

# O Uso de Informações Semânticas para Recomendação de Recursos Educacionais Usando Grafo Bipartido

Hugo Guércio

Pós-graduação em Ciência da  
Computação

Universidade Federal de Juiz de Fora  
Juiz de Fora – MG - Brasil

hugo.guercio@ice.ufjf.br

Victor Ströele

Pós-graduação em Ciência da  
Computação

Universidade Federal de Juiz de Fora  
Juiz de Fora – MG - Brasil

victor.stroele@ice.ufjf.br

Fernanda Campos

Pós-graduação em Ciência da  
Computação

Universidade Federal de Juiz de Fora  
Juiz de Fora – MG - Brasil

fernanda.campos@ufjf.edu.br

## ABSTRACT

The Informal Education has grown in recent years due to technological advances and the large volume of educational content on the internet. People, including researchers, students and teachers looking for educational content in different repositories to supplement their knowledge. In this scenario, the Recommendation Systems are used to help those people find the most relevant resources for them, considering, for example, profile, context and semantic information about these users. This paper proposes a recommendation architecture based on bipartite semantic graphs that aims to explore the information present in different data repositories to identify educational resources that are more adherent to users, particularly researchers. The evaluation of the model was performed using a proof of concept and the results indicate that solution is feasible.

## RESUMO

A Educação Informal tem crescido nos últimos anos em função dos avanços tecnológicos e do grande volume de conteúdos educacionais disponíveis na Internet. As pessoas, incluindo pesquisadores, estudantes e professores, buscam por conteúdos educacionais em diferentes repositórios para complementar seus conhecimentos. Neste cenário, os Sistemas de Recomendação são utilizados para auxiliar essas pessoas a encontrarem os recursos mais relevantes para elas, considerando, por exemplo, perfil, contexto e informações semânticas sobre esses usuários. O presente trabalho propõe uma arquitetura de recomendação baseada em grafos semânticos bipartidos que visa explorar as informações presentes em diferentes repositórios para identificar recursos educacionais que sejam mais aderentes aos usuários, em particular pesquisadores. A avaliação do modelo foi realizada através de provas de conceitos e os resultados indicam que a solução é viável.

## Palavras Chaves

Sistemas de recomendação, extração semântica, e-learning, educação informal

## 1. INTRODUÇÃO

A crescente quantidade de dados disponíveis é um resultado direto da expansão e do uso, cada vez maior, da internet [28]. Essa ampliação do acesso a diversos tipos de informação deu origem a uma nova forma de aprendizagem, a *Educação Informal* [14].

Atualmente, as pessoas buscam por conteúdos educacionais em diferentes repositórios, seja este repositório voltado para o ensino ou não. Com isso, tais conteúdos estão sendo criados e disponibilizado, sendo apresentados nos mais diversos formatos com o intuito de atender às necessidades de pessoas com perfis e interesses variados, como alunos, professores, pesquisadores, dentre outros.

O desenvolvimento e uso de tecnologias, que dão apoio à interação social, também crescem rapidamente, com isso, novas formas de comunicação, trabalho, estudo e entretenimento são usadas por bilhões de pessoas gerando complexas coleções de dados. Exemplificando pode ser citado o compartilhamento de diversos tipos de mídias como áudio, vídeo e fotos, bem como e-mails, grupos de discussão, blogs, wikis e as redes sociais.

Dada essa gama de ferramentas e conteúdos distribuídos por vários repositórios, saber precisamente as necessidades dos usuários é uma tarefa complexa. Os sistemas de recomendação buscam solucionar esse problema apresentando ao usuário recursos que sejam aderentes aos seus interesses, baseando-se no conhecimento extraído de dados históricos desses usuários e dos recursos que serão recomendados.

Com o conhecimento das informações relevantes e dos interesses dos usuários, torna-se possível recomendar itens que satisfaçam às necessidades de um usuário específico [16] ou de grupos de usuários [2]. Neste sentido, este trabalho avança em relação às pesquisas anteriores do Projeto BROAD [16, 17, 18, 19], com o intuito de recomendar recursos educacionais de interesse dos pesquisadores de diferentes áreas, usando informações semânticas e baseando-se nas suas redes de colaboração.

Assim, o objetivo deste trabalho é propor uma arquitetura para recomendação de recursos educacionais baseado em grafos bipartidos. Esses grafos serão compostos por pesquisadores (professores e alunos interessados em um mesmo tema educacional), vídeos, artigos e informações semânticas extraídas através da análise dos atores e dos recursos. Com apoio desse grafo semântico bipartido será desenvolvido um sistema de recomendação de recursos educacionais para auxiliar os pesquisadores na obtenção de conteúdos aderentes aos seus interesses de pesquisa.

Este trabalho está organizado da seguinte forma: a primeira seção apresenta uma introdução e é seguida pela Seção 2 que apresenta a metodologia do trabalho. A Seção 3 tem como objetivo fundamentar os conceitos utilizados neste trabalho, seguida pela Seção 4 que apresenta alguns trabalhos relacionados. Na Seção 5

é apresentada a arquitetura proposta para o sistema de recomendação. A Seção 6 apresenta alguns detalhes do desenvolvimento do protótipo utilizado na avaliação da arquitetura. A Seção 7 mostra alguns resultados e é seguida pela Seção 8 que faz uma avaliação dos resultados por meio de uma Prova de Conceito. Por fim, a Seção 9 apresenta as considerações finais e os trabalhos futuros.

## 2. Metodologia

A metodologia deste trabalho é composta por (1) uma pesquisa bibliográfica para identificar trabalhos relacionados ao tema abordado; (2) a proposta de uma arquitetura para sistemas de recomendação baseada em grafos semânticos bipartidos; (3) o desenvolvimento de um protótipo baseado na arquitetura proposta; (4) e a avaliação das recomendações geradas pelo protótipo desenvolvido através de uma prova de conceito (PoC).

Para dar embasamento teórico ao trabalho é realizado um levantamento bibliográfico dos principais temas e conceitos relacionados ao problema e sua solução. Após a pesquisa bibliográfica é proposta uma arquitetura conceitual seguida pelo desenvolvimento de um protótipo que foi dividido em duas etapas principais. Inicialmente é feita a extração de informações nas fontes de dados selecionadas e, em seguida, o grafo semântico bipartido é construído para produzir a lista de recomendações dos recursos educacionais.

No grafo bipartido os indivíduos são pesquisadores de uma rede social científica construída a partir de informações da DBLP<sup>1</sup> (Bibliografia de Ciência da Computação). As recomendações são feitas para os pesquisadores considerando informações semânticas extraídas a partir dos artigos publicados e disponibilizados nesta base. Os itens recomendados são vídeos compartilhados no Youtube e artigos presentes no repositório arXiv<sup>2</sup>.

Ao final do desenvolvimento da arquitetura, com base no grafo semântico, foi possível extrair os relacionamentos entre os objetos, gerando assim uma lista de recomendação cuja qualidade foi avaliada através da PoC.

## 3. RECOMENDAÇÃO DE CONTEÚDO COM BASE EM REDES SOCIAIS

Nesta seção serão apresentados os fundamentos teóricos necessários para o entendimento deste trabalho.

Os Sistemas de Recomendação são normalmente divididos em três categorias: baseada em conteúdo, colaborativa e híbrida [1]. No presente trabalho será utilizada a abordagem baseada em conteúdo, onde as recomendações são feitas baseadas nas características dos indivíduos e objetos a partir das informações semânticas extraídas dos mesmos. A abordagem baseada em conteúdo não apresenta a limitação de partida a frio. Esta limitação ocorre quando um novo item ou indivíduo é inserido na base sem que ele possua um número significativo de relações. Como foram utilizados repositórios que não se relacionam de forma direta, a abordagem colaborativa não foi adotada, visto que não existe conhecimento prévio das relações entre indivíduos e itens.

Os Sistemas de Recomendação representam uma área de pesquisa importante devido a abundância de aplicações práticas, que ajudam os usuários a lidar com a grande quantidade de

informações presentes em diversas plataformas. Esses sistemas fazem recomendações, como conteúdos, pessoas e serviços, que sejam de interesse dos indivíduos de maneira personalizada ou em grupos [1].

Exemplos de aplicações podem incluir recomendações de diversos produtos em sites de vendas como Amazon.com<sup>3</sup> ou Ebay.com<sup>4</sup>; vídeos no Youtube.com<sup>5</sup>; músicas no Spotify<sup>6</sup>; pessoas no Facebook<sup>7</sup> e no LinkedIn<sup>8</sup>; ou *web services* para desenvolvedores como proposto por [11].

### 3.1 Filtragem Baseada em Conteúdo

Essa abordagem tem suas raízes nos processos de recuperação da informação [4]. Neste tipo de filtragem é realizada uma descrição dos usuários e dos itens com base em suas características. A definição das características que descrevem itens ou usuários pode ser feita considerando informações descritivas, tais como gênero, tipo de mídia, duração, dentre outros; ou por meio de técnicas de extração de informação para que sejam obtidas informações semânticas, identificando características implícitas dos itens ou usuários.

Após essa caracterização é realizada uma comparação entre as descrições, verificando as relações entre elas. Nos casos onde usuários e itens possuem características semelhantes, infere-se que o item é relevante ao usuário [5].

A descrição de um usuário pode ser obtida através de informações fornecidas por ele próprio ou por suas ações na busca por recursos de seu interesse. Uma maneira de trabalhar com esse tipo de filtragem é solicitar ao usuário que ele avalie um conjunto de itens com características distintas. Após a avaliação, o sistema considera que os itens relevantes são os que se assemelham aos itens bem avaliados pelo usuário, da mesma forma, o sistema considera como não relevantes os itens pelos quais o usuário não demonstrou interesse.

### 3.2 Filtragem Colaborativa

A filtragem colaborativa é a técnica mais usada em recomendações como verificado no trabalho de [24]. Essa técnica baseia-se no pressuposto que usuários que demonstraram interesses similares no passado irão compartilhar interesses em comum no futuro [9]. Essa abordagem de filtragem se diferencia da filtragem baseada em conteúdo por não exigir a descrição dos itens a serem recomendados, baseando-se somente na semelhança entre os usuários.

Os sistemas de recomendação que utilizam a filtragem colaborativa possuem coleções de pontuações dadas aos itens pelos usuários, assim os itens avaliados pelos usuários possuem uma informação que indica o quanto o item é relevante à sua necessidade. Desta forma, um usuário pode receber recomendações a partir das pontuações fornecidas por usuários similares a ele.

---

<sup>3</sup><http://www.amazon.com/>

<sup>4</sup><http://www.ebay.com/>

<sup>5</sup><http://www.youtube.com/>

<sup>6</sup> <http://www.spotify.com/>

<sup>7</sup> <http://www.facebook.com>

<sup>8</sup> <https://www.linkedin.com/>

---

<sup>1</sup><http://dblp.uni-trier.de/>

<sup>2</sup><http://arxiv.org/>

### 3.3 Filtragem Híbrida

A abordagem de filtragem híbrida procura combinar as duas técnicas de recomendação descritas anteriormente com o intuito de minimizar as falhas de cada uma delas. Existem diferentes formas de combinar os métodos baseados em conteúdo e colaborativos em um sistema de recomendação híbrido, essas formas são classificadas da seguinte maneira segundo [1]:

- Implementando filtros colaborativos e baseados em conteúdo separadamente e então combinando suas recomendações;
- Incorporando algumas características da filtragem baseada em conteúdo na abordagem colaborativa;
- Incorporando algumas características da filtragem colaborativa na abordagem baseada em conteúdo;
- Construindo um modelo que unifique as características dos filtros colaborativos e baseados em conteúdo.

### 3.4 Redes Sociais

O grande desafio dos sistemas de recomendação é identificar corretamente as necessidades dos usuários. Por isso, as redes sociais têm sido cada vez mais utilizadas no âmbito dos sistemas de recomendação, pois, através delas, é possível extrair diversos tipos de informação dos usuários e definir o perfil dos mesmos.

O assunto Redes Sociais pode ser compreendido como o estudo das relações entre indivíduos e as implicações dessas relações [27]. Os indivíduos são conhecidos como atores, que por sua vez podem representar organizações, conjuntos de pessoas ou um único indivíduo. Os atores estão conectados de acordo com suas afinidades, criando laços sociais.

As redes sociais podem ser definidas como um conjunto finito de atores, ou grupos de atores, e as relações definidas entre eles. Diferentes tipos de redes sociais, com finalidades diversas, podem ser observados segundo o tipo de relacionamento que ela representa.

Existem diversos tipos de relações sociais, mas todas têm como objetivo conectar atores. Conjuntos de atores com relações em comum podem caracterizar grupos de atores com elevado grau de afinidade.

As redes sociais podem ser classificadas em dois tipos: Homogêneas e Heterogêneas [7]. Redes sociais homogêneas são as redes que possuem apenas um tipo de relação entre os objetos. Já nas redes heterogêneas existem diversos tipos de relações entre os objetos e esses objetos também podem ser de tipos diferentes. As redes heterogêneas, que possuem apenas um tipo de objeto, também são conhecidas como redes sociais multi-relacionais.

O uso de redes sociais, em conjunto com os sistemas de recomendação, pode acrescentar informações na forma de relacionamentos entre os usuários. Em geral o relacionamento entre os usuários pode ser implícito ou explícito. De acordo com [3], é possível extrair informações mais complexas como similaridade e a relevância da ligação. Isso é feito a partir da análise da estrutura das ligações e dos padrões comuns entre os usuários, que podem ser obtidos extraindo informações semânticas dos mesmos.

No presente trabalho será utilizada uma rede social científica como base para obtenção das informações semânticas. Este tipo de rede social pode ser caracterizada, por exemplo, redes nas quais os atores são representações de autores de artigos e as ligações são estabelecidas por relacionamentos de coautoria entre os mesmos [10]. Esses artigos podem então ser recomendados como recursos educacionais. O modelo proposto neste trabalho

será então representado por uma rede social científica heterogênea, visto que os atores são os pesquisadores e os recursos educacionais.

## 4. TRABALHOS RELACIONADOS

Como verificado em [24] as abordagens baseadas em grafos vêm sendo bem estudadas e muito utilizadas, principalmente na predição de links em redes sociais, que é uma estratégia que pode ser utilizada para a recomendação de pessoas.

Diversos trabalhos atacam o problema de recomendações a partir de uma abordagem que utiliza grafos bipartidos aplicando com diferentes propósitos. Em [12] o objetivo é especializar algoritmos de recomendação para o caso de grafos bipartidos. De acordo com os estudos, dentre os métodos de predição avaliados, os modelos de recomendação algébricos, que levam todo o grafo em consideração, podem ser aplicados. Segundo o autor, em relação aos algoritmos que se baseiam nos vizinhos imediatos, somente o modelo *preferential attachment* pode ser utilizado em redes bipartidas.

Outros trabalhos utilizam a estratégia de arestas com peso em grafos bipartidos, em [23] as arestas possuem peso referente à qualificação das relações entre os atores. Nesse trabalho uma rede de alocação dos recursos foi utilizada para explicitar a similaridade entre pares de usuários e pares de objetos, que são utilizados para predições de recomendações e qualificações. Os autores concluíram que as predições de recomendação são afetadas de maneira considerável pela acurácia da função de similaridade entre as entidades.

Já em [26] o conceito de peso nas arestas foi utilizado para representar o interesse dos usuários pelos objetos que serão recomendados. O objetivo é aumentar a viabilidade do uso de algoritmos para recomendação em grafos bipartidos.

A abordagem de redes semânticas também pode ser vista em [13] onde o objetivo é recomendar itens de exibição em museus. A rede semântica é construída utilizando apenas as características dos itens, possibilitando os usuários a entender o contexto que engloba o item, despertando o interesse dos mesmos.

Em [25], uma abordagem colaborativa é escolhida para recomendação de imagens a partir de um grafo semântico. O grafo semântico é construído a partir da redução de grafos bipartidos entre autores e termos. Nesta abordagem, o grafo bipartido é reduzido a um grafo unidimensional onde as arestas possuem um peso e os nós representam os autores.

O uso de redes sociais apoiando os sistemas de recomendação também são explorados, como em [2] onde as redes sociais auxiliam a definição de perfis de grupos de usuários a fim de auxiliar a busca e seleção de recursos educacionais de interesse comum de seus membros.

O trabalho apresentando por [8] apresenta um sistema de recomendação de material acadêmico através de uma filtragem híbrida. Os autores integram os modelos de filtragem colaborativa, por conteúdo e por conhecimento. Os materiais recomendados são inseridos na plataforma por professores e são avaliados por outros professores da mesma disciplina.

Em [27] os sistemas de recomendação são explorados com o objetivo de auxiliar novos programadores a entender a lógica de programação através da prática. Um modelo é criado utilizando analisadores de código para recomendação de recursos educacionais presentes em um repositório.

Em [22] o uso de sistemas de recomendação, para auxílio na busca de recursos educacionais, foi avaliado por alunos durante a execução de tarefas de busca em um repositório de educacionais.

A utilização de sistemas de recomendação no contexto educacional também é avaliada em [15], onde deseja-se explorar um Ambiente Virtual de Aprendizagem (AVA) para que um sistema de recomendação especialista seja capaz de responder às perguntas dos usuários do AVA.

Diferentemente de [12, 15, 22, 24, 27] a abordagem deste trabalho utiliza grafos bipartidos com pesos em suas arestas para auxiliar o processo de recomendação. Em [22, 26] também é adotado o peso nas arestas, mas esse peso não é definido com base nas informações semânticas dos atores e dos objetos.

Neste trabalho as arestas possuem um peso que representa o interesse que um pesquisador possui pelo recurso com qual ele está conectado. Diferentemente de [13], esse interesse é definido pela análise semântica de ambos os atores e objetos. Ao final tem-se um grafo semântico bipartido utilizado para recomendação de objetos aos atores, o que torna essa abordagem diferente de [25]. Com relação ao trabalho [2], deseja-se realizar recomendações para usuários de maneira personalizada, contrastando com a solução baseada em grupos apresentada no trabalho.

Como principais diferenças do trabalho [8] destacam-se o tipo de filtragem utilizada, a ausência da utilização de redes sociais além da diferença na coleta dos materiais recomendados. Em [8] os recursos educacionais foram inseridos e avaliados pelos usuários, já na proposta deste trabalho os objetos são coletados de fontes externas e avaliados semanticamente por uma ferramenta. No trabalho realizado em [22] não são usadas informações semânticas e a arquitetura da proposta não é apresentada, assim como algum detalhe de implementação.

Vale ressaltar que a busca por recursos educacionais em repositórios que não são voltados, exclusivamente, para o ensino é um grande desafio da Educação Informal e, consequentemente, do modelo proposto neste trabalho.

## 5. ARQUITETURA PROPOSTA

Nesta seção serão apresentadas as etapas que compõem a arquitetura desenvolvida para o problema de recomendações com filtragem baseada em conteúdo. A Figura 1 mostra uma visão geral da arquitetura conceitual que representa a solução proposta neste trabalho.

A arquitetura foi decomposta em quatro camadas, responsáveis por diferentes etapas para que o processo de recomendação possa ser realizado corretamente. As camadas serão detalhadas nas subseções a seguir.

### 5.1 Extração de Informações

A camada de Extração de Informações, que contempla o processo de aquisição de informações, serve como base para todas as outras etapas do processo de recomendação. Para esta aquisição podem ser utilizadas diversas fontes de informação, com múltiplas perspectivas, com o intuito de recuperar diferentes tipos de informação que podem subsidiar as outras camadas. As fontes de informação são repositórios que contêm dados para extração de conteúdos diversos.

Neste trabalho foram utilizadas três fontes de informação com objetivos distintos. A primeira fonte de informação utilizada tem como objetivo popular uma rede social científica. Para isso foi utilizada a DBLP, pois a partir do serviço é possível ter acesso a uma vasta coleção de dados com diversos tipos de informações relativas às publicações da área de ciência da computação.

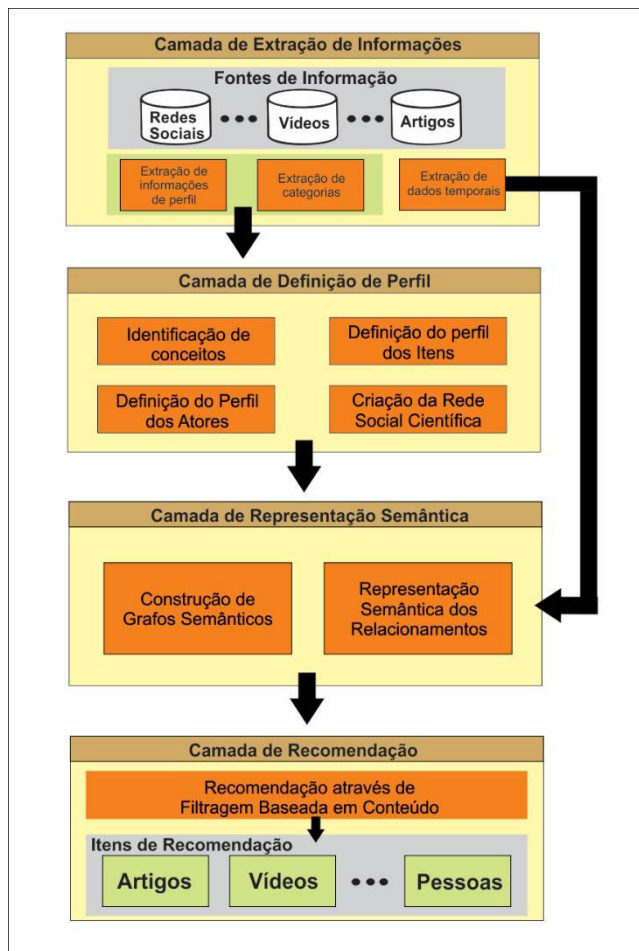


Figura 1: Arquitetura conceitual.

As outras duas fontes de informação utilizadas neste trabalho tinham como propósito fornecer diferentes recursos educacionais em mídias diferenciadas. Para isso foram utilizados o YouTube, onde foram extraídos os vídeos, e o repositório arXiv, que possibilitou a extração de artigos publicados pelos pesquisadores. O serviço é mantido pela biblioteca da universidade americana de Cornell e possui artigos de diferentes áreas, como matemática, ciência da computação, estatística e outras. O objetivo do projeto é disponibilizar um servidor para armazenamento e distribuição de artigos científicos.

### 5.2 Definição de Perfil

Depois da extração e tratamento das informações presentes nas fontes de informação, torna-se necessário realizar uma filtragem dos dados para identificar quais as características de cada recurso encontrado. Esse é o objetivo da camada de Definição de Perfil.

Para a análise do conteúdo e extração de informações semânticas foi utilizada a AlchemyAPI<sup>9</sup>. A API realiza uma análise de textos, estruturando-os e tornando-os legíveis aos computadores. Essa API é capaz de ler e entender rapidamente textos com grande velocidade, utilizando algoritmos linguísticos, estatísticos e de redes neurais. Após o processamento da ferramenta é possível

<sup>9</sup> <http://www.alchemyapi.com/>

extrair entidades, o idioma utilizado no texto, palavras-chave, conceitos relacionados e diversas outras informações.

Essa API foi aplicada para a extração de informações semânticas relacionadas aos vídeos, artigos e pesquisadores. Os conceitos identificados são as ideias discutidas no texto, independente se elas são mencionadas explicitamente ou não. Além dos conceitos de cada item e autor, um coeficiente de relevância é gerado por essa API indicando a importância de cada conceito para um dado item ou autor. De acordo com estudos realizados por [20, 21] foi observado que os resultados da extração de conceitos apresentados pela plataforma são bons quando comparados a outras soluções, fundamentando o uso neste trabalho.

### 5.3 Representação Semântica

A representação semântica dos perfis dos itens e dos pesquisadores foi realizada a partir de um grafo semântico bipartido. O grafo semântico bipartido foi construído através da união dos grafos bipartidos que relacionam os conceitos aos atores (pesquisadores) e recursos educacionais (vídeos/artigos). As arestas dos grafos bipartidos possuem um peso que foi fornecido pela camada de definição de perfil. O grafo é ilustrado na Figura 2.

O grafo semântico que relaciona os atores e os recursos educacionais também possui um valor nas arestas indicando o peso dessas relações, que foi calculado de acordo com a Equação 1.

$$R = \frac{\sum_{i \in C} (CP_i + CI_i)}{1 + \sum_{j \in C} CI_j} \quad (1)$$

Onde  $CP$  e  $CI$  são os conjuntos das relevâncias dos conceitos extraídos dos pesquisadores e dos itens, respectivamente, e  $C = \{CP \cap CI\}$  é o conjunto das relevâncias dos conceitos comuns aos pesquisadores e aos itens. Assim,  $R$  representa o quão relevante um item é para determinado pesquisador.

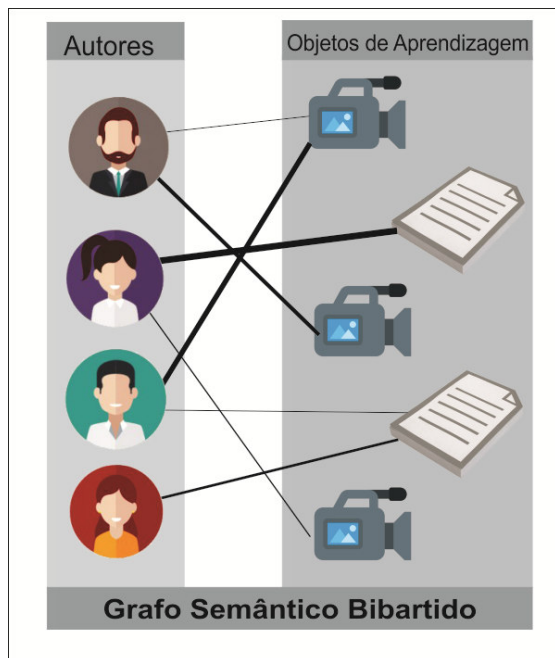


Figura 2: Grafo Semântico Bipartido

Nesta equação, no numerador são somadas as relevâncias referentes aos conceitos presentes nos dois nós (pesquisadores e itens). O denominador é a soma de todos os conceitos que apenas

o item possui, este denominador tem como objetivo penalizar a relevância do relacionamento para os conceitos que não são comuns aos dois nós. Como as relevâncias fornecidas pela API variam entre 0 e 1, onde os valores próximos a 1 indicam maior importância, o intervalo da equação é  $[0, 2n]$  onde  $n$  é a quantidade de conceitos comuns. Desta forma, os valores fornecidos pela equação que estão mais próximos a 0 indicam uma menor aderência entre os nós e valores mais distantes de 0 indicam maior aderência entre os nós.

Somente os conceitos presentes nos dois grafos gerados foram considerados. Desta forma somente os conceitos que aparecem nos dois grafos podem ser utilizados para reduzir a relevância dos relacionamentos.

### 5.4 Recomendação

A camada de recomendação é responsável por realizar as buscas no grafo semântico bipartido para identificar quais são os recursos educacionais são mais aderentes ao perfil do usuário. Além de identificar os recursos que se relacionam ao perfil do usuário este módulo define a prioridade de recomendação a partir da aderência entre os itens do grafo e os usuários, indicada pela Equação 1.

O uso de fontes compostas por diferentes tipos de recursos possibilita atender às necessidades diversas dos usuários, visto que cada indivíduo possui seu próprio interesse por determinado tipo de mídia (vídeos, áudios, textos, etc.). Além disso, por se tratar de repositórios abertos serão recomendados recursos com enfoques educacionais diferentes, possibilitando uma maior abrangência dos interesses dos usuários.

## 6. DESENVOLVIMENTO DO PROTÓTIPO

As próximas subseções apresentam alguns aspectos do desenvolvimento do protótipo, realizado de acordo com a arquitetura apresentada.

A Figura 3 detalha a arquitetura conceitual apresentada anteriormente, mostrando as principais tecnologias utilizadas em cada camada.

### 6.1 Extração de Informações

As fontes de informação são repositórios que contém dados para extração de conteúdos, eles devem possuir dados que contemplam as necessidades de informação de um usuário ou grupos de usuários. Neste trabalho serão usadas como fontes de informações a DBLP, arXiv e o YouTube. Essas fontes de informação foram escolhidas pelo fato de existir um grande volume de dados e os repositórios fornecem algum tipo de suporte para extração dos dados disponíveis. Repositórios mais específicos podem ser utilizados, como, por exemplo, a Biblioteca Digital Brasileira de Computação, BDBComp<sup>10</sup> ou o Banco Internacional de Objetos Educacionais<sup>11</sup>.

#### 6.1.1 Base de Pesquisadores

A fonte de informação DBLP foi selecionada, pois, a partir dos dados disponibilizados pelo serviço, é possível construir uma rede social científica. Esta rede contém os atores que terão suas informações semânticas extraídas na camada de definição de perfil.

Os dados estão disponíveis no formato XML e foram mapeados para um banco de dados relacional. Esse mapeamento foi feito com auxílio da biblioteca *JAVA SAXParser* permitindo a análise

<sup>10</sup> <http://www.lbd.dcc.ufmg.br/bdbcomp/>

<sup>11</sup> <http://objetoseducacionais2.mec.gov.br/>

das diferentes *tags* encontradas para realizar a identificação da estrutura do documento.

Em uma etapa posterior ao processamento das informações do arquivo XML foi realizada a carga dos dados em um banco de dados *PostgreSQL*. Após a consolidação dos dados foi possível selecionar as informações dos autores e dos artigos publicados por eles para o módulo de definição de perfil.

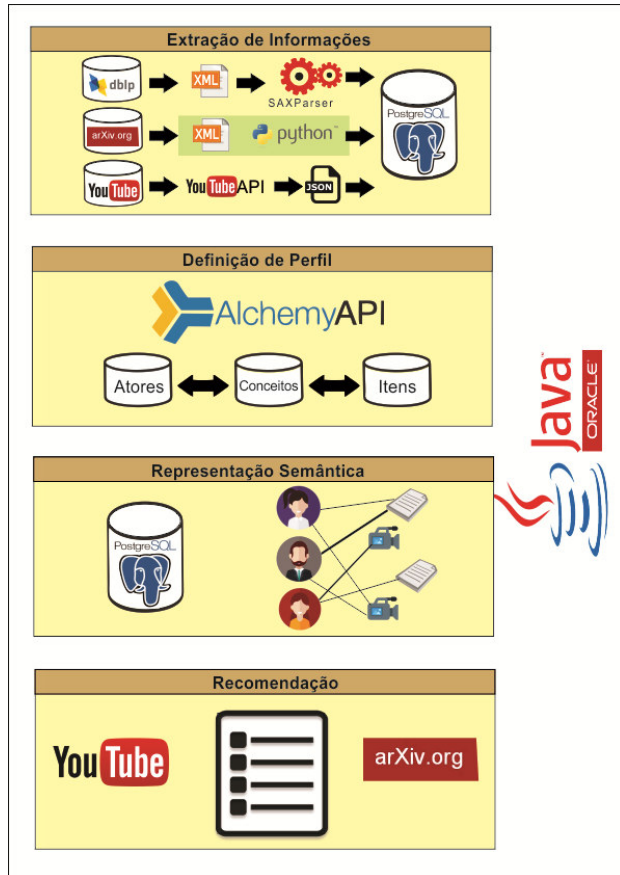


Figura 3: Visão da arquitetura proposta com as tecnologias utilizadas.

### 6.1.2 Base de Vídeos

Conforme dito anteriormente, um dos tipos de recursos a serem recomendados pelo sistema de recomendação são os vídeos. Para isso, é necessário utilizar um repositório com esse tipo de objeto e, neste trabalho, é utilizado o YouTube.

O conteúdo do YouTube pode ser coletado a partir de consultas no site ou utilizando a *YouTube Data API*<sup>12</sup>. Neste trabalho a coleta dos dados foi realizada, exclusivamente, através da API citada anteriormente. A linguagem JAVA foi utilizada para realizar as chamadas a essa API. Os atributos presentes na tabela Tabela 1 foram selecionados dentre aqueles que estão disponíveis por essa API.

Tabela 1: Atributos extraídos dos vídeos.

Atributo	Tipo	Descrição
Id	<i>string</i>	Identificador do vídeo

Title	<i>string</i>	Título do vídeo
Description	<i>string</i>	Descrição textual do vídeo
View_count	<i>long</i>	Total de visualizações
Like_count	<i>long</i>	Quantidade de likes
Dislike_count	<i>long</i>	Quantidade de Dislikes
Published_at	<i>datetime</i>	Data da publicação

Com o propósito de realizar a coleta foi necessário criar uma conta no Google<sup>13</sup> para acessar o *Console de APIs*. Após receber o acesso ao console de desenvolvedores é necessário criar um projeto para o qual são apresentadas as APIs do Google.

O Google fornece 3 APIs que consomem dados do YouTube, sendo elas: *YouTube Data API*, *YouTube Analytics API* e *Youtube Reporting API*. Como dito anteriormente a coleta foi realizada utilizando a *YouTube Data API*. Essa API provê acesso aos dados do Youtube como vídeos, *playlists* e canais. Antes de realizar as requisições à API é necessário criar uma chave de API, que identifica o projeto para verificar as cotas de utilização e acesso.

Com a API pronta para receber as requisições foi possível iniciar o processo de extração dos vídeos, as respostas da API são transmitidas no formato JSON e convertidas para classes JAVA. Para extração dos dados foi realizada uma chamada no método *search* retornando 50 resultados. Este método teve como parâmetro o perfil semântico dos autores, formado a partir dos conceitos identificados. A partir do retorno obtido por essa busca foi possível realizar a coleta dos dados pertinentes de cada vídeo (Tabela 1). Os dados coletados foram transcritos para um banco de dados *PostgreSQL*.

### 6.1.3 Base de Artigos

Os artigos a serem recomendados para os pesquisadores foram extraídos do repositório arXiv. Para a extração dos metadados dos artigos, um algoritmo na linguagem *PYTHON* foi utilizado, conforme pseudocódigo apresentado na Figura 4.

```

1 parametros = (ano, área de pesquisa, token);
2 while token != NULL do
3     requisiçãoAoRepositório(parametros);
4     if máximo de tentativas por hora esgotado then
5         | Aguarde 60 minutos;
6     end
7     else if xml válido then
8         resultados = acrescentar.Leitura(xml);
9         if token != NULL then
10            | parametros.acrescentar(token);
11        end
12 end
13 armazena(resultados);

```

Figura 4: Pseudocódigo.

Esse algoritmo realiza as chamadas ao serviço, recuperando as respostas em XML e armazenando-as em um banco de dados *PostgreSQL*. Foram coletados identificadores dos trabalhos,

<sup>12</sup> <http://developers.google.com/youtube/v3/>

<sup>13</sup> <http://www.google.com>

categorias, data de submissão, título e abstract de todos os artigos presentes na plataforma.

## 6.2 Definição de Perfil

Para a análise do conteúdo e extração de informações semânticas foi utilizada a *AlchemyAPI*. Na definição do perfil dos vídeos, a extração de conceitos ocorreu a partir da análise do título e descrição dos mesmos. Os artigos tiveram seus conceitos extraídos após a avaliação dos títulos e resumos. A especificidade de um conceito para um dado item também foi determinada pela *AlchemyAPI*. Desta forma, foram armazenados os pares compostos por conceitos e itens, assim como sua relevância.

A identificação dos conceitos associados aos autores ocorreu através da análise de todos os títulos dos trabalhos publicados por esses autores. Assim como foi feito para os itens, também foram armazenados os pares de conceitos e autores com suas respectivas relevâncias.

Após a identificação dos conceitos foram construídos dois grafos bipartidos. Ambos possuem dois tipos de nós, no primeiro estão representados os autores e seus conceitos, enquanto no segundo grafo são representados os itens e os conceitos. A aresta dos grafos bipartidos gerados tem como peso o valor da relevância entre as suas extremidades, essa medida foi extraída pelo extrator de conteúdo e representa a importância do conceito aos dados analisados pela API.

## 6.3 Representação e Recomendação

Conforme dito anteriormente, o grafo semântico foi construído unindo os dois grafos bipartidos gerados a partir da análise do conteúdo da camada anterior. O peso das relações entre os usuários e os objetos educacionais foram calculados segundo a Equação 1, que foi calculada através da linguagem SQL e armazenada em seguida no banco de dados utilizado.

De acordo com o perfil identificado, a Equação 1 indica a aderência existente entre os pesquisadores e recursos educacionais. A lista de recomendação foi gerada analisando os valores das arestas do grafo semântico ordenadas de forma decrescente, visto que os maiores valores indicam a maior aderência entre um pesquisador e um item. Nos casos onde ocorreu empate entre os valores de duas arestas diferentes foi utilizado como critério de desempate, para os vídeos, a diferença entre a quantidade de *likes* e *dislikes* do item em questão. No caso dos artigos, o critério de desempate foi a data de publicação.

## 7. RESULTADOS

Após o desenvolvimento e execução da solução proposta foram identificados os dois grafos gerados no processo descrito na Subseção 6.2.

No geral, foram extraídos 86.425 autores, 3.548 vídeos e 1.153.003 artigos. A dimensão dos grafos construídos pode ser vista detalhadamente na Tabela 2. Devido ao grande número de resultados retornados pela base de artigos, foram avaliados somente aqueles que possuíam ao menos uma categoria da área de computação e data de submissão entre os anos de 2015 e 2016.

**Tabela 2: Informações quantitativas dos grafos bipartidos**

	Instâncias avaliadas	Conceitos Identificados	Relacionamentos
<b>Autor</b>	86.425	19.785	615.898

<b>Vídeo</b>	3.548	2.553	10.547
<b>Artigo</b>	24.385	12.310	170.149

Para a construção do grafo semântico, nem todos os autores possuíam conceitos relacionados aos itens. Mesmo assim, a quantidade de autores e itens que compõem o grafo contemplou a maioria dos atores e itens classificados. Aproximadamente 78% dos autores e 72% dos vídeos estão presentes no grafo semântico. Com relação aos artigos, 99.86% deles estão presentes no grafo semântico. Esse alto número já era esperado, pois os autores e artigos são referentes a área de ciência da computação.

**Tabela 3: Conceitos mais comuns.**

Tipo	Conceito	Ocorrências
Vídeo	Vector space	215
	Mathematics	202
	Group	144
Artigo	Mathematics	4.399
	Algorithm	2.612
	Scientific method	2.016
Autor	Mathematics	8.448
	Algorithm	8.413
	Logic	4.215

A Tabela 3 mostra os conceitos que são compartilhados pelo maior número de atores e itens. Pode ser observado que a quantidade dos conceitos mais identificados para os autores e artigos foi bem maior, visto que o conjunto de instâncias de vídeos analisadas foi inferior às avaliadas para autores e artigos.

A seguir seguem alguns dados coletados de dois autores presentes no grafo semântico obtidos através do protótipo desenvolvido neste trabalho. A Tabela 4 mostra os conceitos identificados para o autor 22, ou seja, os conceitos que definem o perfil semântico do pesquisador 22. Os conceitos apresentados são apenas aqueles em que existe algum vídeo que também possui o conceito.

**Tabela 4: Conceitos identificados para o autor 22.**

Autor	Conceito	Relevância
22	Logic	0.778654
	Reasoning	0.719578
	Problem solving	0.689492

Na Tabela 5 são apresentados os conceitos associados a alguns vídeos que se relacionam com o pesquisador 22. Com as informações referentes à relevância dos conceitos para os autores o grafo bipartido é construído. Da mesma forma, é construído um grafo bipartido que representa os vídeos e seus conceitos. Nos dois grafos construídos o peso das arestas representa a relevância encontrada pelo Analisador de Conteúdo, descrito na Subseção 6.2.

**Tabela 5: Conceitos identificados para os vídeos relacionados ao autor 22.**

Vídeo	Conceito	Relevância
DIKkujAIeTY	Logic	0.988
	Reasoning	0.638043
	Inductive reasoning	0.618041
	Fallacy	0.617589
	Deductive reasoning	0.608903
	Inference	0.596085
	Analogy	0.593086
	Abductive reasoning	0.55106
ervHbKa7R5g	Problem solving	0.92607
hq1bUM2tyg0	Problem solving	0.865606
kpqSeXpiM6k	Problem solving	0.895888
SpDIXJ2I2D4	Problem solving	0.954089
	Problem	0.470192
Zyq6TmQVBxk	Problem solving	0.895888

A Tabela 6 representa a lista de recomendações para o autor 22. Essa lista de recomendação foi obtida a partir da construção do grafo semântico, que representa a união dos dois grafos bipartidos construídos.

**Tabela 6: Lista de recomendação para o autor 22.**

Vídeo	Relevância
ervHbKa7R5g	1,615562
kpqSeXpiM6k	1,585380
Zyq6TmQVBxk	1,585380
hq1bUM2tyg0	1,555098
SpDIXJ2I2D4	1,117936
DIKkujAIeTY	0,681447

## 8. AVALIAÇÃO

Este trabalho foi avaliado com base em duas provas de conceitos (PoC) desenvolvidas segundo o método GQM [6]. As provas de conceito foram realizadas com usuários reais, professores e pesquisadores do departamento de ciência da computação de uma instituição de ensino superior. A avaliação foi feita através de um questionário onde esses professores avaliaram diferentes questões sobre a abordagem proposta neste trabalho. A legenda usada no questionário foi concordo plenamente, concordo parcialmente, não concordo nem discordo, discordo parcialmente, discordo plenamente.

Neste primeiro momento foram avaliadas apenas as recomendações feitas por meio do repositório de vídeos. Os artigos recomendados ainda não foram avaliados visto que a

extração dos mesmos foi feita após a análise das considerações dos pesquisadores participantes dessa primeira avaliação.

### 8.1 Primeira PoC: definição de perfil por uma rede social científica

Nesta primeira prova de conceito é analisada a definição do perfil dos professores para avaliar a viabilidade de definição de perfil utilizando uma rede social científica. O objetivo é verificar se as informações extraídas estão corretas, levando em consideração a opinião do usuário.

Através dos conceitos extraídos para cada um dos cinco professores foi possível identificar o interesse de pesquisa deles. Para tal, conforme descrito anteriormente, foram analisadas as publicações de artigos desses professores nos últimos quatro anos. Dos cinco professores convidados a participar desta avaliação quatro deles desenvolvem seus trabalhos em linhas de pesquisa distintas, embora todos sejam da grande área da computação. Em média foram extraídos sete interesses de pesquisa (conceitos) de cada professor.

Após identificar esses interesses os professores foram convidados a avaliá-los respondendo à seguinte pergunta: *As informações extraídas da DBLP representam de fato seus interesses de pesquisa?*



**Figura 5: Avaliação dos interesses extraídos dos professores.**

Quando são analisadas as avaliações na Figura 5, verifica-se que quatro professores concordam que os interesses identificados pelo processo de extração são aderentes aos seus perfis de pesquisa. Nenhum professor discordou do perfil identificado tenha relação com seus interesses de pesquisa, mas houve um professor que indicou ser neutro em relação a esse questionamento.

Esta primeira prova de conceito mostra a viabilidade de extrair informações relacionadas aos interesses de pesquisa dos professores através de uma rede social científica.

Têm-se indícios de que o uso das redes sociais científicas é viável no processo de definição dos interesses de pesquisa dos professores e que, tais interesses, são compatíveis com a realidade dos professores.

### 8.2 Segunda PoC: aderência das recomendações aos interesses dos professores

O objetivo da segunda prova de conceito é analisar a viabilidade da geração de recomendações e avaliar a qualidade das mesmas com base na aderência aos interesses de pesquisa dos professores considerando a opinião do usuário.

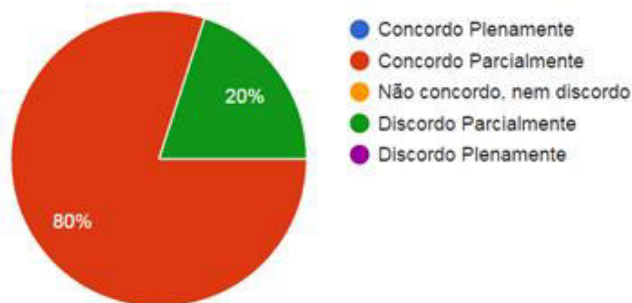
Um sistema de recomendação busca satisfazer as necessidades dos usuários apresentando recursos que atendam às suas necessidades. Neste trabalho foram recomendados vídeos com o intuito de levar aos professores recursos que pudessem complementar suas



pesquisas. Com isso, é necessário saber se as recomendações atenderam a esse objetivo.

Para a realização desta prova de conceito foi solicitado ao usuário que ele avaliasse uma lista de vídeos recomendados a ele. Essa lista foi ordenada segundo a relevância dos termos extraídos para cada autor, calculada de acordo com a Equação 1. Foram recomendados entre 7 e 10 vídeos para cada autor.

Os cinco professores foram convidados a avaliar as recomendações geradas para os conteúdos extraídos para cada um deles. Para realizar a avaliação eles responderam às seguintes perguntas: (1) *O tema dos vídeos recomendados a você está coerente com as informações extraídas da DBLP?* (2) *Você compartilharia o vídeo recomendado para algum aluno ou professor da mesma área?*

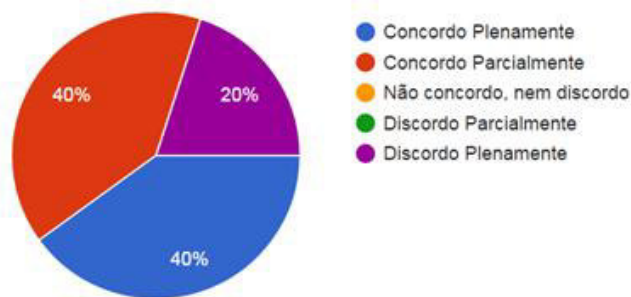


**Figura 6: O tema dos vídeos recomendados a você está coerente com as informações extraídas da DBLP?**

O gráfico da Figura 6 apresenta as avaliações dos professores com relação a primeira pergunta. Pode-se observar que quatro, dentre os cinco professores, concordam parcialmente que as recomendações são coerentes com as informações extraídas da DBLP. Um professor discordou parcialmente que haja uma relação entre as recomendações e os conceitos extraídos.

Também foi solicitado aos professores que eles deixassem seus comentários e críticas para que o sistema de recomendação pudesse ser aprimorado. Nesses comentários foi possível compreender melhor as avaliações feitas por eles. Em geral, sempre há algum vídeo que não é condizente com os interesses de pesquisa dos professores e, por isso, nenhum deles concordou plenamente com as recomendações. O professor que discordou parcialmente recebeu recomendação de vídeos de um canal de humor que abordava, de alguma maneira, os conceitos extraídos da DBLP. Em um trecho dos seus comentários o professor diz que:

“...acredito que possam incluir algum tipo de filtro para selecionar canais relacionados a educação/pesquisa, já que o foco é identificar vídeos associados ao interesse de pesquisa do indivíduo.”



**Figura 7: Você compartilharia o vídeo recomendado para algum aluno ou professor da mesma área?**

O gráfico da Figura 7 consolida as respostas dos professores à segunda pergunta. O objetivo dessa pergunta é identificar se existe alguma recomendação que o professor considera relevante ao ponto de repassar o vídeo sugerido. Observa-se que quatro professores consideram os vídeos interessantes e recomendariam os mesmos para outras pessoas. Um professor, o mesmo que discordou parcialmente na questão anterior, discorda plenamente, ou seja, ele não recomendaria os vídeos a outras pessoas. Este professor considerou os vídeos superficiais aos seus temas de pesquisa e, por isso, não os recomendaria.

Os resultados da primeira prova de conceito indicam que a estratégia de definição do perfil de pesquisadores a partir de uma rede social científica foi bem avaliada pelos professores. Dentre os professores que avaliaram essa estratégia apenas um não concordou com o perfil definido para ele.

Da mesma forma, os resultados da segunda prova de conceito também trouxeram indicativos de que as recomendações feitas aos professores foram, de certa forma, coerentes aos perfis identificados. Como era esperado, o professor que não concordou com a definição de seu perfil também não concordou com os vídeos recomendados.

Assim sendo, os resultados observados mostram a viabilidade das estratégias de extração de características em redes sociais científicas e de recomendação de vídeos.

### 8.3 Ameaças a validade

Existem algumas ameaças à validade deste estudo. O estudo simula que os pesquisadores estavam precisando dos itens que foram recomendados. Portanto, é importante que esse estudo seja expandido para que os professores e pesquisadores possam acessá-lo quando realmente estiverem necessitando de alguma recomendação de um objeto de aprendizagem.

A qualidade dos termos extraídos para identificar o perfil dos professores também é uma ameaça a validade. Há um risco de que, quando outros professores forem selecionados, termos muito genéricos ou ambíguos sejam identificados, impactando diretamente na qualidade das recomendações.

Outra ameaça é a quantidade de professores convidados a fazerem a avaliação das estratégias apresentadas neste trabalho. Embora tenham sido selecionados professores de diferentes áreas de pesquisa, sabe-se que é necessário avaliar a proposta com um maior quantitativo de pessoas.

## 9. CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi propor uma arquitetura de recomendação baseada em informações semânticas e, a partir do desenvolvimento dessa arquitetura, gerar recomendações de

recursos educacionais para auxiliar a educação informal de professores e pesquisadores. Uma pesquisa bibliográfica foi realizada para fundamentar a abordagem da solução proposta e identificar os trabalhos relacionados.

Após a proposta da arquitetura, foi realizada a extração dos dados e, de posse dos mesmos, foram extraídos conceitos com o objetivo de descrever semanticamente os recursos educacionais e atores utilizados no sistema de recomendação. De posse dos atores e itens, com seus respectivos perfis, foi possível quantificar a relação entre eles e criar um grafo semântico bipartido. A partir do grafo semântico construído foi possível extrair uma lista de recomendações ordenada pela relevância entre os nós.

De posse da lista de recomendação, os professores e pesquisadores podem contar com o auxílio do protótipo desenvolvido para selecionar os recursos educacionais que se adequem às suas reais necessidades, reduzindo o tempo gasto na procura por recursos educacionais e aumentando a eficiência na preparação de materiais para auxílio ao ensino e aprendizado.

Com o intuito de avaliar a viabilidade da solução proposta foram realizadas duas provas de conceitos. Um questionário foi desenvolvido e cinco professores o responderam. Com base nas respostas obtidas verificou-se que há indícios da viabilidade do modelo de recomendação baseado em grafos semânticos. Além disso, avaliando as sugestões desses professores, algumas evoluções podem ser aplicadas ao modelo para aprimorar as recomendações, como a recomendação de artigos científicos.

Outra questão importante é a modificação do modelo proposto adicionando o *feedback* fornecido pelo usuário após a recomendação para auxiliar as próximas recomendações. A partir do *feedback* pode ser utilizada uma abordagem de filtragem híbrida para fazer recomendações utilizando, além das informações semânticas, a opinião dos usuários com perfil semelhantes.

Outra oportunidade para trabalhos futuros é abordar o uso de ontologias na etapa de análise do conteúdo para agrupar os conceitos dentro de um domínio, podendo assim modificar a qualificação entre os relacionamentos e alterar a lista de recomendação. Essa evolução irá auxiliar a eliminar recomendações que não tenham relações diretas com o interesse dos pesquisadores.

## 10. AGRADECIMENTOS

Esta pesquisa está sendo parcialmente apoiada pela CAPES, CNPq, FAPEMIG e UFJF.

## 11. REFERÊNCIAS

- [1] Adomavicius, G., and Tuzhilin, A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *Knowledge and Data Engineering, IEEE Transactions on*, 17(6):734-749, 2005.
- [2] Almeida, F.R., Pereira, C. K., Campos, F., and Ströele, V. Recomendação de recursos educacionais para grupos: buscando soluções em redes sociais. In *Anais do Simpósio Brasileiro de Informática na Educação*, 2016.
- [3] Backstrom, L. and Leskovec, J. Supervised random walks: predicting and recommending links in social networks. In *Proceedings of the fourth ACM international conference on Web search and data mining*, pages 635-644. ACM, 2011.
- [4] Baeza-Yates, R., Ribeiro-Neto, B. *Modern Information Retrieval*, volume 463. ACM press New York, 1999.
- [5] Balabanović, M. and Shoham, Y. Fab: content-based, collaborative recommendation. *Communications of the ACM*, 40(3):66-72, 1997.
- [6] Basili, V. R. Software modeling and measurement: The goal/question/metric paradigm. 1992
- [7] Cai, D., Shao, Z., He, X., Yan, X., and Han, J. Community mining from multi-relational networks. In *Knowledge Discovery in Databases: PKDD 2005*, pages 445-452. Springer, 2005.
- [8] M. de Meira Góis, S. J. Rigo, and J. L. Barbosa. Rodac – um Sistema de recomendação de material acadêmico baseado em conhecimento. *XX Conferência Internacional sobre informática Educativa – TISE. Nuevas Ideas en Informatica Educativa*, 2015.
- [9] Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12):61-70, 1992.
- [10] Goldenberg, J., Libai, B., Muller, E., and Stremersch, S. Database submission – the evolving social network of marketing scholars. *Marketing Science*, 29(3):561-567, 2010
- [11] Jiang, B., Zhang, X. X., Pan, W. F., and Hu, B. Bigsir: A bipartite graph based service recommendation method. In *Services (SERVICES), 203 IEEE Ninth World Congress on*, pages 363-369. IEEE, 2013.
- [12] Kunegis, J., De Luca, E. W., and Albayrak, S. The link prediction problem in bipartite networks. In *Computational intelligence for knowledge-based systems design*, pages 380-389. Springer, 2010.
- [13] Maehara, C., Yatsugi, K., Kim, D., and Ushiyama, T. An exhibit recommendation system based on semantic networks for museum. In *Innovations in Intelligent Machines-2*, pages 131-141. Springer, 2012.
- [14] Moreira, J. D. C., Neto, F. M. M., de Medeiros Valentim, R. A., da Costa, A. A. L., Sombra, Ê. L., and de Aliança Neto, A. S. Conhecendo o usuário através de enriquecimento semântico para apoio à aprendizagem informal em saúde. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 25, page 307, 2014.
- [15] Nobrega Santos, S. I. and de Barros Costa, E. Um modelo de recomendação sensível ao context para auxiliary estudantes em ambientes virtuais de aprendizagem. *XIX Conferência Internacional sobre informática Educativa – TISE. Nuevas Ideas en Informatica Educativa*, 2014.
- [16] Pereira, C. K, Campos, F., Ströele, V., Braga, R., David, J. M., and Almeida, R. Extração de características de perfil e de context em redes sociais para recomendação de recursos educacionais. *Revista Brasileira de Informática na Educação*, 23(03):25, 2015
- [17] Pereira, C. K, Campos, F., Ströele, V., David, J. M., Braga, R. Explorando dados ligados através de um Sistema de recomendação educacional. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 26, page 1042, 2015.
- [18] Pereira, C. K, Campos, F., Ströele, V., David, J. M., Braga, R. Extração de características de perfil e context em redes

- sociais para recomendação de objetos de aprendizagem. In *Anais do Simpósio Brasileiro de Informática na Educação*, volume 25, page 506, 2014.
- [19] Rezende, P. A. A., Pereira, C. K., Campos, F., David, J. M. N., and Branga, R. Personna: proposta de ontologia de context e perfil de alunos para recomendação de objetos de aprendizagem. *Revista Brasileira de Informática na Educação*, 23:70, 2015.
- [20] Rizzo, G. and Troncy, R. Nerd: evaluating a named entity recognition tools in the web of data. In: *(ISWC'11) workshop on Web Scale Knowledge Extraction, Bonn, Germany*. 2011
- [21] Saif, H., He, Y., and Alani, H. Semantic sentiment analysis of twitter. In *The Semantic Web-ISWC 2012*, pages 508-524. Springer, 2012.
- [22] Santos, M. B. and Silva, J. M. C. Avaliação de um Sistema de recomendação para objetos de aprendizagem obaa. *XIX Conferência Internacional sobre informática Educativa – TISE. Nuevas Ideas en Informatica Educativa*, 2014.
- [23] Sawant, S. Collaborative filtering using weighed bipartite graph projection: a recommendation system for yelp. *CS224W: Social and Information Network Analysis (December 10, 2013)*, 2013
- [24] Shi, Y., Larson, M. and Hanjalic, A. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Computing Surveys (CSUR)*, 47(1):3, 2014.
- [25] Shieh, J. R., Yeh, Y. T., Lin, C. H., Lin, C. Y., and Wu, J. L. Collaborative knowledge semantic graph image search. In *Proceedings of the 17<sup>th</sup> international conference on World Wide Web*, pages 1055-1056. ACM, 2008.
- [26] Ting, Y., Yan, C., and Xiang-wei, M. Personalized recommendation system based on web log mining and weighted bipartite graph. In *Computational and Information Sciences (ICCIS), 2013 Fifth International Conference on*, pages 587-590. IEEE, 2013.
- [27] Vera, J., Mamani, A. O., and Villalba, K. Modelo de sistema de recomendación de Objetos de Aprendizaje en dispositivos móviles, caso: Desarrollo del pensamiento computacional. *XX Conferência Internacional sobre informática Educativa – TISE. Nuevas Ideas en Informatica Educativa*, 2015
- [28] Wasserman, S. Social network analysis: Methods and applications, volume 8. *Cambridge university press*, 2014.
- [29] Zhang, G. Q., Zhang, Q. G., Yang, Q. F., Cheng, S. Q., and Zhou, T. Evolution of the internet and its cores. *New Journal of Physics*, 10(12):123027, 2008.