Projeto de Trabalho de Conclusão de Curso

Sistema para Exibição, Busca e Recomendação de Vídeos

Lohan Rodrigues Narcizo Ferreira

JUIZ DE FORA

JULHO, 2019

Sistema para Exibição, Busca e Recomendação de Vídeos

Lohan Rodrigues Narcizo Ferreira

Universidade Federal de Juiz de Fora

Instituto de Ciências Exatas

Departamento de Ciência da Computação

Bacharelado em Ciência da Computação

Orientador: Eduardo Barrére

JUIZ DE FORA

JULHO, 2019

Aos meus amigos.

Aos pais, pelo apoio e sustento.

Resumo

Este trabalho aborda os principais aspectos arquiteturais e o desenvolvimento de um sistema de exibição e busca de vídeos com mecanismo de navegação sobre o conteúdo e recomendação de mídias relacionadas. Tal sistema tem como objetivo melhorar a extração de conhecimento da mídia vídeo, ao oferecer uma melhor categorização e formação de conjuntos de metadados além de funcionalidades para interação com o conteúdo da mídia buscada. Para tal, são utilizadas ferramentas de segmentação de vídeo em cenas e algoritmos de processamento de linguagem natural como o ranqueamento de termos quanto à relevância em relação ao contexto.

Palavras-chave: multimídia, recuperação de informação, processamento de linguagem natural, segmentação baseada em tópicos

Conteúdo

Lista de Figuras

1 Introdução

Diversos sistemas estão produzindo, a todo momento, novos conjuntos de dados, novas informações e novos conceitos. Acessá-los também deixou de ser um problema, com a expansão da inclusão digital e o crescente avanço da tecnologia. Meios não faltam pra se procurar uma informação, seja em buscadores na internet, redes sociais, bibliotecas on-line ou até mesmo em aplicações de conversa.

Mas em meio a tanta informação, resta-nos a questão de como convertê-la em conhecimento, como organizar e categorizar conjuntos de informações, que por vezes podem ser ambíguas, de forma a dá-las sentido. A busca de vídeos, por exemplo, é constantemente afetada por esta má organização, seja por título mal elaborado ou disponibilização em local inadequado, levando a resultados de pouca relevância em relação ao que foi buscado.

A área de multimídia tem importante papel neste problema, visto que dela surgem inúmeras formas de acesso a informação que hoje possuímos. O processamento de um vídeo, a qualidade de uma imagem, de um fragmento de audio são tratadas neste ramo, e no projeto aqui apresentado, será feito uso deste conceito de multimídia no que diz respeito ao processamento de um vídeo e como exibir, seja no próprio vídeo ou em mídias adicionais paralelas ao vídeo, os conhecimentos pertinentes daquele conteúdo de forma mais clara e palpável.

Para tornar tal objetivo possível, não só a área de multímidia tem grande importância neste projeto, mas também a área de recuperação de informação que trata da busca de informações relevantes sobre um dado conteúdo, dados ou metadados, através de todo material disponível: textos, sons, imagens etc. E tal como definido sobre esta área de pesquisa, neste projeto é feito o uso de algumas técnicas de extração de informações de vídeos e processamentos sobre estas informações de forma a identificar o que pode ser relevante, o que é importante ser notado, e como organizar tal informação que será exibida a alguem que deseje buscá-la.

1.1 Problema

Neste projeto, é feito o uso da ferramenta desenvolvida no aritgo (23) capaz de fazer a distribuição das informações disponibilizadas em um vídeo, através da transcrição do áudio contido no mesmo utilizando ASR (Automatic Speech Recognition) e de processos de anotação semântica, em conjuntos denominados cenas, onde toda a informação contida em uma dada cena está fortemente relacionada. Apesar de ser utilizado neste projeto a extração de informações via áudio, existem projetos que tratam de outros elementos do vídeo para adquirir informações, por exemplo o encontrado em (7) que utiliza uma ferramenta de OCR (Optical Character Recognition) para extrair informações de textos contidos nas imagens do vídeo, e utiliza destes dados para fazer categorização e indexações.

Tomando como base os dados de divisão em cenas e transcrição de áudio obtidos pela ferramenta citada anteriormente, o problema em questão tratado aqui é, dado um conjunto de informações extraídas de um vídeo, tais como o contexto de uma determinada cena, o que pode ser feito para disponibilizar esse material ao usuário de uma forma melhor do que simplesmente um vídeo. Que tipos de mecanismos adicionais podem ser feitos de modo a melhorar a interação do usuário com este conjunto de dados, obtidos com a ferramenta de extração(23), para adquirir o máximo possível de conhecimento, informação útil para seu objetivo? Inerente a este problema, vem o fato de que tais respostas são obtidas através do processamento de linguagem natural, ser capaz de transformar em código o entendimento de palavras e frases para definir contextos e significados, tarefa que apesar de já ser desenvolvida a algum tempo e possuir algoritmos eficientes, ainda possui barreiras a serem superadas.

1.2 Justificativa

Seja no meio acadêmico ou fora dele, a busca por informação através deste tipo de mídia, o vídeo, é constante, porém nem sempre a forma como o mesmo está sendo disponibilizado realmente ajuda aqueles que o procuram a atingir seus objetivos. Um título inadequado para um vídeo, categorização ou palavras chaves do vídeo que não condizem com o conteúdo contido no mesmo, entre outros metadados errados podem dificultar bas-

tante a busca e identificação de informações relevantes, o que torna necessário uma melhor análise sobre tais materiais, uma melhor distribuição e definição sobre seus conteúdos.

Mesmo no caso onde a mídia disponibilizada já possui uma boa definição e categorização, alguns mecanismos possuem falhas quanto a fornecer uma informação ao usuário. Por exemplo, tomando como contexto o meio acadêmico, é complicado identificar um tópico específico em uma vídeo aula de longa duração, exigindo que muito mais tempo seja gasto com a busca da informação do que com o aprendizado que vem com a mesma, o que pode impactar diretamente no rendimento de um aluno. Um mecanismo que forneça buscas específicas em mídias deste tipo não só tornaria mais agradável a busca pela informação como também poderia melhorar o desempenho acadêmico a longo prazo.

1.3 Objetivo

Considerando a utilização da transcrição automática de áudio (ASR) para transcrição do áudio de vídeos, ferramentas de segmentação de vídeo em cenas, algoritmos de ranqueamento de palavras e mecanismos de busca automática de conteúdos através da internet, o objetivo deste projeto é desenvolver uma ferramenta que seja capaz de suprir esta lacuna nas mídias, capaz de fornecer para vídeos um mecanismo de busca eficiente que permita ao usuário acesso rápido a informação, seja capaz de relacionar de forma coerente o assunto abordado em um vídeo à outras mídias: imagens, textos ou outros vídeos; e disponibilizá-las em paralelo para um melhor aproveitamento de tempo e de conteúdo. Para alcançar tal objetivo, alguns objetivos específicos são importantes:

- Identificar trabalhos relacionados e quais ferramentas foram utilizadas.
- Avaliar diferentes ferramentas de processamento de linguagem natural e sua capacidade de integração com o sistema.
- Avaliar diferentes ferramentas de busca automática de informação na internet.
- Desenvolvimento do sistema de disponibilização de mídias com o auxílio das ferramentas encontradas.
- Coleta de uma base de dados de vídeos para testes.
- Identificação de pontos fortes e fracos do sistema.

2 Revisão Bibliográfica

Este capitulo apresenta uma breve análise da literatura sobre os conceitos aplicados no desenvolvimento deste projeto, evolução das áreas pertinentes, estado de desenvolvimento das ferramentas existentes utilizadas no sistema e também outros projetos com funcionalidades ou objetivos similares onde as mesmas são utilizadas.

2.1 Multimídia

As formas de comunicação utilizadas pela humanidade são alguns dos aspectos mais marcantes quando se quer analisar a evolução histórica da civilização, de tochas a cartas, estes mecanismos utilizados em qualquer era sempre estiveram em constante modificação. Na atualidade, com todo o avanço da tecnologia disponível não seria diferente e algumas formas de comunicação vão perdendo seu espaço em prol de outras (8).

A digitalização da informação facilitou o desenvolvimento de um nova forma de comunicação, que combina de uma só vez varias informações que anteriormente estariam separadas, como imagens e sons, com a utilização de um computador. Atualmente esta ideia é aplicada com frequência nas tarefas mais simples de uma pessoa em sua rotina, por exemplo o uso de vídeos que podem simultaneamente disponibilzar informações visuais e auditivas, ou as diversas interfaces gráficas de programas que fornecem informação junto a interação.

Frente a nova forma de disponibilização de informação, novos estudos se originaram para pesquisar a fundo seu funcionamento, formando a área de pesquisa Multimídia. Segundo (8), o conceito de multimídia diz respeito a utilização de múltiplas técnicas de transmissão de informação num diálogo homem-máquina como sons e animações, textos junto a figuras estáticas que, em geral, estão distribuídas e disponibilizadas de forma não sequencial e interativa, as informações conseguem ser acessadas livremente e independentemente.

2.2 Extração de informações em vídeos

A facilidade de produção e transmissão de dados com o uso das tecnologias mais atuais que aplicam esse conceito de multimídia iniciaram uma nova fase e um novo problema, o Big Data.

Big Data é uma área que trabalha com grandes quantidades de dados, geralmente não estruturados, que visam, através de inúmeras tecnicas chamadas soluções de big data, extrair informações e conhecimento dessas massas de dados.

> "Big Data não é apenas um produto de software ou hardware, mas um conjunto de tecnologias, processos e práticas que permitem às empresas analisarem dados a que antes não tinham acesso e tomar decisões ou mesmo gerenciar atividades de forma muita mais eficiente."(20)

Big Data também é definido como cinco características importantes dessa área : volume, variedade, velocidade, veracidade e valor (20), que representam as formas de dados com que se trabalham (volume e variedade), e os desafios ao desenvolver técnicas que processem estes dados.

Seguindo os conceitos de Big Data, atualmente existem muitas técnicas de extração de informações para um dos mecanismos de transmissão de informações mais famoso, o vídeo. Visto sua composição, atualmente existem muitas técnicas e ferramentas visando a extração de cada um destes dados que são utilizados para muitos fins.

Focando inicialmente na parte visual do vídeo, muitos trabalhos podem ser encontrados na literatura que extraem os dados das imagens e frames que ocorrem pelo vídeo e fazem processamento com estes dados.

O artigo (18) trabalha com o desenvolvimento de um algoritmo capaz de extrair, dentre os frames dos vídeos, aqueles considerados principais ou que façam melhor subdivisões no vídeo em questão, informação que pode ser muito útil por exemplo para sumarização do conteúdo de um vídeo. De maneira similar, também trabalhando com a extração de frames, mas com objetivos diferentes temos (6), que foca em combinar as informações de uma curta sequência de frames de um vídeo em um único frame que contenha todas as informações unicas e importantes de forma a manter a resolução espacial observada ao assistir os frames em sequência.

De cada frame de um vídeo é possivel extrair inúmeros dados, com isso qualquer tipo de objetivo pode ser aplicado na extração das imagens que compoem um vídeo, por exemplo o artigo (4) trata um problema de extração das imagens de um video focando na identificação de carros, independente do tipo, que estejam se movendo ao longo de uma rua mantendo o mínimo de erros possível. De maneira similar, temos algo muito comum nas câmeras atualmente que utilizam técnicas de normalização na imagem para identifiação de rostos, como o desenvolvido em (3).

Um exemplo interessante, e que mais se aproxima do foco deste projeto, visto que sua motivação também é a organização de um conjunto de mídias, é (5) que desenvolve um mecanismo de busca num banco de imagens e vídeos onde as buscas são feitas utilizando outras imagens e vídeos dos quais padrões de cores, textura e formato são extraídos e utilizados como parâmetros na busca por informações relacionadas.

Uma ferramenta importante neste ramo de extração de informação de vídeos é o OCR (*Optical Character Recognition*), que tem como função reconhecer e transcrever os textos contidos em uma imagem. Existem vários trabalhos focados no desenvolvimento e análise de ferramentas OCR como em (15) que faz uma avaliação de um dos OCRs mais famosos e gratuitos disponíveis, o Tesseract[1] desenvolvido pela Google; e (2) que faz um relato histórico sobre o funcionamento, as técnicas de reconhecimento e os diferentes sistemas de OCR já desenvolvidos.

Também existem trabalhamos que aplicam o OCR em problemas similares aos já anteriormente citados como o de sumarização de vídeos em (10), que processa os textos extraidos de imagens ou vídeos para realizar segmentações de partes do vídeo, que segundo o autor é de grande utilidade para indexação de conteúdo da mídia.

A lista de trabalhos desenvolvidos que utiliza desta ferramenta é extensa provando que só a imagem de um vídeo já oferece um mar de informações que abre portas para muito desenvolvimento. Mas um vídeo não possui somente imagens, temos uma outra grande fatia de informações que pode ser extraída de seu áudio e, assim como a imagem, o som tem sido amplamente pesquisado. Como dito anteriormente, este projeto faz uso de uma ferramenta de reconhecimento de fala e transcrição, e tal ferramenta realiza seu

[1] https://opensource.google.com/projects/tesseract

trabalho sobre o som extraído de um vídeo.

Na literatura é possivel encontrar muitos trabalhos que tratam desse reconhecimento como (16) e (1) que se preocupam com o tratamento dos ruídos durante o processamento, elemento que frequentemente causa erros em sistemas de diálogo. Indo além do simples reconhecimento do que foi falado em um vídeo, é possível encontrar trabalhos focados na identificação de sensações e emoções transmitidas pelo som do vídeo, em (24) o projeto desenvolvido foca na identificação das emoções de um discurso utilizando diretamente os sinais de som, evitando a fase de transcrição que poderia acumular erros e dificultar o processo.

Além do reconhecimento da fala, existem trabalhos concentrados na extração direta de características dos áudios no intuito de criar melhores categorizações e organizações, objetivo compartilhado com o deste projeto. Em (13) é desenvolvido um framework para facilitar a extração de características dos áudios através de programação genética de modo a permitir integração entre vários métodos famosos de extração de características.

2.3 Processamento de linguagem natural

A utilização de ferramentas de transcrição, seja de som ou de imagem como as citadas na seção anterior, nos fornece uma grande quantidade de informações que podem ser extraídas de um vídeo e convertidas em texto, porém interpretá-las é outro desafio. Compreender o sentido e o contexto do que está sendo falado para tomar conclusões e definir características sobre o que está sendo dito implica em, de alguma forma, ensinar a máquina a entender a lingua humana. Este ramo da ciência da computação tem o nome de processamento de linguagem natural. Segundo Oliveira (12):

> "A tarefa de processar uma linguagem natural permite que os seres humanos comuniquem-se com os computadores da forma mais 'natural' possível, utilizando a linguagem com a qual mais estão habituados. Elimina-se, desta maneira, a necessidade de adaptação a formas inusitadas de interação, ou mesmo o aprendizado de uma linguagem artificial, cuja sintaxe costuma ser de difícil

aprendizado e domínio, a exemplo das linguagens de consulta a bancos de dados."

O autor também descreve o processo de interpretação de uma linguagem natural sendo divida em três partes:

- Análise morfológica - Que identifica os elementos básicos de uma linguagem natural, suas palavras, sentenças e símbolos.
- Análise sintática - Se baseia na gramática da linguagem natural para relacionar os elementos identificados na análise morfológica através de árvores de derivação.
- Análise semântica - Fase focada em atribuir sentido e significado às relações formadas na análise sintática.

Esta área está intimamente ligada a recuperação de informação, e para melhor organização do processamento dos dados, alguns algoritmos estatísticos precisam ser usados para dar peso a termos ou sentenças conforme suas relações nos textos ou documentos. Em (11) é apresentado mais a fundo este relacionamento entre processamento de linguagem natural e recuperação de informação caracterizando e apresentando prolemas em sistemas de recuperação de informação e como estes problemas poderiam ser solucionados com o uso de técnicas de processamento de linguagem natural, além da apresentação de alguns métodos estatísticos mais famosos.

Não se limitando a métodos estatísticos, é possível encontrar trabalhos que utilizam de redes neurais no processamento de linguagens naturais para melhorar a identificação de termos e significados como pode ser visto em (17).

3 Metodologia

Este trabalho tem como natureza de pesquisa a pesquisa qualitativa visto que é focado na aplicação de um conjunto de módulos para o desenvolvimento de uma ferramenta na qual a qualidade de uso e a opinião de seus usuários é material principal para moldar suas funcionalidades e como são oferecidas.

Os tipos que definem este projeto são o de pesquisa bibliográfica, experimental (ou de laboratório) e aplicada pois trata do desenvolvimento de um sistema com o objetivo de resolver um problema específico da área, para tal a busca por informações atualizadas são essenciais no desenvolvimento da ferramenta assim como a interpretação dos módulos que serão integrados: como funcionam, quais suas vantagens em relação a outros, métodos para melhor utilizá-los e combiná-los. O projeto utilizará de testes e simulações das partes desenvolvidas constantemente para avaliar o funcionamento do sistema, quais partes atingem ou não as expectativas e quais elementos devem ser inseridos, alterados ou removidos para melhores resultados.

A coleta de dados se deu por meio de questionários e entrevistas visando obter informações sobre o comportamento de um usuário ao interagir com o sistema e sua opinião qualitativa : o que espera encontrar, que objetivos alcançou ou não, se cumpriu com o que foi proposto, entre outras informações.

O elemento de estudo mais importante neste projeto é o vídeo, portanto o primeiro passo no caminho da extração de informações do mesmo é buscar entendimento de quais informações estão disponíveis dentro de cada vídeo, se é possível separá-las, qual a dependência entre suas partes e o que é possível fazer com cada parte, além de encontrar ferramentas que de fato permitam a extração de um dado.

Após estudo da estrutura do vídeo, é necessário decidir quais partes do mesmo usar, selecionar qual a ferramenta mais recomendada para a extração destes dados e como eles são gerados a fim de projetar como serão futuramente processados. Além da preocupação com a ferramenta, deve se saber da origem dos vídeos que serão usados no sistema, qual a qualidade dos mesmos e como suas características afetam a ferramenta a

ser utilizada para extração. Para este projeto a escolha tomada foi de utilização da parte de audio e a divisão temporal do vídeo como informações basicas a serem extraídas. O próximo capítulo explica a ferramenta escolhida e seu funcionamento.

As informações extraídas do vídeo permitem uma melhor caracterização do assunto que ele trata, porém alcançar esse conhecimento exige algum tipo de processamento sobre os dados extraídos seja através de equações ou redes neurais. Uma vez adquirido esse conhecimento nosso objetivo é exibir mais mídias sobre este mesmo assunto de modo a melhorar, por exemplo, o aprendizado com o vídeo. Novamente é necessário estudar qual o melhor tipo de mídia inserir, analisar como integrar tais mídias junto ao vídeo de maneira a não atrapalhar quem o assiste e, por último, como obtê-las.

Conhecer o assunto de um vídeo permite, além da aquisição de mídias adicionais, a chance de criar um relacionamento entre um conjunto de vídeos de modo a decidir se seus assuntos tem ideias em comum e utilizar desta informação para gerar recomendações entre eles. Independente do processamento utilizado para definir a relação entre dois vídeos, é de se esperar que o tempo de execução nem sempre seja rápido, logo para evitar futuros transtornos pela parte do usuário é importante que estes dados sejam previamente obtidos e armazenados em algum lugar de fácil acesso para uso posterior, em resumo um banco de dados que facilite o manuseio da informação.

Uma vez definido todo o conjunto de funcionalidades e elementos que o sistema deve possuir, o próximo passo é decidir como melhor disponibilizá-los aos usuários. Projetar uma interface que permita acesso ágil e agradável ao que deseja ser mostrado, decidir quais tecnologias usar no desenvolvimento da mesma, tendo em mente que ela deverá se comunicar com outros softwares (banco de dados, por exemplo). Por fim, realizar testes sobre esta interface para avaliar quais são seus pontos fortes e fracos, quão impactante foi a escolha do tipo de informação a ser extraida no entendimento dos assuntos dos vídeos, quão util e/ou precisa foi a obtenção e apresentação de mídias adicionais, o que precisa ser melhorado e até onde é possível melhorar. O sistema desenvolvido neste projeto foi nomeado S.E.R. (Search, Exibition and Recomendation of Videos).

4 Estrutura do sistema

Este capítulo visa explicar com mais detalhes os elementos que compôem o sistema desenvolvido, como foram desenvolvidos, como se relacionam e quais as escolhas importantes tomadas ao utilizar cada uma.

4.1 Informação do vídeo e Segmentação

Como dito anteriormente, a primeira escolha a ser tomada sobre o vídeo foi de quais elementos seriam extraídos e utilizados no processo de interpretação do vídeo. No contexto deste projeto a decisão tomada foi de utilizar o áudio e divisão temporal do vídeo por se tratarem de elementos cuja ferramenta de extração já se encontrava em desenvolvimento em outro projeto (23) e portanto de fácil acesso e aprendizado de seu funcionamento.

A ferramenta é composta por três módulos principais, um módulo de reconhecimento de fala ASR (Automatic Speech Recognition), que foca na transcrição de todo audio contido em um vídeo para pequenos arquivos texto; um módulo de anotação semântica que utiliza uma base de dados previamente desenvolvida para definir conceitos, ou assuntos, a que se referem os textos obtidos pelo módulo de transcrição e por último um framework de geração de cenas que utiliza as informações resultantes do módulo anterior e define intervalos de tempo, contendo assuntos específicos, chamados cenas ou tópicos. Estes tópicos são os elementos principais do sistema S.E.R, sendo usado como base para todos os processamentos a serem feitos posteriormente. A figura 4.1 mostra, de maneira simplificada, a estrutura da ferramenta utilizada.

4.2 Termos de Busca

Em posse dos arquivos resultantes da ferramenta de extração do vídeo, é necessario obter o contexto de cada tópico, qual exatamente é o assunto tratado em cada fragmento do vídeo. O método utilizado aqui foi a aplicação de um algoritmo de processamento de

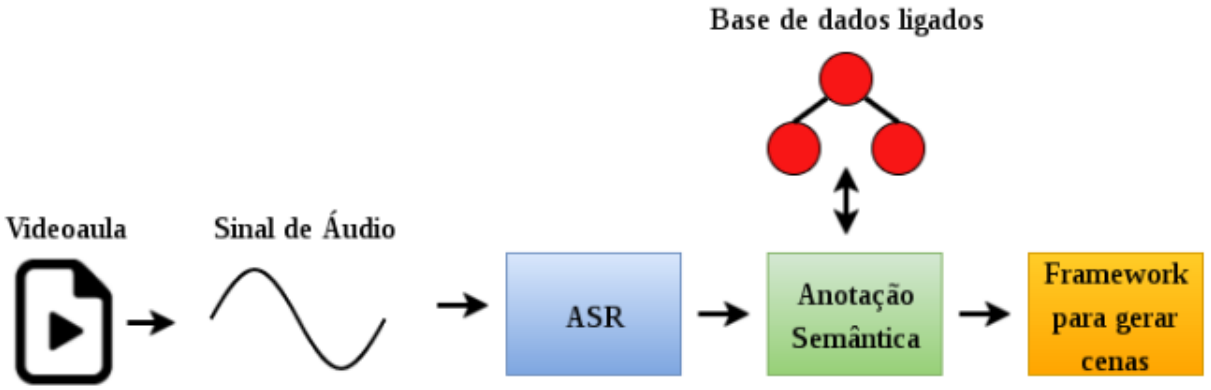

Figura 4.1: Representação esquemática do funcionamento da ferramenta de segmentação de vídeo em tópicos(cenas). Imagem extraida da referência (23)

linguagem natural: o TF-IDF (do inglês Term Frequency - Inverse document frequency) que consiste no cálculo da frequência com que os termos aparecem em textos e na contagem dos documentos para definir os termos mais importantes, conforme a equação 4.1, onde a é um termo ou palavra contido no conjunto de documentos, t é um fragmento de transcrição do áudio obtido através da ferramenta de extração, $f_{a,t}$ representa quantas vezes o termo a aparece no fragmento t, N representa o número de fragmentos de texto e n_a o número de fragmentos (documentos) em que o termo a aparece.

$$TFIDF(a,t) = (1 + \log(f_{a,t})) \times \log\left(\frac{N}{n_a}\right) \tag{4.1}$$

Sendo nosso objetivo entender o assunto de cada cena, utilizamos a divisão temporal, obtida com a segmentação, para definir qual conjunto de fragmentos de texto representa cada tópico, e cada um desses conjuntos passa então pelo calculo TF-IDF a fim de se definir as palavras mais relevantes ou que melhor representem aquela cena.

Essas palavras são essenciais para várias funcionalidades do sistema, uma delas é a navegaçao por termos de busca. A ideia desta funcionalidade é fornecer uma interface aos usuários para que através de uma busca simples com palavras, seja possível navegar entre as bases do sistema e encontrar vídeos relevantes ao tema desejado.

O algoritmo de busca consiste de duas etapas: Na primeira o conjunto de palavras utilizadas para realizar a busca é enviada para um módulo de tokenização que separa termo a termo a frase de busca e entao passa por um processo de *Stemming* (11), que consiste na extração do radical de uma palavra que pode ser flexionada ou gera palavras derivadas com

sentido similar (exemplo aumentativos e diminutivos), e por uma avaliação de stopwords, remoção de elementos que pouco contribuem para os algoritmos de comparação e busca como artigos e preposições.

Terminada a primeira fase, a segunda consiste da formação de *n-grams* (uma sequência de n termos retirada da frase de busca original) de tamanhos variando de um elemento (o próprio token) até o número de elementos que a frase continha (que é a propria frase). Em seguida uma busca é feita comparando estes *n-grams* com os conjuntos de palavras chave de cada cena para identificar quais cenas melhor atendem a solicitação do usuário. O uso de stemização e stopwords reduzem o tamanho dos termos finais que serão usados nas comparações durante a fase de busca, acelerando portanto a velocidade com que as comparações são feitas para que não impactem no tempo de espera do sistema uma vez que este mecanismo foi projetado pra ser frequentemente utilizado.

4.3 Mídias Adicionais

Além do mecanismo de busca, a interpretação do assunto das cenas, proporcionado pelas palavras-chave, permite a busca por outros elementos que possam auxiliar a compreensão do usuário em relação ao que é falado no vídeo.

A ideia das mídias adicionais é utilizar destas palavras-chave e realizar buscas em outras bases de dados a fim de encontrar outras formas de informações estáticas (outras mídias), que possam ser apresentadas em paralelo ao vídeo sem atrapalhar o foco no mesmo, e fornecer uma melhor explicação sobre algum(ns) elemento(s) importante(s) relacionado às cenas.

A implementação das mídias adicionais no sistema consiste de unir estas palavras chaves mais importantes de cada cena e, conforme ocorrem mudanças de cenas no decorrer do vídeo, realizar requisições para duas bases de pesquisas famosas, a Wikipedia e o Google. As requisições são feitas de maneira diferenciadas para cada uma das bases conforme a existência ou não de APIs que facilitem este processo.

No caso do Google, utilizamos a Google Custom Search Engine, que fornece meios para criar uma guia de busca configurável a ser incorporada em páginas HTML. Através desta guia requisições são feitas cena a cena com o uso das palavras chaves e os resultados

são mostrados aos usuários em um visual similar à uma busca realizada no próprio site do Google.

No caso da Wikipedia, a requisição é feita através de uma API disponibilizada pela mesma para acessar diretamente a base de dados do MediaWiki utilizando HTTP GET. As requisições também são feitas cena a cena utilizando as palavras chaves, porém os resultados obtidos são reorganizados antes de serem mostrados aos usuários uma vez que a resposta da requisição é um arquivo no formato JSON contendo um conjunto de informações (conforme solicitado na montagem da requisição HTTP GET) sobre páginas da wikipedia que tenham a ver com os termos de busca utilizados.

4.4 Recomendação de Vídeos

O último módulo importante do sistema é o de recomendação de vídeos. Um mecanismo para aprimorar a navegação do usuário no sistema sugerindo vídeos com assunto potencialmente similar ao do vídeo que está sendo assistido no momento.

Diferente dos módulos anteriores, este módulo não se baseia somente nas palavras-chave, mas em toda a transcrição das cenas obtido pela ferramenta de extração e, enquanto os outros módulos processam durante a utilização do sistema, este é majoritariamente processado durante a construção da base de vídeos uma vez que necessita de cálculos mais complexos e demorados.

O processamento utilizado aqui para calcular o relacionamento entre os vídeos é feito através de um método de processamento de linguagem natural chamado Word2Vec (21). Word2Vec é um grupo de modelos relacionados que são usados para produzir integrações de palavras. Esses modelos são compostos de redes neurais de duas camadas que são treinadas para reconstruir o contexto ou semântica das palavras. O Word2vec recebe como entrada um grande corpus de texto do qual produz um espaço vetorial, cujo tamanho das dimensões podem variar conforme a precisão desejada, onde cada palavra única no corpo é atribuída a um vetor de valores correspondente no espaço. Os vetores são posicionados no espaço vetorial de modo que as palavras que compartilham contextos comuns no corpus estejam localizadas próximas umas das outras no espaço.

Existem duas arquiteturas de modelo diferentes sobre a qual o Word2Vec funci-

ona, o CBOW (do inglês Continuous bag of words) e Skip-gram que processam de maneira diferente o conjunto de palavras apesar de possuirém o mesmo objetivo final.

De forma resumida, o modelo CBOW define conjuntos de palavras onde a ordem não importa e dá probabilidade de prever os termos conforme o contexto, através dos resultados define os valores de similaridade entre textos. O modelo Skip-Gram segue uma topologia semelhante ao do CBOW porém a forma de predição é invertida, dada uma palavra qualquer o modelo tenta reconstruir o contexto ao qual ela pertence. Mais detalhes sobre a eficiência e desenvolvimento destes modelos podem ser encontrados em (21).

Para o cálculo de relacionamento dos vídeos da base do sistema, utilizamos um modelo previamente treinado disponibilizado pelo Google, contendo 3 milhões de palavras e frases em inglês sendo cada um destes relacionado a um vetor de dimensão 300.

Tendo em mãos um modelo pronto, o processamento de similaridade acontece sempre que um novo vídeo é inserido na base de vídeos, neste momento os arquivos de transcrição correspondentes a cada cena do vídeo são enviados ao modelo para serem comparados cena a cena com todos os outros vídeos ja existentes na base e os valores de similaridade entre cada par de cenas é armazenado em um banco de dados. Uma vez guardado todos estes valores, enquanto os usuários estiverem assistindo a uma determinada cena de um vídeo no sistema, uma requisição ao banco de dados é feita para descobrir quais outras cenas possuem valor de similaridade acima de um valor específico e todos os que satisfazem esta condição são então listados e recomendados.

4.5 Construção das bases de dados

Esta seção visa explicar melhor o funcionamento da conexão entre os módulos citados anteriormente visto que alguns deles exigem comunicações entre si e com bancos de dados para armazenamento de informações coletadas.

O primeiro componente é um servidor capaz de armazenar e processar vídeos que vão sendo inseridos na base visto que após a obtenção de informações através da ferramenta de segmentação, estes dados ainda precisam passar por vários outros módulos, cada um gerando um resultado diferente que deve ser armazenado em algum lugar.

O servidor foi desenvolvido utilizando a linguagem Python com um conjunto de bibliotecas que permite o desenvolvimento de uma API com arquitetura REST (Representational State Transfer) por ter conceitos simples, de fácil entendimento e acesso quanto aos serviços fornecidos. A arquitetura REST (9) trabalha com a ideia de um conjunto de recursos do servidor, acessados através da montagem de URIs que refletem diretamente no que deseja ser feito com tal recurso (Ex: localhost/api/upload , quando se deseja enviar um arquivo).

No caso deste servidor, a API conta com um único serviço responsável por pegar um vídeo junto às informações extraídas do vídeo e realizar todo o pré-processamento necessário, incluindo a montagem das bases de dados. Este serviço faz uma verificação sobre a base onde o vídeo deseja ser inserindo, criando uma nova caso necessário e então utiliza os módulos anteriormente citados para fazer o processamento sobre o vídeo: gerar palavras-chave, fazer os cálculos de relacionamento entre as cenas do vídeo inserido e as cenas de todos os vídeos já existentes na base desejada, guardar os videos e enviar os dados de pré-processamento para o banco de dados.

A figura 4.2 mostra de maneira simplificada o processo do servidor desde o recebimento de um vídeo segmentado até a distribuição dos dados.

Como visto na figura 4.2, após processado todos os dados, estes são separados entre dois locais diferentes. A base de vídeos é um conjunto de arquivos locais que permite ao sistema acessar, por exemplo, os arquivos de vídeo MP4 já enviados para posterior disponibilização.

O outro local é um banco de dados orientado a grafos, o Blazegraph(22). Seguindo uma premissa similar ao de um banco relacional, porém com meios de armazenamento e obtenção de dados diferente, este banco foca no armazenamento de triplas em uma estrutura de sujeito-predicato-objeto (14) seguindo o modelo RDF, um padrão definido pela W3C (World Wide Web Consortium) para processamento de metadados. Considerando os tipos de dados que seriam guardados com a necessidade de que seu acesso fosse mais rápido e simples, sendo estes as palavras chaves de cada cena, a divisão de cenas dos vídeos e o relacionamento entre cenas, observou-se que a estrutura do blazegraph fornecia os mecanismos para suprir esta necessidade. A organização em triplas torna mais

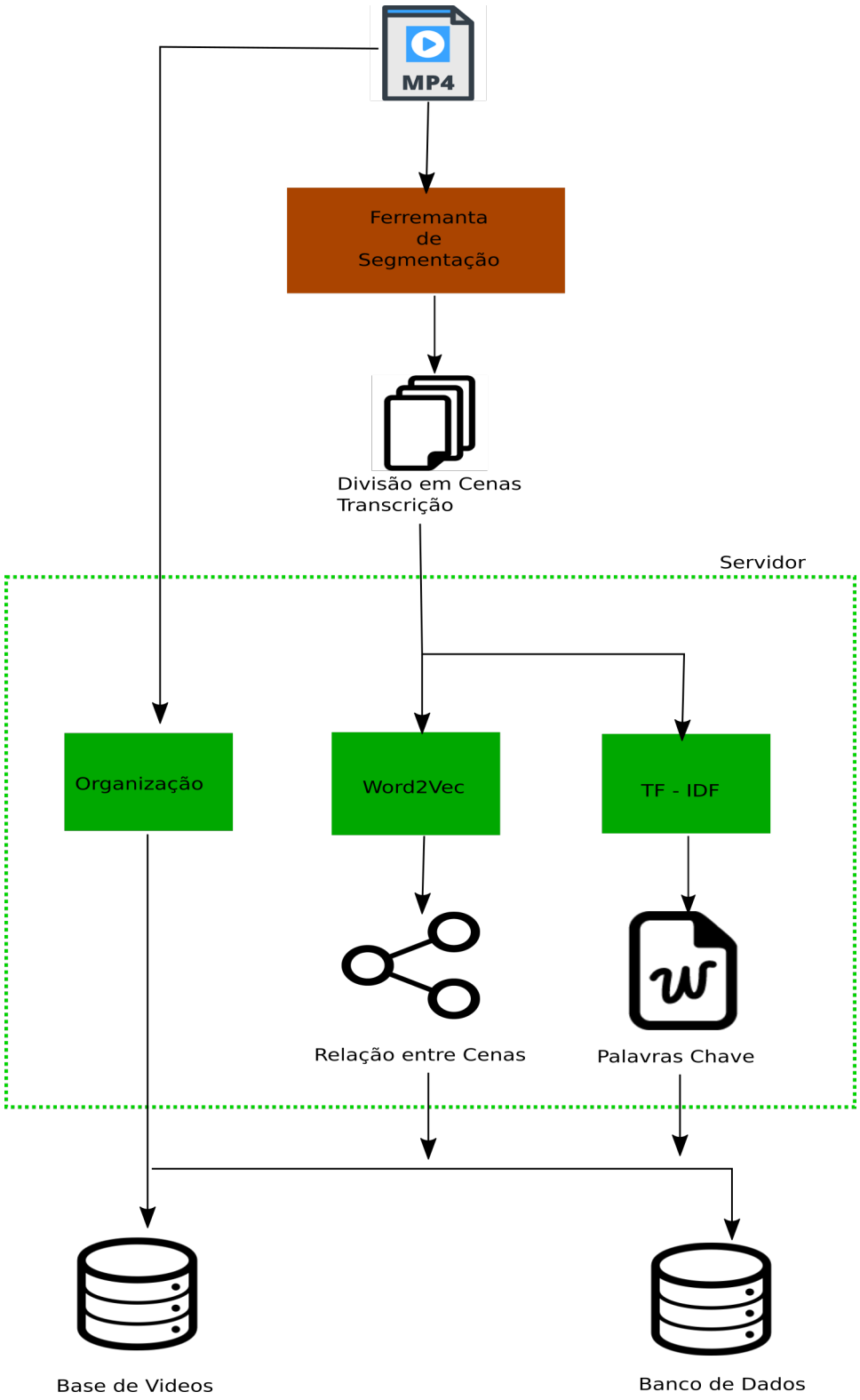

Figura 4.2: Diagrama mostrando a aplicação e organização feita pelo servidor

intuitivos o armazenamento de dados de relacionamento dos vídeos e outros atributos que são necessários na exibição e a aquisição destas triplas é feita através do SPARQL (19), uma linguagem de consulta para dados no modelo RDF em bancos de dado, também padronizado pela W3C.

4.6 A Interface Web

O último elemento que compõem este sistema é a interface com a qual o usuário poderá interagir com o sistema, depois de processado todos os vídeos e feitos seus relacionamentos todas estas informações tem que estar acessíveis ao usuário.

O desenvolvimento desta aplicação web foi feita utilizando Python CGI (Common Gateway Interface) que permite o desenvolvimento de páginas web não somente com scripts HTML previamente escritos, mas através de outros executáveis e outras linguagens que geram páginas dinamicamente.

A ideia de usar CGI parte principalmente da necessidade de fazer alguns processamentos no próprio front-end do sistema. Como a página gerada para cada vídeo possui elementos únicos referentes somente àquele vídeo, alguns cálculos e processamentos são feitos imediatamente ao acessar a página e os resultados obtidos são utilizados para gerar, dinamicamente, um script HTML que será a página visualizada pelo usuário.

A interface web é divida em duas partes importantes, ambas fazendo bastante uso dos termos de buscas que foram previamente gerados e armazenados no blazegraph. A primeira parte é a busca inicial por um vídeo através destes termos de busca que faz requisições diretamente ao banco de dados, com os termos de busca inseridos pelo usuário e verifica quais vídeos, de uma base previamente selecionada, possui dentre suas palavras chaves, partes dos termos utilizados. Os termos utilizados pelo usuário passam primeiramente por um processo de stemização (recuperação de radicais ou parte principal da palavra) e em seguida são combinadas numa requisição e enviadas ao blazegraph ,que retorna uma lista com os vídeos que melhor se adequarem aos termos solicitados. Esta lista de vídeos é mostrada ao usuário que, ao selecionar um deles é então direcionado a segunda parte da interface. A imagem 4.3 mostra a página inicial do sistema (que também pode ser acessada através do link http://35.198.36.227/), após uma requisição utilizando

a palavra de busca "Software"sobre um dos bancos de dados atualmente definidos no sistema. A tela possui design simples contando com uma guia de busca no centro da página junto a seleção de base de dados e os resultados obtidos são exibidos na região inferior da tela.

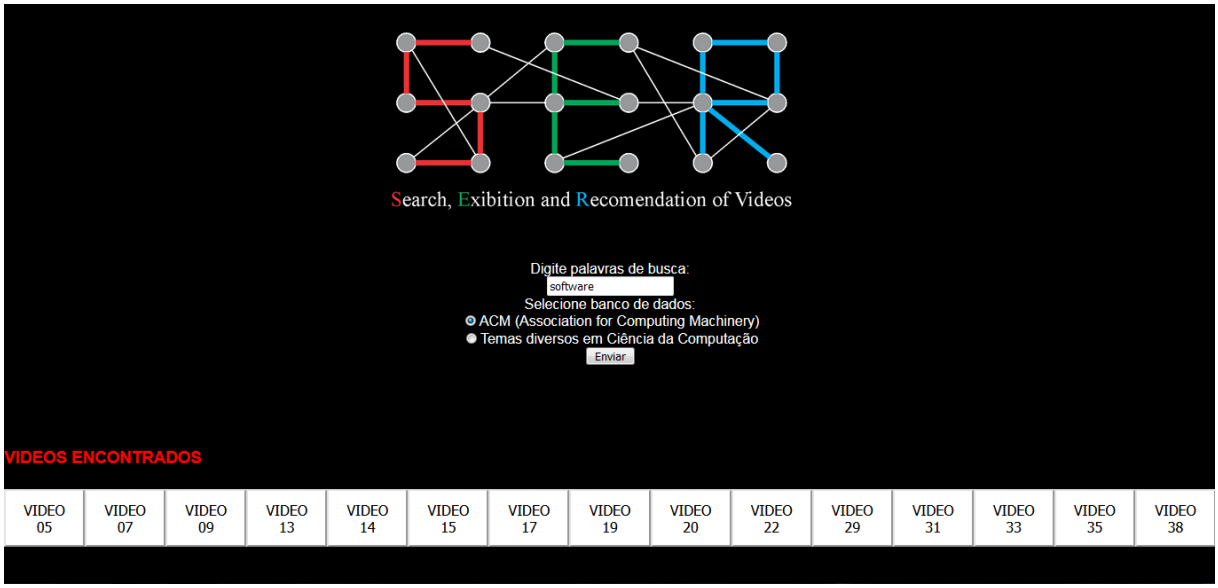

Figura 4.3: Tela inicial do sistema S.E.R.

Esta parte utiliza de todos os dados obtidos em processamentos prévios nos módulos explicados anteriormentes. Nesta página existe uma guia de busca, similar à guia de busca existente na primeira parte, esta também usa os termos de busca e o processo de stemização, porém desta vez a busca nao é sobre todos os vídeos da base, ela é feita somente para o conjunto interno de dados daquele vídeo. A intenção desta guia de busca é permitir uma melhor navegabilidade sobre o vídeo uma vez que os vídeos podem ser longos e nem toda a informação contida nele é o que deseja ser buscado. Esta guia de busca quando utilizada com um conjunto de termos de busca, faz uma comparação com as palavras-chave de cada cena e, para toda cena que possuir palavras coincidentes com as utilizadas na busca, um botão é gerado na própria pagina que permite ao usuário navegar diretamente ao momento em que a cena começa no vídeo.

A página possui guias que permitem ao usuário acessar as mídias adicionais obtidas através dos mecanismos de busca do Google e do Wikipedia como citado anteriormente na seção de Mídias Adicionais. As requisições para obtenção de vídeos são feitas seguindo a temporização do vídeo e sempre que uma mudança de cena é detectada, incluindo as

mudanças de cena forçadas pela navegação do usuário, uma atualização é feita nas guias de mídia adicional com novo conteúdo possivelmente relevante para a cena em questão.

O último elemento importante a ser falado desta página é a aba de recomendação de vídeos. Quando a página é gerada, além da obtenção dos termos de busca via requisição ao Blazegraph, uma outra requisição é feita para obter os valores de relacionamento das cenas do vídeo atual com as cenas dos outros vídeos pertencentes à mesma base. Em posse destes valores, uma vez que uma cena é iniciada durante o decorrer do vídeo, uma análise é feita sobre estes valores comparando com um valor mínimo e todos que forem superiores a este são considerados boas recomendações para a cena em questão e são exibidos numa lista de tópicos relacionados que permite ao usuário acessar diretamente outro vídeo, na cena específica que foi considerada relacionada, de maneira rápida.

A figura 4.4 mostra a tela da segunda parte do sistema, alcançada após uma busca realizada na tela inicial. Ocupando uma parte central da tela, encontramos o player do vídeo propriamente dito, seguido imediatamente abaixo por uma guia de busca que fornece a navegação entre cenas com uso de palavras chave ateriormente explicado e os resultados das cenas obtidas são exibidos imediatamente abaixo na forma de botões que quando selecionados fazem saltos na duração do vídeo até o ponto onde aquela cena começa. Abaixo da guia de busca temos duas grandes abas colapsáveis de mídias adicionais contendo informações sobre as duas bases de busca anteriormente citadas no desenvolvimento deste módulo, o Google e a Wikipedia. A figura 4.5 mostra a busca de mídias adicionais em funcionamento.

Finalmente ocupando a parte direita da tela temos a lista de vídeos relacionados que é obtida cena a cena utilizando os valores de relacionamento pré processado, como detalhado na seção de recomendação de vídeos, ao selecioná-los o usuário é enviado ao vídeo em questão na cena considerada mais relacionada a cena do vídeo que estava sendo assistido e , por motivos de praticidade, uma busca na guia de navegação da nova página é realizada utilizando o último termo buscado no vídeo anterior.

Com a utilização do Python-CGI todas as tarefas relacionadas a obtenção de dados (requisições ao blazegraph e acesso a arquivos) e processamento dos mesmos, como os mecanismos de busca, são feitos pelo script incial em Python desta página. O script

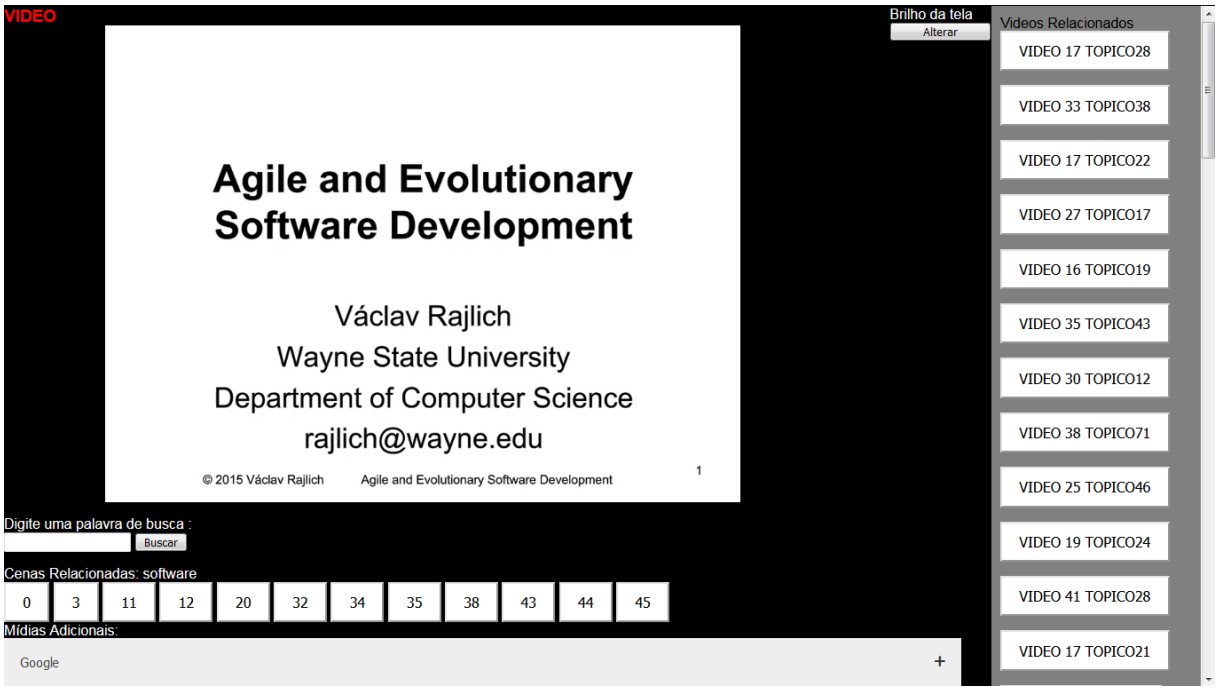

Figura 4.4: Tela de um vídeo do sistema

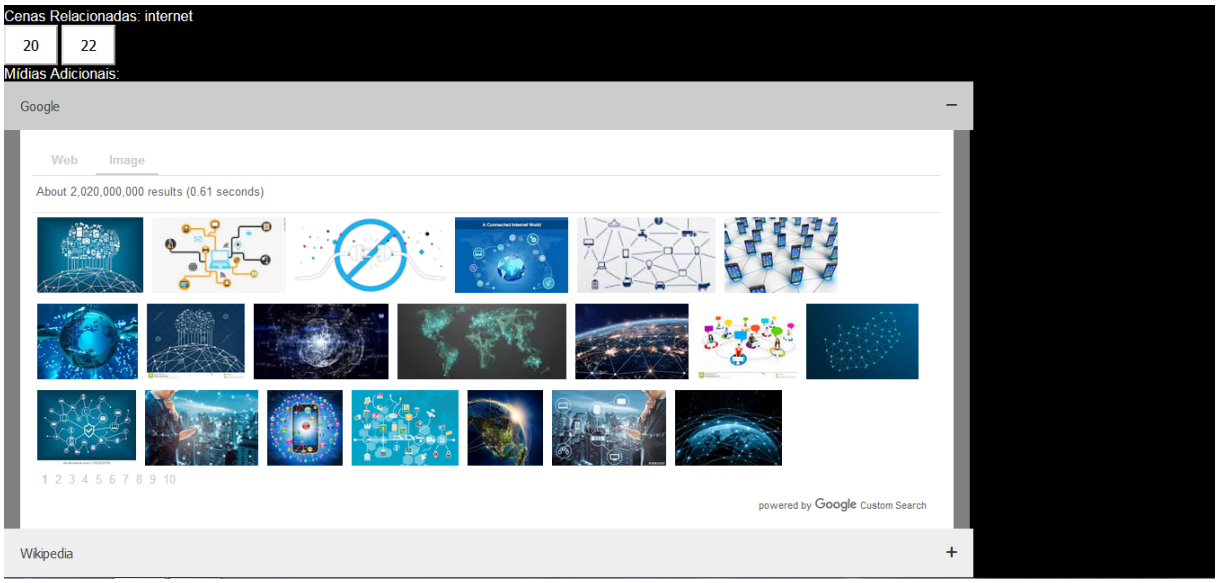

Figura 4.5: Tela de mídias adicionais

Python então gera um Script HTML, o próprio front-end, com fragmentos de Javascript que são responsáveis por alterações dinâmicas da página, a atualização de mídias adicionais e da lista de recomendação de tópicos conforme ocorrem as trocas de cenas, cujo desenvolvimento é facilitado pela simplicidade da transferência de dados entre os dois scripts, proporcionado pelo CGI.

5 Testes

Para avaliação do sistema S.E.R. um teste foi realizado com estudantes da UFJF do curso de Licenciatura em Computação (EAD, Ensino a Distância) enquanto cursavam a disciplina de Sistemas Multimídia no primeiro semestre de 2019 (2019/1), no total 37 alunos participaram do teste. O teste consistiu em fornecer acesso à interface do sistema para o conjunto de alunos após preencher o sistema com duas bases de vídeos distintas contendo vídeos sobre a área de Ciência da Computação. As bases consistiam de vídeos informativos tais como vídeo aulas e vídeos contendo seminários sobre problemas e temas variados da área em questão. Após o contato durante alguns dias com o sistema, para se familiarizarem com as funcionalidades e interfaces, um questionário foi aplicado contendo perguntas sobre as funcionalidades e usabilidade do sistema. Gráficos e discussões sobre as perguntas e resultados obtidos serão apresentados a seguir. O questionário aplicado pode ser visto no Apêndice A .

A primeira pergunta busca mostrar quanto a ideia do sistema parece ser atrativa em seus conceitos mais básicos, o gráfico da figura 5.1 mostra que a maior parte dos alunos demonstrou grande interesse sobre a premissa do sistema.

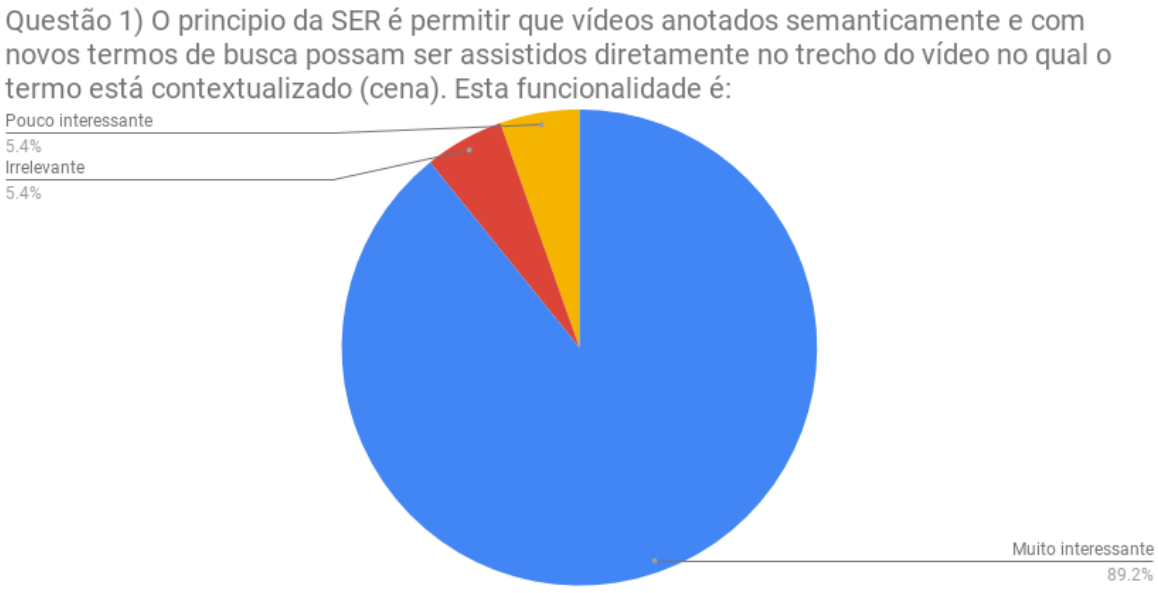

Figura 5.1: Resultados para primeira questão do questionário

A segunda pergunta, refinando melhor para o contexto no qual o sistema foi planejado mostra, como pode ser observado no gráfico da figura 5.2, que os alunos acreditam no potencial do sistema para aprimorar a vida dos estudantes em seus meios de estudo.

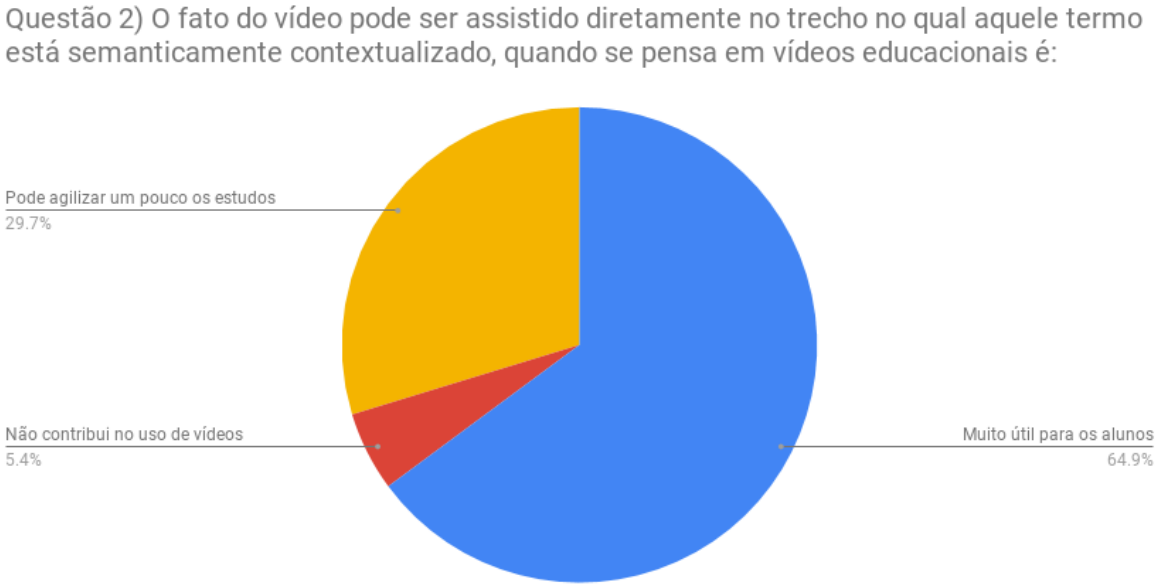

Figura 5.2: Resultados para segunda questão do questionário

A terceira pergunta foca na percepção dos participantes quanto a uma das funcionalidades importantes do sistema que é a navegabilidade com o uso de palavras-chave. Boa parte dos participantes conseguiu notar que as buscas feitas com palavras-chave levavam à cenas dos vídeos que de fato tratavam do termo buscado, como pode ser visto no gráfico da figura 5.3, mostrando a eficiência das ferramentas de segmentação semântica e definição de palavras-chave por cena utilizados na construçao do sistema.

Diferente das perguntas anteriores, a quarta pergunta foca na opinião dos participantes sobre a interface do sistema, vide gráfico da figura 5.4. As opiniões aqui divergem bastante mostrando que apesar de possuir boas funcionaldiades o sistema é falho no que diz respeito à apresentação das mesmas, exigindo portanto melhorias na parte visual para tornar a interface mais atrativa e agradável.

A quinta pergunta busca novamente avaliar a clareza com que as funcionalidades são apresentadas e quão intuitiva é sua utilização. Os resultados obtidos no gráfico da figura 5.5 mostram que das três funcionalidades principais do sistema : exibição de mídias adicionais, busca e navegabilidade; a maior parte dos participantes conseguiu identificar

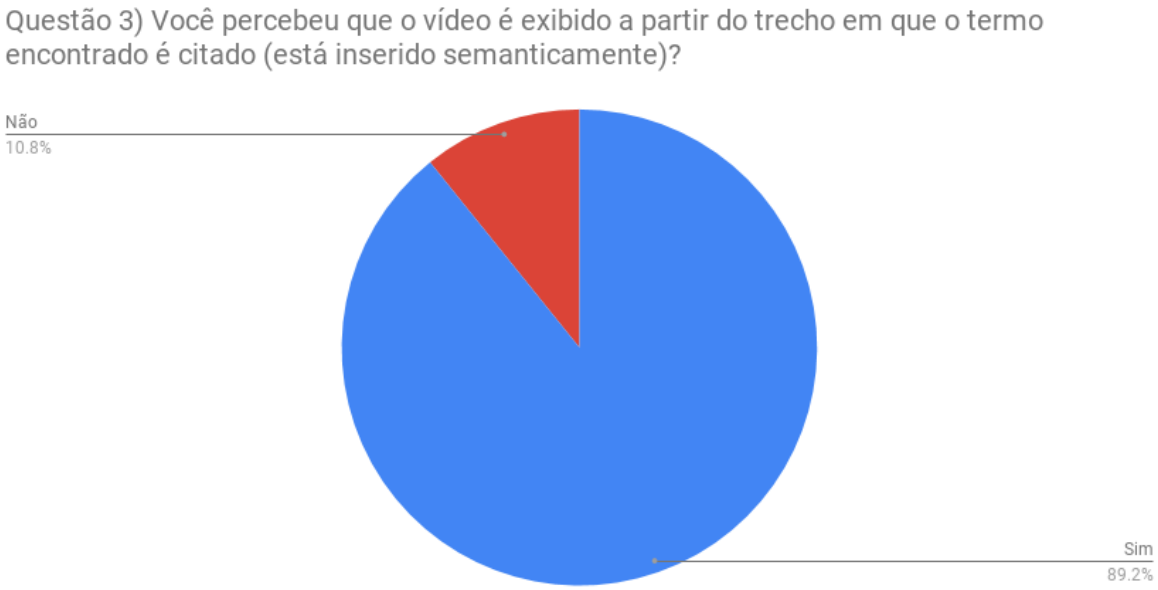

Figura 5.3: Resultados para terceira questão do questionário

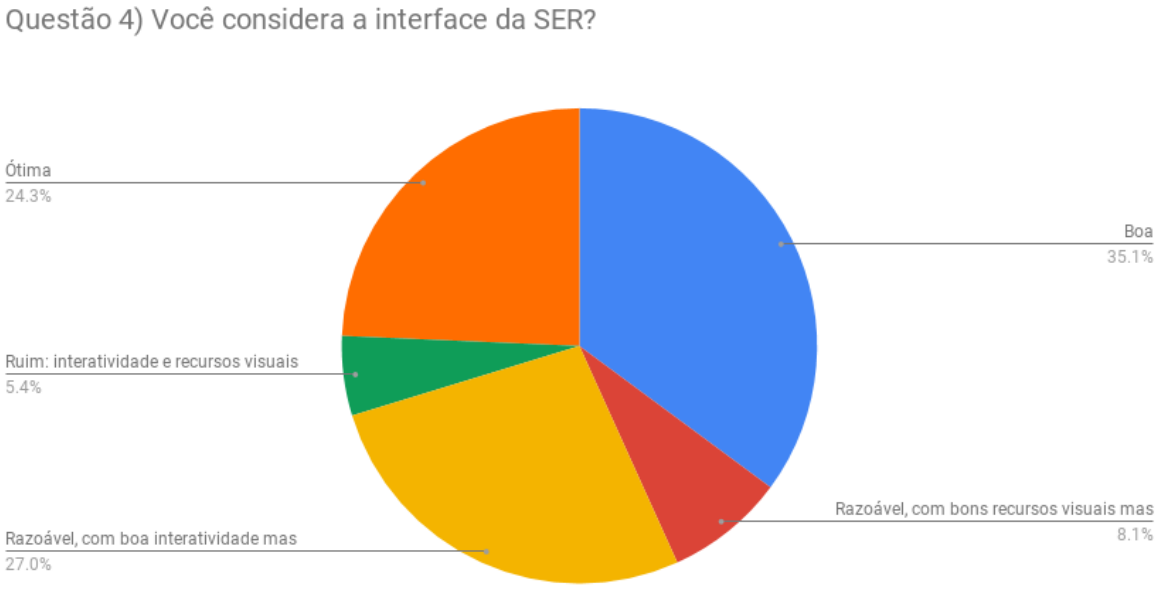

Figura 5.4: Resultados para quarta questão do questionário

pelo menos duas das funcionalidades sendo que mais da metade teve acesso a todas elas. Estes resultados concordam com os obtidos na questão anterior no que diz respeito às funcionalidades exigirem uma melhor apresentação para tornar mais fácil sua identificação e posterior utilização no sistema.

A apresentação dos dados da sexta pergunta difere das perguntas anteriores, como pode ser visto no gráfico da figura 5.6, para exibir de forma mais clara quão interessante seria para os participantes ver mudanças nos elementos levantados pela questão (1 -

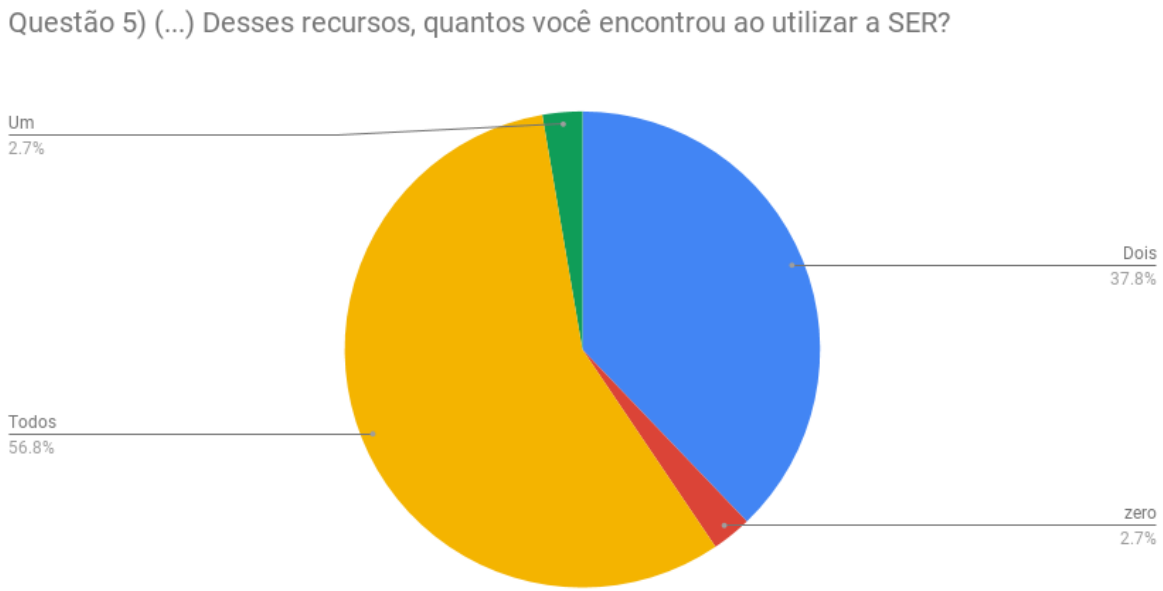

Figura 5.5: Resultados para quinta questão do questionário

apresentação das mídias adicionais, 2 - exibição dos videos relacionados, 3 - exibição dos trechos de videos com termo de busca solicitado, 4 - forma de busca inicial dos termos). Algo importante a ser notado aqui é que novamente poucos alunos mudariam funcionalidades do sistema (elemento 4) mas boa parte deles gostaria de melhorar, de alguma maneira, a forma como as mesmas são acessadas ou exibidas aos usuários do sistema.

Como a última pergunta é uma questão aberta sobre a opinião geral dos participantes em relação ao sistema não há como gerar gráficos sobre a mesma, porém a resposta de cada um dos participantes pode ser encontrada no Apêndice B, note que por ser uma pergunta aberta e opcional nem todos os participantes responderam à esta pergunta. Algumas opiniões são mais detalhadas descrevendo situações de uso e outras são mais breves simplesmente apontando pontos positivos ou negativos do sistema, podemos novamente notar alguns pontos, que já foram levantados na análise dos gráficos das questões anteriores, através das respostas individuais tais como o interesse dos alunos sobre a ideia do sistema e seu potencial como ferramenta para aprimorar a vida acadêmica dos alunos. Também vemos novamente sugestões de aprimoramento da interface algumas inclusive fazendo comparação a outras ferramentas já existentes que possuam alguma funcionalidade semelhante e, além das sugestões de interface, existem algumas sugestões quanto

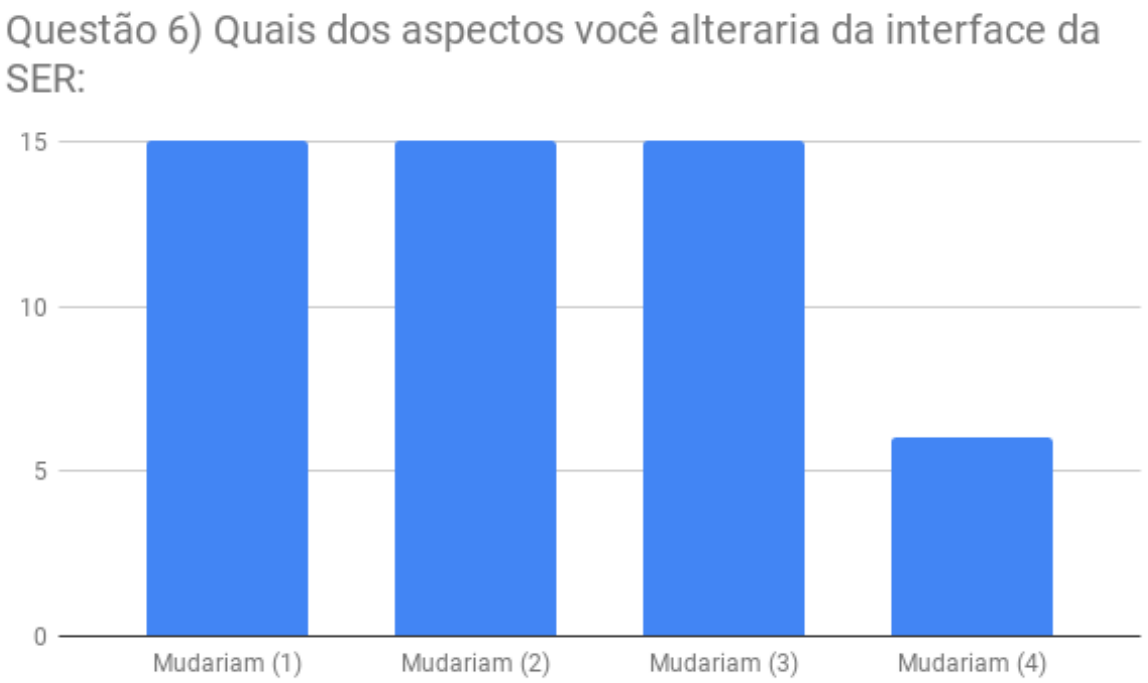

Figura 5.6: Resultados para sexta questão do questionário

a melhoria nos algoritmos que envolvem a navegação e mídias adicionais que as vezes apresentavam resultados incoerentes, o que já era esperado visto que o sistema depende de ferramentas cujos resultados são sensíveis a qualidade dos vídeos utilizados em seu processamento, o que pode ocasionar alguns erros de interpretação durante a geração de palavras-chave que são usadas na navegação e mídias adicionais.

6 Conclusão e Perspectivas

Neste trabalho apresentamos uma proposta de sistema para exibição de vídeos junto a funcionalidades adicionais tais como navegação interativa dentre vídeos e apresentação de mídias adicionais semânticamente relacionadas ao conteúdo dos vídeos em exibição, no intuito de ser uma ferramenta de estudos que aprimore o meio acadêmico. Como preparação para o desenvolvimento, analisamos vários trabalhos já presentes na literatura com ferramentas que tivessem alguma utilidade para o sistema ou funcionalidades similares, estes trabalhos serviram de guia para a tomada de algumas decisões durante o desenvolvimento do sistema tais como uso de técnicas de processamento de linguagem natural apresentadas em (11) .

A partir da revisão bibliográfica uma estrutura para o sistema foi definida utilizando diversas ferramentas previamente encontradas tais como a ferramenta de transcrição de vídeos para texto e segmentação semântica de vídeos em cenas, algoritmos de processamento de linguagem natural (TF-IDF) para interpretação das cenas e posterior definição de palavras-chave das mesmas e ferramentas de calculo de correlação entre cenas para recomendação de vídeos.

Desenvolvida uma primeira versão do sistema com interface, uma avaliação foi feita fornecendo acesso à um conjunto de alunos para que experimentassem as funcionalidades do mesmo e, em seguida, um questionário foi aplicado para que dessem sua opinião sobre diversos pontos, tanto na parte da interface desenvolvida quanto no estado das funcionalidades prometidas.

A partir dos resultados do questionário obtidos pudemos concluir diversas características sobre o sistema desenvolvido. A maioria dos alunos demonstrou grande interesse e expectativa sobre o potencial do sistema como facilitador da vida acadêmica, enquanto as funcionalidades pareciam claras e evidentes para a maioria alguns alunos tiveram dificuldades em identificá-las seja por problemas na funcionaldiade em si ou na forma como foi apresentada que foi o ponto mais criticado do sistema. Segundo a maioria dos participantes, mudanças na interface gráfica do sistema e até no modo como é disponibilizado

(aplicação web) poderiam ser modificados.

É importante lembrar que o teste foi realizado em uma amostra pequena de alunos e idealmente mais testes como este deverão ser feitos para tomadas de decisões quanto a futuras atualizações do sistema. Porém, baseado nos resultados obtidos ate agora, como perspectivas de evoluções no sistema podemos citar a utilização de frameworks para desenvolvimentos de interfaces mais agradáveis e intuitivas ao usuário, modificando a forma como as funcionalidades são exibidas aos usuários por exemplo; e utilização de outros algoritmos de processamento de linguagem natural, na intenção de obter palavras-chaves mais precisas em relação ao contexto dos vídeos que, por consequência, trariam melhoras para a ferramenta de navegação entre vídeos e entre cenas de vídeo, e também melhoras para a ferramenta de recomendação de mídias adicionais que poderiam fornecer informações mais precisas em relação ao contexto da cena.

Outra perspectiva importante é a expansão das bases de dados utilizadas no sistema além da adaptação do algoritmo para diferentes línguas para que possa incluir uma variedade maior de vídeos de diferentes áreas e seja mais acessível a um conjunto maior de pessoas.

Bibliografia

[1] Dendrinos, M.; Bakamidis, S. ; Carayannis, G. Speech enhancement from noise: A regenerative approach. **Speech Communication**, v.10, n.1, p. 45–57, 1991.

[2] Mori, S.; Suen, C. Y. ; Yamamoto, K. Historical review of ocr research and development. **Proceedings of the IEEE**, v.80, n.7, p. 1029–1058, 1992.

[3] Akutsu, A.; Tonomura, Y. **Video tomography: An efficient method for camerawork extraction and motion analysis**. In: Proceedings of the second ACM international conference on Multimedia, p. 349–356. ACM, 1994.

[4] Nakanishi, T.; Shio, A. ; Ishii, K.-I. Automatic vehicle image extraction based on spatio-temporal image analysis. **Systems and computers in Japan**, v.26, n.12, p. 71–82, 1995.

[5] Flickner, M.; Sawhney, H.; Niblack, W.; Ashley, J.; Huang, Q.; Dom, B.; Gorkani, M.; Hafner, J.; Lee, D.; Petkovic, D. ; others. Query by image and video content: The qbic system. **computer**, v.28, n.9, p. 23–32, 1995.

[6] Schultz, R. R.; Stevenson, R. L. Extraction of high-resolution frames from video sequences. **IEEE transactions on image processing**, v.5, n.6, p. 996–1011, 1996.

[7] Lienhart, R. **Automatic text recognition for video indexing**. In: Proceedings of the fourth ACM international conference on Multimedia, p. 11–20. ACM, 1997.

[8] de Pádua Paula Filho, W. **Multimídia: conceitos e aplicações**. Livros Técnicos e Científicos, 2000.

[9] Fielding, R. T.; Taylor, R. N. **Architectural styles and the design of network-based software architectures**, volume 7. University of California, Irvine Doctoral dissertation, 2000.

[10] Lienhart, R.; Wernicke, A. Localizing and segmenting text in images and videos. **IEEE Transactions on circuits and systems for video technology**, v.12, n.4, p. 256–268, 2002.

[11] Gonzalez, M.; Lima, V. L. S. **Recuperação de informação e processamento da linguagem natural**. In: XXIII Congresso da Sociedade Brasileira de Computaçao, volume 3, p. 347–395, 2003.

[12] Oliveira, F. A.; NAVAUX, P. O. A. Processamento de linguagem natural: princípios básicos e a implementação de um analisador sintático de sentenças da língua portuguesa. **Rio Grande do Sul**, 2004.

[13] Mierswa, I.; Morik, K. Automatic feature extraction for classifying audio data. **Machine learning**, v.58, n.2-3, p. 127–149, 2005.

[14] Klyne, G.; Carroll, J. J. Resource description framework (rdf): Concepts and abstract syntax. 2006.

[15] Smith, R. **An overview of the tesseract ocr engine**. In: Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on, volume 2, p. 629–633. IEEE, 2007.

[16] Athanaselis, T.; Bakamidis, S.; Giannopoulos, G.; Dologlou, I. ; Fotinea, E. **Robust speech recognition in the presence of noise using medical data**. In: Imaging Systems and Techniques, 2008. IST 2008. IEEE International Workshop on, p. 349–352. IEEE, 2008.

[17] Collobert, R.; Weston, J. **A unified architecture for natural language processing: Deep neural networks with multitask learning**. In: Proceedings of the 25th international conference on Machine learning, p. 160–167. ACM, 2008.

[18] Luo, J.; Papin, C. ; Costello, K. Towards extracting semantically meaningful key frames from personal video clips: from humans to computers. **IEEE Transactions on Circuits and Systems for Video Technology**, v.19, n.2, p. 289–301, 2009.

[19] Segaran, T.; Evans, C. ; Taylor, J. **Programming the Semantic Web: Build Flexible Applications with Graph Data**. "O'Reilly Media, Inc.", 2009.

[20] Taurion, C. **Big data**. Brasport, 2013, 31,32p.

[21] Mikolov, T.; Chen, K.; Corrado, G. ; Dean, J. Efficient estimation of word representations in vector space. **arXiv preprint arXiv:1301.3781**, 2013.

[22] SYSTAP, L. **Blazegraph**, 2015.

[23] Soares, E. R.; Barrére, E. **An approach for automatic segmentation of scenes in educational videos through the use of audio transcription and semantic annotation**. In: Proceedings of the 23rd Brazillian Symposium on Multimedia and the Web, p. 229–235. ACM, 2017.

[24] Lotfidereshgi, R.; Gournay, P. **Biologically inspired speech emotion recognition**. In: Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on, p. 5135–5139. IEEE, 2017.

A Questionário

Questão 1) O principio da SER é permitir que vídeos anotados semanticamente e com novos termos de busca possam ser assistidos diretamente no trecho do vídeo no qual o termo está contextualizado (cena). Esta funcionalidade é:

Escolha uma:

a. Muito interessante

b. Irrelevante

c. Pouco interessante

Questão 2) O fato do vídeo pode ser assistido diretamente no trecho no qual aquele termo está semanticamente contextualizado, quando se pensa em vídeos educacionais é:

Escolha uma:

a. Muito útil para os alunos

b. Pode agilizar um pouco os estudos

c. Não contribui no uso de vídeos educacionais.

Questão 3) Você percebeu que o vídeo é exibido a partir do trecho em que o termo encontrado é citado (está inserido semanticamente)?

Escolha uma:

a. Sim

b. Não

Questão 4) Você considera a interrface da SER:

Escolha uma:

a. Ótima

b. Boa

c. Razoável, com boa interatividade mas com recursos visuais ruins

d. Razoável, com bons recursos visuais mas interatividade ruim

e. Ruim: interatividade e recursos visuais

Questão 5) A SER permite:

1) Exibir/esconder mídias/conteúdos associados ao termo encontrado(Wikipedia e Google).

2) Ir direto para outros trechos do mesmo vídeo que apresentam o termo encontrado.

3) Ir para outros vídeos que possuem o mesmo termo encontrado (vídeos relacionados). Desses recursos, quantos você encontrou ao utilizar a SER?

Escolha uma:

a. zero

b. Um

c. Dois

d. Todos

Questão 6) Quais dos aspectos você alteraria da interface da SER:

1) Forma de apresentação das mídias associadas (Wikipedia e Google).

2) Forma de exibição dos vídeos relacionados.

3) Forma de exibição dos trechos de vídeo que possuem o termo encontrado.

4) Formas de busca inicial dos termos.

Escolha uma:

a. Somente 1

b. Somente 2

c. Somente 3

d. Somente 4

e. Nenhum deles

f. 1, 2 e 3

g. 2 e 3

h. Todos

Questão 7) Faça as observações que achar pertinente.

B Respostas obtidas para a questão 7

É uma ótima ideia e bem complexa para se resolver. Acho que poderia, por questões de usabilidade, agilidade e praticidade, ser uma extensão no navegador, por exemplo. Só pelo fato do usuário não ter q ir até uma página para fazer essa busca, muitos acabam desistindo. Existe uma extensão pra chrome, Invideo for youtube, que encontra o que foi dito em um vídeo baseando-se na sua legenda.

A interface e a identidade visual podem ser melhoradas para se tornarem mais atraentes e fazer com que a experiência do usuário não seja frustrante. Termos do design como UI e UX podem ser aplicados no sistema.

interface muito boa, instrumento que eu não conhecia e vai ser muito util pra mim.

O projeto SER, me pareceu muito interessante, pois propicia agilidade no processo de busca de conteúdos relacionados aos termos que queremos para estudo.

O projeto SER é excelente ferramenta para auxiliar os alunos em assistir os vídeos educacionais.

Achei muito díficil a interface, e poderia ser em português..

Muito interessante o SER no contexto de se poder interagir direto no vídeo com o trecho referente a sua pesquisa.

É uma ferramenta interessante, mas poderia ter uma interface mais intuitiva e amigável. poderia conter recursos em português.

Achei a ferramenta de boa qualidade, com interface gráfica interessante. Parabéns.

Poderia ter uma versão em Português.

Achei a ferramenta um pouco confusa pela apresentação em outro idioma. Também achei um pouco incompreensível as correlações feitas entre o termo procurado e os conteúdos apresentados no Google e Wikipédia.

Esse projeto SER achei muito interessante, pois conseguir assistir um vídeo direto no trecho com o terno selecionado na busca é uma ferramenta interessante pois agiliza muito os estudos.

Achei a ideia realmente muito útil e interessante.

Do ponto de vista de aluno, que muitas vezes precisa pesquisar sobre algum tema, é muito interessante poder fazer esta pesquisa em um vídeo.

O aluno está de parabéns pela sua ideia.

Sei que o software ainda está em fase de elaboração e, por isto, vou relatar minha experiência de uso.

Vou ser bem sincero e crítico não no sentido de denegrir e/ou menosprezar o trabalho do aluno (cuja ferramenta tem um potencial enorme) e, sim, no sentido de provocar melhorias e, assim, poder ver este software ser largamente utilizado por todas as pessoas e ver este aluno colhendo os frutos de seu trabalho.

Achei a interface péssima. Embora simples, ela é muito feia e tenta copiar na tela inicial a interface de busca do Google. Acho que o autor do software poderia ser mais original e fazer uma interface mais bonita e que valorizasse o seu produto.

Com relação ao mecanismo de busca ele não funcionou adequadamente não. Pra falar a verdade ele mais falhou que acertou.

Por exemplo, eu pesquisei a palavra big data e, ao clicar no link da Wikpedia ela me trouxe vários links sobre temas pornográficos e e links com palavras que não posso dizer aqui.

Ao pesquisar sobre internet na base de dados 2 Temas diversos em Ciência da Computação, veio vídeo sobre UML.

E assim por diante.

Deveria ter na página de exibição dos vídeos um link para retornar à página inicial e, também, uma opção para zerar a consulta pois, quando se retorna para a tela inicial os quadradinhos com os vídeos encontrados na busca anterior permanecem na tela.

É um projeto bem interessante, a idéia é boa mas está na fase inicial e precisa ser melhorado em alguns aspectos principalmente visual. Parabéns!

Muito interessante este sistema, acho que seria de grande valia na educação a distância.Ferramenta muito prática de ser usada.Gostei muito.

Achei excelente por conseguir a localização do vídeo através da pesquisa por palavra contida no mesmo..

Recurso que permite a localização de vídeo na busca das palavras.

A ideia de se encontrar o trecho do vídeo que fala sobre o assunto é muito interessante, pois facilita o processo de estudo, e não necessário assistir todo o vídeo para localizar o tópico que se deseja, ele busca o trecho pela palavra, o que pode torna o processo um pouco cansativo, ele buscou temas diferentes do que eu imaginava, por exemplo pesquisei android achando que acharia algo sobre programa android ele buscou algo voltado para Robótica, o de Hardware, eu pensei em partes do computador ele buscou um tema diferente, esse tópico podia ser melhorado para simplificar a pesquisar, como utilizar o método de pesquisa do Netlix, na qual os assuntos aparecem por tópicos na qual são classificados, e de acordo com o perfil do usuário.

O layout da página poderia ser mais instrutivo, indicando melhores formas de se realizar a pesquisar, ou classificar tópicos de pesquisa.

A ideia do site é ótimo e é o futuro da tecnologia, simplificar os processos.

Alguns resultados não fizeram muito sentido com o termo originalmente buscado.

Testando a SER achei todos os processos muito interessantes, porém a forma de pesquisa, ou seja a forma de busca achei um pouco vazia. Poderia ter algumas opções iniciais como por exemplo: quando clicar na janela de busca aparecer algumas opções de pesquisas, como palavras associadas aos

vídeos, ou assuntos associados aos mesmos, entre outros. .Mas gostaria de parabenizar, gostei muito!

Achei a ferramenta de boa qualidade, com interface gráfica interessante. Parabéns.

Apesar de eu ter pouca habilidade no inglês, após alguns testes, percebi que o vídeo não estava parando o termo ou mesmo no contexto da palavra que digitei e, ainda, as sugestões de google e wikipedia, não tinham relação com a palavra que procurei. Diante disso, creio que a interface ainda precise de alguns ajustes no seu sistema de busca. De maneira geral, é uma ideia bastante inovadora, pois atualmente as buscas por assuntos se baseiam em títulos dos vídeos que usuários disponibilizam na internet, ou seja, títulos mal elaborados dificultam a localização de conteúdos. A forma de pesquisa aqui proposta, proporciona agilidade e foco nos resultados esperados em uma busca por termos específicos.

Achei interessante e bastante válido o fato de que o vídeo é exibido a partir do trecho em que o termo encontrado é citado.

Na hora do trecho do vídeo, não consegui ouvir a palavra que eu digitei, no caso software. Na pesquisa do google não era a palavra que eu digitei que ele pesquisou.

Não consegui abrir a definição do wikipedia.

interface muito boa, instrumento que eu não conhecia e vai ser muito util pra mim, usei só para testar e gostei muito vou usar mais veses

Achei bem interessante o sistema, só demorei um pouco para acostumar com a interface, pois achei ela muito pouco intuitiva.

Está faltando colocar o nome(título) nos vídeos.E outra sugestão colocar o brilho da tela abaixo do vídeo, deixando o botão alterar ao lado do nome brilho da tela, para ganhar mais espaço.

As perguntas citadas a cima tratam bastante sobre a interface e gostaria de falar um pouco dela.

A interface da SER poderia ser mais dinâmica, como por exemplo após realizar a consulta, deixar as cenas relacionadas e as mídias adicionais na mesma tela, sempre precisar da utilização da barra de tarefas.

Poderia estar deixando o tamanho do vídeo um pouco meno, ex.: tamanho do youtube, podendo adicionar a mídia adicionais ao lado. Com isso o usuário conseguiria assistir o vídeo e realizar a consulta.

Observações que não interfere na boa proposta que o projeto propõe.

Ao baixar um determinado link de áudio e ele criar automaticamente um título para o seu link. Achei pertinente que após a criação desse link permite ao usuário editar um novo título. Isso proporciona uma comodidade e maior facilidade no momento de busca.

Achei o programa muito interessante para um trabalho em sala de aula com alunos, permitindo que possamos criar vários áudios renomeados e unir todos formando uma edição no final com vários componentes ao mesmo tempo.

Achei muito interessante a SER, mas tive dificuldades de entender os vídeos por conta do inglês.

Pesquisa muito interessante para nós da área da tecnologia.

A interface é bastante interessante. É nítido que direciona justamente para a parte em questão delimitada pelas palavras escolhidas.

Achei a ferramenta de bastante valia, mas a interface gráfica não é muito atrativa e tem poucos recursos para que os usuários possam interagir, mas fora isso, gostei da ferramenta. Existe versão com vídeos com áudio em português?

Base de dados em Português; Layout mais amigável; Forma de voltar al menu principal mais rapidamente.