



Dissertação de Mestrado

**Sistema de Recomendação de Vídeos Educacionais:
Um Estudo de Caso no YouTube**

Roberth Raphael Araújo Pinheiro

Orientadores:

Evandro de Barros Costa

Patrick Henrique da Silva Brito

Maceió, outubro de 2018

Robert Raphael Araújo Pinheiro

Sistema de Recomendação de Vídeos Educacionais: Um Estudo de Caso no YouTube

Dissertação de mestrado apresentada ao Programa de Pós-Graduação em Modelagem Computacional de Conhecimento do Instituto de Computação da Universidade Federal de Alagoas.

Orientadores:

Evandro de Barros Costa
Patrick Henrique da Silva Brito

Maceió, outubro de 2018

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário: Marcelino de Carvalho

P654s Pinheiro, Roberth Raphael Araújo.
Sistema de recomendação de vídeos educacionais : um estudo de caso no YouTube
/ Roberth Raphael Araújo Pinheiro. – 2019.
59 f. : il.

Orientador: Evandro de Barros Costa.

Orientador: Patrick Henrique da Silva Brito.

Dissertação (mestrado em Modelagem Computacional de Conhecimento) –
Universidade Federal de Alagoas. Instituto de Computação. Maceió, 2018.

Bibliografia: f. 50-52.

Apêndices: f. 53-59.

1. Modelagem computacional. 2. Vídeo educativo - Recomendação. 4. YouTube
(Recursos eletrônico). I. Título.

CDU: 004.414.23



UNIVERSIDADE FEDERAL DE ALAGOAS/UFAL
Programa de Pós-Graduação em Modelagem Computacional de Conhecimento
Avenida Lourival Melo Mota, Km 14, Bloco 12, Cidade Universitária
CEP 57.072-900 – Maceió – AL – Brasil
Telefone: (082) 3214-1364/1825



Membros da Comissão Julgadora da Dissertação de Mestrado de Roberth Raphael Araújo Pinheiro, intitulada: “Sistema de Recomendação de Vídeos Educacionais: Um Estudo de Caso no YouTube”, apresentada ao Programa de Pós-Graduação em Modelagem Computacional de Conhecimento da Universidade Federal de Alagoas, em 11 de outubro de 2018, às 09h00min, na sala 15 do Instituto de Computação da Ufal.

COMISSÃO JULGADORA

Prof. Dr. Evandro de Barros Costa

Ufal – Instituto de Computação

Orientador

Prof. Dr. Patrick Henrique da Silva Brito

Ufal – Instituto de Computação

Orientador

Prof. Dr. Arturo Hernández-Domínguez

Ufal – Instituto de Computação

Examinador

Prof. Dr. Cleyton Caetano de Souza

IFPB – Instituto Federal de Educação

Examinador

Maceió, outubro de 2018.

Resumo

Sistemas de Recomendação vêm sendo tema de pesquisas com aplicações em diversas áreas, tais como: comércio, saúde e educação. Aqueles voltados para a educação permitem a recomendação de livros, artigos e até vídeos educacionais, levando em consideração os interesses dos usuários. A quantidade de vídeos educacionais disponíveis na Internet vêm crescendo com o passar do tempo. Tanto as iniciativas privadas como as autônomas disponibilizam, na rede de computadores mundialmente interligada, vídeos sobre diversas áreas do conhecimento. No YouTube, por exemplo, é possível encontrar vários canais educativos, com níveis e temas diversificados. No entanto, com a facilidade de disponibilização desses vídeos, tem-se uma sobrecarga de informações, tornando a escolha dos usuários difícil e árdua. Assim, faz-se necessária a criação de ferramentas que auxiliem na escolha desses vídeos educacionais, levando em consideração aspectos significativos para cada usuário. Este trabalho propõe uma apresentação tanto das funcionalidades como das técnicas, de forma detalhada, utilizadas para a recomendação de vídeos educacionais. Para a prova de conceito, utilizou-se o site do YouTube como repositório de vídeos, visto que este é um dos principais sites de vídeos da Internet.

Palavras-chave: Sistemas de Recomendação, Vídeos Educacionais.

Abstract

Systems of Recommendation have been the subject of researches with applications in several areas, such as: commerce, health and education. Those focused on education allow the recommendation of books, articles and educational videos, taking into account the interests of users. The amount of educational videos available on the Internet has been growing over time. Both private and autonomous initiatives provide videos in the world-wide network of computers with videos on various areas of knowledge. On YouTube, for example, you can find various educational channels, with varying levels and themes. However, with the ease of making these videos available, there is an overload of information, making the choice of users difficult and arduous. Thus, it is necessary to create tools that help in the selection of these educational videos, taking into account significant aspects for each user. This work proposes a presentation of both the functionalities and the techniques, in a detailed way, used for the recommendation of educational videos. For proof of concept, the YouTube site was used as a repository of videos, as this is one of the main video sites on the Internet.

Keywords: Recommendation Systems, Educational Videos.

Conteúdo

Lista de Figuras	5
Lista de Tabelas.....	6
1. Introdução	7
1.1. Motivação e Contextualização da Pesquisa	7
1.2. Abordagem da Pesquisa	9
1.2.1. Questões da Pesquisa.....	9
1.2.2. Objetivos da Pesquisa.....	9
1.2.3. Aspectos Metodológicos	10
1.3. Estrutura da Dissertação.....	11
2. Fundamentação Teórica.....	12
2.1. Sistemas de Recomendação	12
3. Trabalhos Relacionados	15
3.1. A Health Information Recommender System.....	15
3.2. The YouTube Video Recommendation System.....	15
3.3. A Predictive Model for Video Lectures Classification	15
3.4. YouTube Recommender Network (YRN)	16
3.5. VCSR System.....	16
3.6. Comparação entre os Sistemas de Recomendação.....	17
4. Sistema de Recomendação de Vídeos Educacionais: Um Estudo de Caso no YouTube.....	19
4.1. Visão Geral.....	19
4.2. Requisitos e Arquitetura do Sistema	22
4.3. Componentes do Sistema	25
4.4. Extração de Vídeos	27
4.5. Pré-processamento e Classificação dos Vídeos	33
4.6. Engenho de Recomendação	34
4.7. Coletor de <i>Feedback</i>	37
5. Método e Avaliação	39
6. Conclusões e Trabalhos Futuros	48

Lista de Figuras

Figura 1 – Arquitetura do Sistema VCSR.....	17
Figura 4 – Visão Geral do Sistema de Recomendação.	20
Figura 5 – Etapas realizadas pela solução.	21
Figura 6 – Arquitetura em camadas.	23
Figura 7 – Componentes do Sistema.....	25
Figura 8 – Busca de vídeos com campo de texto aberto.	27
Figura 9 – Busca de vídeos através da seleção de temas.	28
Figura 10 – Tipos de credenciais da Google API.....	29
Figura 11 – Parâmetros da documentação.....	30
Figura 12 – Requisição Javascript para obter informações dos vídeos do YouTube.....	31
Figura 13 – Requisição ao recurso “vídeos” para obter detalhes sobre os vídeos.	32
Figura 14 – Algoritmo fusão aplicado para gerar recomendações de vídeos.....	35
Figura 15 – Algoritmo fusão aplicado para gerar recomendações de vídeos.....	38
Figura 16 – Respostas sobre frequência de uso do YouTube	39
Figura 17 - Respostas sobre o uso do YouTube para fins educacionais.	39
Figura 18 – Informações retornadas ao buscar um termo “Herança”.	40
Figura 19 – Informações retornadas ao buscar um termo “Herança”, tendo feito login.	41
Figura 20 – Retorno da API do YouTube com ruídos.	42
Figura 21 – Técnica de Recomendação FBC.....	43
Figura 22 – Técnica Híbrida – Algoritmo Filtragem sem corte.....	44

Lista de Tabelas

Tabela 1 – Algoritmo fusão aplicado para gerar recomendações de vídeos.	37
--	----

1. Introdução

Existem várias plataformas de vídeos, atualmente, cada uma com um volume significativo de vídeos que cresce ainda mais, como YouTube¹, Daily Motion², Meta Café³, Vimeo⁴. Além desses, tem-se popularizado os serviços de VOD (*Video on Demand*). Essa grande sobrecarga de informações deve-se à facilidade de *upload* de novos vídeos nessas plataformas, assim como a popularização da Internet. Segundo o Google (2018), o YouTube possui mais de um bilhão de usuários, que assistem a bilhões de horas de vídeos por dia e fazem *upload* de centenas de milhares de vídeos.

Essa sobrecarga de informações leva o usuário a desprender tempo para encontrar o que de fato procura. Uma maneira de solucionar o problema é através do uso de Sistemas de Recomendação, que selecionam o conteúdo de acordo com o objetivo do usuário.

Sistemas de Recomendação vêm sendo tema de pesquisas com aplicações em diversas áreas, tais como: comércio, saúde e educação. Aqueles voltados para a educação permitem a recomendação de livros, artigos e até vídeos educacionais, levando em consideração fatores como interesses dos usuários, dados demográficos, relacionamentos, dentro outros. Tanto as iniciativas privadas como as autônomas disponibilizam na rede de computadores mundialmente interligada vídeos em diversas áreas do conhecimento. Assim, faz-se necessária a criação de ferramentas que auxiliem na escolha desses vídeos educacionais, levando em consideração aspectos significativos para cada usuário. Neste trabalho, é proposto um Sistema de Recomendação (SR) de vídeos educacionais. Para a prova de conceito, utilizou-se o site do YouTube como repositório de vídeos, visto que é considerado um dos mais populares sites de vídeos da Internet. Segundo Davidson (2010), o YouTube foi fundado em fevereiro de 2005 e cresceu rapidamente, se tornando o site de vídeos mais popular do mundo.

1.1. Motivação e Contextualização da Pesquisa

¹ <http://www.youtube.com/>

² <http://www.dailymotion.com/>

³ <http://www.metacafe.com/>

⁴ <http://vimeo.com>

Uma das motivações gerais para a presente pesquisa vem do cotidiano das pessoas, as quais lidam com tomadas de decisões diariamente, como por exemplo, qual livro ler, qual filme assistir, qual produto comprar e etc. No contexto educacional não é diferente, sendo comum encontrar dificuldades na seleção dos materiais de estudo, sobretudo após o advento da Internet, onde novas oportunidades de aprendizagem surgiram, trazendo consigo grandes desafios (BASO, 1998, p. 1); (CAZELLA, 2010, p. 161).

Um dos principais desafios do mundo virtual, discutido há mais de duas décadas, está na recomendação de informações ou mesmo serviços disponíveis na internet (ALMEIDA, 2016, p. 17). Isso pode ser aplicado ao cenário educacional, no qual recomendações de conteúdos educacionais poderiam trazer mais facilidade aos aprendizes. Nesta perspectiva, existem algumas iniciativas que visam selecionar recursos de aprendizagem como artigos, vídeos e objetos educacionais no geral para recomendação. Essas iniciativas se concebem a partir de ambientes de *softwares* denominados Sistemas de Recomendação que levam em consideração critérios como: preferências de seus usuários, características e conteúdos dos itens a serem recomendados.

O uso de vídeos no ensino não é algo novo, exceto no que diz respeito à variedade de formato⁵ de vídeos, facilidade de aplicação em sala de aula, técnica de vídeos que um instrutor pode utilizar e a investigação sobre aprendizagem multimídia que fornece suporte teórico empírico para a sua utilização como uma ferramenta de ensino eficaz (BERK, 2009, p. 1). Em uma perspectiva ampla, é possível assistir a uma videoaula sobre determinado assunto, ler um livro *online*, sem a necessidade de comprar, ler artigos e tutoriais e uma infinidade de possibilidades. Respeitando a ideia dos múltiplos estilos de aprendizagem, os vídeos podem ser uma ferramenta que submete o aluno a estímulos visuais e sonoros que podem fazer com os mesmos aprendam melhor do que com uma abordagem tradicional, baseada em textos (MATTAR, 2009). McKINNEY *et al* (2009), por exemplo, conseguiram demonstrar que um grupo de alunos que utilizou *podcasts* teve melhor desempenho em provas do que outro grupo, que assistiu a aulas tradicionais em sala.

Apesar dos fatos supracitados, é preciso ter cuidado quando o assunto é aprendizagem. Com a facilidade de publicação de vídeos caseiros, a qualidade didática

⁵ Diversos formatos estão disponíveis na Internet, como MP4, AVI, MPEG.

dos mesmos pode estar comprometida, sobretudo no contexto educacional, visto que qualquer um pode publicar vídeos, inclusive pessoas sem os devidos conhecimentos técnicos de ensino/aprendizagem. Essa grande quantidade de vídeos amadores é um dos motivos de críticas de alguns autores, como KEARSLEY (2007) e COLLINS e BERGE (2000), os quais consideram que o fenômeno estaria comprometendo a qualidade do material produzido para EaD. Neste trabalho, considerou-se que o YouTube pode ser utilizado como uma ferramenta de EaD, a depender dos propósitos dos usuários. Além disso, acredita-se que o problema da sobrecarga relatado por KEARSLEY (2007) e COLLINS e BERGE (2000) ainda persiste nos dias atuais, visto que a quantidade de vídeos disponíveis cresce desproporcionalmente à quantidade de ferramentas para filtrar esses vídeos. Ainda é comum, por exemplo, um aluno que está em busca de uma videoaula se deparar com o questionamento de qual vídeo escolher. Por vezes, perde-se tempo até encontrar uma aula que concilie boa didática com qualidade de imagem e som (SILVA, 2014).

1.2. Abordagem da Pesquisa

Nesta seção, foi descrito o formato da pesquisa realizada na área de Sistemas de Recomendação, assim como uma análise dos problemas e requisitos que foram levados em consideração.

Nas subseções seguintes descreveu-se as questões da pesquisa, seus objetivos e aspectos metodológicos.

1.2.1. Questões da Pesquisa

A questão levantada por esta pesquisa está relacionada à definição de uma estratégia eficaz de recomendação de vídeos educacionais, levando-se em consideração várias características dos vídeos. Dado que é possível obter as propriedades dos vídeos de forma prática e ter acesso às suas características através de metadados, como, por exemplo, autor, categoria, quantidade de curtidas, quantidade de visualizações, comentários, dentre outros indicadores de qualidade, temos a seguinte questão de pesquisa:

QP: Como identificar se um vídeo é educacional e se possui qualidade para que recomendações apropriadas sejam produzidas?

1.2.2. Objetivos da Pesquisa

Este trabalho tem por objetivo geral propor uma solução de Sistemas de Recomendação, extraindo as informações disponíveis nos vídeos para gerar recomendações relevantes. Além disso, pretende-se atingir os seguintes objetivos específicos:

1. Determinar se um vídeo é educacional;
2. Extrair informações dos vídeos para sugerir vídeos adequados;
3. Desenvolver um sistema de recomendação em um domínio específico que auxilie os usuários a selecionarem seus materiais de estudo;
4. Validar o sistema desenvolvido aplicando-o a alunos, observando-se os resultados.

1.2.3. Aspectos Metodológicos

Esta dissertação propõe a concepção, o desenvolvimento e a avaliação de um sistema de recomendação de vídeos educacionais, através do uso de técnicas de recomendação, modelos de referências, técnicas de aprendizagem de máquina e pré-processamento de dados.

A pesquisa surgiu a partir de uma reflexão sobre a grande quantidade de vídeos educacionais disponíveis em repositórios na Internet, não dedicados a esse fim. Isso trouxe uma dificuldade singular na seleção de conteúdos confiáveis e de qualidade. Nesse sentido, levou-se em consideração as limitações existentes nos sistemas de recomendação de vídeos de gerações passadas e os novos recursos que surgiram para combater essas limitações. A título de exemplo, o YouTube, que dispõe atualmente de mais de um bilhão de usuários (YOUTUBE, 2016), possui ferramentas e informações que podem ser utilizadas a favor das recomendações. Por exemplo, o conteúdo do vídeo que pode ser convertido em texto através da legenda, informações como curtidas, visualizações, comentários, descrição, autor do vídeo dentre outros também estão disponíveis. Todas essas informações poderiam ser usadas para avaliar qualidade em um determinado vídeo?

O problema de como conceber SR de vídeos que utilizem as propriedades dos vídeos para gerar as recomendações foi abordado através de uma análise do domínio. Analisou-se a literatura relacionada a SR de vídeos, observando-se o que se pretende para uma nova geração de SR, extraindo-se daí um novo conjunto de aspectos, entre os quais podem-se destacar dificuldades em interoperabilidade entre sistemas, custo de manutenção, compartilhamento de materiais, dentre outros.

1.3. Estrutura da Dissertação

Esta dissertação tem seu conteúdo seccionado em seis capítulos, sendo o primeiro capítulo equivalente a esta introdução e os outros estruturados da seguinte forma:

- **Capítulo 2 - Fundamentação Teórica:** neste capítulo, são descritos os conceitos inerentes às técnicas de recomendação, incluindo as técnicas utilizadas neste trabalho: Filtragem Baseada em Conteúdo (FBC), Filtragem Colaborativa (FC) e abordagem híbrida.
- **Capítulo 3 - Trabalhos Relacionados:** neste capítulo, são apresentados os trabalhos correlatos à dissertação em questão, descrevendo suas características, vantagens, desvantagens e arquitetura. Além disso, é feita uma comparação entre os trabalhos relacionados, dando destaque aos requisitos atendidos de cada ambiente.
- **Capítulo 4 – Sistema de Recomendação de Vídeos Educacionais: Um Estudo de Caso no YouTube:** é apresentada a solução desenvolvida, dando destaque aos componentes implementados e às técnicas de recomendação e de engenharia de *software* utilizadas.
- **Capítulo 5 – Método e Avaliação:** é apresentado um estudo de caso, descrevendo aspectos correspondentes ao método utilizado na concepção do sistema de recomendação proposto, tendo-se escolhido como tema "Programação Orientada a Objetos com foco no conceito de Herança". Após a apresentação do método, uma avaliação é feita sobre os aspectos de qualidade do sistema.
- **Capítulo 6 - Conclusões e Trabalhos Futuros:** as conclusões são apresentadas, descrevendo como os objetivos foram alcançados, além dos resultados obtidos através de indicadores. Por fim, os trabalhos futuros são apresentados.

Além dos capítulos supracitados, esta dissertação apresenta as referências que nortearam a pesquisa e um apêndice que descreve as tecnologias utilizadas.

2. Fundamentação Teórica

Almejando-se um entendimento melhor deste trabalho, pretende-se apresentar, a seguir, alguns conceitos fundamentais. Com esse propósito, neste capítulo são apresentados alguns conceitos de Sistemas de Recomendação.

2.1. Sistemas de Recomendação

Segundo Rolim *et al* (2017), os SRs são uma representação de uma prática comum entre as pessoas, como indicar algo para alguém de acordo com uma necessidade específica. No entanto, os SRs utilizam técnicas de inteligência artificial para recomendar a informação mais adequada para as necessidades do usuário.

Os SRs são ferramentas de software baseadas em um conjunto de tecnologias e técnicas das áreas de Recuperação da Informação (RI) e Inteligência Artificial (IA), que analisam grandes volumes de dados, especialmente de produtos e informações sobre o usuário, e então prevê sugestões relevantes para os usuários, para que eles possam tomar melhores decisões dentre as alternativas disponíveis. (HENRIQUES, 2017).

Alguns SRs são responsáveis por utilizar as preferências dos usuários para gerar sugestões personalizadas de objetos, que podem ser produtos, serviços e até conteúdos. Muito utilizados em sites de *e-commerce*, essas lojas virtuais costumam fidelizar seus usuários clientes, pois eles se sentem valorizados ao ter disponíveis esses objetos sem muito trabalho. Torres (2004, p. 34) descreve as seguintes etapas para personalização de sites na internet: A) Identificação do cliente para, a partir do reconhecimento do usuário, carregar suas preferências. B) Descoberta de hábitos de consumo para descobrir os interesses dos usuários através de interações com o site. Existem duas formas gerais de se extrair as preferências do usuário: implícita ou explícita. A forma implícita leva em consideração as interações do usuário no site sem que ele perceba. A outra forma de gerar o perfil do usuário é de forma explícita, na qual o sistema questiona o usuário sobre suas preferências, categorias de interesse e etc. Esta última peca por fazer com que o usuário tenha o trabalho de responder a diversas perguntas. Idealmente, pode-se mesclar as duas abordagens, monitorando o perfil do usuário e montando-o implicitamente, mas permitindo que o próprio usuário altere essas preferências caso não concorde com elas ou caso queira refiná-las. C) Criação das recomendações através de técnicas de Inteligência Artificial. Dessa forma, o Sistema de Recomendações consegue recomendar os melhores itens aos quais o usuário pode preferir.

Existem várias classificações para Sistemas de Recomendação que foram propostas ao longo dos anos. No entanto, Torres (2004, p. 41) adaptou uma das que considerou mais abrangentes, dividindo-a em três grandes blocos:

- Entrada/saída: Trata-se de como as informações entram e saem de um SR;
- Método de recomendação: Aborda as formas de recomendação existentes;
- Outros aspectos de projeto: Abrange a forma como as recomendações são apresentadas ao usuário e o grau de personalização obtido.

Existem várias técnicas que podem ser utilizadas para gerar recomendação. Filtragem Baseada em Conteúdo (FBC) (BALABANOVIĆ e SHOHAM, 1997) e Filtragem Colaborativa (FC) (SARWAR *et al*, 2001) são algumas das abordagens mais conhecidas. Na FBC, é levada em consideração o conteúdo acessado e avaliado pelos usuários para gerar recomendações de novos conteúdos. Já na FC é realizada uma análise da similaridade entre os usuários, para que recomendações sejam realizadas de acordo com as avaliações de usuários com perfis similares (ROLIM *et al*, 2017).

Outras técnicas como as baseadas em árvores de decisão, sistemas multiagentes, redes bayesianas, raciocínio baseado em casos e regras associativas também são utilizadas. Nesse cenário, cada técnica possui suas vantagens e desvantagens e uma funciona melhor que a outra a depender do domínio em questão (TORRES, 2004, p. 34).

Existe também a abordagem híbrida ou filtragem híbrida, onde a estratégia para tirar proveito das vantagens de mais de uma técnica é, justamente, combinar duas ou mais de modo a desenvolver um sistema que recomende o conteúdo mais adequado para o usuário. Assim pode-se reduzir as desvantagens de uma técnica através das vantagens de outra (ROLIM *et al*, 2017). Um exemplo comum é utilizar a filtragem baseada em conteúdo combinada com a filtragem colaborativa. A combinação dessas duas técnicas possibilita que o sistema seja beneficiado pelos bons resultados para usuários que não possuem perfis similares, pela precisão da recomendação independentemente do número dos usuários, e pela descoberta de similaridades entre os usuários, além da recomendação relacionada com o histórico do usuário (CAZELLA *et al*, 2010).

De acordo com Torres (2004) e Burke (2002), os Sistemas de Recomendação Híbridos, como são chamados os SRs que combinam mais de uma técnica, podem ser classificados como:

- Ponderado: A similaridade de um item é computada a partir de uma combinação de várias técnicas de recomendação, com diferentes pesos para cada uma delas.
- Alternado: O sistema utiliza um critério para alternar a técnica que gera a recomendação.
- Cascata: Uma técnica de recomendação refina as recomendações fornecidas por outra técnica.
- Combinação de *features*: *Features* de diferentes dados são utilizadas em um único algoritmo.
- Aumento de *features*: As recomendações geradas por uma técnica são apresentadas na mesma lista.
- Misto: Recomendações de várias técnicas são apresentadas na mesma lista.
- *Meta-level*: O modelo aprendido por uma técnica de recomendação é utilizado como informação de entrada para outra técnica.

3. Trabalhos Relacionados

Neste capítulo, serão apresentadas algumas propostas para Sistemas de Recomendação que possuem relação com o trabalho proposto. Sendo expostos alguns trabalhos relacionados, que servem como base para este trabalho e outros que são considerados uma tendência para o sistema apresentado nesta dissertação.

3.1. A Health Information Recommender System

Rivero-Rodriguez et al. (2013) propuseram um SR no domínio da saúde para disponibilizar vídeos confiáveis; para isso, utilizaram a informação de canais respeitados de vídeos sobre saúde. Eles disponibilizaram vídeos com enriquecimento de informações de forma confiável, a partir de dados de vídeos do YouTube e de um serviço oferecido pela Biblioteca Nacional de Medicina dos Estados Unidos, chamado Medline Plus.

Eles aplicaram quatro métodos para gerar as recomendações e avaliaram seus resultados. O método que apresentou o melhor resultado foi uma combinação de dois outros métodos existentes, onde o primeiro identificava termos médicos relacionados a um termo do título e submetia ao Medline Plus para obter recomendações. Já o segundo identificava termos médicos relacionados a dois ou mais termos do título, que também eram submetidos ao Medline Plus para obter recomendações. Esta combinação dos dois resultados trouxe mais recomendações de qualidade (boas recomendações) e menos recomendações ruins, quando comparada aos outros métodos.

3.2. The YouTube Video Recommendation System

Davidson et al. (2010) utilizaram os passos digitais do usuário para prover recomendações de vídeos do YouTube. As atividades de visualizar, adicionar aos favoritos, curtir, entre outras, eram armazenadas e analisadas para gerar um conjunto de preferências do usuário. Estima-se que 60% das visualizações do YouTube sejam originadas das recomendações do serviço.

3.3. A Predictive Model for Video Lectures Classification

Silva et al. (2014) propuseram o uso de algoritmos de aprendizagem de máquina para aprender a avaliar vídeos, a partir de um conjunto de informações sobre vídeos do YouTube. Foi realizado um experimento onde a predição de avaliação dos vídeos foi analisada a partir de diversos algoritmos de classificação, dentre eles: Nave Bayers (TAHERI et al, 2011), Support Vector Machine (SVM) (WANG, 2008) e C4.5 Classifier

(QUINLAN, 2014). O que demonstrou maior desempenho foi o SVM e os atributos que se mostraram mais relevantes foram: quantidade de visualizações, quantidade de curtidas e quantidade de não curtidas.

3.4. YouTube Recommender Network (YRN)

Qin et al. (2010) propuseram um SR que apresenta um *ranking* de vídeos do YouTube a partir de informações extraídas de uma rede social dos usuários, mais especificamente uma rede social com resenhas⁶ de vídeos. Os usuários escrevem sobre seus vídeos de interesse e assim este SR consegue ampliar seu leque de recomendações, quando comparado às recomendações do YouTube que são restritas ao tema que está se assistindo e às *tags* de interesse do usuário. Eles utilizaram a API do YouTube para incorporar funcionalidades como buscar vídeos e obter informações dos vídeos e dos usuários em seu próprio sistema.

Apesar da proposta ser interessante, uma vez que busca os interesses dos usuários para recomendar não só sobre o tema do vídeo a que ele está assistindo, mas temas que o mesmo possa se interessar, o YRN não considera critérios de seleção dos vídeos quanto à sua qualidade, o objetivo é apenas recomendar vídeos que possam interessar ao usuário. No entanto, quando um usuário faz uma busca no próprio YouTube e utiliza palavras-chaves, como por exemplo “Programação Orientada a Objetos” ou “Java” o YouTube já traz vídeos relevantes de acordo com quantidade de visualizações, mas não há como saber se o vídeo possui qualidade no que tange didática, áudio ou imagem. O YRN não se preocupa com isso, focando apenas nas recomendações mais abrangentes.

3.5. VCSR System

O artigo “An Automatic Multimedia Content Summarization System for Video Recommendation”, de Yang *et al* (2009), propõe-se a reconhecer a legenda de um vídeo a partir de um módulo denominado de “Optical Character Recognition” (OCR) e assim resumir o conteúdo do mesmo. A proposta é facilitar a inclusão de resumos de vídeos em acervos, de forma automatizada, além de recomendar a usuários vídeos de acordo com seu perfil. Na Figura 1, é possível verificar a arquitetura do projeto VCSR System.

⁶ Produção textual, por meio da qual o autor faz uma breve apreciação, e uma descrição a respeito de acontecimentos.

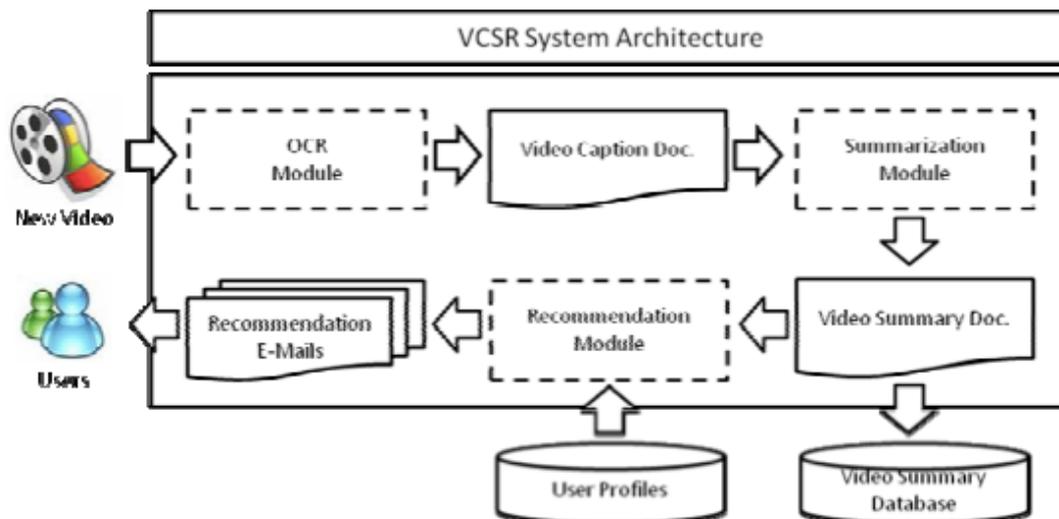


Figura 1 – Arquitetura do Sistema VCSR.

Uma vez que um novo vídeo é recebido, o Módulo OCR reconhece legendas como documentos de legenda do vídeo. Estes documentos são então passados para o módulo de compactação e os documentos de síntese para o vídeo são gerados por extração das palavras-chave. Por fim, os e-mails de recomendação de vídeo são gerados pelo módulo de recomendação, que estima a relevância para cada aluno de acordo com seus perfis. Ao combinar esses três módulos, o sistema pode gerar recomendações automaticamente e enviar e-mails de recomendação de vídeo quando há um novo vídeo de entrada. Em outras palavras, o processo no sistema VCSR é automatizado, sem qualquer intervenção humana.

O trabalho proposto nesta dissertação difere do Sistema VCSR em vários aspectos, dentre eles:

- O VCSR resume vídeos e envia recomendações por e-mail, no entanto não há uma interação maior entre os vídeos.
- O enfoque desta dissertação é utilizar a API do YouTube para obter os arquivos de legenda. Dessa forma, não foi necessário desenvolver um módulo OCR. Hoje em dia, plataformas de TV também possuem recursos para obtenção de legenda e *CloseCaption*.

3.6. Comparação entre os Sistemas de Recomendação

É possível notar que, apesar de VCSR utilizar a recomendação de vídeos, a técnica utilizada para extração da legenda é através de reconhecimento de textos em imagens; o

que não se faz necessário ao utilizar um moderno repositório de vídeos, no qual é possível obter a legenda. Mesmo os vídeos que não possuem legendas podem ser facilmente utilizados com o recurso de reconhecimento de voz que a plataforma utilizada, YouTube, possui. Além disso, o YouTube é uma ferramenta bastante popular com vários canais educativos. Muitos alunos recorrem a ele para estudar ou tirar dúvidas quando necessário. Assim, o trabalho proposto atende à funcionalidade de extração de vídeos do YouTube.

Todos os trabalhos possuem um repositório previamente alimentado com informações sobre os vídeos educativos. No entanto, o trabalho proposto se destaca ao utilizar algoritmos para aprender a qualificar um vídeo.

Cada proposta é interessante para o contexto em que se aplica, mas o trabalho proposto traz uma visão mais moderna e uma aplicação a usuários de todo o mundo, visto que o YouTube hoje está presente em todos os continentes. Outro destaque da proposta trata-se do uso de um algoritmo híbrido de recomendação que busca as vantagens das recomendações baseadas em conteúdo e as vantagens das recomendações baseadas em avaliações.

4. Sistema de Recomendação de Vídeos Educacionais: Um Estudo de Caso no YouTube

Neste capítulo, será apresentada a solução proposta neste trabalho, intitulada “Easy YouTube” (EYT). A apresentação se dará a partir de uma visão geral desta solução, na qual será explicada, em linhas gerais, a ideia da aplicação e seus objetivos.

Posteriormente, a arquitetura geral desta solução, utilizada como referência para sua implementação, será abordada seguida dos métodos usados para extração de dados dos vídeos do YouTube, classificação e o pré-processamento dos dados, essenciais para gerar recomendações de qualidade. Também serão descritas as técnicas de recomendações utilizadas, dentre elas filtragem baseada em conteúdo e uma técnica baseada em avaliações dos usuários. Por fim, serão apresentados detalhes de sua implementação, vantagens e desvantagens das escolhas realizadas, descrevendo-se um cenário de uso a fim de ilustrar a aplicabilidade da solução conforme os requisitos definidos. Cada item será explicado no decorrer desse capítulo.

4.1. Visão Geral

O objetivo principal do sistema proposto é a recomendação de vídeos educacionais. Para recomendar vídeos significativos e garantir a qualidade didática dos mesmos, é imprescindível levar em consideração aspectos educacionais, técnicos e tecnológicos. Esses serão discutidos no decorrer deste capítulo, mas por ora é possível entender de forma genérica o funcionamento do sistema desenvolvido, sem se aprofundar nos componentes internos do Sistema de Recomendação.

A utilização do sistema pelo usuário poderá ser através de um computador conectado à Internet. A partir daí, será possível acessar o Sistema de Recomendação hospedado em um Servidor de Aplicação Web, que disponibilizará uma lista com temas de estudo. Posteriormente, na Figura 4 é possível acompanhar os passos:

- A) O usuário seleciona um tema de estudo – por exemplo, "Programação Orientada a Objetos / Herança" – e essa informação é enviada ao Servidor de Aplicação Web.
- B) O sistema busca vídeos relacionados ao tema de estudo no YouTube, extraindo informações e armazenando para uso posterior;



Figura 2 – Visão Geral do Sistema de Recomendação.

- C) O sistema recupera informações referentes às avaliações dos usuários, classifica e realiza um pré-processamento dos dados obtidos dos vídeos.
- D) O sistema utiliza técnicas de recomendação, cruzando as informações recuperadas no item B e C para gerar recomendações significativas de vídeos educacionais no tema de estudo selecionado.

A solução EYT é um sistema web que pode servir como base na construção de aplicações responsáveis por classificar vídeos do YouTube de um determinado tema e, em seguida, utilizar essas informações como dados de entrada para gerar recomendações de vídeos relevantes ao usuário, só que, desta vez, educacionais. As recomendações geradas auxiliarão os usuários na árdua tarefa de selecionar vídeos educacionais de qualidade, minimizando o tempo de procura por bons materiais. Os benefícios dessa solução não atingem apenas os usuários aprendizes, mas também podem beneficiar professores, que, por vezes, levam tempo preparando ou buscando materiais complementares de estudo.

A solução desenvolvida pode ser executada a partir de navegadores/*browsers* e em qualquer plataforma/sistema operacional. O sistema utiliza o engenho de busca do YouTube para trazer os vídeos, como se o usuário estivesse no próprio site do YouTube, e permite ao usuário executar ações semelhantes, como iniciar a exibição, pausar, adiantar ou retroceder um vídeo e buscar vídeos por palavras-chave ou frases. No entanto, foi implementada também outra forma de buscar vídeos educacionais, a partir de temas de

estudo pré-definidos. A diferença é que este último realizará buscas com enriquecimento de consulta, adicionando automaticamente palavras-chaves relacionadas ao tema.

O processo para recomendação de vídeos educacionais é iniciado com o ato do usuário de selecionar um tema e assunto de estudo que esteja disponível na solução e, a partir daí, a aplicação EYT se encarrega de realizar alguns procedimentos, através de seus módulos, como pode ser acompanhado na Figura 5.

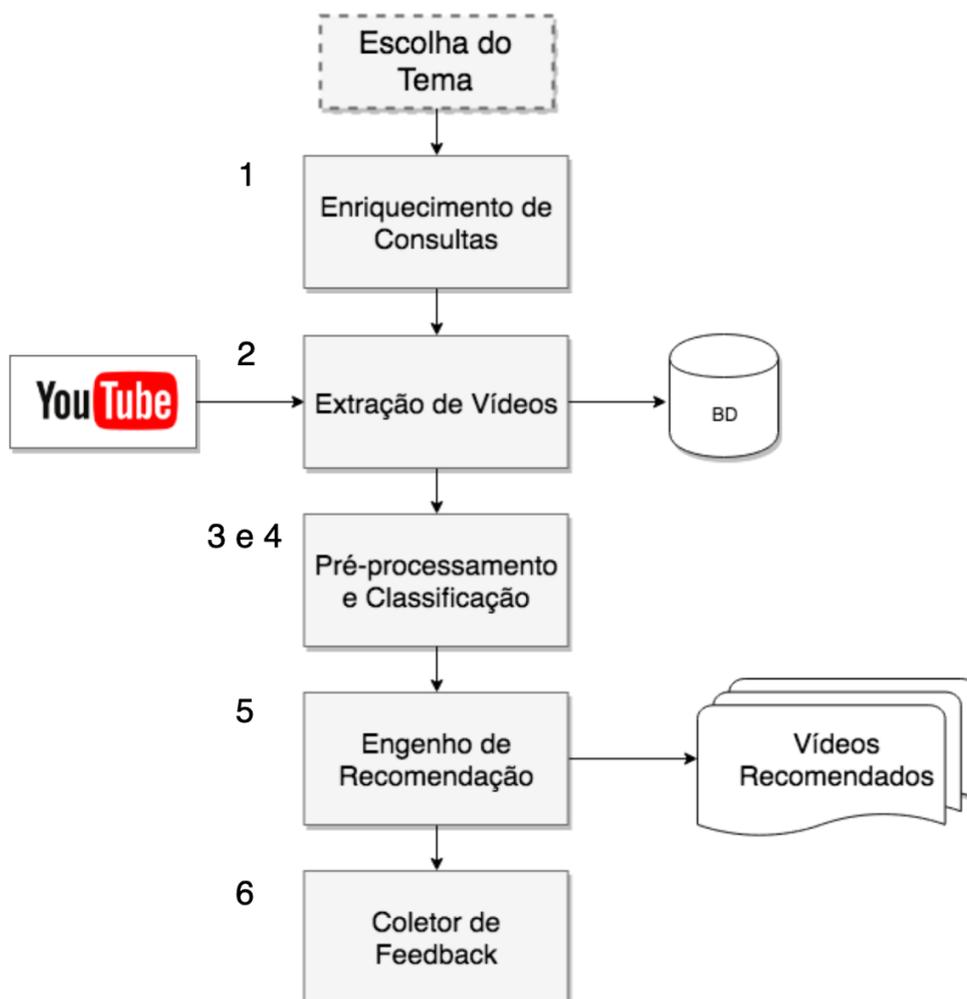


Figura 3 – Etapas realizadas pela solução.

Mais detalhes sobre os módulos citados na Figura 5 serão apresentados nas próximas seções. A seguir, será apresentada uma visão geral dos módulos e dos seus procedimentos:

1. O módulo enriquecimento de consultas requer a interação prévia de um especialista, no qual serão definidas as palavras e termos recorrentes e importantes de um determinado tema. Quando o usuário seleciona um tema pré-definido, o sistema recupera as palavras e termos recorrentes e o próximo módulo utiliza essas informações para extração dos vídeos do YouTube;
2. O módulo de extração de vídeos recupera informações sobre os vídeos do YouTube associados aos termos e palavras recorrentes de um tema pré-definido. Essas informações são armazenadas em uma base de dados própria, para uso posterior;
3. O módulo de pré-processamento é responsável pelo tratamento dos dados obtidos. Funcionará como um filtro, selecionando palavras relevantes e tratando dados para encaminhar ao módulo de classificação e para o engenho de recomendação. Mais detalhes são apresentados na seção 4.5;
4. O módulo de classificação utiliza dados obtidos a partir do YouTube, para classificar os vídeos, levando em consideração os atributos mais relevantes, segundo algoritmos de aprendizagem de máquina utilizados;
5. O engenho de recomendação é responsável por realizar o cálculo de similaridade entre o tema buscado e os vídeos educacionais classificados. Além disso, executa os algoritmos de recomendação, baseando-se na similaridade e nas características de interesse do usuário, gerando as recomendações dos vídeos para o usuário;
6. O coletor de feedback é responsável por coletar as avaliações do usuário referente as recomendações geradas. Os vídeos recomendados são avaliados com uma nota que varia entre 1 e 7.

A coleta do *feedback* do usuário é de fundamental importância para melhorar o algoritmo de aprendizagem de máquina usado na classificação. Assim, a solução pode ser evoluída para uma ferramenta educacional poderosa, que contribua cada vez mais para o processo de aprendizagem de seus usuários.

4.2. Requisitos e Arquitetura do Sistema

Um dos pontos arquiteturais mais relevantes da solução foi a criação de um *design* de código que possibilitasse que os componentes individuais do sistema fossem dissociados uns dos outros, permitindo que fossem compreendidos e depurados isoladamente. Por fazer parte do ecossistema do YouTube, as recomendações precisam

ser resilientes a falhas. Assim, possíveis instabilidades do YouTube serão contornadas com a base de dados própria que o sistema EYT possui.

Um outro ponto, não menos importante, diz respeito à interoperabilidade da solução. Outras aplicações poderão integrar a solução “Easy YouTube” e promover interação, independente da plataforma de desenvolvimento. Graças ao conceito de *WebServices* e de *API REST*, o EYT pode executar tarefas simples ou complexas para outras aplicações que a integrem. Assim, tem-se um nível maior de reutilização das funcionalidades da solução, permitindo que essas aplicações terceiras se preocupem apenas com o processo de negócio a qual pretendem realizar.

Outras vantagens da utilização da tecnologia de *WebServices* e de *API REST* podem ser destacadas, como o fato de ser descrita em linguagem natural, com termos próximos aos utilizados pela aplicação.

A arquitetura da solução segue o padrão já conhecido “arquitetura em camadas”. No topo, a camada de interface representa a camada de mais alto nível. Os itens nessa camada podem variar; hoje, trabalha-se com Interface Web para navegadores / *browsers* de computadores, no entanto, é possível adicionar outros dispositivos clientes.

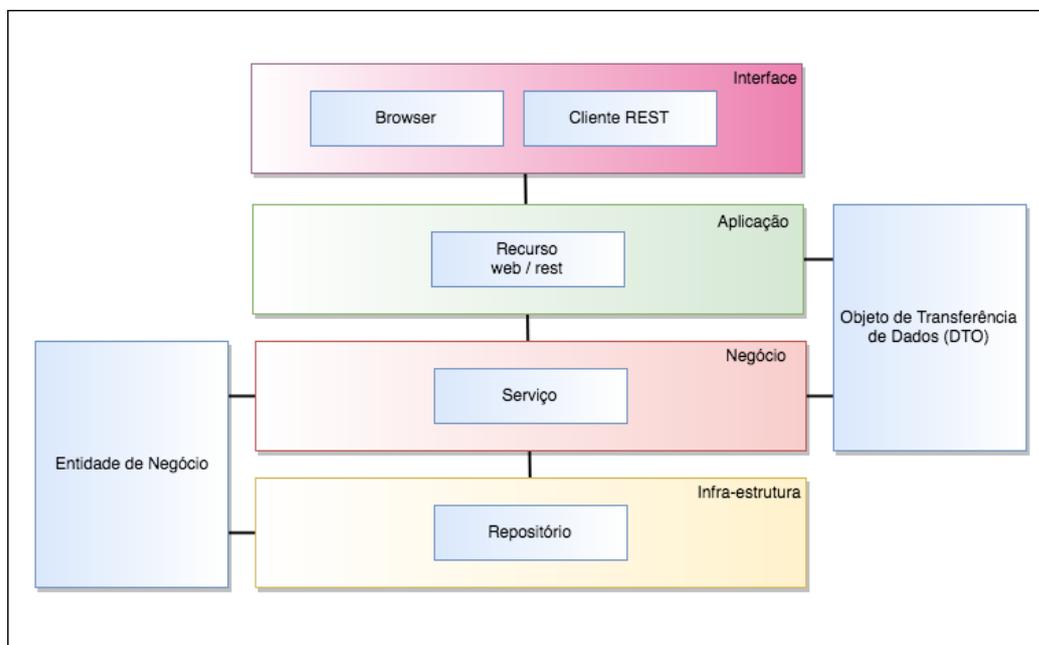


Figura 4 – Arquitetura em camadas.

Há várias maneiras de dividir um sistema de *software*, no entanto, as convenções levam à Arquitetura em Camadas (SCHMIDT, 2013). O princípio fundamental dessa

arquitetura é isolar os elementos em uma camada, na qual cada elemento depende apenas de elementos da mesma camada ou de elementos das camadas abaixo dela. A ideia é que cada camada se especialize em determinadas características do *software*, permitindo um *design* mais coeso.

No projeto atual, adotou-se a seguinte estrutura para as camadas conceituais:

- Interface ou o usuário ou camada de apresentação: responsável por apresentar informações ao ator e interpretar seus comandos. O ator pode ser um usuário humano ou outro sistema de computador.
- Camada de Aplicação: define as funcionalidades existentes no *software* e direciona as requisições para a camada de negócio resolver os problemas. Ela não contém regras de negócio, mas coordena tarefas e delega trabalhos, funcionando como uma centralizadora de requisições.
- Camada de Negócio: é a principal camada do *software*, representa os conceitos do negócio, a situação do negócio e regras de negócio. Tem-se o controle do estado que reflete a situação do negócio, embora detalhes de sua armazenagem sejam delegados à infraestrutura. Além disso, é de responsabilidade da camada converter objetos complexos em objetos simples e vice-versa.
- Camada da Infraestrutura: fornece recursos técnicos genéricos que suportam as camadas mais altas – envio de mensagens para o aplicativo, persistência do domínio, desenho de *widget* para a UI, e assim por diante.
- Entidade de Negócio: representa o modelo do domínio; pode possuir regras de negócio e juntamente com a camada de domínio representa conceitos do negócio.
- Objeto de Transferência de Dados: representa os objetos de transferência de valor, ou DTO (*Data Transfer Objects*). Esses objetos diferem das entidades de negócio por não possuírem o modelo de domínio do *software*; são objetos que contêm tipos primitivos e não há regras de negócio encapsuladas. Dessa forma, o modelo do domínio não é exposto aos atores que consumirem os serviços disponibilizados pelo *software*.

4.3. Componentes do Sistema

Para aumentar a sua flexibilidade, o sistema foi dividido em componentes, tendo sido construído de tal forma que permite a troca de componentes com diferentes técnicas para posterior avaliação.

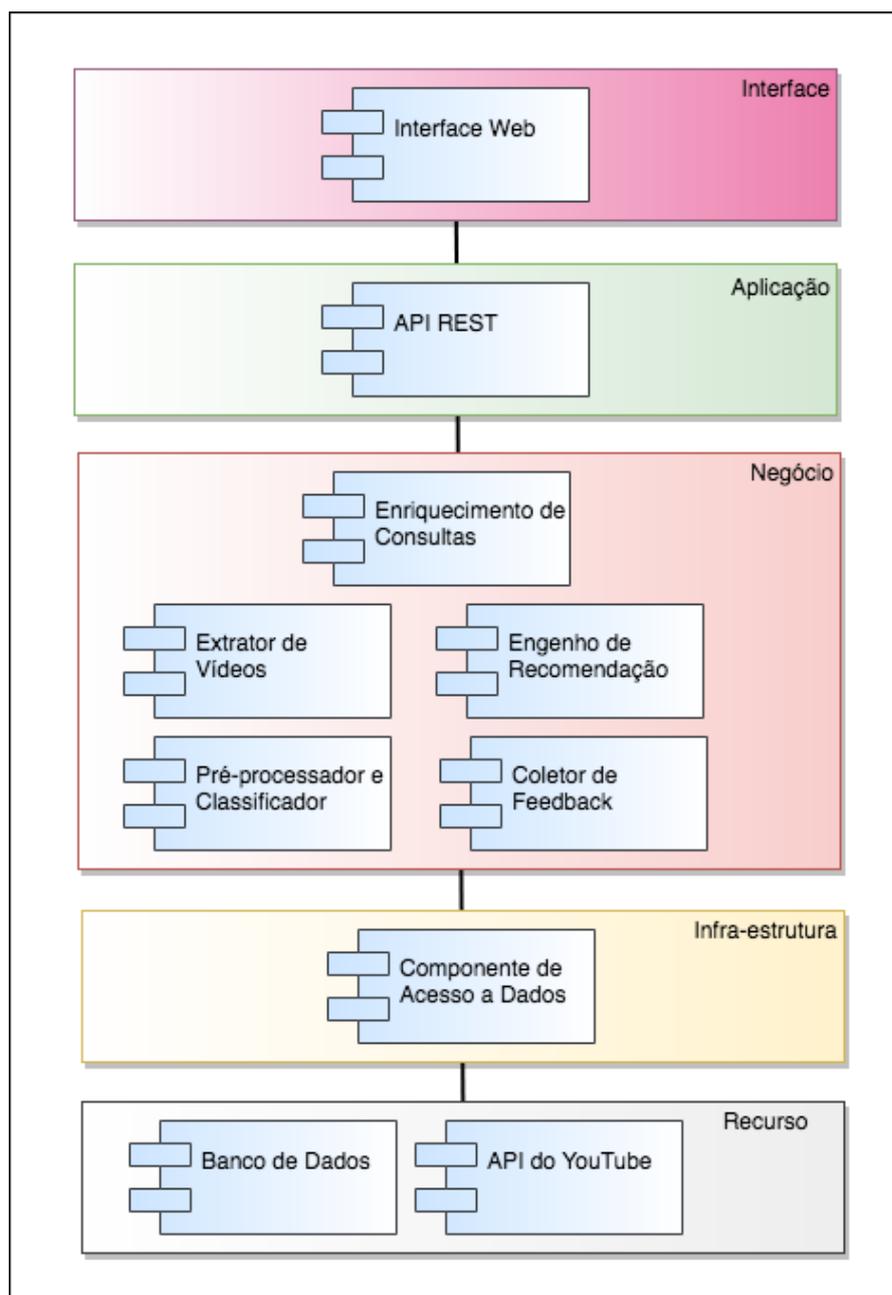


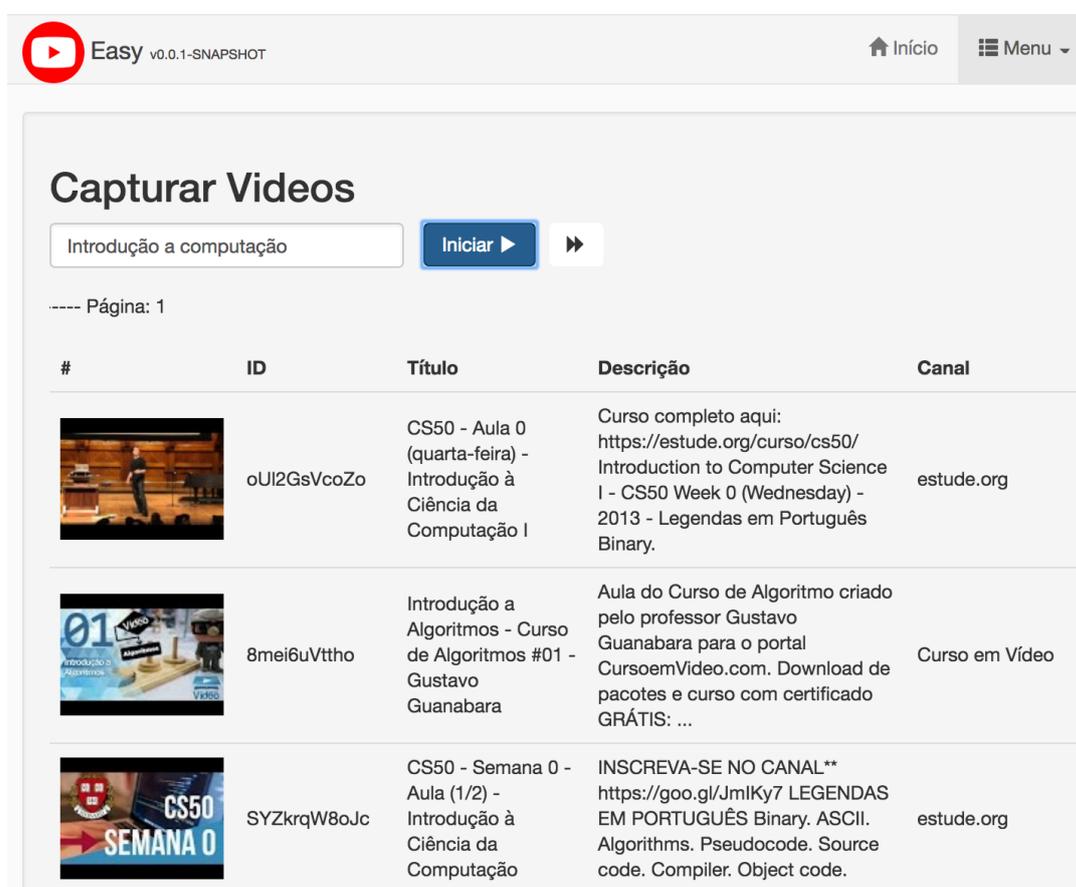
Figura 5 – Componentes do Sistema.

Os componentes podem ser visualizados juntamente com as camadas definidas na arquitetura. A saber, constam:

- Interface: contém os componentes cliente que acessam o sistema. Podem ser interfaces específicas para web, aplicações para *smartphones*, TV conectada, etc.
- Aplicação: O componente API REST gerenciará a requisição recebida e como o sistema deverá respondê-la. Para isso, deverá ter acesso aos componentes Enriquecimento de Consulta, Extrator de Vídeos, Pré-processador e Classificador, Engenho de Recomendação e Coletor de *Feedback*.
- Negócio: contém os componentes ou serviços responsáveis pela lógica do negócio.
 - Enriquecimento de Consulta: é capaz de incrementar termos de um domínio pesquisado com o objetivo de melhorar os resultados da busca;
 - Extrator de Vídeos: responsável por capturar as informações dos vídeos do YouTube e armazená-las em banco de dados próprio;
 - Pré-processador e Classificador: se responsabiliza por processar as informações e classificar os vídeos com propriedades que facilitarão o trabalho do Engenho de Recomendação.
 - Engenho de Recomendação: responsável por recomendar recursos educacionais a partir de palavras-chave, implementando a estratégia de recomendação baseada em conteúdo. Para isso, o componente acessa, através da camada de persistência, uma base de dados de recursos educacionais.
 - Coletor de feedback: permite armazenar as avaliações das recomendações e aprimorar o processo de recomendação.
- Componentes de Acesso a Dados: responsável pela comunicação com recursos externos como, por exemplo, banco de dados e API do YouTube.
- Recurso: abrange os dados de negócio e recursos externos como serviços e sistemas.

4.4. Extração de Vídeos

A primeira etapa tem o objetivo de buscar informações sobre os vídeos no YouTube e armazenar em banco de dados local, para facilitar a manipulação nas etapas de classificação e pré-processamento dos dados. A busca de vídeos se dá de duas formas: a primeira é através de busca simples, na qual o sistema apresenta um campo de busca e o usuário digita as palavras-chave que achar pertinente. Neste caso, o usuário fica livre para digitar quaisquer palavras-chave, conforme apresentado na Figura 8.



Easy v0.0.1-SNAPSHOT Início Menu

Capturar Videos

Introdução a computação Iniciar ▶▶

---- Página: 1

#	ID	Título	Descrição	Canal
	oUI2GsVcoZo	CS50 - Aula 0 (quarta-feira) - Introdução à Ciência da Computação I	Curso completo aqui: https://estude.org/curso/cs50/ Introduction to Computer Science I - CS50 Week 0 (Wednesday) - 2013 - Legendas em Português Binary.	estude.org
	8mei6uVttho	Introdução a Algoritmos - Curso de Algoritmos #01 - Gustavo Guanabara	Aula do Curso de Algoritmo criado pelo professor Gustavo Guanabara para o portal CursoemVideo.com. Download de pacotes e curso com certificado GRÁTIS: ...	Curso em Vídeo
	SYZkrqW8oJc	CS50 - Semana 0 - Aula (1/2) - Introdução à Ciência da Computação	INSCREVA-SE NO CANAL** https://goo.gl/JmIKy7 LEGENDAS EM PORTUGUÊS Binary. ASCII. Algorithms. Pseudocode. Source code. Compiler. Object code.	estude.org

Figura 6 – Busca de vídeos com campo de texto aberto.

A outra maneira é a partir da seleção de temas pré-definidos, que já possuem enriquecimento de consultas com palavras-chaves adicionais. Por exemplo, na escolha do tema "Conceitos da Orientação a Objetos" e do subtema "Herança", outras palavras-chave podem ser adicionadas automaticamente à busca, como "Classe", "Extends" e "Java", permitindo que a busca seja enriquecida com palavras frequentemente encontradas em aulas do tema selecionado, como pode ser visualizado na Figura 9.

#	ID	Título	Descrição	Canal	Publicação	Duração	Caption	Dislii
	800QkZkK164	Aprenda Java Fácil : Aula14 - Herança	Aula do blog Aprendajavafacil.blogspot.com sobre Herança. Acesse : www.aprendajavafacil.blogspot.com ...	Bruno César Barelli Bet	23/01/2013	PT14M21S	true	13
	3Dq4z791J7U	POO (parte 7) - Programação Orientada a Objetos com Java / Polimorfismo - parte 1	Série de vídeo-aulas sobre Programação Orientada a Objetos (POO) usando a linguagem de programação JAVA. Post no blog dessa matéria: Polimorfismo ...	Douglas Del Frari	04/04/2013	PT12M45S	true	2

Figura 7 – Busca de vídeos através da seleção de temas.

Este último método requer análise prévia de um "professor" para que os termos associados façam sentido. Foi adicionada ao sistema uma funcionalidade para facilitar a adição desses termos.

Qualquer que tenha sido o método escolhido, a solução se utilizará da API do YouTube disponibilizada pelo Google, que permite a incorporação de funções executadas no site do YouTube em outro site ou aplicativo. Para interagir com a API, fez-se necessário obter credenciais de autorização através do Googler Developers Console, disponível em <https://console.developers.google.com/>. Na Figura 10, é possível verificar que existem três tipos disponíveis de credenciais, e a escolhida nesse trabalho foi “Chave de API”. Essa chave de API é utilizada a cada requisição e serve tanto para identificar a aplicação que está realizando a requisição, como também limitar o número de requisições a fim de não sobrecarregar a plataforma.

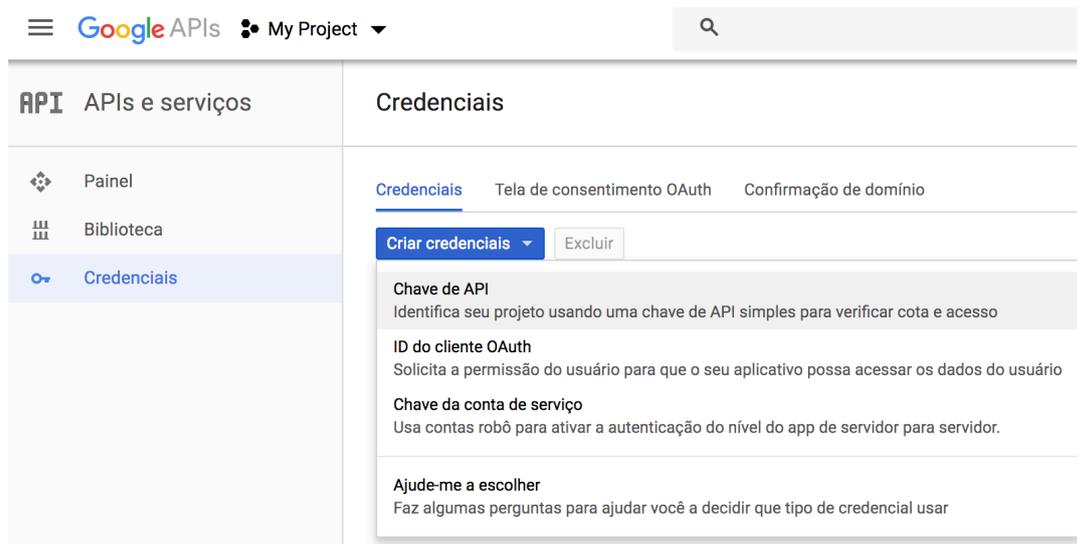


Figura 8 – Tipos de credenciais da Google API.

É disponibilizada uma lista de tipos de recursos que podem ser recuperados usando a API, mas para o objetivo dessa etapa, foi necessário o uso do método *Search* que disponibiliza um resultado de pesquisa contendo várias informações sobre um vídeo, canal ou *playlist*. Para utilizar esse recurso, é necessário informar um método disponível; para este caso, foi utilizado o método *list*, que retorna um conjunto de resultados de pesquisa que correspondem a parâmetros de consulta especificados na solicitação da API. Por padrão, um conjunto de resultados de pesquisa identifica recursos de vídeo, *channel* e *playlist* correspondentes, mas também é possível configurar consultas para recuperar somente um tipo específico de recurso.

Para realizar a solicitação HTTP, utiliza-se o método GET através da URL "https://www.googleapis.com/youtube/v3/search" e é necessário informar um parâmetro literal chamado "*part*" (Figura 11), que especifica uma lista separada por vírgulas de uma ou mais propriedades de recurso *search* que serão incluídas pela resposta da API. Os nomes *part* que podem ser incluídos no valor do parâmetro são *id* e *snippet*.

Parâmetros	
Parâmetros obrigatórios	
part	<p>string</p> <p>O parâmetro part especifica uma lista separada por vírgulas de uma ou mais propriedades de recurso search que serão incluídas pela resposta da API. Os nomes part que podem ser incluídos no valor do parâmetro são id e snippet.</p> <p>Se o parâmetro identificar uma propriedade que contenha propriedades criança, estas serão incluídas na resposta. Por exemplo, em um resultado search, a propriedade snippet contém outras propriedades que identificam o título, a descrição, entre outros, do resultado. Se você definir part=snippet, a resposta da API também conterá todas essas propriedades aninhadas.</p>

Figura 9 – Parâmetros da documentação.

No trabalho em questão, foi definido "*id,snippet*" para o campo parâmetro "*part*". Cada vídeo no YouTube é identificado de forma única por um tipo literal de onze caracteres, chamado de "ID", ao informar o parâmetro *part* como *id*, a resposta da API incluirá este identificador único. Em "*fields*", é possível especificar todas as informações que precisarão ser retornadas, como título, descrição, imagem com tamanho padrão e título do canal. Além disso, foi necessário definir o parâmetro "*maxResults*", que limita a quantidade de vídeos retornados. A API do YouTube permite que se realize a busca até no máximo de 50 em 50 registros. Nesse caso, as informações sobre a próxima página e página anterior são disponibilizadas para que os próximos 50 registros sejam recuperados, como pode ser observado no trecho de código abaixo:

```

$http.get('https://www.googleapis.com/youtube/v3/search', {
  params: {
    key: "AIzaSyBqLJ58XJDoys4WUXeioh__sNYg0ZdcmMQ",
    type: 'video',
    maxResults: '50',
    pageToken: pagina,
    part: 'id,snippet',
    fields: 'items/id,' +
    'items/snippet/title,' +
    'items/snippet/description,' +
    'items/snippet/thumbnails/default,' +
    'items/snippet/channelTitle,' +
    'nextPageToken,' +
    'prevPageToken',
    q: searchText,
    videoCaption: 'closedCaption'
  }
})

```

Figura 10 – Requisição Javascript para obter informações dos vídeos do YouTube.

Percebe-se que a chave criada também precisa ser informada no campo “*key*”. Armazenar as informações em uma base de dados específica foi de fundamental importância para o desenvolvimento do trabalho, pois ao utilizar os dados diretamente do YouTube, essas informações poderiam sofrer alterações pelo dinamismo existente na ferramenta; além disso, alguns dos dados capturados precisaram ser tratados e pré-processados para o sucesso dos resultados.

Além do ID, foi preciso capturar outras informações que o método *list* do recurso “*Search*” não disponibilizava. Assim, precisou-se fazer, para cada vídeo, uma outra chamada à API do YouTube, mas agora ao recurso “vídeos”, que permite recuperar mais informações sobre um determinado vídeo. Para tanto, foi necessário informar como parâmetro o ID do vídeo, assim como definir os campos *part* e *fields*, explicados anteriormente, com os dados desejados, como demonstrado na Figura 13.

```

$http.get('https://www.googleapis.com/youtube/v3/videos', {
  params: {
    key: "AIzaSyBqLJ58XJDoys4WUXeioh__sNYg0ZdcmMQ",
    type: 'video',
    maxResults: '1',
    part: 'snippet,contentDetails,statistics',
    fields: 'items/snippet/publishedAt,' +
            'items/snippet/channelId,' +
            'items/snippet/categoryId,' +
            'items/snippet/tags,' +
            'items/contentDetails/duration,' +
            'items/contentDetails/caption,' +
            'items/statistics/viewCount,' +
            'items/statistics/likeCount,' +
            'items/statistics/dislikeCount',
    id: video.id.videoId
  }
})

```

Figura 11 – Requisição ao recurso “vídeos” para obter detalhes sobre os vídeos.

Dessa forma, foi possível capturar outras informações como: data de publicação, ID do canal, ID da categoria, *tags*, duração, indicador de legenda, quantidade de visualizações, quantidade de *likes* e quantidade de *dislikes* (não curtidas).

A data de publicação fez-se importante, pois um vídeo publicado há muito tempo pode conter informações defasadas e o sistema poderá inferir isso, barrando que esse vídeo seja recomendado. Outras informações importantes como canal e categoria trazem também significado, já que a categoria pode filtrar apenas vídeos educacionais e mostrar que alguns canais são bem respeitados, podendo ser considerados uma fonte segura. Rivero-Rodriguez *et al.* (2013) propuseram um Sistema de Recomendação no domínio da saúde para disponibilizar vídeos confiáveis; para isso, utilizaram a informação de canais respeitados de vídeos sobre saúde. Outra propriedade importante são as *tags*, bastante utilizadas em outros Sistemas de Recomendação para inferir similaridade entre vídeos. A duração do vídeo pode trazer informações implícitas sobre como um tema é tratado num vídeo, se é um resumo do tema ou um vídeo mais detalhado, por exemplo. Já o indicador “*caption*” informa se o vídeo possui legenda, característica que será utilizada em trabalhos futuros, nos quais pretende-se acessar o texto das legendas para transformá-las em palavras-chave com o objetivo de aumentar a acurácia entre a similaridade dos vídeos a serem recomendados e o tema de interesse. A partir da

quantidade de visualizações, pode-se inferir que um vídeo tem despertado interesse dos usuários; no entanto, para ganhar confiança se é um vídeo bom ou ruim, as propriedades “quantidade de curtidas” e “quantidade de não curtidas” serão bastante úteis.

Neste módulo, a intenção não é tratar os dados capturados, e sim extraí-los do YouTube, armazenando seu estado no banco de dados. Algumas informações obtidas no momento da captura provavelmente sofrerão alterações com o passar do tempo (quantidade de curtidas, visualizações dentre outras), mas outras medidas podem ser tomadas para atualizar as informações na base local e tratá-las em um trabalho futuro. Para o trabalho em questão, foram capturados 1000 vídeos do YouTube para posterior classificação e avaliação.

4.5. Pré-processamento e Classificação dos Vídeos

Como mencionado na sessão anterior, os dados foram armazenados sem tratamentos prévios, ou seja, em base de dados local, da mesma forma que a API do YouTube os retornou. No entanto, para o processo de classificação dos vídeos e de recomendação, faz-se necessário tratar esses dados de forma conveniente ao processo. Por exemplo, para realizar comparação dos temas de estudo com os vídeos em questão, os títulos precisaram ser tratados, dividiu-se a sentença em palavras, eliminando pontos, espaços e outros caracteres. Os resultados são armazenados como em uma matriz, com uma palavra em cada célula. Além disso, as conhecidas “*stop words*” são removidas dessa matriz. Essas palavras irrelevantes são aquelas que não fornecem significado ao conteúdo, como preposições, artigos ou verbos comuns. Como resultado, temos todas as palavras significativas.

Outra operação realizada é o tratamento de algumas informações como a propriedade “*duration*” que, além de retornar o tempo do vídeo, traz a língua em que o vídeo foi gravado no formato “PT21M14S”, indicando que o vídeo está na língua portuguesa e que possui 21 minutos e 14 segundos. Esta informação, da maneira como foi retornada, dificultaria a classificação dos vídeos, pois há duas informações distintas em um campo só. O pré-processamento irá extrair a linguagem, atribuindo-a a uma classe e armazenando o tempo em um tipo de dados mais coerente. Outra informação que precisa ser tratada são as *tags*, ou palavras-chave de um vídeo, que são retornadas como uma literal de várias palavras separadas por vírgula. Essas *tags* se repetem muito entre um vídeo e outro, então é fundamental que esses termos sejam tratados de forma

individualizada e única, permitindo a contagem de sua utilização e melhorando o cálculo da similaridade entre termos de um vídeo.

Após o pré-processamento dos dados, cabe ao algoritmo de aprendizagem de máquina classificar, através de um modelo preditivo, se o vídeo é um vídeo educacional e se pode ser considerado de qualidade ou não. Para isso, utilizou-se um conjunto de treinamento de 100 vídeos, contendo avaliações realizadas por especialistas e alunos que já tiveram em sua grade curricular o assunto “Orientação a Objetos / Herança”.

Um dos parâmetros que indica se um vídeo é educacional foi a propriedade “Categoria”, retornada pela API do YouTube. O YouTube define 15 categorias, uma delas chama-se “Educativa” que indica que o vídeo é educacional.

4.6. Engenho de Recomendação

Após a eliminação dos vídeos considerados inadequados, por não serem educacionais ou não possuírem os critérios mínimos de qualidade, são enviados ao engenho de recomendação os vídeos que restaram. O engenho de recomendação é responsável por receber a lista de vídeos considerados bons e ordená-los de acordo com sua relevância para o usuário ativo.

Nesta proposta, duas técnicas foram combinadas com o objetivo de melhorar as recomendações, reduzindo as possíveis desvantagens da utilização de uma das técnicas isoladas: a FBC e a FC. No entanto, existiram algumas particularidades referentes à FC que serão explicadas no decorrer desta seção. Quando há a combinação de duas ou mais técnicas, tem-se os chamados Sistemas de Recomendação Híbridos, que podem ser classificados de várias formas, como pôde ser acompanhado no Capítulo 2. Nesta proposta, a classificação mais adequada de acordo com Burke (2002) seria abordagem mista, na qual as técnicas utilizadas são apresentadas em uma mesma lista.

O algoritmo utilizado para gerar as recomendações é chamado de “fusão”, conhecido como o único algoritmo do tipo misto desenvolvido. Nele, as técnicas são executadas em paralelo e as recomendações são geradas tanto a partir da FBC como a partir da FC, conforme apresentado na Figura 14.

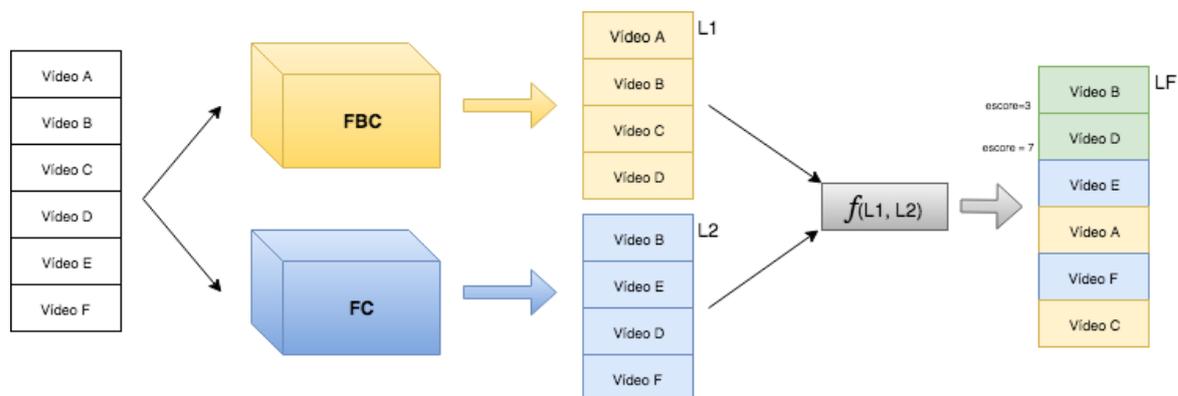


Figura 12 – Algoritmo fusão aplicado para gerar recomendações de vídeos.

Os vídeos recebidos pelo engenho de recomendação são submetidos, em paralelo, aos algoritmos de FBC e ao de FC; cada algoritmo disponibiliza sua lista de recomendações que são submetidas a uma função, em que as recomendações que aparecem nas duas listas recebem prioridade na lista de recomendação final. Para tanto, cada item presente em ambas as listas recebem uma pontuação, que é calculada a partir da soma de suas posições nas listas de origem. A lista final, então, é ordenada de forma crescente, onde quanto menor a pontuação de um vídeo, mais posicionado ao topo da lista final ele estará. Os vídeos que não estiverem presentes em ambas as listas serão adicionados ao final da lista, respeitando a seguinte regra: um item da lista de filtragem colaborativa e outro item da lista de filtragem baseada em conteúdo, de acordo com suas posições nas listas originais. Na Figura 14, apresentada acima, é possível acompanhar o seguinte exemplo: um “Vídeo B” que aparece na segunda posição na lista de FBC e que aparece na primeira posição da lista de FC terá sua pontuação igual a três, já que $(3 = 2 + 1)$, ocupando o primeiro lugar da lista final (LF) por ter a menor pontuação. Já o “Vídeo D” é o segundo elemento da LF e obteve pontuação sete, já que $(7 = 4 + 3)$. Esses dois elementos foram os únicos presentes em ambas as listas; depois disso, o “Vídeo E” foi o próximo selecionado, por estar mais acima, somente, na lista de filtragem colaborativa. O próximo item será o “Vídeo A”, presente, mais acima, somente na lista de filtragem baseada em conteúdo, e assim por diante.

Ao se gerar a lista de recomendação L2, foi utilizada a técnica Filtragem Colaborativa, que é baseada nas recomendações boca a boca, ou seja, que pessoas com gostos similares ao do usuário ativo poderiam ter indicado. Para que isso ocorra, é necessário armazenar as opiniões dos usuários sobre os vídeos. Depois disso, baseado nessas opiniões, pessoas com perfis semelhantes são agrupadas e os vídeos mais bem

avaliados por usuários com preferências similares são recomendados ao usuário ativo. Uma das principais desvantagens dessa abordagem é o problema do primeiro avaliador, também chamado de “*Cold Start*”, uma vez que novos usuários, que não possuem avaliações, não poderão ser comparados a outros, tendo que partir para um fluxo alternativo. Nesta proposta, como estamos utilizando o repositório de vídeos do YouTube, teremos dados que poderão ser utilizados para gerar as recomendações, mesmo sem avaliações prévias. A quantidade de visualizações, de curtidas e de não curtidas é levada em consideração e, apesar de não tratar a similaridade entre usuários semelhantes, pode-se conseguir informações relevantes, já que o YouTube é uma plataforma acessada por milhões de usuários que interagem tanto ao visualizar um vídeo, como também ao classificar um vídeo como bom ou ruim, trazendo informações demográficas importantes. Assim, se um usuário não possui avaliações suficientes ou usuários semelhantes, será possível recomendar utilizando esses dados extraídos dos vídeos do YouTube. Para tanto, utilizou-se a seguinte fórmula:

$$(\% \text{Curtidas} * 1) + (\% \text{ Visualizações}) * 2 / 3$$

A quantidade de visualizações de vídeos do YouTube pode ser exageradamente grande, mas quando se trata de inferir a qualidade do vídeo, este dado isoladamente pode gerar dúvidas sobre a qualidade, pois um usuário que visualizou pode não ter gostado do vídeo. Assim, o primeiro passo foi verificar a quantidade de pessoas que avaliaram um determinado vídeo (quantidade de curtidas + quantidade de não curtidas) e gerar o percentual de avaliações positivas desse vídeo (% curtidas). Após esse passo, calculou-se o percentual de visualização do vídeo baseado na soma de visualizações de todos os vídeos. Após essa etapa, calculou-se a média ponderada dos valores, sendo que para a quantidade de curtidas atribuiu-se um peso 1 e para quantidade de visualizações atribuiu-se um peso de 0,66. Considerou-se que a quantidade de curtidas deveria receber uma importância maior que a quantidade de visualizações, pois muitas vezes um vídeo recente pode não ter grande quantidade de visualizações, mas ser um vídeo muito bem avaliado por pessoas que o assistiram.

Conforme apresentado na Tabela 1, é possível verificar que o Vídeo D possui apenas quatro visualizações, e que duas pessoas, 50% das visualizações, curtiram. Uma solução para o problema das poucas avaliações é utilizar um meio de atribuir um “peso” aos coeficientes de similaridade, de tal forma que se o número de vídeos avaliados for

muito pequeno, esse peso reduzirá o valor da média. No projeto, verificou-se que se o número de visualizações for menor que 50, o valor da média será multiplicado por $n/50$, onde $n=50$.

Vídeos	Quantidade de Curtidas	Quantidade de Não Curtidas	Quantidade de Visualizações	Likes (%)	Visualizações (%)	Média Ponderada
A	1900	6	41073	0,9968520462	0,3134625658	0,5412590593
B	4700	13	89933	0,997241672	0,6863542700	0,7899834040
D	2	0	4	1	0,0000305274	0,3333536849
C	8	2	20	0,8	0,0001526368	0,2667684245

Tabela 1 – Algoritmo fusão aplicado para gerar recomendações de vídeos.

Já na FBC, o objetivo é medir a similaridade entre os vídeos e o tema selecionado. A principal técnica para medir similaridade entre textos é o TF-IDF (TORRES, 2004, p. 78). Neste trabalho, foram usados o título dos vídeos e a descrição, mas pretende-se, no futuro, utilizar também o conteúdo das legendas, a fim de obter melhores resultados com as palavras mais frequentes no conteúdo do vídeo.

4.7. Coletor de *Feedback*

O Coletor de *Feedback* faz parte de uma das últimas etapas do Sistema de Recomendação. Após os vídeos recomendados serem apresentados com o motivo pelo qual o vídeo foi recomendado, é chegada a hora de obter uma avaliação rápida das recomendações. O usuário ativo poderá dar uma nota ao vídeo, através da marcação de estrelas, que vai de 1 a 7, e essa nota servirá como entrada para as recomendações colaborativas e para analisar se as recomendações estão sendo satisfatórias.

Conforme apresentado na Figura 15, é possível acompanhar o momento em que o sistema apresenta uma recomendação trazendo 7 estrelas. Quando o usuário escolhe, as estrelas acendem, indicando que o usuário realizou a avaliação.

Easy v0.0.1-SNAPSHOT

Início Menu Conta Administração Idioma

Avaliar Vídeos



Curso Orientação a Objetos - #1 Abstração

DX

#01

ABSTRAÇÃO

Este vídeo foi recomendado na posição #1 pela enorme quantidade de curtidas e visualizações que obteve no YouTube, unido ao alto grau de similaridade com o tema buscado.

Descartar

Easy YouTube!

Figura 13 – Algoritmo fusão aplicado para gerar recomendações de vídeos.

5. Método e Avaliação

Antes da etapa de desenvolvimento do sistema proposto, foi submetido um formulário de pesquisa *online* a um grupo de 41 pessoas – dentre elas alunos e professores de uma instituição de ensino federal –, no qual se buscou compreender como a plataforma YouTube era utilizada, incluindo suas características de uso. Assim, percebeu-se que uma parcela alta dos entrevistados informou que usavam o YouTube todos os dias, como mostra a Figura 16.

Com que frequência você acessa o YouTube?

41 respostas

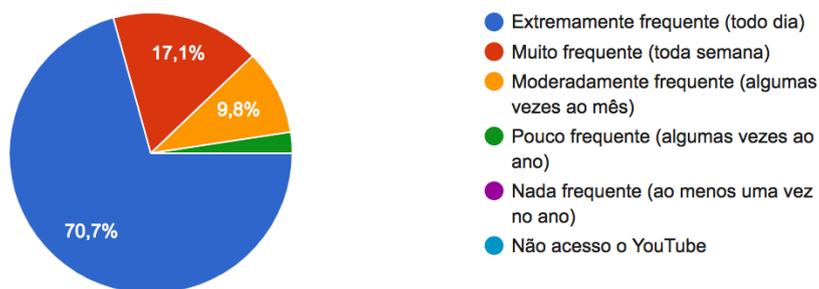


Figura 14 – Respostas sobre frequência de uso do YouTube

Ao questionar se o público da pesquisa utilizava o YouTube para fins educacionais, percebeu-se que mais da metade dos entrevistados usam pelo menos algumas vezes por mês com essa finalidade, como mostra a Figura 17.

Com que frequência você assiste video aulas no YouTube?

41 respostas



Figura 15 - Respostas sobre o uso do YouTube para fins educacionais.

Outras informações obtidas referem-se às principais características que são levadas em consideração pelos usuários na hora de escolher um vídeo. A qualidade da imagem e do som foram as características que mais receberam grau de importância, seguida do título do vídeo e da imagem da capa. Os gráficos contendo todas as informações podem ser encontrados no Apêndice A deste trabalho.

Já no que tange ao sistema proposto, para avaliá-lo, o primeiro passo foi realizar uma busca de vídeos diretamente na plataforma do YouTube, que já é conhecida pela grande quantidade de vídeos e pelo seu engenho de busca. É importante destacar que, no momento da primeira busca, não havia usuário autenticado para não interferir nas preferências do usuário. Conforme pode ser observado na Figura 18, ao buscar um termo, como “Herança”, vários vídeos foram recomendados, a exemplo de clipes de filmes, músicas, vídeos de humor, videoaulas sobre herança genética e, na posição 38ª, um vídeo sobre herança, teoria da programação orientada a objetos. Isso demonstra a imensa quantidade de informações que podem interferir na busca por vídeos educacionais. Essas informações, que não fazem parte do interesse do usuário, são os chamados ruídos.

Figura 16 – Informações retornadas ao buscar um termo “Herança”.

Conforme apresentado na Figura 19, é possível verificar que a mesma busca foi efetuada, utilizando o mesmo termo, só que dessa vez com o usuário autenticado. Um detalhe importante é que com este usuário alguns vídeos sobre a teoria da orientação a objetos foram assistidos. Diante dessas características, obteve-se na 5ª posição o vídeo que anteriormente estava na 38ª.

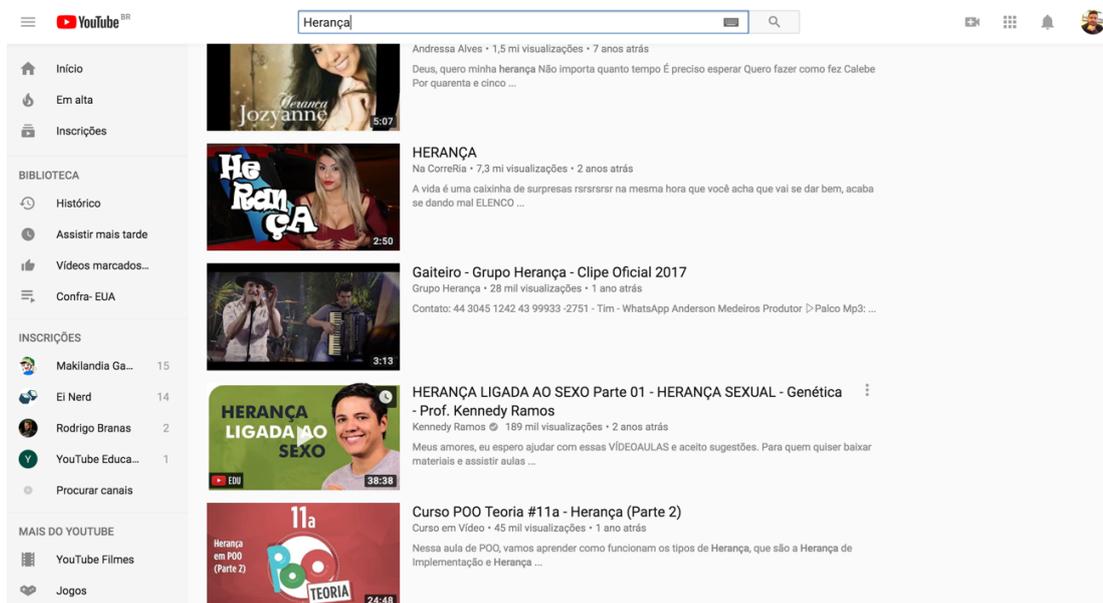


Figura 17 – Informações retornadas ao buscar um termo “Herança”, tendo feito login.

É possível inferir que o YouTube mantém dados de navegação, como ações que indicam que um usuário assistiu a um vídeo, por exemplo. Esses dados são utilizados para melhorar suas recomendações, mas ainda assim é possível visualizar ruídos, fazendo com que o usuário se empenhe para encontrar o vídeo que solucionará seus problemas. Uma opção é melhorar os termos a serem buscados até encontrar o vídeo ideal; mesmo assim, pode haver esforço. Outra opção é utilizar um SR que filtre os vídeos educacionais, descartando os vídeos referentes a clipes musicais, vídeos de humor ou até mesmo vídeos educacionais irrelevantes. Assim, recomendando apenas os vídeos mais significativos baseados no tema buscado e nas avaliações dos usuários.

Ao se utilizar a API do YouTube para buscar vídeos, mesmo enriquecendo as buscas com palavras de mesmo sentido, a lista de vídeos retornada pela API não é satisfatória. Existem muitos vídeos em outras línguas, vídeos ligados a outros temas, que nada querem dizer com programação orientada a objetos, vídeos de baixa qualidade, dentre outros. Isso corrobora a ideia de que enriquecer a busca com termos sinônimos pode ajudar a encontrar bons resultados. Isso é feito em casos como quando se busca um termo “POO” e é buscado também, implicitamente, “Programação Orientada a Objetos” para ampliar o leque de opções ou até mesmo especializar uma busca. Outro exemplo é quando se busca um termo como “Herança” e o sistema busca também “Generalização / Especialização”. Enriquecimento de consultas é uma etapa importante do processo, mas

não basta só encontrar boas opções de resultados, é importante eliminar as opções ruins, para minimizar os ruídos que são levados ao usuário.

Imagem	Título ↑↓	Id Vídeo ↑↓	Descricao ↑↓	Canal ↑↓
	Curso de Java #01 - História do Java - Gustavo Guanabara	sTX0UEpIF54	Aula Completa em http://www.cursoemvideo.com/lesson/historia-java/ A História da Linguagem Java se inicia em 1990, quando a Sun Microsystems criou uma ...	Curso em Vídeo
	Curso de Java #02 - Como Funciona o Java - Gustavo Guanabara	v_ZCtgwbS3o	Aula completa em http://www.cursoemvideo.com/lesson/funcionamento-java-curso-de-java/ Como funciona o Java? Você sabe o que é JVM, JRE, JDK, JavaC, ...	Curso em Vídeo
	Curso de Java #04 - Primeiro Programa em Java - Gustavo Guanabara	ijXXaYsVx4	Aula completa disponível em http://www.cursoemvideo.com/lesson/primeiro-programa-java/ Como criar um tradicional "Olá, Mundo!" (Hello World) em Java ...	Curso em Vídeo
	Curso de Java Passo a Passo para Iniciantes Aula 0001	NGUjvuYFbdE	http://www.videoinformatica.com.br Neste curso de java online para iniciantes você vai aprender a criar seus próprios programas de computador como um ...	Gilson Alves de Oliveira
	Java: Learn Java Programming For Beginners (Fast & Easy!)	00RxteR1oGQ	Get My BEST-SELLING Book, The Complete Software Developer's Career Guide For FREE < https://simpleprogrammer.com/yt/career-guide-free Java: Learn ...	Bulldog Mindset
	What is Java? - Java Programming Tutorial	2Xa3Y4xz8_s	http://www.guru99.com/java-platform.html This Beginners Tutorial gives introduction to JAVA Programming Language. It explains the use of JAVA and basic ...	Guru99

Figura 180 – Retorno da API do YouTube com ruídos.

Utilizando a API do YouTube para retornar vídeos relacionados ao tema, dessa vez com enriquecimento de consulta, foi possível verificar, conforme Figura 18, que a API do YouTube retornou em sua 5ª e 6ª posição vídeos sobre programação java, que correspondem à língua inglesa, o que consideramos ruído neste trabalho, pois a língua nativa dos usuários é o português, além desses vídeos referentes a direito de herança e a herança genética, que também são considerados ruídos, foram retornados e apoiam a ideia de que os resultados do YouTube podem e devem ser melhorados no contexto educacional.

Com base nesses indícios, o sistema proposto foi desenvolvido utilizando técnicas de Inteligência Artificial, para recomendar vídeos educacionais significativos e que sejam referentes ao tema selecionado. Como já foi mencionado, o engenho de recomendação dessa proposta utiliza o algoritmo fusão para combinar as técnicas “filtragem baseada em conteúdo” e “filtragem baseada em avaliações”; no entanto, antes de apresentar os

resultados desta técnica combinada, é possível verificar, na Figura 21, as recomendações geradas pela técnica FBC aplicada isoladamente.

Recomendações filtragem baseada em conteúdo							
#	Score	Título	Tags	Score	Avaliações	Algoritmo	
1	1(+)	 POO Conceito de Herança .		1.6875149		FBC	
2	2(+)	 Vídeo Aula- Conceitos básicos de Herança (POO)		1.2656362		FBC	
3	3(+)	 Curso POO Teoria #10a - Herança (Parte 1)	Curso de Programação a Objetos, Herança em POO, Gustavo Guanabara,	1.039565		FBC	
4	4(+)	 Curso POO Teoria #11a - Herança (Parte 2)	Curso de Programação a Objetos, Curso de POO, Gustavo Guanabara, Herança de Implementação, Herança para Diferença, Teoria de Herança ,	1.0338778		FBC	
5	5(+)	 Programação Orientada a Objetos (Parte 7): herança múltipla, o conceito de interface.	Java, POO, Netbeans IDE, Herança múltipla, Interface,	0.9362775		FBC	
6	6(+)	 Herança de classes - Programação Java - s01e09	POO, herança , conceito de herança , classe pai, classe filha, orientação a objetos, programação Java, Java,	0.88136184		FBC	
7	7(+)	 Herança em Java	java herança , Java (Programming Language), herança , mielinet,	0.85079193		FBC	
8	8(+)	 C# Aula 22 POO Conceito de Classe e Herança	Código logo, C#, programação, visual studio, jarvis, projeto, assistente,	0.84375745		FBC	
9	9(+)	 Aula Java 106 - Conceito de Herança (Superclasses e Subclasses)		0.82473457		FBC	
10	10(+)	 poo - Herança		0.7596836		FBC	

Figura 19 – Técnica de Recomendação FBC.

O sistema recomendou os vídeos com maior grau de similaridade com os termos buscados, ignorando as avaliações dos usuários do YouTube. Observou-se que o primeiro vídeo recomendado é relativamente recente, com apenas 1 (uma) curtida e apenas 17 (dezessete) visualizações. Apesar do título informar que se trata de conceitos de herança, o interlocutor menciona várias vezes a tecnologia C# (C Sharp), o áudio do vídeo possui ruídos e ele foi adicionado ao YouTube em 24 de maio de 2017. Assim, podemos inferir

que esta não seria uma boa recomendação, visto que com mais de um ano de disponibilidade na plataforma o vídeo recebeu apenas uma curtida. Percebe-se também que as 10 (dez) primeiras recomendações possuem o termo “Herança”, devido ao algoritmo utilizado que compara os termos buscados aos encontrados no “título” e nas “tags”. As recomendações 3 e 4 possuem boa qualidade de imagem e som e fazem parte da mesma aula que foi dividida em 2 (duas) partes.

Já na Figura 22, as recomendações foram geradas a partir da técnica híbrida, misturando as duas técnicas, através do algoritmo Fusão.

Recomendações - Fusão							
#	Score		Título	Tags	Score	Avaliações	Algoritmo
1	26(4+22)		Curso POO Teoria #11a - Herança (Parte 2)	Curso de Programação a Objetos, Curso de POO, Gustavo Guanabara, Herança de Implementação, Herança para Diferença, Teoria de Herança.	1.0338778	2751	Fusão
2	28(3+25)		Curso POO Teoria #10a - Herança (Parte 1)	Curso de Programação a Objetos, Herança em POO, Gustavo Guanabara,	1.039565	2480	Fusão
3	34(13+21)		Curso POO Teoria #12a - Conceito Polimorfismo (Parte 1)	Curso de Programação a Objetos, Curso de POO, Gustavo Guanabara, Polimorfismo, POO Polimorfismo, Sobreposição,	0.64959943	2824	Fusão
4	54(23+31)		Curso POO Teoria #13a - Conceito Polimorfismo (Parte 2)	Curso de Programação a Objetos, Curso de POO, Polimorfismo, Sobrecarga, Sobreposição, Override, Ad-hoc, Gustavo Guanabara,	0.5344189	1927	Fusão
5	59(29+30)		Curso POO Java #10b - Herança (Parte 1)	Curso de Programação a Objetos, Curso de POO, Herança em Java, POO em Java, Gustavo Guanabara,	0.41083857	1947	Fusão
6	63(30+33)		Curso POO Java #11b - Herança (Parte 2)	Curso de Programação a Objetos, Curso de POO, POO Java, Gustavo Guanabara, Herança em Java,	0.41083857	1819	Fusão
7	80(31+49)		Curso POO PHP #10b - Herança (Parte 1)	Curso de Programação a Objetos, Curso de POO, POO PHP, Herança em PHP, Gustavo Guanabara,	0.41083857	790	Fusão
8	87(37+50)		Curso POO PHP #11b - Herança (Parte 2)	Curso de Programação a Objetos, Curso de POO, POO Java, POO PHP, Gustavo Guanabara,	0.37542439	785	Fusão
9	92(12+80)		Entenda o Conceito de Herança #4	estudo na web, aprenda uml, como criar diagramas, html5, css3, php, desenvolvimento web, analise de sistema, modelagem de sistemas, engenharia de software, herança, conceito de herança, aprendendo herança, estudo sobre herança.	0.69331855	246	Fusão
10	107(11+96)		Genética - Aula 06 - Teoria da Herança Cromossômica	Univesp, Biologia, Genética,	0.7257506	88	Fusão

Figura 20 – Técnica Híbrida – Algoritmo Filtragem sem corte.

Percebe-se que o vídeo que havia ficado na 4ª (quarta) colocação na filtragem, baseada em conteúdo, agora assume o primeiro lugar na fusão. Isso aconteceu porque ao calcular a pontuação da técnica, somou-se a pontuação da FBC que foi 4 mais os pontos da filtragem baseada nas avaliações que foi 22, assim o vídeo ficou com score 26, tendo sido o menor do cálculo realizado pelo algoritmo de fusão. Percebeu-se também que mais vídeos de um mesmo canal, intitulado de “Curso em Vídeo”, apareceram nas sete próximas colocações. Esses vídeos aparentam ter sido editados de forma profissional, percebe-se melhor qualidade de som e imagem, o que pode ter acarretado nas avaliações positivas que recebeu. Apesar dos vídeos que ficaram na 3ª (terceira) e 4ª (quarta) posição não terem a palavra-chave “Herança” nos títulos ou *tags*, as recomendações foram pertinentes, uma vez que polimorfismo é um assunto complementar a herança. Através do conceito de herança, pode-se chegar ao de polimorfismo. Além disso, foram as únicas avaliações entre as 10 primeiras que não aparecem o termo “Herança”. O número de recomendações desse tipo teria sido maior se não houvesse sido incluído um limite de corte para a pontuação da técnica FBC. As pontuações abaixo do limite de corte, 0.29, foram descartadas e não foram usadas na técnica de fusão. Isso permitiu que recomendações mais relevantes e mais bem avaliadas fossem vistas nas primeiras posições.

Ainda na Figura 22, pode-se perceber, na posição 10, uma recomendação sobre “teoria da herança cromossômica”, que nada tem a ver aos termos pesquisados, exceto pela palavra “herança”. Isso aconteceu pelas boas avaliações que o vídeo recebeu e pelo fato de ter tido uma boa pontuação na técnica FBC: 0.72. No entanto, por se tratar de um tema com termo homônimo, com pronúncia e escrita idêntica, mas alheio ao do domínio, considera-se que essa não foi uma boa recomendação. Apesar de ter sido apresentada numa colocação distante das primeiras, ainda é possível utilizar técnicas para descartar os termos homônimos encontrados. Uma delas se obtém a partir das próprias avaliações das recomendações. Também pode-se usar ontologias para resolver o problema, no entanto considerou-se solucioná-lo nos trabalhos futuros.

As avaliações apresentadas a seguir demonstraram que o sistema proposto conseguiu obter índice de aceitação satisfatório por parte dos seus usuários. Para tanto, foram submetidas listas pré-processadas e disponibilizadas *online* a um grupo de 20 usuários, na maioria alunos de computação – total de 17 –, e mais 3 profissionais da computação, e foi pedido que analisassem a qualidade das recomendações dos vídeos, levando em consideração o tema “Conceitos da Programação Orientada a Objetos /

Herança” e os ruídos (vídeos não educacionais, vídeos em outras linguagens, vídeos sobre outros temas, vídeos de má qualidade) apresentados. Foram geradas 4 listas, cada uma com características diferentes:

- A) Lista de vídeos retornadas pelo YouTube ao se buscar o termo “Herança”;
- B) Lista de vídeos retornadas pelo YouTube ao se buscar o termo “Herança”, mas estando autenticado, tendo o usuário realizado recentes visualizações de vídeos ligados ao tema;
- C) Lista de vídeos retornados pela API do YouTube, tendo os termos buscados sofrido melhora através de enriquecimento de consulta;
- D) Lista de vídeos selecionados pelo sistema proposto, utilizando além do enriquecimento de consulta, a técnica de recomendação híbrida com filtragem baseada em conteúdo e avaliações do próprio YouTube.

A lista **A** obteve rejeição de 80%, sendo que 90% dos usuários informaram que tiveram que realizar um esforço grande para encontrar algum vídeo interessante nessa listagem. Já a lista **B** teve 75% de rejeição, e 55% dos usuários informaram que tiveram que realizar algum esforço para encontrar um vídeo interessante. A lista **C** recebeu 60% de rejeição, e os usuários informaram ter que realizar algum esforço para encontrar um vídeo interessante. Por último, a lista **D**, que foi resultado do sistema proposto, obteve 85% de aprovação, e os usuários informaram que atingiram seu objetivo ao escolher algum dos 3 primeiros itens recomendados.

Além do enriquecimento de consulta aplicado na última listagem, várias técnicas foram utilizadas, inclusive de Inteligência Artificial, como é o caso do algoritmo fusão, bastante comentado na abordagem híbrida de Sistemas de Recomendação (TORRES, 2014). Esses fatores certamente contribuíram para as avaliações positivas recebidas. Destaca-se, dentre outros, para as avaliações positivas, a classificação dos vídeos em “Educacionais” ou “Não Educacionais”, que conseguiu eliminar boa parte dos vídeos. Além disso, a utilização de outras informações sobre o vídeo, como linguagem, duração e avaliações pelos usuários do YouTube serviram para classificar os vídeos e gerar outra pilha de descarte, como foi o caso dos vídeos narrados em outra língua diferente da língua nativa do usuário.

Os vídeos remanescentes foram submetidos a duas técnicas de recomendação, a primeira denominada filtragem baseada em conteúdo que gera um cálculo de similaridade entre os termos do tema buscado e o título e descrição do vídeo. A saída dessa primeira técnica é um conjunto de vídeos ordenados por relevância segundo seu conteúdo. Já a segunda técnica, baseada em filtragem colaborativa, realiza um cálculo a partir das avaliações dos usuários do YouTube e retorna um conjunto de vídeos ordenados por relevância segundo suas avaliações. Esses dois conjuntos de vídeos são comparados e os que aparecem nas duas listagens são priorizados na lista final. Os vídeos restantes são adicionados alternadamente, de acordo com sua posição nos conjuntos gerados.

A combinação da FBC e FC tem sido realizada amplamente e tem conseguido bons resultados. A título de exemplo, pode-se mencionar o trabalho realizado por Torres (2004), que testou vários algoritmos em um sistema de recomendação de artigos e obteve 78% de boas recomendações numa categoria que chamou de “*Top-All*” e 28% numa categoria que chamou de “*Top-1*”, quase o dobro do segundo melhor algoritmo.

6. Conclusões e Trabalhos Futuros

Neste trabalho, apresentou-se um Sistema de Recomendação com o objetivo de obter vídeos educacionais baseado em um tema escolhido e nas avaliações que os vídeos receberam. Este sistema foi aplicado à plataforma de vídeos do YouTube, que possui uma diversidade de vídeos e avaliações de seus usuários. Tal aplicação serviu como um estudo de caso e se prestou inclusive como prova de conceito.

As principais contribuições deste trabalho referem-se à implementação de um Sistema de Recomendação que potencialmente poderiam ser utilizados como solução para vários domínios de aplicação, já que suas funcionalidades permitem o cadastro e gerenciamento de quaisquer temas. Sua arquitetura também pode ser usada como referência para outras aplicações, assim como também suas próprias funcionalidades que podem ser consumidas a partir de serviços REST. Além disso, serviu como prova de conceito do uso das técnicas utilizadas para aprimorar e melhorar as recomendações, tais como a utilização das avaliações dos usuários do YouTube, a linguagem nativa do vídeo, o descarte de vídeos que não são educacionais, além de uma proposta para o enriquecimento de consultas.

Dentre as características dos vídeos, destacou-se o uso das avaliações dos usuários do YouTube. Essas avaliações foram utilizadas para melhorar as recomendações geradas pelo uso isolado do algoritmo de filtragem baseada em conteúdo. Isso foi possível ao se utilizar a técnica de recomendação híbrida, na qual os vídeos são submetidos, em paralelo, aos algoritmos de FBC e ao de Filtragem Baseada em Avaliação (FBA). Posteriormente, uma pontuação é calculada de acordo com a aparição dos vídeos nas duas listas.

As avaliações apresentadas demonstraram que o sistema proposto conseguiu obter índice de aceitação satisfatório por parte dos seus usuários. No entanto, ainda existem algumas melhorias previstas que foram remetidas para trabalhos futuros.

Diante de tantas possibilidades de ação que surgiram durante a pesquisa, foi preciso focar em algumas características para o experimento, para que fosse possível concluir o trabalho a tempo, não sendo descartada a hipótese de utilizar as outras características no futuro. É possível, por exemplo, realizar análise de sentimentos nos comentários para identificar aspectos de qualidade do vídeo, eleger canais confiáveis de publicação de vídeos, inferir sobre vídeos não relevantes, dentre outros. Pretende-se, no futuro, resolver o problema dos termos homônimos que fazem com que vídeos não relacionados ao tema estudado sejam apresentados nas listas de recomendações. Além

disso, um dos problemas presentes na recomendação de vídeos, sobretudo ao se utilizar FBC, encontra-se na limitação de acesso ao conteúdo. Muitas vezes, o título e descrição não retratam o conteúdo do vídeo. Para esse problema, pretende-se obter a legenda dos vídeos a fim de utilizar seu conteúdo na técnica FBC e assim melhorar ainda mais as recomendações. Percebeu-se que há, ainda, uma gama de possibilidades que podem ser incrementadas, dentre elas a utilização dos metadados dos vídeos e a utilização de algoritmos de aprendizagem de máquina para analisar e descobrir padrões de qualidade nos vídeos.

Referências

ALMEIDA, Rafael Ferreira de *et al.* Recomendação de recursos educacionais para grupos em redes sociais. 2016. Disponível em: <https://repositorio.ufjf.br/jspui/>. Acesso em 10/11/2018.

BALABANOVIĆ, Marko; SHOHAM, Yoav. Fab: content-based, collaborative recommendation. **Communications of the ACM**, v. 40, n. 3, p. 66-72, 1997.

BASU, Chumki *et al.* Recommendation as classification: Using social and content-based information in recommendation. In: **Aaai/iaai**. 1998. p. 714-720.

BERK, Ronald A. Multimedia teaching with video clips: TV, movies, YouTube, and mtvU in the college classroom. In: **International Journal of Technology in Teaching and Learning**, v. 5, n. 1, p. 1-21, 2009.

BURKE, R. Hybrid recommender systems: survey and experiments. **User Modeling and User-adapted Interaction**, 2002; 12(4):331-370.

CAZELLA, Sílvio César; NUNES, M. A. S. N.; REATEGUI, Eliseo. A Ciência da Opinião: Estado da arte em Sistemas de Recomendação. André Ponce de Leon F. de Carvalho; Tomasz Kowaltowski..(Org.). Jornada de Atualização de Informática-JAI, p. 161-216, 2010.

COLLINS, Mauri; BERGE, Zane L. Technological minimalism in distance education. **The Technology Source**, 2000.

Computer-Based Medical Systems (CBMS), IEEE 26th International Symposium on. IEEE, 2013. p. 257-261.

DAVIDSON, James *et al.* The YouTube video recommendation system. **Proceedings of the fourth ACM conference on Recommender systems**. ACM, 2010. p. 293-296.

DIAS, Alessandro da Silveira. Aumentando a Acurácia de Predição de Avaliação de Sistemas de Recomendação de Vídeo com o Uso de Pontos de Interesse. 2013. 79 f. Dissertação (Mestrado em Computação) – Instituto de Informática, Universidade Federal do Rio Grande do Sul.

FACELI, K. *et al.* **Inteligência Artificial: Uma Abordagem de Aprendizado de Máquina**. 1º. ed. Rio de Janeiro: LTC, 2011.

GOOGLE INC., “YouTube Data API”. 2018. Disponível em: <https://developers.google.com/youtube/v3/>.. Acesso em 12/11/2018.

GOOGLE INC., Estatísticas do YouTube, 2018. Disponível em: <https://www.youtube.com/intl/pt-BR/yt/about/press/>. Acesso em: out./2018.

HENRIQUES, Artur de Carvalho Montenegro. Rec-Vest: um sistema de recomendação de vestuário em lojas de comércio eletrônicos. 2017.

KEARSLEY, Greg; MOORE, M. G. **Educação a distância: uma visão integrada**. São Paulo: Thomson Learning, 2007.

MATTAR, João. YouTube na educação: o uso de vídeos em EaD. São Paulo: Universidade Anhembi Morumbi, 2009.

McKINNEY, Dani; DYCK, Jennifer L.; LUBER, Elise S. iTunes University and the classroom: Can podcasts replace Professors? *Computers & Education* 52 (2009) 617–623. Disponível em: <http://www.fredonia.edu/departament/psychology/pdf/CAE1263.pdf>. Acesso em: 13/11/2018.

PARK, D. H., KIM, H.K., CHOI, I.Y., KIM, J.K. A literature review and classification of recommender Systems research. **Expert Systems with Applications**, 2012. p. 10059 - 10072.

QIN, S., Menezes, R., and Silaghi, M. 2010. A recommender system for YouTube based on its network of reviewers. Proceedings of the IEEE 2nd International Conference on Social Computing (SocialCom). 323-328.

QUINLAN, J. Ross. **C4. 5: programs for machine learning**. Elsevier, 2014.

RIVERO-RODRIGUEZ, A. *et al.* A health information recommender system: Enriching YouTube health videos with Medline Plus information by the use of SnomedCT terms. Disponível em: <https://ieeexplore.ieee.org/document/6627798/>. Acesso em 20/10/2018.

ROLIM, Vitor et al. Um Estudo Sobre Sistemas de Recomendação de Recursos Educacionais. In: **Anais dos Workshops do Congresso Brasileiro de Informática na Educação**. 2017. p. 724.

SARWAR, Badrul et al. Item-based collaborative filtering recommendation algorithms. In: **Proceedings of the 10th international conference on World Wide Web**. ACM, 2001. p. 285-295.

SCHMIDT, Douglas C. et al. **Pattern-Oriented Software Architecture, Patterns for Concurrent and Networked Objects**. John Wiley & Sons, 2013.

SILVA, Priscylla & Costa, Evandro & Pinheiro, Roberth. (2014). A Predictive Model for Video Lectures Classification. 21. 10.5753/cbie.sbie.2014.21.

TYBEL, D. O. **TCC, uma escrita em blocos**. 1^o. ed. São Mateus: King Cross Publicações, 2015.

SILVA, Priscylla; COSTA, Evandro; PINHEIRO, Roberth. A Predictive Model for Video Lectures Classification. Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). 2014. p. 21.

TAHERI, Sona; MAMMADOV, Musa; BAGIROV, Adil M. Improving naive Bayes classifier using conditional probabilities. In: **Proceedings of the Ninth Australasian Data Mining Conference-Volume 121**. Australian Computer Society, Inc., 2011. p. 63-68.

TORRES, R. **Personalização na Internet**. 1º. ed. São Paulo: Novatec Editora LTDA, 2004.

WANG, Guosheng. A survey on training algorithms for support vector machine classifiers. In: **Networked Computing and Advanced Information Management, 2008. NCM'08. Fourth International Conference on**. IEEE, 2008. p. 123-128.

YANG, Jie Chi et al. An automatic multimedia content summarization system for video recommendation. **Journal of Educational Technology & Society**, v. 12, n. 1, p. 49, 2009.

YOUTUBE. Estatísticas. 2016. Disponível em: <https://www.youtube.com/yt/press/pt-BR/statistics.html>. Acesso em 03/10/2018.

APÊNDICE A – PESQUISA REALIZADA COM ALUNOS E PROFESSORES DE UMA INSTITUIÇÃO DE ENSINO FEDERAL

PERGUNTAS RESPOSTAS 41

41 respostas

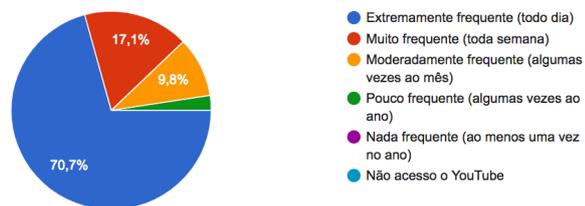
RESUMO INDIVIDUAL

Aceitando respostas

Informações Gerais

Com que frequência você acessa o YouTube?

41 respostas



Com que frequência você assiste video aulas no YouTube?

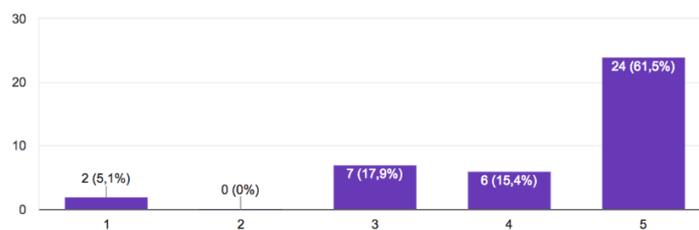
41 respostas



Em uma escala de 0 a 5, qual a relevância das características de vídeos do YouTube que você leva em consideração na hora de selecionar uma vídeo aula para assistir?

Título do vídeo?

39 respostas



Descrição do vídeo?

39 respostas

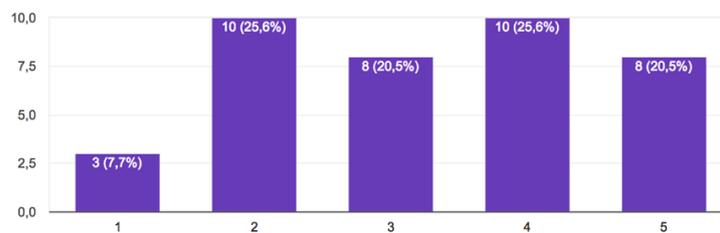
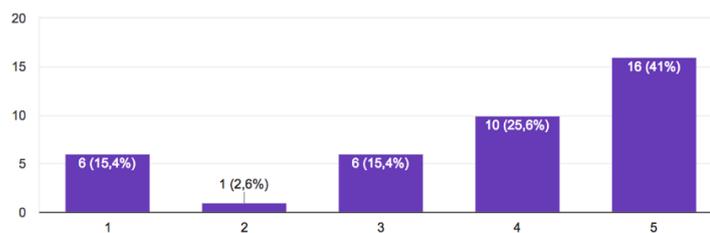


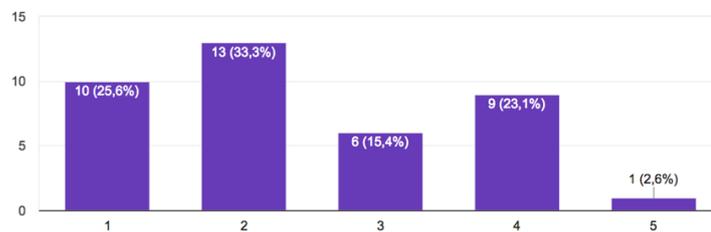
Imagem de capa do vídeo (Thumbnail)?

39 respostas



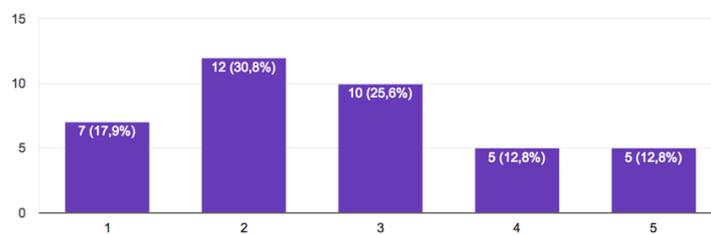
Quantidade de Comentários do vídeo?

39 respostas



Conteúdo dos Comentários do vídeo?

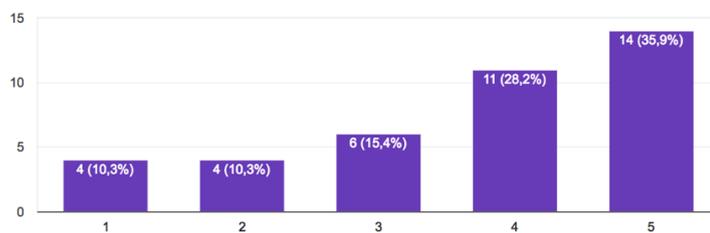
39 respostas



Duração do vídeo?

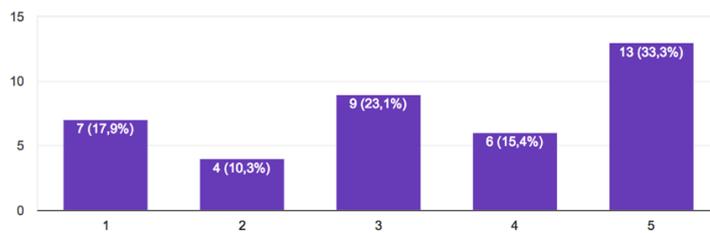


39 respostas



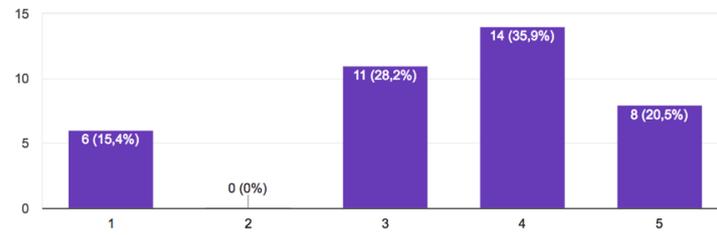
Data da postagem do vídeo?

39 respostas

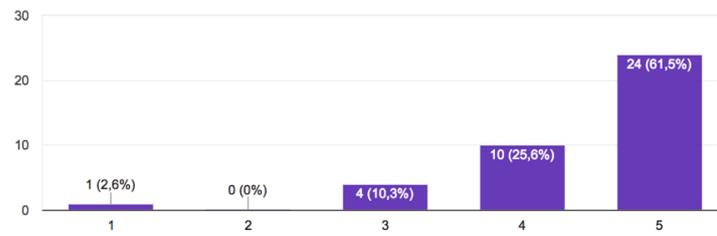


Autor/canal do video?

39 respostas

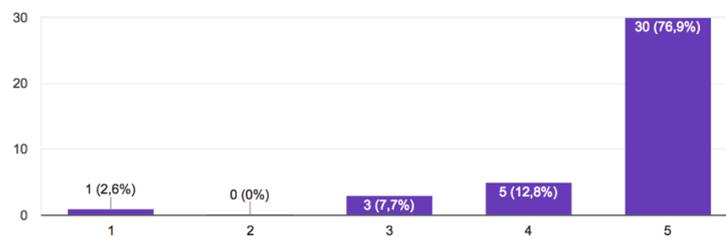
**Qualidade do som?**

39 respostas



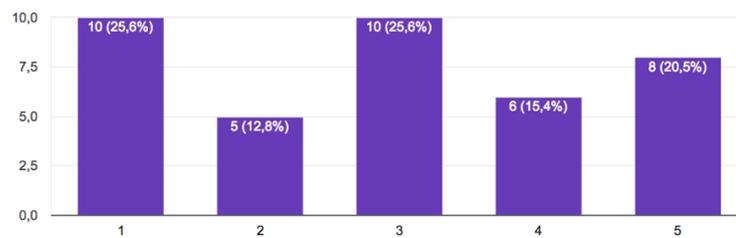
Qualidade da imagem?

39 respostas



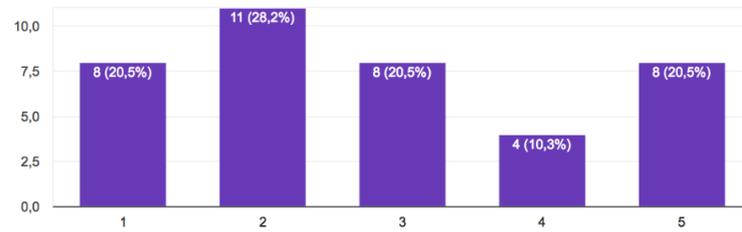
Quantidade de curtidas?

39 respostas



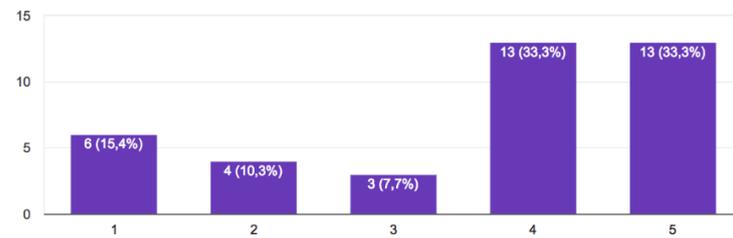
Quantidade de "não gostei" (dislikes)?

39 respostas



Quantidade de visualizações?

39 respostas



Se existir alguma outra característica que você leva em consideração ao selecionar uma video aula para assistir no YouTube, por favor, informe aqui:

3 respostas

No decorrer de poucos minutos levo em consideração se a aula é clara e objetiva, não ser explicada repetidas vezes e com bons exemplos conta bastante.

eu gosto de vídeo aulas que possuem um conteúdo didático e de fácil absovição :)

se ele tem legenda em outro idioma