

UNIVERSIDADE FEDERAL DE JUIZ DE FORA  
INSTITUTO DE CIÊNCIAS EXATAS  
BACHARELADO EM SISTEMAS DE INFORMAÇÃO

# **Desenvolvimento de um Modelo Semântico para Recomendações Baseado em Grafos**

**Hugo Guércio Fernandes**

JUIZ DE FORA  
DEZEMBRO, 2015

# Desenvolvimento de um Modelo Semântico para Recomendações Baseado em Grafos

HUGO GUÉRCIO FERNANDES

Universidade Federal de Juiz de Fora  
Instituto de Ciências Exatas  
Departamento de Ciência da Computação  
Bacharelado em Sistemas de Informação

Orientador: Victor Ströele de Andrade Menezes

JUIZ DE FORA  
DEZEMBRO, 2015

## Resumo

A grande quantidade de alternativas pode dificultar a identificação da informação mais relevante. Os sistemas de recomendação tem como objetivo auxiliar um usuário, ou grupos de usuários, na identificação das informações mais relevantes à sua pesquisa. Com a crescente quantidade de dados, os sistemas de recomendação vem sendo cada vez mais estudados, pois é cada vez mais difícil encontrar a informação desejada. O presente trabalho tem como objetivo modelar e desenvolver um sistema de recomendações baseado em grafos bipartidos, com informações semânticas dos usuários e itens de recomendação.

**Palavras-chave:** Sistemas de Recomendação, Extração Semântica, Grafo Bipartido

## Abstract

The large amount of alternatives can hamper the recognition of which information is more relevant. Recommendation systems aim at identification of the most relevant result to a given search. With growing data, recommendation systems have been studied hard and the interest in this area remains since it's getting harder to retrieve the desired information. This paper aims at modeling and developing a recommendation system based on a bipartite network with semantic information about users and items.

**Keywords:** Recommendation Systems, Semantic Extraction, Bipartite Graph.

## AGRADECIMENTOS

A Deus por ter me dado força para superar todas as dificuldades, à minha família por todo o apoio, oportunidades, amparo e carinho. Em especial aos meus pais, Sandra e Moisés, e minha irmã Kamila, pelo grande apoio pois sem eles nada disso seria possível

Gostaria de agradecer aos meus amigos por se mostrarem sempre dispostos a me animar, ensinar e auxiliar em todos os problemas e a Ludmila, minha companheira e futura esposa, por sempre estar ao meu lado me enchendo de amor, carinho e motivação.

Ao professor Victor, pela paciência e dedicação, sem as quais este trabalho não se realizaria e aos professores do Departamento de Ciência da Computação por todos os ensinamentos durante todos esses anos.

A todos, muito obrigado!

# Sumário

<b>Lista de Figuras</b>	<b>5</b>
<b>Lista de Tabelas</b>	<b>6</b>
<b>1 Introdução</b>	<b>7</b>
1.1 Apresentação do Problema . . . . .	8
1.1.1 DBLP . . . . .	8
1.1.2 Youtube . . . . .	9
1.2 Motivação . . . . .	9
1.3 Objetivos . . . . .	10
1.4 Metodologia . . . . .	10
1.5 Estrutura do Trabalho . . . . .	11
<b>2 Revisão Bibliográfica</b>	<b>12</b>
2.1 Redes Sociais . . . . .	12
2.2 Recomendações . . . . .	13
2.2.1 Filtragem Baseada em Conteúdo . . . . .	13
2.2.2 Filtragem Colaborativa . . . . .	14
2.2.3 Filtragem Híbrida . . . . .	14
2.3 Grafos Bipartidos . . . . .	15
2.4 Trabalhos Relacionados . . . . .	15
<b>3 Solução Proposta</b>	<b>18</b>
3.1 Workflow . . . . .	18
3.2 Fontes de Informação . . . . .	18
3.2.1 Base de Pesquisadores . . . . .	19
3.2.2 Base de Vídeos . . . . .	21
3.3 Análise do Conteúdo . . . . .	22
3.4 Construção do Grafo Semântico . . . . .	24
3.5 Recomendação . . . . .	26
<b>4 Resultados</b>	<b>28</b>
<b>5 Considerações Finais</b>	<b>31</b>
<b>Referências</b>	<b>33</b>

## Lista de Figuras

3.1	Workflow implementado . . . . .	19
3.2	Modelo de banco de dados da DBLP por Duque (2015) . . . . .	20
3.3	Construção do grafo semântico. . . . .	24
3.4	Grafo semântico. . . . .	26
4.1	Autores e vídeos presentes no grafo semântico. . . . .	28

## Lista de Tabelas

3.1	Atributos extraídos dos vídeos. . . . .	21
3.2	Funções da AlchemyAPI. . . . .	23
4.1	Informações quantitativas dos grafos bipartidos. . . . .	28
4.2	Conceitos mais comuns. . . . .	29
4.3	Conceitos identificados para os autores 22 e 56. . . . .	29
4.4	Conceitos identificados para os vídeos relacionados ao autor 56. . . . .	29
4.5	Conceitos identificados para os vídeos relacionados ao autor 22. . . . .	30
4.6	Listas de recomendação para os autores 22 e 56. . . . .	30



# 1 Introdução

A crescente quantidade de dados disponíveis é um resultado direto da ampliação, cada vez maior, da internet (Zhang *et al.*, 2008). O aumento de fontes de dados livres intensifica a necessidade de transformar estes dados em conhecimento.

O desenvolvimento e uso de tecnologias, que dão apoio à interação social, cresce rapidamente. Com isso novas formas de comunicação, trabalho e entretenimento são usadas por bilhões de pessoas gerando complexas coleções de dados. Exemplificando podemos citar o compartilhamento de diversos tipos de mídia como áudio, vídeo e fotos. Também podemos mencionar emails, grupos de discussão, blogs, wikis e ferramentas de redes sociais. Além disso, o crescente uso de dispositivos móveis possibilita a captura de dados geográficos, aumentando cada vez mais as possibilidades de análise.

Em muitas situações o usuário pode ficar desorientado com tantas informações, por outro lado, essa excepcional quantidade de dados pode subsidiar diversas pesquisas para a geração de conhecimento e recomendações. O conhecimento descoberto pela análise dos dados pode ser utilizado em diferentes cenários como: análise de mercado, recomendação de produtos, controle de produção ou retenção de consumidores (Han & Kamber, 2006).

Com o conhecimento das informações relevantes dos usuários e itens, torna-se possível recomendar os itens mais interessantes, seja a um grupo de usuários ou a um usuário específico.

As recomendações, na maioria das vezes, são feitas a partir do conhecimento extraído de dados históricos dos usuários e itens. Casos bem conhecidos de recomendações podem ser exemplificados pela sugestão de itens no Amazon.com como analisado em (Oestreicher-Singer & Sundararajan, 2010) e em filmes na Netflix.com (Amatriain, 2013). O caso de recomendações na Netflix tornou-se bastante popular desde o lançamento de uma competição que visava aumentar a acurácia do seu sistema de recomendações (Hallinan & Striphas, 2014).

## 1.1 Apresentação do Problema

Com a grande quantidade de dados disponíveis, muitas vezes as pessoas encontram muitas alternativas. Com isso, a identificação de qual informação é mais relevante pode se tornar uma tarefa árdua. Os sistemas de recomendação tem o propósito de fornecer ao indivíduo as informações mais relevantes de acordo com sua necessidade.

Os sistemas de recomendação são normalmente divididos em três categorias: baseada em conteúdo, colaborativa e híbrida. No presente trabalho utilizaremos a abordagem baseada em conteúdo, onde as recomendações são feitas baseadas nas características dos indivíduos e objetos. As principais características serão informações semânticas extraídas dos indivíduos e objetos.

Neste trabalho, os indivíduos serão pesquisadores de uma rede social científica. As recomendações serão feitas para pesquisadores a partir dos artigos escritos e disponibilizados na DBLP (Bibliografia de Ciência da Computação). Os itens recomendados serão vídeos compartilhados em um canal especializado em conteúdo educacional.

Nas subseções a seguir são apresentadas as fontes dos dados utilizados no trabalho.

### 1.1.1 DBLP

O serviço DBLP provê informações bibliográficas dos principais trabalhos científicos. Inicialmente DBLP representava *DataBase Systems and Logic Programming* (Lei & Reuther, 2006). O acrônimo original perdeu seu sentido e hoje deve-se aceitar DBLP como o nome do serviço oferecido.

Segundo Lei & Reuther (2006), o propósito das publicações científicas são os de documentar e compartilhar novos conhecimentos e resultados. Para isso faz-se necessário um serviço confiável para armazenar, organizar e possibilitar o acesso às publicações. O DBLP foi criado com esse propósito. Os indivíduos analisados neste trabalho são os autores dos trabalhos contidos na DBLP.

### 1.1.2 Youtube

Fundado em fevereiro de 2005, o YouTube transformou-se em sinônimo para compartilhamento de vídeos online, tornando-se o site mais popular com este propósito.

A imensa popularidade do YouTube pode ser verificada nas estatísticas fornecidas pela empresa: o serviço possui mais de um bilhão de usuários e, a cada dia, são assistidas centenas de milhares de horas de vídeos gerando bilhões de visualizações <sup>1</sup>.

Como descrito em (Hansen *et al.*, 2010) o YouTube é baseado em duas camadas, diferenciando vídeos (conteúdo) e usuários (comunidade). Com isso pode-se analisar tanto redes de vídeos ou usuários, seja de forma independente ou combinada.

Com a abundante quantidade de vídeos presentes na plataforma, existe uma grande necessidade de apresentar os resultados mais relevantes para cada usuário. Para realizar esta tarefa, o serviço utiliza um complexo sistema de recomendação fornecendo sugestões para os usuários.

O YouTube fornece sugestões majoritariamente em três locais: sua página inicial (<http://www.youtube.com>), a página “Browse” (<http://www.youtube.com/videos>) e sugestões de vídeos relacionados ao vídeo assistido no momento.

No presente trabalho foram coletados vídeos do YouTube de um canal específico. O canal escolhido pertence ao Khan Academy<sup>2</sup>, um site educacional que provê educação de graça para qualquer pessoa.

Foram extraídas informações dos vídeos do canal. Dentre elas estão o título, descrição, quantidade de visualizações e outras. A partir dos dados coletados dos vídeos foram extraídas as informações semânticas associadas a eles.

## 1.2 Motivação

Os sistemas de recomendação estão cada vez mais inseridos no cotidiano das pessoas, podendo muitas vezes passar despercebidos. Com a onipresença das tecnologias, a quantidade de dados é cada vez maior, isso demanda transformação desses dados em informações úteis. Os sistemas de recomendação tornam-se cada vez mais importantes visto a crescente

---

<sup>1</sup><https://www.youtube.com/yt/press/pt-BR/statistics.html>

<sup>2</sup><https://www.khanacademy.org/>

necessidade dessa transformação.

Com o estudo de sistemas de recomendação, é possível adquirir uma percepção melhor da forma como as tecnologias podem influenciar as decisões das pessoas. Esse poder de sugestão é cada vez mais pesquisado pela academia e por instituições públicas ou privadas e, por isso, todo estudo que tenha o intuito de acrescentar informações é válido.

## 1.3 Objetivos

O objetivo geral deste trabalho é apresentar um workflow de recomendação de itens à indivíduos baseada em grafos bipartidos. Esses grafos serão compostos por pesquisadores, vídeos e as informações semânticas extraídas através da análise dos atores e itens.

Os indivíduos serão pesquisadores que possuem trabalhos publicados no repositório de Ciência da Computação denominado DBLP. Os itens recomendados serão vídeos extraídos da plataforma de compartilhamento de vídeos mais difundida na internet, o YouTube.

Como resultado do workflow proposto serão geradas recomendações de itens para indivíduos específicos. Para isso planeja-se extrair informações semânticas da rede social científica composta pelos relacionamentos de coautoria entre os pesquisadores presentes na DBLP. Também pretende-se reconhecer as informações semânticas presentes nos vídeos. A partir dessas informações semânticas pretende-se propor um modelo para identificar as arestas de maior relevância presentes nos grafos. Ao fim desse trabalho espera-se extrair, do grafo gerado, recomendações de itens para indivíduos específicos.

## 1.4 Metodologia

A pesquisa realizada neste trabalho será de natureza aplicada, buscando identificar uma solução para o problema de recomendação baseada em conteúdo. O objetivo deste trabalho é explicativo pois ele visa identificar os fatores que possibilitam a recomendação. Com relação ao procedimento técnico, será realizada uma pesquisa bibliográfica e a abordagem será quantitativa, analisando os atributos dos itens e atores.

Após a pesquisa bibliográfica uma solução para o problema foi desenvolvida e separada em duas etapas principais. A primeira parte compreende a extração da informação nas fontes de dados selecionadas. Em seguida foi necessário armazenar e tratar o conteúdo extraído. Com as características dos objetos de estudo identificadas foi possível realizar uma análise do conteúdo através de extração das informações semânticas. Após a extração das informações semânticas foi possível diferenciar dois grafos bipartidos, entre os objetos de estudo e as informações semânticas identificadas, onde a relação é representada pela relevância entre os objetos e as informações semânticas.

A segunda etapa constrói, a partir dos grafos gerados anteriormente, um grafo semântico, relacionando os objetos de estudo através de uma qualificação das relevâncias entre eles e as informações semânticas identificadas. Após a criação do grafo semântico foi possível extrair os relacionamentos entre os objetos, gerando assim listas de recomendação.

## 1.5 Estrutura do Trabalho

O restante do trabalho está organizado em 5 capítulos. O Capítulo 2 tem como objetivo fundamentar os conceitos utilizados no trabalho descrevendo cada um de maneira sucinta, além disso são apresentados alguns trabalhos que possuem relação com o objetivo deste trabalho.

No Capítulo 3 é apresentada a solução proposta para o problema, solucionado através da metodologia estabelecida anteriormente.

No Capítulo 4 temos a apresentação dos resultados encontrados de acordo com a solução proposta no trabalho, seguido pelo Capítulo 5 que apresenta as considerações finais, possibilitando uma visão geral do trabalho assim como possibilidades de aperfeiçoamento em trabalhos futuros.

## 2 Revisão Bibliográfica

Este capítulo tem como objetivo fundamentar os conceitos utilizados nesse trabalho bem como apresentar alguns trabalhos relacionados ao tema de pesquisa apresentado. O capítulo será dividido em seções que devem agrupar os conhecimentos envolvidos.

### 2.1 Redes Sociais

O assunto de rede sociais pode ser compreendido como o estudo das relações entre indivíduos e as implicações dessas relações (Wasserman, 1994). Para definir corretamente redes sociais alguns conceitos serão expostos a seguir.

Os indivíduos são conhecidos como atores, esses atores podem representar organizações, conjuntos de pessoas ou um único indivíduo. Os atores estão conectados por laços relacionais.

Existem diversos tipos de laços, mas todos tem como objetivo conectar pares de atores. Conjuntos de atores com laços comuns podem ser classificados como grupos. Conjuntos de laços entre membros de um grupo são denominados relações.

Com os conceitos acima, podemos definir redes sociais como um conjunto finito de atores, ou grupos de atores, e as relações definidas entre eles. Diferentes tipos de redes sociais podem ser observadas com finalidades diversas.

As redes sociais podem ser classificadas em dois tipos: Homogêneas e Heterogêneas (Cai *et al.*, 2005). 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. As redes heterogêneas também são conhecidas como redes sociais multi-relacionais.

No presente trabalho iremos utilizar uma rede social científica. Este tipo de rede social refere-se a redes onde os atores são representações de autores e as relações são estabelecidas por trabalhos realizados em co-autoria entre os autores (Goldenberg *et al.*, 2010).

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 Backstrom & Leskovec (2011), 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.

## 2.2 Recomendações

Os Sistemas de Recomendação são uma área de pesquisa importante devido a abundância de aplicações práticas, que ajudam os usuários a lidar com a grande quantidade de dados presentes em diversas plataformas. Esses sistemas fazem recomendações, como conteúdos e serviços, interessantes a cada indivíduo de maneira personalizada (Adomavicius & Tuzhilin, 2005). 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 em Youtube.com<sup>5</sup>, músicas em Spotify<sup>6</sup> ou *web services* para desenvolvedores como proposto por Jiang *et al.* (2013)

De acordo com Adomavicius & Tuzhilin (2005) os sistemas de recomendação vem sendo separados em três categorias: baseada em conteúdo, colaborativa e híbrida. Nas seções a seguir, os principais tipos de filtragem de informação são descritos. De acordo com Belkin & Croft (1992), a filtragem de informação é descrita como a variedade de processos para entrega da informação às pessoas que a necessitam.

### 2.2.1 Filtragem Baseada em Conteúdo

Essa abordagem tem suas raízes nos processos de recuperação da informação (Baeza-Yates *et al.*, 1999). Neste tipo de filtragem é realizada uma descrição dos usuários e itens com base em suas características. Após essa caracterização é realizada uma comparação entre as descrições, verificando as relações entre elas. Nos casos onde usuários e itens possuem as mesmas descrições pode-se inferir que o item é relevante ao usuário (Balabanović &

---

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

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

<sup>5</sup>[www.youtube.com](http://www.youtube.com)

<sup>6</sup>[www.spotify.com](http://www.spotify.com)

Shoham, 1997).

De acordo com Reategui *et al.* (2005), a descrição de um usuário pode ser obtida através de informações fornecidas por ele próprio ou por suas ações. Uma maneira de trabalhar com esse tipo de filtragem é solicitar ao usuário uma análise de um conjunto de itens. Após a avaliação, o sistema considera que os itens relevantes são os que se assemelham aos itens escolhidos pelo usuário, da mesma forma, o sistema considera como não relevantes os itens pelos quais o usuário não demonstrou interesse.

### 2.2.2 Filtragem Colaborativa

A filtragem colaborativa é a técnica mais usada em recomendações como verificado no trabalho de Shi *et al.* (2014). Essa técnica se baseia no pressuposto que usuários que demonstraram interesses similares no passado irão compartilhar interesses em comum no futuro (Goldberg *et al.*, 1992). Essa abordagem de filtragem se diferencia da filtragem baseada em conteúdo por não exigir a descrição dos itens a serem recomendados, pois baseia-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 com sua necessidade. Desta forma, um usuário pode receber recomendações a partir das pontuações fornecidas por usuários similares a ele.

### 2.2.3 Filtragem Híbrida

A abordagem de filtragem híbrida procura combinar as duas técnicas de recomendação descritas anteriormente, com isso essa abordagem visa evitar as limitações das técnicas originais.

Existem diferentes formas de combinar os métodos baseados em conteúdo e colaborativos em um sistema de recomendação híbridos, essas formas são classificadas da seguinte maneira segundo Adomavicius & Tuzhilin (2005):

- 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.

## 2.3 Grafos Bipartidos

Grafos são modelos versáteis que possibilitam a análise de um grande leque de problemas onde, os pontos e suas conexões possuem alguma interpretação, seja ela física ou conceitual.

Como definido em (Gross & Yellen, 2005), um grafo  $G = (V, E)$  é uma estrutura matemática finita composta por dois conjuntos  $V$  e  $E$ . Os elementos de  $V$  são denominados nós, e os elementos  $E$  são denominados arestas. Cada aresta tem um conjunto de um ou dois nós associados. Esses nós são denominados extremidades da aresta.

Ainda segundo Gross & Yellen (2005), um grafo bipartido pode ser definido onde, em um grafo  $G$ , o conjunto de vértices  $V$  pode ser particionado em dois subconjuntos,  $U$  e  $W$ . Neste grafo  $G$ , cada aresta possui uma extremidade em  $U$  e a outra em  $W$ .

Os grafos bipartidos vem sendo muito usados na teoria da codificação da informação, podendo enunciar o grafo de Tanner como exemplo de utilização. Neste trabalho eles serão utilizados para modelar a relação semântica entre pessoas e objetos. Outros exemplos de uso de grafos bipartidos podem ser encontrados no Seção 2.4.

## 2.4 Trabalhos Relacionados

Como verificado por Shi *et al.* (2014) as abordagens baseadas em grafos vem sendo bem estudadas e muito utilizadas, principalmente na predição de links em redes sociais. Um exemplo típico é o uso de passeios aleatórios e suas variâncias (Tong *et al.*, 2006).

Diversos trabalhos atacam o problema de recomendações a partir de uma abor-

dagem que utiliza grafos bipartidos. Essa abordagem pode ser aplicada com diferentes propósitos. O trabalho realizado por Kunegis *et al.* (2010) tenta especializar algoritmos de recomendação para o caso de grafos bipartidos. Devido às características dos grafos bipartidos, técnicas de recomendação não podem ser generalizadas sem algum tipo de especialização. 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 baseam-se nos vizinhos imediatos, somente o modelo *preferential attachment* pode ser utilizado em redes bipartidas.

Alguns trabalhos também utilizam a ideia de processos difusos em redes bipartidas. No trabalho realizado por Liu *et al.* (2010), um algoritmo foi apresentado modificando outros algoritmos de recomendação para obter as relações usuário-usuário através de um processo difuso. Já no trabalho executado por Ju & Xu (2014) um algoritmo baseado no processo de difusão heterogênea é criado para ser utilizado em um grafo bipartido de usuários e objetos. No algoritmo proposto, a disseminação dos recursos entre os nós é influenciada pelo grau de interesse dos usuários e objetos.

Outros trabalhos utilizam a ideia de arestas com peso em grafos bipartidos, no trabalho realizado por Sawant (2013), as arestas possuem peso referente à qualificação. Neste 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. Eles concluíram que as predições de recomendação são afetadas de maneira considerável pela acurácia da função de similaridade entre as entidades. Em (Ting *et al.*, 2013) o conceito de peso nas arestas para demonstrar o interesse entre usuários e objetos também é utilizado. O objetivo é aumentar a viabilidade do uso de algoritmos para recomendação em grafos bipartidos.

A abordagem de redes semânticas também pode ser vista no trabalho de Maehara *et al.* (2012) onde o objetivo é recomendar itens de exibição em museus. A rede semântica é construída a partir das características dos itens possibilitando os usuários a entender o contexto que engloba o item, despertando o interesse dos usuários. Em (Shieh *et al.*, 2008), 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, reduzindo a grafos unidimensionais onde as arestas possuem um peso.

## 3 Solução Proposta

Neste capítulo serão apresentadas as etapas que compõe a solução para o problema de recomendações com filtragem baseada em conteúdo. A solução visa apresentar o modelo de recomendação proposto neste trabalho.

### 3.1 Workflow

O workflow implementado segue a apresentada na Seção 1.4. Na Figura 3.1 podemos ter uma visão geral do workflow que representa a solução proposta neste trabalho. O primeiro passo é realizar a coleta dos dados nas fontes de informação selecionadas, nesta fase também são realizados tratamentos nos dados extraídos com o objetivo de aumentar a eficácia do analisador de conteúdo.

Após a extração e tratamento dos dados é possível fornecer esses dados tratados ao analisador de conteúdo. Depois da análise do conteúdo ter sido realizada é possível identificar as informações semânticas extraídas dos dados fornecidos ao analisador de conteúdo, essas informações estão relacionadas aos atores e itens. Em seguida o quantificador de relacionamentos é responsável por qualificar o relacionamento entre os atores e itens a partir das informações semânticas identificadas anteriormente. A partir do grafo obtido pela relação entre os atores e itens obtidas no quantificador de relacionamentos é possível realizar as recomendações. De posse do grafo semântico o recomendador deve ordenar os itens que apresentam mais importância ao ator desejado e fornecer essa informação a ele.

As próximas seções deste capítulo explicam cada etapa implementada com um detalhamento maior.

### 3.2 Fontes de Informação

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

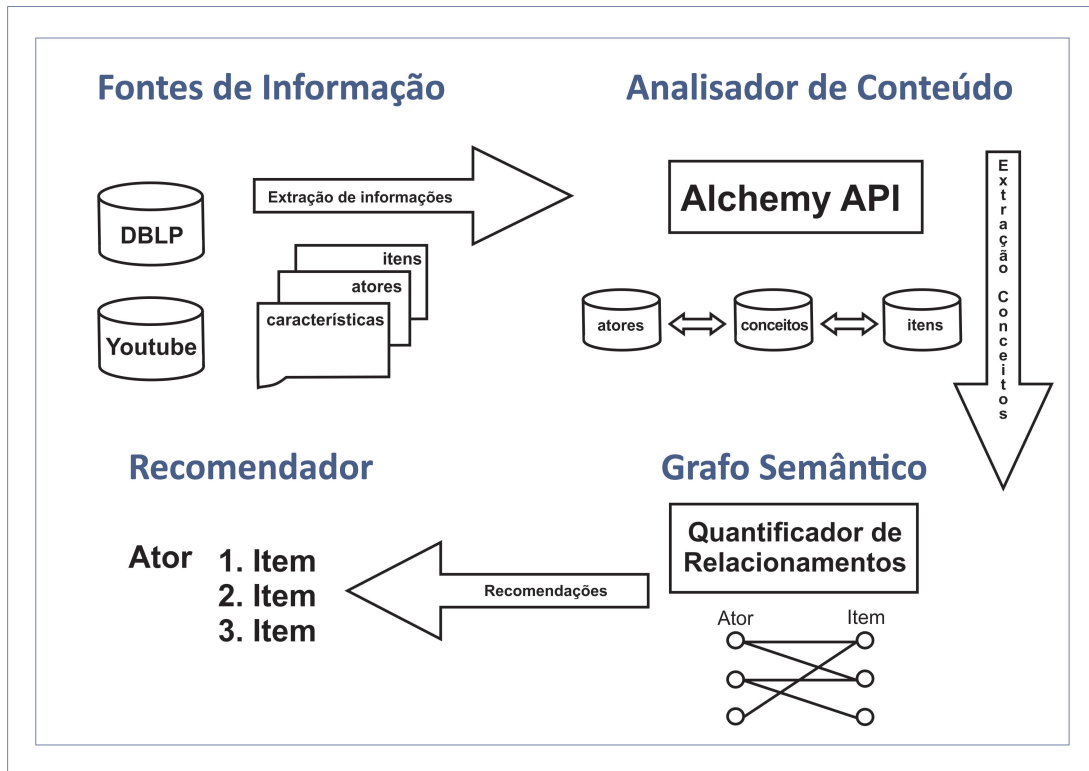


Figura 3.1: Workflow implementado

ou grupos de usuários. Neste trabalho serão usadas como fontes de informações a DBLP e YouTube. A fonte de informação DBLP foi selecionada pois a partir dos dados disponibilizados pelo serviço, é possível identificar uma rede social científica. Esta rede contém os atores que terão suas informações semânticas extraídas em uma fase posterior. A outra fonte de informação utilizada foi o YouTube, por conter uma vasta coleção de objetos possibilitando a recomendação de uma grande quantidade de itens aos atores selecionados. As subseções a seguir descrevem os passos para a extração das informações utilizadas.

### 3.2.1 Base de Pesquisadores

Como dito anteriormente, neste trabalho serão extraídas informações de redes sociais científicas. A DBLP possui uma vasta coleção de dados com diversos tipos de informação relativas às publicações da área de ciência da computação, como artigos, *journals*, livros e outros.

Os dados estão disponíveis no formato XML e foram mapeados para um banco de dados relacional. Esse mapeamento foi realizado por Duque (2015) que realizou a extração utilizando a biblioteca *JAVA SAXParser* permitindo a análise das diferentes *tags*

encontradas para realizar a identificação da estrutura do documento. Após a interpretação das *tags* um modelo entidade-relacionamento foi realizado. Este modelo pode ser visto na Figura 3.2.

Em uma etapa posterior ao processamento das informações do arquivo foi realizada a carga em um banco de dados *PostgreSQL* que foi utilizado para a extração dos dados relevantes a este trabalho. Os autores analisados no presente trabalho foram os que possuíram ao menos 5 publicações nos últimos 4 anos.

Essa base tem como objetivo fornecer dados ao analisador de conteúdo. Após a atuação do analisador de conteúdo sobre os dados extraídos dessa base será possível obter um grafo que relaciona os autores identificados na base com informações semânticas

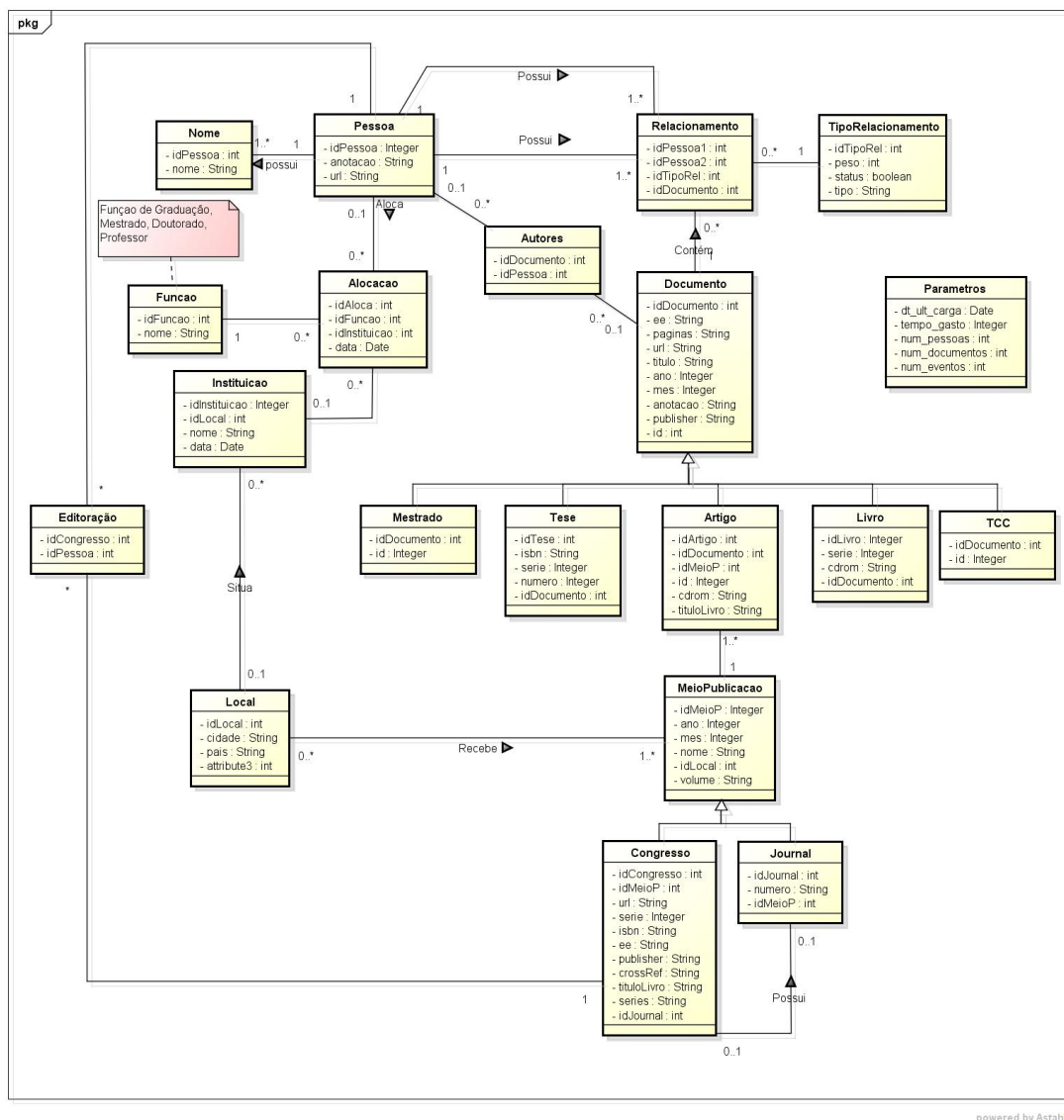


Figura 3.2: Modelo de banco de dados da DBLP por Duque (2015)

extraídas pelo analisador de conteúdo.

### 3.2.2 Base de Vídeos

De acordo com as seções anteriores, os objetos a serem recomendados são vídeos e para isso um repositório com esses objetos fez-se necessário. A base de vídeos escolhida nesse trabalho para coleta dos objetos foi o YouTube. O conteúdo do YouTube pode ser coletado a partir de varreduras no site ou atualizando a *YouTube Data API*<sup>7</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 API. Os atributos presentes na tabela 3.1 foram selecionados entre os dados disponíveis.

Tabela 3.1: 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 conteúdo do vídeo
view_count	<i>unsigned long</i>	Quantidade de visualizações
like_count	<i>unsigned long</i>	Quantidade de likes
dislike_count	<i>unsigned long</i>	Quantidade de dislikes
published_at	<i>datetime</i>	Data de publicação

Com o propósito de realizar a coleta foi necessário criar uma conta do Google<sup>8</sup> para acessar o Console de APIs. Após receber o acesso ao console de desenvolvedores é necessário criar um projeto. Com o projeto criado 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*, esta 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, esta chave 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 coleta, as respostas da API são transmitidas no formato JSON e convertidas para classes JAVA. O primeiro passo para realizar a extração dos dados foi a identificação do canal

<sup>7</sup><https://developers.google.com/youtube/v3/>

<sup>8</sup>[www.google.com](http://www.google.com)

escolhido. Como nem todos os canais do YouTube têm um nome de usuário exclusivo é necessário identificar de forma única o canal desejado. Uma requisição a API foi realizada para identificação do canal e retornou alguns identificadores referentes à chave de busca utilizada. A chave utilizada foi "Khan Academy", além da chave de busca foi definido que seriam retornados os id's e quantidade de vídeos de todos os objetos do tipo "channel". Essa requisição retornou 10 resultados, o canal foi identificado de acordo com a quantidade de vídeos carregados.

Após a identificação do canal foi necessário coletar as informações de cada *playlist* para colher os vídeos do canal identificado de forma única no passo anterior. Feito isso foi possível extrair os identificadores de cada *playlists* do canal. Com as *playlists* identificadas foi possível identificar os vídeos presentes em cada uma. Algumas *playlists* possuíam mais itens que o valor máximo de resposta em uma requisição. Para coletar todos os itens das *playlists* foi necessário utilizar um *PageToken*. O *PageToken* identifica uma página específica no conjunto de resultados que será retornado. Desta forma, após a coleta de todos os dados fornecidos pela requisição era realizada uma nova requisição passando o parâmetro *PageToken* referenciando a próxima página. Isso foi realizado até que todas as páginas fossem visitadas e seus dados coletados.

Finalmente, após os vídeos terem seus identificadores foi possível realizar as requisições a API para coleta dos atributos selecionados. Estes atributos estão descritos na tabela 3.1. A coleta foi realizada de forma individual para cada vídeo. Os dados coletados foram transcritos para um banco de dados *PostgreSQL*.

### 3.3 Análise do Conteúdo

Para a análise do conteúdo e extração de informações semânticas foi utilizada a *AlchemyAPI*<sup>9</sup>. Essa API realiza uma análise de textos, estruturando o texto e tornando-o legível aos computadores. Essa API é capaz de ler e entender rapidamente textos com grande velocidade, utilizando algoritmos linguísticos, estatísticos e de redes neurais. A API oferece 12 funções como parte da análise de texto, essas funções podem ser vistas na Tabela 3.2.

---

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



Função	Descrição
<i>Entity Extraction</i>	Identifica as entidades discutidas explicitamente no texto
<i>Sentiment Analysis</i>	Identifica se as palavras carregam conotação positiva ou negativa
<i>Keyword Extraction</i>	Identifica palavras chaves no texto analisado
<i>Concept Tagging</i>	Identifica ideias abstratas discutidas no texto, mesmo quando não discutidas explicitamente no texto
<i>Relation Extraction</i>	Identifica os sujeitos, verbos e objetos nos textos analisados
<i>Taxonomy Classification</i>	Realiza uma classificação do texto em uma hierarquia taxonômica
<i>Author Extraction</i>	Identifica o autor do texto analisado
<i>Language Detection</i>	Identifica o idioma do texto analisado
<i>Text Extraction</i>	Identifica informações relevantes de sites removendo códigos HTML
<i>Microformat Parsing</i>	Adicionam informações semânticas a sites facilitando a categorização e indexação
<i>Feed Detection</i>	Identifica feeds em páginas retornando links.
Linked Data	Habilita conteúdos a serem inseridos na web semântica

Tabela 3.2: Funções da AlchemyAPI.

Essa API foi aplicada para a extração de informações semânticas relacionadas aos vídeos e pesquisadores. Ela realiza análises semânticas de texto, permitindo a extração de conjuntos de conceitos relevantes. Os conceitos identificados são as ideias discutidas no texto analisado, independente se elas são mencionadas explicitamente ou não.

Além dos conceitos de cada vídeo e autor, um coeficiente de relevância foi gerado indicando a importância de cada conceito para um dado vídeo ou autor. De acordo com estudos realizados por Saif *et al.* (2012) e Rizzo & Troncy (2011) foi observado que os resultados da extração de conceitos apresentados pela plataforma são bons quando comparados a outras soluções, fundamentando o uso neste trabalho.

Para os vídeos, a extração de conceitos ocorreu a partir da análise do título do vídeo e de sua descrição. A especificidade de um conceito para um dado vídeo foi determinada pela *AlchemyAPI*. Desta forma, foram armazenados os pares compostos por conceitos e vídeos, assim como sua relevância.

A identificação dos conceitos associados aos autores ocorreu de forma análoga a realizada para extração dos conceitos dos vídeos. No caso da identificação dos conceitos associados aos atores, a *AlchemyAPI* realizou a análise de todos os títulos de documentos que possuem o ator como autor do documento.

Após a identificação dos conceitos foram construídos dois grafos bipartidos. O primeiro possui dois tipos de nós que representam conceitos e autores, o segundo grafo possui nós representando vídeos e conceitos. A aresta dos grafos gerados é 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.

### 3.4 Construção do Grafo Semântico

O grafo semântico foi contruído utilizando os dois grafos bipartidos gerados a partir da análise do conteúdo como pode ser visto na Figura 3.3. Os nós do grafo semântico representam os atores e itens e as arestas possuem a relevância entre os nós. A Equação 1 foi utilizada para representar a relevância entre dois nós atuando como um quantificador dos relacionamentos.

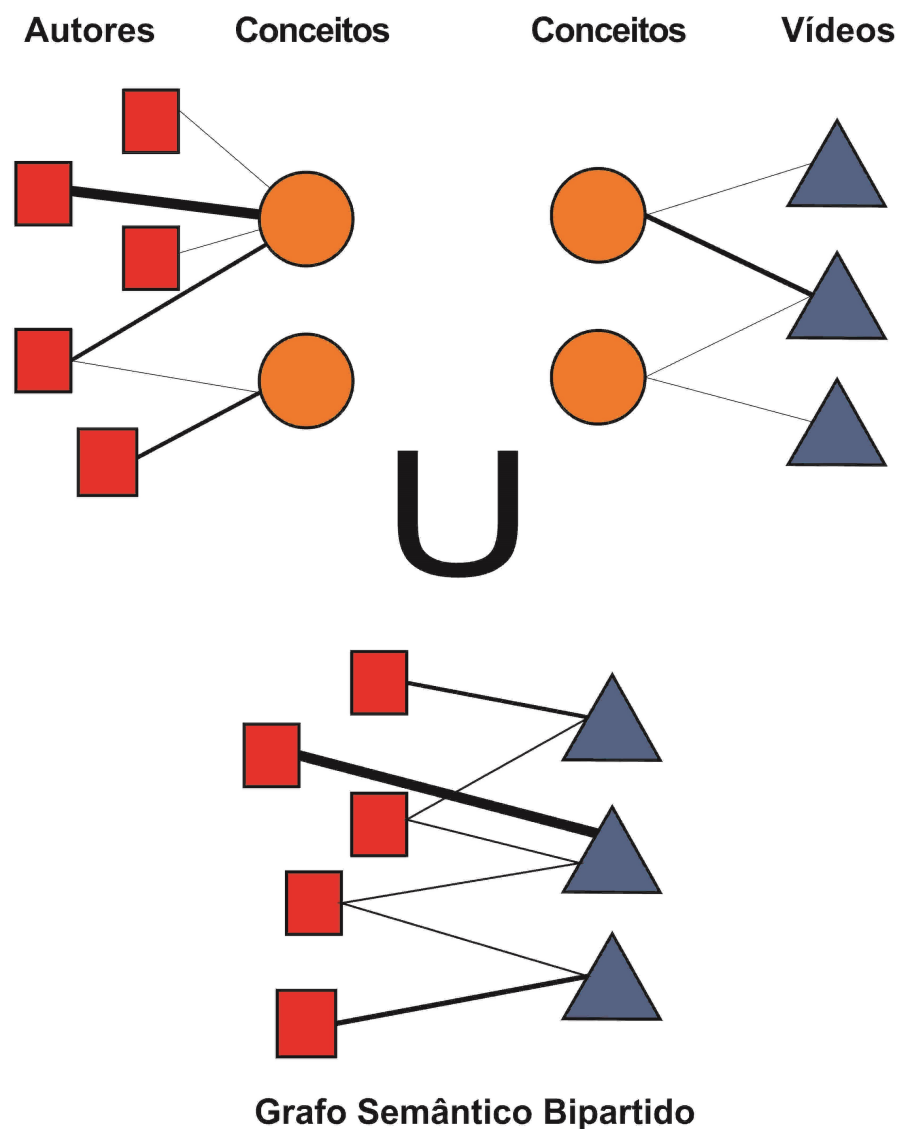


Figura 3.3: Construção do grafo semântico.

$$\frac{\sum_{i=1}^n RCCA_i + RCCI_i}{1 + \sum_{j=1}^o RCDA_j + \sum_{k=1}^p RCDI_k} \quad (1)$$

A parte superior da equação soma as relevâncias referentes aos conceitos presentes nos dois nós. O denominador da equação é a soma de todos os conceitos que apenas um dos nós 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 os 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.

Para exemplificar podemos observar a Figura 3.4. Na figura temos um ator com os conceitos *Mathmatics* com relevância de 0,82 e *Logic* com relevância de 0,69. Deseja-se quantificar a relevância entre este autor e um vídeo com os conceitos *Mathmatics*, *Logic* e *Algorithm* com as relevâncias de 0,7, 0,73 e 0,69, respectivamente. Substituindo os valores na Equação 1 temos a Equação 2 onde o resultado representa o peso da aresta do grafo semântico que liga esses dois nós.

$$\frac{0,82 + 0,7 + 0,69 + 0,73}{1 + 0,69} = 1,74 \quad (2)$$

Como um outro exemplo, deseja-se identificar o peso da aresta no grafo semântico entre o mesmo vídeo e um autor com os conceitos *Mathmatics*, *Logic* e *Group* com as relevâncias de 0,85, 0,72 e 0,62, respectivamente. Substituindo os valores na Equação 1 temos a Equação 3

$$\frac{0,85 + 0,7 + 0,72 + 0,73}{1 + 0,62 + 0,69} = 1,30 \quad (3)$$

Como podemos observar, mesmo com as relevâncias dos conceitos comuns do segundo autor tendo valores mais próximos de um a aresta que liga este autor ao vídeo teve um valor menor, visto que a soma dos conceitos não comuns penalizou a relação.

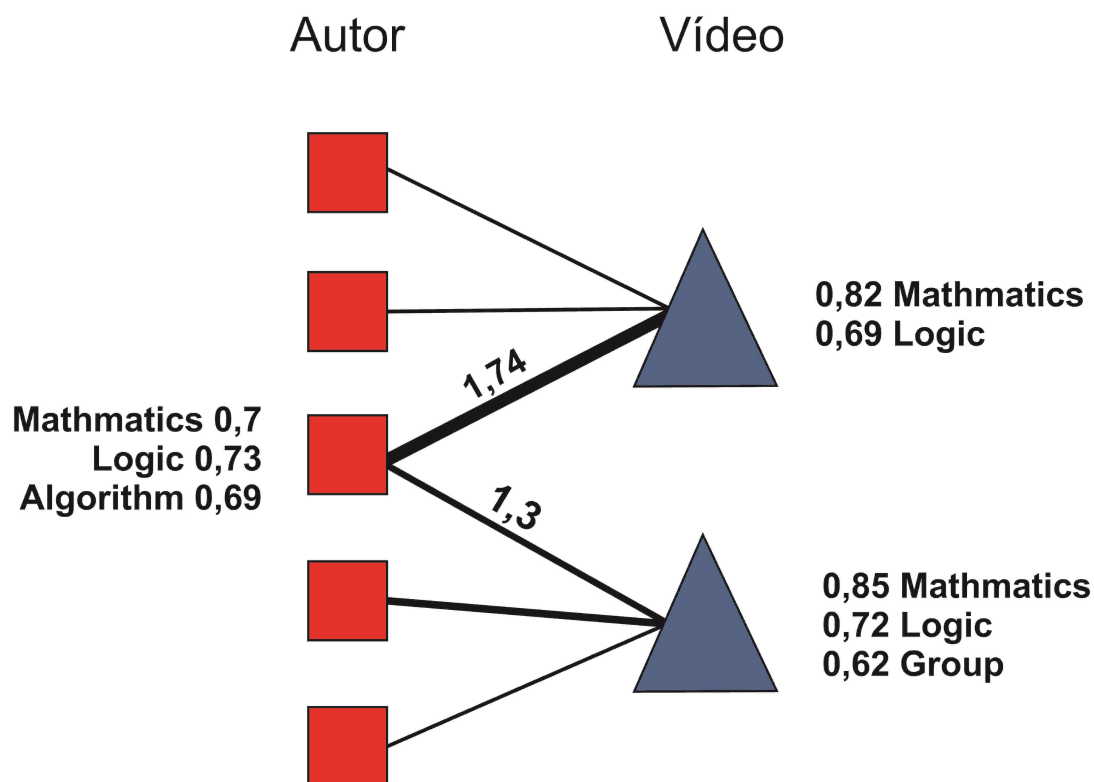


Figura 3.4: Grafo semântico.

### 3.5 Recomendação

Como dito anteriormente, o modelo de recomendação utilizado neste trabalho foi o baseado em conteúdo. O conteúdo dos atores e itens foi utilizado em uma análise semântica com a finalidade de representar o perfil dos atores e itens, que neste trabalho são, respectivamente, autores e vídeos.

De acordo com este perfil identificado, a Equação 1 indica a aderência existente entre os autores e objetos. A lista de recomendação foi gerada analisando os valores das arestas do grafo semântico obtidas através da Equação 1 ordenados de forma decrescente, visto que os maiores valores indicam a maior aderência entre um autor e um vídeo. Nos

---

casos onde ocorreu empate entre os valores de duas arestas diferentes foi utilizado como critério de desempate a diferença entre a quantidade de *likes* e *dislikes* do item em questão. Alguns resultados da recomendação serão apresentados no próximo capítulo.

## 4 Resultados

Após o desenvolvimento e execução da solução proposta foram identificados os dois grafos gerados no processo descrito na Seção 3.3. A dimensão dos grafos construídos pode ser vista na Tabela 4.1.

	Instâncias Avaliadas	Conceito Identificado	Relacionamentos
<b>Autor</b>	86425	19785	615898
<b>Vídeo</b>	3548	2553	10547

Tabela 4.1: Informações quantitativas dos grafos bipartidos.

Para a construção do grafo semântico, nem todos os nós atenderam os requisitos. Mesmo assim, a quantidade de autores e vídeos que compõem o grafo contemplou a maioria dos atores e itens classificados. Os gráficos apresentados na Figura 4.1 mostram a proporção dos autores e vídeos que estão presentes no grafo semântico. Esses nós se relacionaram através de 1598 conceitos.

O grafo semântico tem o formato do grafo apresentado na Figura 3.4. Os nós representados por triângulos azuis são autores e os quadrados vermelhos são vídeos. A largura da aresta simboliza a relevância entre os nós. Quanto maior a largura da aresta, maior a relevância do vídeo ao autor. A relevância foi calculada como descrito na Equação 1.

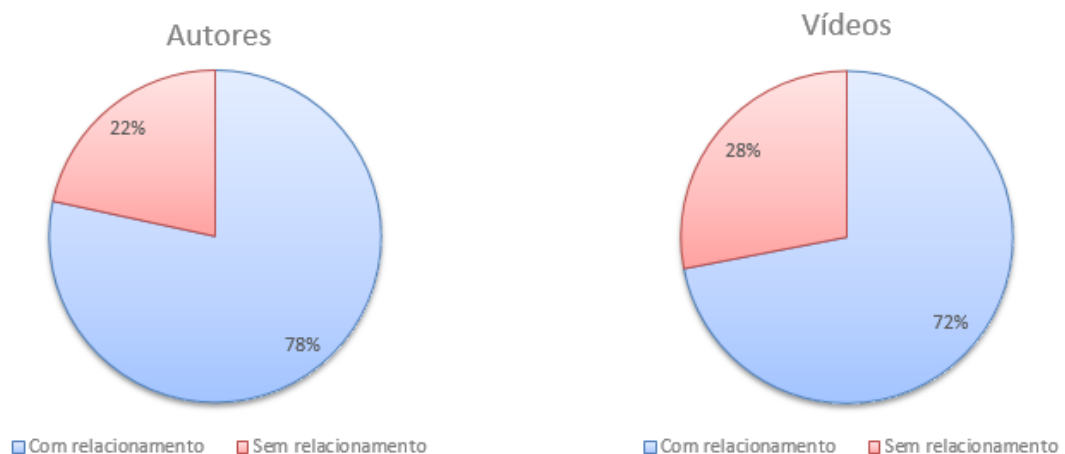


Figura 4.1: Autores e vídeos presentes no grafo semântico.

A Tabela 4.2 mostra os conceitos que são compartilhados pelo maior número de autores e vídeos. Podemos ver que a quantidade dos conceitos mais identificados para os

autores foi bem maior, isto aconteceu pois o conjunto de instâncias de autores analisados foi bem maior, como pode ser visto na Tabela 4.1.

<b>Vídeo</b>		<b>Autor</b>	
<b>Nome do conceito</b>	<b>Quantidade</b>	<b>Nome do conceito</b>	<b>Quantidade</b>
Vector space	235	Mathematics	8448
Mathematics	202	Algorithm	8413
Group	144	Logic	4215
Calculus	126	Artificial intelligence	3979
Integral	122	Computer	3725

Tabela 4.2: Conceitos mais comuns.

A seguir seguem dados coletados de dois autores presentes no grafo semântico obtido através do workflow desenvolvido neste trabalho. A Tabela 4.3 mostra todos os conceitos identificados para os autores 22 e 56. Os conceitos apresentados são apenas aqueles em que existe algum item que também possui o conceito.

<b>Autor</b>	<b>Conceito</b>	<b>Relevância</b>
22	Logic	0.778654
22	Reasoning	0.719578
22	Problem solving	0.689492
56	Knowledge	0.806061
56	Linguistics	0.6453
56	Psychology	0.624365
56	Mind	0.61308

Tabela 4.3: Conceitos identificados para os autores 22 e 56.

Na Tabela 4.4 podemos ver os conceitos associados a alguns vídeos que se relacionam com o autor 56. Da mesma forma, a Tabela 4.5 mostra os conceitos identificados em alguns vídeos associados ao autor 22.

<b>Vídeo</b>	<b>Conceito</b>	<b>Relevância</b>
4xfOq00BzJA	Mind	0.895686
l3OkPYhDi9w	Mind	0.862412
LUtcigWSBsw	Psychology	0.86184
W4JfDtfwOug	Linguistics	0.845194
69rDtSpshAw	Mind	0.838456
Tx7Cwn5pgQo	Mind	0.837225

Tabela 4.4: Conceitos identificados para os vídeos relacionados ao autor 56.

Com as informações referentes à relevância dos conceitos aos autores podemos construir um grafo bipartido. Da mesma forma, podemos construir um grafo bipartido

que representa os vídeos e conceitos. Nos dois grafos construídos o peso das arestas representa a relevância encontrada pelo Analisador de Conteúdo, descrito na Seção 3.3.

A Tabela 4.6 representa a lista de recomendação para o autor 56 e para o autor 22. Essas listas de recomendação foram concebidas a partir da construção do grafo semântico, que representa a união dos dois grafos bipartidos construídos.

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

Tabela 4.5: Conceitos identificados para os vídeos relacionados ao autor 22.

Autor	Vídeo	Relevância	Autor	Vídeo	Relevância
22	ervHbKa7R5g	0.646682	56	4xfOq00BzJA	0.490540
22	Zyq6TmQVBxk	0.634601	56	W4JfDtfwOug	0.489729
22	kpqSeXpiM6k	0.634600	56	LUtcigWSBsw	0.484984
22	hq1bUM2tyg0	0.622479	56	l3OkPYhDi9w	0.479722
22	DIKkujAIeTY	0.592363	56	69rDtSpshAw	0.471933
22	SpDIXJ2I2D4	0.553688	56	Tx7Cwn5pgQo	0.471533

Tabela 4.6: Listas de recomendação para os autores 22 e 56.



## 5 Considerações Finais

O objetivo deste trabalho foi desenvolver um workflow de recomendação baseada em informações semânticas e a partir do workflow proposto gerar recomendações. Para isso, foi realizada uma pesquisa bibliográfica afim de fundamentar a abordagem da solução proposta e identificar os trabalhos relacionados.

Após a modelagem, foi realizada a extração dos dados e, de posse dos dados, foram extraídos conceitos com o objetivo de descrever os itens 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.

Uma das maiores dificuldades na realização do trabalho foi a extração dos dados dos itens, a API utilizada pra extração possuía um alto nível de complexidade e a documentação não era tão completa como esperado. Além disso, no período em que a extração foi realizada ocorreu a migração entre duas versões fazendo com que a documentação ficasse desatualizada e modificando os métodos utilizados.

Outro problema encontrado foi o volume dos dados a serem trabalhados. Para solucionar esse problema um escopo de atores a serem atacados foi definido assim como um processo de otimização nas consultas utilizadas.

Como trabalhos futuros tem-se a modificação do workflow proposto adicionando o *feedback* fornecido pelo usuário após a recomendação. A partir do *feedback* fornecido pode-se utilizar essa informação para atualização dos relacionamentos similares ao avaliado. Também planeja-se realizar a validação do workflow desenvolvido, essa validação poderá ser realizada de forma qualitativa, através da identificação dos conceitos referentes aos colaboradores e, a partir dessa identificação das informações semânticas realizar a recomendação de itens e avaliar os resultados obtidos através da análise das informações fornecidas pelos colaboradores através de questionários. Outra forma para realizar a validação é aplicá-la a cenários já avaliados por outras soluções e realizar a comparação dos

---

resultados encontrados.

Outra oportunidade para trabalhos futuros é abordar o uso de ontologias na etapa de análise do conteúdo afim de agrupar os conceitos dentro de um domínio, podendo assim modificar a qualificação entre os relacionamentos e alterar a lista de recomendação.

## Referências

- Adomavicius, G.; 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**, v.17, n.6, p. 734–749, 2005.
- Amatriain, X. **Big personal: data and models behind netflix recommendations**. In: Proceedings of the 2nd International Workshop on Big Data, Streams and Heterogeneous Source Mining: Algorithms, Systems, Programming Models and Applications, p. 1–6. ACM, 2013.
- Backstrom, L.; 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, p. 635–644. ACM, 2011.
- Baeza-Yates, R.; Ribeiro-Neto, B. ; others. **Modern information retrieval**, volume 463. ACM press New York, 1999.
- Balabanović, M.; Shoham, Y. Fab: content-based, collaborative recommendation. **Communications of the ACM**, v.40, n.3, p. 66–72, 1997.
- Belkin, N. J.; Croft, W. B. Information filtering and information retrieval: Two sides of the same coin? **Communications of the ACM**, v.35, n.12, p. 29–38, 1992.
- Cai, D.; Shao, Z.; He, X.; Yan, X. ; Han, J. **Community mining from multi-relational networks**. In: Knowledge Discovery in Databases: PKDD 2005, p. 445–452. Springer, 2005.
- Duque, A. d. P. Extração de dados científicos para construção e análise de redes sociais. **Departamento de Ciência da Computação - UFJF**, 2015.
- Goldberg, D.; Nichols, D.; Oki, B. M. ; Terry, D. Using collaborative filtering to weave an information tapestry. **Communications of the ACM**, v.35, n.12, p. 61–70, 1992.

- Goldenberg, J.; Libai, B.; Muller, E. ; Stremersch, S. Database submission-the evolving social network of marketing scholars. **Marketing Science**, v.29, n.3, p. 561–567, 2010.
- Gross, J. L.; Yellen, J. **Graph theory and its applications**. CRC press, 2005.
- Hallinan, B.; Striphas, T. Recommended for you: The netflix prize and the production of algorithmic culture. **new media & society**, p. 1461444814538646, 2014.
- Han, J.; Kamber, M. **Data Mining, Southeast Asia Edition: Concepts and Techniques**. Morgan kaufmann, 2006.
- Hansen, D.; Shneiderman, B. ; Smith, M. A. **Analyzing social media networks with NodeXL: Insights from a connected world**. Morgan Kaufmann, 2010.
- Jiang, B.; Zhang, X.-x.; Pan, W.-f. ; Hu, B. **Bigsir: A bipartite graph based service recommendation method**. In: Services (SERVICES), 203 IEEE Ninth World Congress on, p. 363–369. IEEE, 2013.
- Ju, C.; Xu, C. Personal recommendation via heterogeneous diffusion on bipartite network. **International Journal on Artificial Intelligence Tools**, v.23, n.03, 2014.
- Kunegis, J.; De Luca, E. W. ; Albayrak, S. **The link prediction problem in bipartite networks**. In: Computational intelligence for knowledge-based systems design, p. 380–389. Springer, 2010.
- Ley, M.; Reuther, P. **Maintaining an online bibliographical database: The problem of data quality**. In: EGC, p. 5–10, 2006.
- Liu, J.-G.; Zhou, T.; Che, H.-A.; Wang, B.-H. ; Zhang, Y.-C. Effects of high-order correlations on personalized recommendations for bipartite networks. **Physica A: Statistical Mechanics and its Applications**, v.389, n.4, p. 881–886, 2010.
- Maehara, C.; Yatsugi, K.; Kim, D. ; Ushiyama, T. **An exhibit recommendation system based on semantic networks for museum**. In: Innovations in Intelligent Machines–2, p. 131–141. Springer, 2012.

- Oestreicher-Singer, G.; Sundararajan, A. Recommendation networks and the long tail of electronic commerce. **Available at SSRN 1324064**, 2010.
- Reategui, E. B.; Cazella, S. C. **Sistemas de recomendação**. In: XXV Congresso da Sociedade Brasileira de Computação. Universidade do Vale do Rio dos Sinos (UNISINOS). São Leopoldo. Citeseer, 2005.
- Rizzo, G.; Troncy, R. Nerd: evaluating named entity recognition tools in the web of data. 2011.
- Saif, H.; He, Y. ; Alani, H. **Semantic sentiment analysis of twitter**. In: The Semantic Web–ISWC 2012, p. 508–524. Springer, 2012.
- Sawant, S. Collaborative filtering using weighted bipartite graph projection: a recommendation system for yelp. **CS224W: Social and Information Network Analysis (December 10, 2013)**, 2013.
- Shieh, J.-R.; Yeh, Y.-T.; Lin, C.-H.; Lin, C.-Y. ; Wu, J.-L. **Collaborative knowledge semantic graph image search**. In: Proceedings of the 17th international conference on World Wide Web, p. 1055–1056. ACM, 2008.
- Shi, Y.; Larson, M. ; Hanjalic, A. Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. **ACM Computing Surveys (CSUR)**, v.47, n.1, p. 3, 2014.
- Ting, Y.; Yan, C. ; 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, p. 587–590. IEEE, 2013.
- Tong, H.; Faloutsos, C. ; Pan, J.-Y. Fast random walk with restart and its applications. 2006.
- Wasserman, S. **Social network analysis: Methods and applications**, volume 8. Cambridge university press, 1994.

---

Zhang, G.-Q.; Zhang, G.-Q.; Yang, Q.-F.; Cheng, S.-Q. ; Zhou, T. Evolution of the internet and its cores. **New Journal of Physics**, v.10, n.12, p. 123027, 2008.