibliotecas digitais e portais de universidades para que os acadêmicos possam realizar buscas por obras acadêmicas, contudo essas ferramentas raramente empregam abordagens de acesso personalizado e com recomendação.

Para contextualizar, a Universidade de São Paulo (USP), por exemplo, possui o portal de teses [USP 2017]. Já a Universidade Federal de Minas Gerais (UFMG) possui a Biblioteca Digital de Teses e Dissertações [UFMG 2017]. O Portal de Periódicos CAPES/MEC [MEC 2017] disponibiliza ferramentas de busca de obras acadêmicas e periódicos em repositórios diversos, entretanto nenhum dos ambientes de pesquisa oferece recomendação de obras acadêmicas. Adicionalmente, a UEA possui um sistema Web que permite acesso ao acervo da biblioteca [UEA 2017], livros e revistas, contudo as obras acadêmicas produzidas na própria universidade como TCCs e artigos não estão disponibilizadas.

Desta forma, diante do problema supracitado relacionado à Sobrecarga de Informação no âmbito acadêmico, especialmente na UEA, foi realizado um projeto de pesquisa que visou desenvolver uma ferramenta de acesso a trabalhos acadêmicos. Tal ferramenta foi implementada por meio um Sistema de Recomendação de obras acadêmicas que emprega técnicas de filtragem de informação. Este artigo está organizado da seguinte forma: Na Seção 2 são descritos os principais conceitos e técnicas empregadas no trabalho; A Seção 3 apresenta uma análise dos trabalhos relacionados; A Seção 4 detalha o desenvolvimento da solução proposta; Na Seção 5 são descritos os testes e resultados obtidos; e na Seção 6 é descrito a conclusão do artigo.

2. Fundamentação Teórica

Nesta seção são descritos os principais conceitos e técnicas que serviram de base para o desenvolvimento deste trabalho.

2.1. Sistemas de Recomendação

Segundo [Adomavicius and Tuzhilin 2005], os Sistemas de Recomendação podem ajudar o usuário a lidar com o problema da sobrecarga de informação, sendo este um tema que tem despertado grande atenção da comunidade acadêmica que se reúne em importantes eventos relacionado ao tema como [Geva and Trotman 2014] e [Kobsa and Zhou 2014]. Além disso, grandes empresas como Amazon (em [Linden et al. 2003]), Netflix (em [Zhou et al. 2008]) e Google (em [Das et al. 2007]), estão investindo no emprego de sistemas de recomendação para obter diferencial competitivo em seus serviços.

Sistemas de recomendação tradicionais filtram itens relevantes de acordo com os interesses de um usuário ou de um grupo de usuários levando em consideração a similaridade entre os perfis de usuários ou entre os conteúdos e perfil do usuário. Tais sistemas estão sendo utilizados em vários domínios de aplicações, onde os itens recomendados podem ser programas de TV, livros, músicas, filmes, notícias, vídeos, anúncios, serviços, links patrocinados, páginas da web, produtos de uma loja virtual, etc [Jiangshan et al. 2010], [Silva 2011].

2.2. Técnicas de Filtragem de Informação

Os Sistemas de Recomendação são baseados em técnicas de filtragem de informação oriundas da área de Recuperação de Informação (RI). Segundo [Silva 2011], as principais técnicas são: Filtragem Baseada em Conteúdo (FBC), Filtragem Colaborativa (FC) ou Híbrida (FH).

2.2.1. Filtragem Baseada em Conteúdo (FBC)

Esta é uma técnica de filtragem de informação que consiste em comparar o conteúdo das descrições de itens com o conteúdo das preferências especificadas em um perfil do usuário, com o objetivo de selecionar os itens relevantes para o usuário baseado na similaridade entre eles. Em uma variação da técnica é possível filtrar itens para o usuário baseado na similaridade entre os itens pois os usuários tendem a interessar-se por itens semelhantes aos que foram anteriormente demonstrado interesse. As preferências podem ser fornecidas explicitamente pelo próprio usuário ou obtidas implicitamente por meio do monitoramento das ações do usuário enquanto o mesmo interage com o sistema.

A FBC pode ser implementada por meio da técnica de Indexação de Frequência dos Termos (IDF) e, também, por meio do Modelo do Espaço Vetorial (MEV) descritos em [Baeza-Yates and Ribeiro-Neto 1999], onde os itens e o perfil do usuário são representados por meio de vetores com uma dimensão para cada termo que ocorre no documento. Adicionalmente, uma vez definidas as dimensões, é necessário atribuir valores para eles. Assim, cada componente do vetor corresponde a frequência que um termo (TF) ocorre em um documento ou em um perfil de usuário e, deste modo, é possível empregar o cálculo do *Cosseno* entre o ângulo dos vetores para obter o grau de similaridade. Os itens cujos conteúdos forem mais similares ao conteúdo de um determinado perfil de usuário são considerados relevantes [Adomavicius and Tuzhilin 2005].

2.2.2. Filtragem Colaborativa (FC)

De acordo com [Chen 2005], a FC trabalha combinando as opiniões de pessoas que expressaram inclinações similares no passado para prever o que poder ser de interesse agora para uma determinada pessoa. Mais minuciosamente, conforme [Zanette 2008], estes tipos de sistemas tentam prever itens aos usuários de acordo com itens avaliados por outros usuários. Basicamente, o sistema utiliza informações da base de dados para encontrar os pares (pessoas com interesses similares) para assim gerar a recomendação de acordo com os pares. Desta forma, os usuários passam a colaborar com os sistemas para serem beneficiados com recomendações a partir das informações que os próprios usuários forneceram ao ambiente.

Segundo [Silva 2011], os Sistemas de Recomendação baseados em FC se diferenciam dos demais, justamente porque não exigem um processamento das características (descrições) dos conteúdos. Em contraste, esta técnica procura explorar o intercâmbio de experiências entre usuários e usa as opiniões para prever os interesses de outros. Para obter os *vizinhos mais próximos*, ou seja calcular a similaridade entre os perfis de usuários podem ser empregadas métricas como *Pearson* e *Cosseno*. Uma descrição detalhada da FC pode ser encontrada em [Zhou et al. 2008].

2.2.3. Filtragem Híbrida (FH)

A técnica de FH procura combinar os pontos fortes das técnicas de FBC e FC visando superar as limitações de cada técnica de filtragem e apresentar melhores resultados que possam atender às expectativas dos usuários. Segundo [Adomavicius and Tuzhilin 2005],

na literatura é possível encontrar algumas propostas para combinação das técnicas FBC e FC como segue:

- combinação linear dos resultados da FBC e FC que consiste em usar as técnicas de forma separada de forma que os resultados possam ou não serem combinados;
- incorporar na FC algumas características da FBC e vice-versa; e
- construção de um modelo unificado que incorpora as características de ambas técnicas.

Para o desenvolvimento da recomendação híbrida empregada no sistema descrito neste artigo foi empregada a combinação linear do resultados obtidos pela FC e FBC.

3. Trabalhos Relacionados

O objetivo desta seção consiste em apresentar uma análise de alguns trabalhos encontrados na literatura que abordam o tema Sistema de Recomendação de itens acadêmicos. Os trabalhos analisados empregam diferentes abordagens para recomendar itens acadêmicos e se correlacionam com o Sistema de Recomendação apresentado neste artigo, buscando assim posicionar este trabalho em relação aos demais.

[Cazella et al. 2008] apresentam um Sistema de Recomendação de artigos acadêmicos baseado em FC para ambientes móveis. Sistema que recomenda somente artigos científicos e emprega uma única técnica de filtragem. Já em [Cazella et al. 2009], apresentaram um Sistema de Recomendação de artigos acadêmicos que utiliza a técnica de FBC, também para ambientes móveis. Da mesma forma que o primeiro trabalho, o sistema recomenda somente artigos científicos e emprega uma única técnica de filtragem. Do mesmo modo, em [Barcellos et al. 2007] foi construído um sistema de recomendação de artigos acadêmicos integrado a um sistema de busca que utiliza a técnica de FBC. Neste trabalho, além de ser aplicada uma única técnica de filtragem, a recomendação é fornecida somente quando o usuário realiza uma pesquisa por uma determinada obra.

Em [Tantanasiriwong 2011] algoritmos de clusterização foram comparados e empregados em um Sistema de Recomendação de artigos extraídos do repositório *SCI corpus*. A recomendação consiste em agrupar os artigos que possuem características similares. Em contraste à solução proposta neste artigo, neste sistema não foi empregada nenhuma técnica de filtragem de informação no processo de recomendação, porém o mesmo foi baseado somente em clusterização.

Em [Casagrande et al. 2015] foi feita a aplicação de um sistema de recomendação em um repositório digital de obras literárias. Neste trabalho foi empregada a técnica de FH incorporando características da FC na FBC, com o agrupamento de perfis de usuário, porém sem a construção explícita de um perfil de usuário o qual não o permite determinar ou modificar as suas preferências e também sem a combinação linear dos resultados obtidos. Da mesma forma, em [Costa et al. 2011] foi construído um Sistema de Recomendação personalizada de artigos científicos para Bibliotecas Digitais. Nesse trabalho também foi aplicada a técnica de FH, porém o sistema não permite a especificação e modificação do perfil pelo usuário.

O Sistema de Recomendação proposto neste artigo se difere com relação aos demais trabalhos apresentados por empregar uma abordagem de recomendação híbrida que combina as duas técnicas de filtragem de informação. Como também, o perfil de usuário

pode ser modificado pelo mesmo para personalizar a sua recomendação. Além disso, o sistema recomenda não somente artigos acadêmicos, mas também TCCs, dissertações e teses, bem como possibilita que os próprios usuários cadastrem suas obras públicas.

4. Tenebris: Solução Proposta

Esta seção apresenta uma visão geral sobre o Sistema de Recomendação Híbrida, denominado Tenebris, e descreve a sua arquitetura de sistema, aspectos de implementação dos componentes e o sistema implementado.

4.1. Descrição do Sistema

A solução proposta consiste em um sistema de recomendação de obras acadêmicas que emprega a técnica de Filtragem Híbrida para apoio à pesquisa científica. Desta forma, o sistema proverá acesso personalizado aos trabalhos acadêmicos gerados na instituição, viabilizando que os alunos e professores possam receber recomendações de trabalhos acadêmicos de acordo com os seus interesses e necessidades de informação.

Para isso, é necessário que os usuários se cadastrem no sistema e especifiquem de forma explícita as suas áreas de pesquisas, autores, instituições e temas de interesse. Além disso, é possível também que as obras acadêmicas sejam cadastradas pelos próprios usuários, o que viabiliza um ambiente colaborativo de pesquisa científica. As obras acadêmicas apresentadas no sistema podem ser avaliadas de forma explícita pelos usuários.

O Sistema de Recomendação foi desenvolvido para plataforma Web. Deste modo, pode ser acessado por meio de navegadores tradicionais e/ou navegadores de dispositivos móveis, pois as suas páginas foram desenvolvidas de forma responsiva.

4.2. Arquitetura do Sistema

A Figura 1 apresenta a arquitetura do Sistema de Recomendação. A arquitetura é distribuída entre cliente e servidor, e baseada em dois módulos *front-end* e *back-end*. O módulo *front-end* possui os componentes responsáveis pela interação com o usuário e monitoramento da interação. Já no módulo *back-end* ficam localizados os componentes de suporte a filtragem de informação, gerenciamento de perfis de usuários e seus trabalhos acadêmicos.

A comunicação entre os dois módulos é baseada em [Richardson and Ruby 2007], e no formato JSON (*JavaScript Object Notation*) [Valet-Harpe 2013], que é uma formatação leve de troca de dados entre sistemas.

4.3. Componentes do Front-end

Estes componentes são responsáveis por viabilizar as interações do usuário com o sistema e a comunicação entre o *front-end* e *back-end*. Por meio de tais componentes o sistema gera e apresenta a recomendação híbrida de obras para o usuário, e também obtém de forma explícita o perfil do usuário e as suas avaliações de obras.

4.3.1. Componente Gerenciador de Recomendação Híbrida

Este componente é responsável por executar a FH que consiste em gerar uma lista final de recomendação de obras por meio da combinação linear das lista de obras filtradas

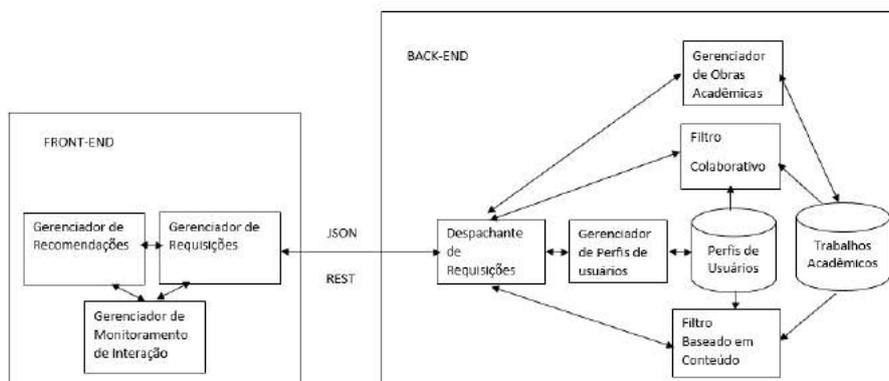


Figura 1. A arquitetura do Sistema de Recomendação.

pela ação dos componentes de FC e FBC. Um problema detectado durante a criação da lista de obras recomendadas foi a existência da recomendação de obras iguais, pois cada componente de filtragem trabalha de forma independente na geração das lista de obras. Com isso, foi preciso adotar uma solução que removesse essas repetições, foi então, que adotou-se um algoritmo denominado “misto” para solucionar o problema.

Baseado no trabalho de [Burke 2002] utilizamos o mesmo princípio para construir um algoritmo denominado como misto, onde neste trabalho, aplicamos uma técnica conhecida da matemática, a teoria dos conjuntos. Ele funciona da seguinte maneira: inicialmente temos um conjunto vazio de obras recomendadas, então quando o Gerenciador de Recomendação recebe a lista única de obras, automaticamente, o componente percorre toda a lista de obras verificando obra por obra adicionando ao conjunto apenas as obras que não pertencem ao conjunto. Portanto, é mantido no conjunto apenas aquelas obras cujo o ID ainda não pertence ao conjunto, eliminando assim as obras iguais. A Figura 2 ilustra o seu funcionamento.

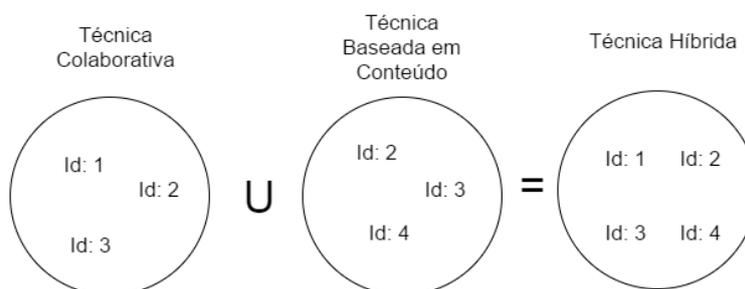


Figura 2. Funcionamento do Algoritmo Misto.

4.3.2. Componente Gerenciador de Monitoramento de Interação

Tudo que compete a interação entre o usuário e sistema, acontece por meio deste componente. Ele é responsável por capturar as ações do usuário, monitorar as obras acessadas e

avaliadas. Além disso, gerencia o fluxo de navegação entre telas e operações permitindo solicitações de recursos ao sistema, tais como: recomendações de obras, perfil do usuário, interesses, e detalhes de obras.

4.3.3. Componente Gerenciador de Requisições

Todo o fluxo de comunicação de dados entre *front-end* e *back-end* acontece por meio do componente gerenciador de requisição responsável por estabelecer a conexão com o servidor (*back-end*) e encaminhar qualquer requisição de recurso. O *back-end* é conectado ao *front-end* por meio de uma conexão de dados via Internet baseada em REST e no formato JSON para troca de dados.

4.4. Componentes do Back-end

Estes componentes são responsáveis pelo gerenciamento dos dados dos perfis do usuário e das obras acadêmicas. Além disso, também são os responsáveis pela implementação das técnicas de filtragem de informação.

4.4.1. Componente Despachante de Requisição

Este é o primeiro componente de integração do *back-end* com o *front-end*. É o componente responsável por receber as requisições oriundas do *front-end* e acionar o componente apropriado para atender a requisição recebida. Além disso, também controla o processo de recomendação, pois determina a sequência de execução dos componentes de filtragem de informação.

4.4.2. Componente Gerenciador de Perfis de Usuários

Este é o componente responsável pelo acesso, armazenamento, atualização e exclusão de dados do perfil do usuário armazenados em uma tabela específica do banco de dados. O perfil do usuário é cadastrado de forma explícita pelo usuário por meio do *front-end* do sistema de recomendação que envia os dados no formato JSON para o componente Gerenciador de Perfis de Usuários.

4.4.3. Componente Gerenciador de Obras Acadêmicas

Este é o componente responsável pelo acesso, armazenamento, atualização e exclusão dos dados de obras acadêmicas armazenados em uma tabela específica do banco de dados. A obra acadêmica é cadastrada de forma explícita pelo autor do trabalho por meio do *front-end* do sistema de recomendação que envia os dados no formato JSON para o componente Gerenciador de Obras Acadêmicas.

4.4.4. Componente Filtro Colaborativo

Este componente é responsável por implementar a FC. Por meio do ID do usuário, o componente acessa uma tabela específica do banco de dados, que contém avaliações do próprio usuário e do restante da comunidade de usuários. É importante frisar que o componente de FC utiliza um limitante numérico para a quantidade de usuários, já pré-estabelecido, o que chamamos de vizinhos.

Quanto mais avaliações de usuários diferentes, maiores serão os vizinhos. Com isso em mente, o componente cria a rede de colaborações, verificando as avaliações dos usuários para com os artigos no banco de dados e então, assemelhando artigos para serem mostrados para o usuário; artigos esses já avaliados pelos usuários com perfis similares ao usuário que se deseja fazer recomendação, mas ainda assim não avaliados pelo mesmo. Este componente utiliza do coeficiente de *Pearson* para o cálculo de semelhança entre perfis do usuário.

A quantidade de artigos que será recomendada ao usuário é limitada por um parâmetro numérico passado anteriormente pelo componente de requisição, com isso, a lista final para artigos da FC é organizada de forma a retornar essa quantidade fixa, dando prioridade para artigos com maior grau de similaridade com o perfil do usuário que se deseja fazer a recomendação. Para implementação do filtro foi empregado *API Mahout* [Owen et al. 2012] que fornece um conjunto de classes de suporte ao Modelo Espaço Vetorial e cálculo de *Pearson* supracitado.

4.4.5. Componente Filtro Baseado em Conteúdo

Este componente é o responsável por executar a FBC. Por meio do ID do usuário, o componente acessa um perfil de usuário armazenado em uma tabela específica do banco de dados para obter os interesses do usuário que serão, posteriormente, comparados as descrições de obras acadêmicas também obtidas de uma tabela específica do banco de dados. Esta comparação visa filtrar um conjunto de obras que possuem as descrições mais similares aos interesses do usuário.

Para isso, o componente emprega a Técnica de Indexação de Frequência dos Termos, o Modelo Vetorial e o cálculo do cosseno do ângulo entre os vetores descritos em [Baeza-Yates and Ribeiro-Neto 1999] para obter o valor da similaridade entre o vetor que representa o perfil do usuário e seus interesses e os vetores que representam as obras acadêmicas localizados dentro de espaço vetorial. Cada componente do vetor corresponde a frequência que um termo ocorre em uma obra ou em um perfil de usuário. Após executar o cálculo de similaridade, as obras que forem mais similares para um determinado perfil de usuário são consideradas relevantes.

O resultado da filtragem consiste em uma lista de obras ordenadas de forma decrescente de acordo com o valor de similaridade. Para implementação do filtro foi empregado *API Lucene* [McCandless et al. 2012] que fornece um conjunto de classes de suporte às técnicas TF-IDF e ao Modelo Espaço Vetorial.

4.5. Sistema Implementado

O sistema implementado, denominado Tenebris, foi desenvolvido para a plataforma Web com o principal objetivo de gerar recomendações de obras acadêmicas para o usuário com base no seu perfil de interesses e pelas avaliações por ele realizadas através do sistema.

A Figura 3(a) apresenta a tela inicial do modelo onde são apresentados os campos login e senha para acesso ao sistema. Para o caso de novos usuários é necessária a criação de um perfil para o mesmo através do botão “Não sou cadastrado”, no qual logo em seguida é solicitada a sua instituição de ensino. A Figura 3(b) mostra a tela de obras acadêmicas recomendadas para certo usuário baseado em seu perfil. Nota-se que, caso o usuário esteja realizando seu primeiro acesso, a lista de obras recomendadas é baseada na instituição de ensino que foi selecionada no cadastro de seu perfil.

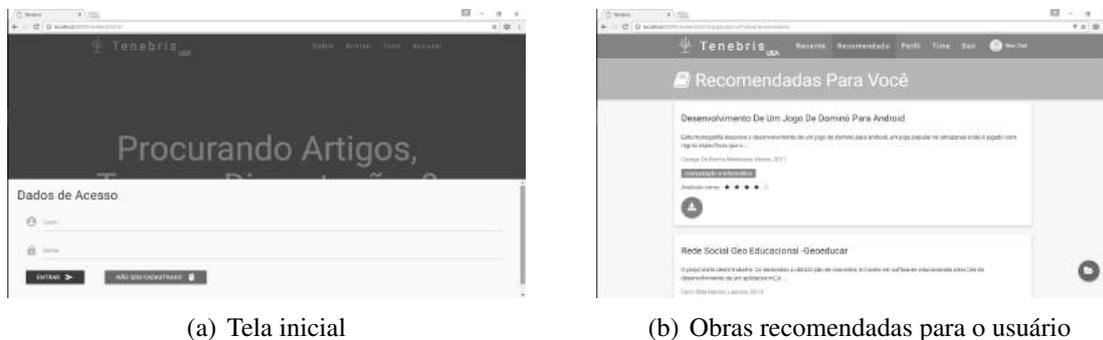


Figura 3. Telas do sistema Tenebris

A Figura 4(a) apresenta a tela com informações de uma obra e a avaliação do mesmo. Observa-se que, além do título, é exposto o resumo, o autor, a área de estudo, o tipo de obra acadêmica, a instituição de ensino, a avaliação média de outros usuários e a opção de avaliação pelo usuário para aquela obra, como também a opção de baixá-la em formato PDF. A fim de atualizar os interesses do usuário, há dentro do sistema a aba “Perfil”, onde é possível acrescentar ou excluir *tags* de interesse, instituições de ensino e autores, como também há possibilidade de atualizar uma área de estudo como apresentado na Figura 4(b). Desta forma, a lista de obras recomendadas torna-se mais especializada ao gosto do usuário.



Figura 4. Telas do sistema Tenebris

5. Testes e resultados

Nesta seção são descritos os testes que foram realizados com o sistema de recomendação e os resultados que foram obtidos.

5.1. Dados

Para que seja possível analisar a qualidade das recomendações e, conseqüentemente, o desempenho do sistema, foram estipulados dados a serem analisados. São eles: quantidade de obras recomendadas pelo sistema, quantidade de obras avaliadas (por cada usuário) e quantidade de obras recomendadas pelo sistema e que o usuário julga relevantes.

5.2. Métricas

A eficiência de um sistema de recomendação pode ser mensurada por meio das métricas: precisão, revocação e medida F como constam em [Jiangshan et al. 2010]. É possível encontrar inúmeros trabalhos que empregam tais técnicas para análise e estudo de desempenho de um sistema de recomendação.

- **Precisão:** segundo [Jiangshan et al. 2010], conforme o seu nome indica, mensura o quão precisa foi a resposta, em uma proporção de itens recomendados que seja realmente relevantes. Obedecendo, logo, a fórmula:

$$Pr = \frac{(Artigos_{relevantes}) \cap (Artigos_{retornados})}{(Artigos_{relevantes})} \quad (1)$$

- **Revocação:** segundo [Jiangshan et al. 2010], a revocação é usada para mensurar a fração de todos os itens relevantes que o sistema conseguiu recomendar dentre o número total de itens relevantes disponíveis na coleção. Segue a fórmula.

$$Rev = \frac{(Artigos_{relevantes}) \cap (Artigos_{retornados})}{(Artigos_{retornados})} \quad (2)$$

- **Medida-F:** Pode-se afirmar que a medida F é precisamente uma métrica harmônica. De acordo com [Jiangshan et al. 2010], quanto mais próximo de 1 (um) for o valor da medida F mais eficiente será o sistema.

$$M_F = \frac{2 * (Pr * Rev)}{Pr + Rev} \quad (3)$$

5.3. Descrição dos Testes

Com o objetivo de avaliar a qualidade da recomendação gerada pelo sistema, testes foram realizados durante uma semana com um grupo de alunos do curso de Engenharia da Computação (EC) e Sistemas da Informação (SI) da Escola Superior de Tecnologia (EST)

da UEA. O período de testes ocorreu durante o período de férias, onde os alunos que desenvolvem projetos frequentam a EST com um horário mais flexível e costumam passar o dia na universidade.

A EST disponibilizou um computador para este trabalho, no laboratório de projetos, onde os alunos convidados para os testes também frequentam. Na semana anterior à realização da semana de testes, foram instalados nesta mesma máquina todos os programas necessários e foi carregado o banco de dados com TCCs provenientes da própria EST, onde foram recolhidos em sites de professores da instituição que fazem o *upload* desses arquivos que são, em suma, de seus orientandos. Os usuários que testaram o sistema foram no total 8 alunos, sendo 5 provenientes de EC e 3 de SI. Como os usuários não tinham aulas e passavam o dia na universidade, eles podiam utilizar o sistema no horário que lhes fosse melhor. Cada usuário utilizou o Tenebris 1 vez por dia, durante 5 dias, no decorrer de uma semana. Em seu primeiro contato com o sistema o usuário era explicado sobre suas funcionalidades, funcionamento e aplicabilidade. Prosseguindo, cada usuário criou seu próprio login com senha e indicou a sua faculdade de origem e, após isso, alimentou o banco de dados ao relacionar breves *tags* relacionadas aos seus interesses com relação ao conteúdo dos artigos. A tela dessa aplicação consta na Figura 4(b).

Após a especificação de suas *tags*, o usuário avaliava 5 obras, ou seja, usava de seus próprios critérios para classificar uma obra entre 1 até 5 estrelas, sendo 1 estrela um artigo que não é de seu interesse e 5 estrelas um artigo que é de seu interesse e muito relevante. Para essa avaliação, o usuário estava livre para navegar e escolher os artigos que quisesse e utilizasse dos critérios que bem preferisse para avaliar – é importante ressaltar que embora os membros da equipe estivessem por perto para sanar possíveis dúvidas que pudessem surgir sobre o sistema, os mesmos buscaram por não influenciar a avaliação de nenhum usuário para com obra nenhuma para que se tivesse uma avaliação clara e coesa, que representasse o interesse do usuário da melhor possível. Também é necessário esclarecer o número estabelecido de 5 avaliações por dia para cada usuário, isto ocorreu porque se viu necessário estabelecer um número fixo de obras a serem avaliadas porque o sistema constava de somente 50 TCCs e 5 trabalhos variados. Se os usuários avaliassem várias obras, poderia ocorrer um cenário indesejado onde a FC não colabora, pois não recomenda uma obra para o usuário sendo esta uma obra que ele mesmo já informou ao sistema explicitamente que é de seu interesse ou não, sendo então desnecessária a sua recomendação. Após a avaliação das 5 obras, o usuário solicitava uma recomendação ao sistema que assim o recomendava. Para cada obra retornada pelo sistema era necessário que o usuário respondesse se aquela obra recomendada seria de seu interesse, ou seja, se ele a leria e se concordava que o sistema havia lhe retornado uma obra relevante. Esses dados são importantes para mensurar a qualidade das recomendações naquele dia e em diante, o que possibilita o estudo da curva de aprendizado do sistema que será alvo de discussão na próxima subseção.

Assim, cada usuário utilizou o Tenebris uma vez por dia, no horário que lhe fosse desejado, sempre acompanhado de um membro da equipe desenvolvidora e sem tempo pré-estabelecido, pois cada usuário tinha seu tempo e sua forma de escolher para avaliar obras. Inicialmente, a ideia de se estipular uma determinada quantidade de tempo foi concebida, mas descartada após ser constatado que era desnecessária uma vez que todos os usuários poderiam utilizar o sistema no mesmo dia, em qualquer horário.

Ao final da semana, os usuários foram convidados a preencher um formulário para *feedback*, elogios e sugestões de melhoria. Alguns resultados obtidos foram, por exemplo, elogios à interface, usabilidade, aplicabilidade, também como sugestões de melhorias, como alguns *bugs* e, principalmente, a incapacidade de, na página de acervo principal, não ser possível abrir uma nova guia para abrir um artigo. Na época em questão, era necessário que o usuário abrisse o artigo na mesma página e então voltasse para a página de acervo para que continuasse e buscasse outro artigo. Este detalhe foi o principal alvo de sugestões de melhorias, e posteriormente, nas versões atuais, foi corrigido.

5.4. Resultados Obtidos

Os gráficos a seguir apresentam o desempenho do sistema para cada usuário durante o período de uma semana de testes. Os gráficos são divididos em dois eixos, sendo o eixo das abscissas (x) correspondente aos dias da semana, e o eixo das coordenadas (y) correspondente ao desempenho do sistema mensurado pelas métricas de avaliação. Além disso, cada linha traçada representa o desempenho do sistema de acordo com o teste realizado por um usuário. Torna-se importante destacar que quanto maior a porcentagem, melhor o desempenho do sistema.

Nos dois primeiros dias de testes o desempenho do sistema foi regular. A razão para isto é que o sistema estava gerando recomendações somente por meio da FBC. A FC necessita de uma determinada quantidade de dados para funcionar, e o sistema ainda não dispunha do número suficiente de avaliações de obras para executar a FC. Esta limitação da FC é conhecida na literatura como o problema do *cold-start* [Chen 2005]. Vale ressaltar que apesar dos resultados regulares, os testes iniciais foram importantes, porque viabilizaram o treinamento do sistema por meio das avaliações de obras recomendadas, o que gerou os dados necessários para execução da FC.

Por meio dos gráficos de resultados (Figuras 5 e 6) é possível constatar que o sistema começa a apresentar melhor desempenho para a maioria dos usuários a partir dos dias seguintes, pois é quando a FC começa a atuar. Foram necessárias cerca de 25 avaliações no total, vindas dos 8 usuários, para que a FC iniciasse a geração de listas de obras, uma vez que quanto mais avaliações de obras o sistema recebe, melhor será a recomendação. Com isso, ambas as técnicas de filtragem começaram a gerar listas de obras que foram combinadas de forma linear viabilizando a FH do sistema. Os resultados confirmam que o desempenho do sistema é muito sensível ao número de avaliações de obras que podem estar associadas com um determinado usuário. Adicionalmente, observa-se a melhora dos valores das métricas de revocação, precisão e, principalmente, da constante harmônica da medida-F (Figura 7) cujos os seus valores estão mais próximo de 1, o que demonstra uma evolução no desempenho do sistema.

Não existe uma forma de *benchmark* para análise dos resultados, sendo a conclusão, de que um sistema de recomendação é eficiente, muito conectado ao ambiente, ferramentas, testes e técnicas aplicadas. O Tenebris, em uma semana, foi capaz de reconhecer os gostos de vários usuários, utilizando duas técnicas de recuperação de informação e combinando-as, atingindo uma Medida-F acima de 85% para todos os usuários que o utilizaram, e inclusive beirar 90% para alguns. Levando em consideração este cenário, é favorável a ideia de que o sistema de recomendação Tenebris é um sistema de recomendação eficiente, o que demonstra que, ao final da semana de testes, os usuários que o utilizaram ficaram satisfeitos com as recomendações feitas.

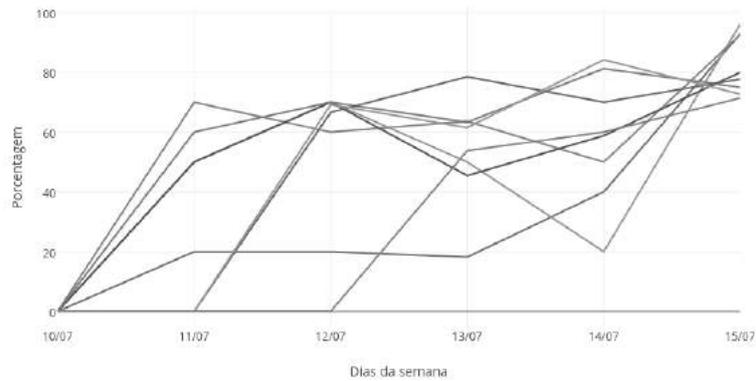


Figura 5. Precisão

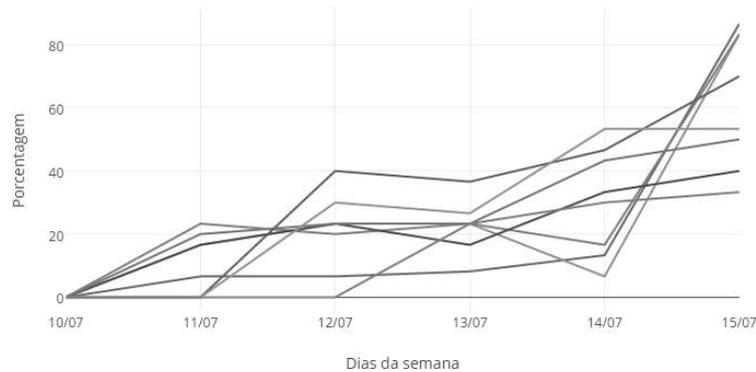


Figura 6. Revocação

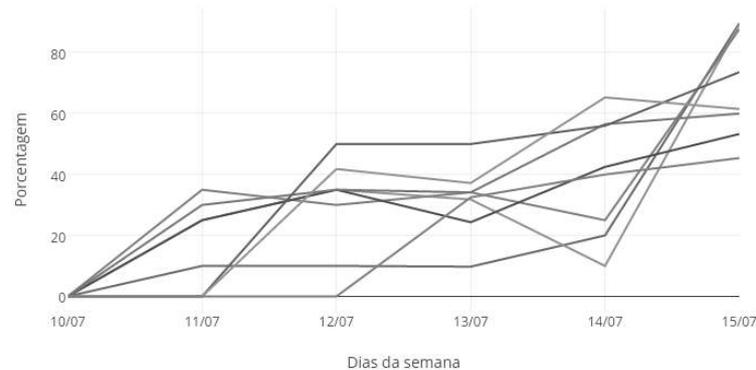


Figura 7. Medida-F

6. Conclusões e Trabalhos Futuros

Este artigo apresentou a proposta de um Sistema de Recomendação de obras acadêmicas baseado em uma abordagem de Recomendação Híbrida por meio da combinação linear dos resultados obtidos das Filtragens Colaborativa e Baseada em Conteúdo. O artigo apresentou a arquitetura do sistema, seus principais componentes e também os testes que foram realizados para avaliar a qualidade das recomendações geradas. Por meio dos resultados obtidos dos testes, foi possível constatar que o desempenho do sistema, de acordo com os valores das métricas de desempenho, foi satisfatório.

Torna-se importante ressaltar que a ferramenta construída é um projeto aberto. Logo, pode ser implementado e utilizado não somente pela UEA, mas por qualquer instituto de ensino superior do Brasil. Como trabalhos futuros, está o estudo e desenvolvimento de um novo componente de suporte à Filtragem Baseada em Conhecimento que utiliza técnicas de Aprendizagem Máquina e, desta forma, amplia o sistema visando melhorar cada vez mais a qualidade das recomendações.

Agradecimentos

Os autores agradecem a Luiz E. F. Bentes e a Elloá B. Guedes pelas discussões técnicas ao longo deste trabalho. O autor L. G. C. Evangelista agradece ao apoio financeiro provido pela Fundação de Amparo à Pesquisa do Amazonas (FAPEAM) pela bolsa de iniciação científica concedida pelos PAIC/UEA 2015-2016 e 2016-2017.

Referências

- Adomavicius, G. and Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. In Lin, X. and Chen, L., editors, *IEEE Transactions on Knowledge and Data Engineering*, pages 734–749. IEEE.
- Baeza-Yates, R. A. and Ribeiro-Neto, B. (1999). *Modern Information Retrieval*. PRENTICE HALL, 1th edition.
- Barcellos, C. D., Musa, D. L., Brandão, A. L., and Warpechowski, M. (2007). Sistema de recomendação acadêmico para apoio a aprendizagem. *RENOTE*, 5(2).
- Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370.
- Casagrande, M. F. R., Kozima, G., and Willrich, R. (2015). A recommendation technique based on metadata for digital repositories oriented to learning. *Brazilian Journal of Computers in Education*, 23(2):69–70.
- Cazella, S. C. and das Chagas, I. C., Barbosa, V., J. L., and Reategui, E. B. (2008). Um modelo para recomendação de artigos acadêmicos baseado em filtragem colaborativa aplicado à ambientes móveis. In Tarouco, L. M. R., editor, *RENOTE – Revista Novas Tecnologias na Educação*, pages 12–21. UFRGS.
- Cazella, S. C., Reategui, E., Machado, M., and Barbosa, J. (2009). Recomendação de objetos de aprendizagem empregando filtragem colaborativa e competências. In Crespo, S., editor, *Anais do SBIE 2009*. SBC.
- Chen, A. (2005). Context-aware collaborative filtering system: predicting the user’s preferences in ubiquitous computing. In *CHI’05 Extended Abstracts on Human Factors in Computing Systems*, pages 1110–1111. ACM.
- Costa, E. B., Oliveira, T. T., Silva, A. P., and Bittencourt, I. I. (2011). Sistema de recomendação híbrido para bibliotecas digitais que suportam o protocolo oai-pmh. In *Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE)*, volume 1.
- Das, A. S., Datar, M., Garg, A., and Rajaram, S. (2007). Google news personalization: scalable online collaborative filtering. In Williamson, C. and Zurko, M. E., editors, *16th international conference on World Wide Web Pages*, pages 271–280. ACM.

- Geva, S. and Trotman, A., editors (2014). *The Thirty-Seventh Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM.
- Jiangshan, X., Zhang, L., Lu, H., and Li, Y. (2010). The development and prospect of personalized tv program recommendation system. In Dutta, A. and Paul, S., editors, *2010 IEEE 4th International Conference on Internet Multimedia Services Architecture and Application*. IEEE.
- Kobsa, A. and Zhou, M., editors (2014). *RecSys '14: Proceedings of the 8th ACM Conference on Recommender Systems*, New York, NY, USA. ACM. 609146.
- Linden, G., Smith, B., and York, J. (2003). Amazon.com recommendations: item-to-item collaborative filtering. In Blake, M. B., editor, *IEEE Internet Computing – Volume 3, Issue 1*, pages 76–80. ACM.
- McCandless, M., Hatcher, E., and Gospodnetic, O. (2012). *Lucene in action*. Manning Publication, 1th edition.
- MEC (2017). Portal Periódico Capes/MEC. URL: www-periodicos-capes-gov-br Access in 2017-03-27.
- Owen, S., Anil, R., Dunning, T., and Friedman, E. (2012). *Mahout in action*. Manning Publication, 2th edition.
- Richardson, L. and Ruby, S. (2007). *RESTful Web Services-Web services for the Real World. 2007*. O'Reilly Media, 1th edition.
- Silva, F. S. (2011). *PersonalTVware: uma infraestrutura de suporte a sistemas de recomendação sensíveis ao contexto para TV Digital Personalizada*. PhD thesis, Universidade de São Paulo.
- Tantanasiriwong, S. (2011). A comparison of clustering algorithms in article recommendation system. In Zeng, Z. and Li, Y., editors, *Volume 8349 Fourth International Conference on Machine Vision (ICMV 2011): Machine Vision, Image Processing, and Pattern Analysis*. SPIE.
- UEA (2017). Sistema Integrado de Bibliotecas – UEA. URL: <http://biblioteca.uea.edu.br/biblioteca/index.php> Access in 2017-03-20.
- UFMG (2017). Sistemas de Bibliotecas da UFMG. URL: <https://www.bu.ufmg.br/bu/> Access in 2017-03-20.
- USP (2017). Biblioteca Digital de Teses e Dissertações da USP. URL: <http://www.theses.usp.br/> Access in 2017-03-20.
- Valet-Harpe, I. (2013). *The JSON Data Interchange Format*. ECMA International, 1th edition.
- Zanette, L. R. (2008). Sistema de recomendação de itens baseado na rede de confiança do usuário. Master's thesis, Universidade Federal do Rio de Janeiro, Rio de Janeiro. Programa de Pós-Graduação em Informática.
- Zhou, Y., Wilkinson, D., Schreiber, R., and Pan, R. (2008). Large-scale parallel collaborative filtering for the netflix prize. In *International Conference on Algorithmic Applications in Management*, pages 337–348. Springer.