

Um Serviço Web para Geração Automática de Audiodescrição de Vídeos Digitais

Wesnydy Lima Ribeiro



CENTRO DE INFORMÁTICA
UNIVERSIDADE FEDERAL DA PARAÍBA

João Pessoa, 2018

Wesnydy Lima Ribeiro

Um Serviço Web para Geração Automática de Audiodescrição de Vídeos Digitais

Monografia apresentada ao curso de Bacharelado em ciência da computação
do Centro de Informática, da Universidade Federal da Paraíba,
como requisito para a obtenção do grau de Bacharel em ciência da computação

Orientador: Tiago Maritan Ugulino de Araújo

Novembro de 2018

Catálogo na publicação
Seção de Catalogação e Classificação

R484s Ribeiro, Wesnydy Lima.

Um Serviço Web para Geração Automática de
Audiodescrição de Vídeos Digitais / Wesnydy Lima
Ribeiro. - João Pessoa, 2018.
62 f. : il.

Orientação: Tiago Maritan Ugulino de Araújo.
Monografia (Graduação) - UFPB/CI.

1. Audiodescrição. 2. Serviço Web. 3. Vídeo. 4.
Acessibilidade. I. Araújo, Tiago Maritan Ugulino de.
II. Título.

UFPB/BC



CENTRO DE INFORMÁTICA
UNIVERSIDADE FEDERAL DA PARAÍBA

Trabalho de Conclusão de Curso de Bacharelado em ciência da computação intitulado *Um Serviço Web para Geração Automática de Audiodescrição de Vídeos Digitais* de autoria de Wesnydy Lima Ribeiro, aprovada pela banca examinadora constituída pelos seguintes professores:

Thaís Gaudencio do Rêgo

Prof^ª. Dr^ª. Thaís Gaudencio do Rêgo

Centro de Informática/UFPB

Tiago Maritan U. de Araújo

Prof^º. Dr^º. Tiago Maritan Ugulino de Araújo

Centro de Informática/UFPB

Virginia Pinto Campos

MSc. Virginia Pinto Campos

Pesquisadora do Núcleo LAVID/UFPB

João Pessoa, 9 de novembro de 2018

AGRADECIMENTOS

Agradeço a Deus, por conceder a mim a capacidade de lutar pelos meus sonhos e objetivos de vida.

Aos meus pais Leandra Souza e Adriano Ribeiro, pelo apoio moral e incentivo as minhas decisões que me guiaram até aqui, além de todos os esforços que me permitiram estudar.

A minha irmã, avó e parentes que diretamente ou indiretamente, contribuíram com a minha jornada.

Ao meu orientador, prof. Tiago Maritan, não só pela orientação mas também por todas as oportunidades que me fizeram crescer profissionalmente durante o período que estive na universidade.

Aos meus colegas do LAViD, pelas trocas de conhecimento, pelos momentos de descontração em momentos de tensão dos nossos projetos e principalmente pelo companheirismo.

A Virginia Campos e toda a equipe do GTAaaS 2.0 pelas contribuições fundamentais que resultaram neste trabalho.

Finalmente, agradeço a todos os amigos que torceram por mim e me incentivaram a continuar.

RESUMO

A audiodescrição é um recurso de acessibilidade desenvolvido para beneficiar as pessoas com deficiência visual no contexto de compreensão de informações visuais. De acordo com Szarkowska (2011), alguns dos maiores obstáculos à popularização e disponibilização da audiodescrição são o tempo de produção geralmente demorado e os altos custos envolvidos durante o processo de geração desse conteúdo. Com o objetivo de reduzir esses problemas, o presente trabalho tem como proposta o projeto e o desenvolvimento de um *web service RESTful* para a geração automática de trilhas de audiodescrição de vídeos digitais. A solução desenvolvida utiliza estratégias de geração automática de roteiro de audiodescrição baseada em informações extraídas do vídeo e do seu roteiro, além usar ferramentas *Text-To-Speech* (TTS) em conjunto com técnicas de sincronização para gerar as trilhas de audiodescrição. Os resultados mostraram que a solução foi capaz de reduzir consideravelmente o tempo de criação de recursos de audiodescrição quando comparado com a metodologia tradicional. Adicionalmente, considerando que a maioria dos vídeos disponíveis na atualidade não possuem audiodescrição, a proposta pode ser considerada uma alternativa viável quando a audiodescrição humana não estiver disponível ou não for viável.

Palavras-chave: Audiodescrição, Serviço Web, Vídeo, Acessibilidade.

ABSTRACT

Audio description is an accessibility feature designed to benefit visually impaired people in the context of visual information comprehension. According to Szarkowska (2011), some of the major obstacles to the popularization and availability of audio description are the generally time consuming production and the high costs involved during the process of generating this content. In order to reduce these problems, the present work proposes the design and development of a RESTful web service for the automatic generation of audio description tracks of digital videos. The solution developed utilizes strategies of automatic generation of audio description script based on information extracted from the video and its script, besides using Text-To-Speech (TTS) tools in conjunction with synchronization techniques to generate audio description tracks. The results showed that the solution was able to considerably reduce the time of audio creation of audio description resources when compared to the traditional methodology. Additionally, considering that most of the videos currently available do not have audio description, the proposal can be considered a viable alternative when human audio description is not available or is not feasible.

Key-words: Audio description, Web Service, Video, Accessibility.

LISTA DE FIGURAS

1	Descrição de imagem (Exemplo 1).	19
2	Descrição de imagem (Exemplo 2).	19
3	Visão alto nível de uma arquitetura <i>RESTful</i> [13].	21
4	Arquitetura proposta em [14].	23
5	Sistema proposto em [3].	24
6	Visão geral da plataforma em desenvolvimento.	26
7	Visão Esquemática do Serviço Proposto.	30
8	Visão esquemática do processo de geração automática de AD.	32
9	Avaliação dos usuários sobre a facilidade de entendimento da AD gerada pelo protótipo.	36
10	Avaliação dos usuários sobre contribuição da AD gerada pelo protótipo na compreensão dos vídeos.	37
11	Tempo de geração da trilha de AD com roteiro.	38
12	Tempo de geração da trilha de AD sem roteiro.	39
13	Tempo de geração da trilha de AD variando a resolução do vídeo C.	40
14	Tempo de geração da trilha de AD variando a resolução do vídeo D.	40
15	Tempo de processamento de 100 requisições com 1, 2 e 4 <i>workers</i> e com roteiro.	42
16	Tempo de processamento de 100 requisições com 1, 2 e 4 <i>workers</i> e sem roteiro.	43

LISTA DE TABELAS

1	Características gerais dos trabalhos relacionados.	28
2	Características dos Vídeos da Primeira Parte dos Testes.	34
3	Características dos Vídeos da Segunda Parte dos Testes.	37
4	Sumarização do tempo de processamento de 100 requisições com roteiro disponível.	41
5	Sumarização do tempo de processamento de 100 requisições com ausência do roteiro.	42
6	Tempo de processamento de 100 requisições com roteiro disponível. . . .	59
7	Tempo de processamento de 100 requisições com ausência do roteiro. . . .	61

LISTA DE ABREVIATURAS

AD – Audiodescrição

API – *Application Programming Interface*

FTP – *File Transfer Protocol*

GUI – *Graphical User Interface*

HTTP – *Hypertext Transfer Protocol*

IBGE – Instituto Brasileiro de Geografia e Estatística

IIOP – *Internet Inter-Orb Protocol*

REST – *Representational State Transfer*

RMI – *Remote Method Invocation*

SMTP – *Simple Mail Transfer Protocol*

SOAP – *Simple Object Access Protocol*

SRT – *SubRip Text*

TTS – *Text-To-Speech*

URI – *Uniform Resource Locator*

XML – *Extensible Markup Language*

W3C – *World Wide Web Consortium*

WHO – *World Health Organization*

WSDL – *Web Services Description Language*

Sumário

1	INTRODUÇÃO	15
1.1	Definição do Problema	16
1.2	Objetivo geral	16
1.3	Objetivos específicos	16
1.4	Estrutura da monografia	16
2	FUNDAMENTAÇÃO TEÓRICA	18
2.1	Audiodescrição	18
2.2	<i>Web Services</i>	19
2.2.1	REST	20
2.2.2	Princípios do REST	21
3	TRABALHOS RELACIONADOS	23
3.1	Descrevendo vídeos <i>on-line</i> com síntese de fala para narração	23
3.2	Anotação colaborativa para acessibilidade de vídeo	24
3.3	Descrição de vídeo acessível sob demanda	25
3.4	Troca de Vídeo Descritivo	26
3.5	Descrição remota em tempo real	27
3.6	Características gerais dos trabalhos	27
4	SOLUÇÃO PROPOSTA	29
4.1	Serviço Web de Geração de AD	29
4.2	Web Service	30
4.3	Geração Automática de AD	31
5	RESULTADOS E DISCUSSÕES	34
5.1	Compreensão Geral dos Vídeos	34
5.2	Tempo de Geração das Trilhas de AD	37
5.2.1	Mensuração do Tempo de Geração Variando a Duração dos Vídeos	38
5.2.2	Mensuração do Tempo de Geração Variando a Resolução dos Vídeos	39

5.2.3	Mensuração do tempo de geração para um conjunto de cem requisições	41
6	CONSIDERAÇÕES FINAIS	44
	REFERÊNCIAS	45
	APÊNDICE A – DOCUMENTAÇÃO DA API DO SERVIÇO	49
	APÊNDICE B – TABELAS COMPLETAS DOS RESULTADOS	59

1 INTRODUÇÃO

As pessoas com deficiência visual possuem dificuldades para captar e compreender informações essenciais em ambientes que dispõem de experiências audiovisuais, como cinema, teatro, sala de aula, internet, dentre outros. Como consequência disso, essas pessoas possuem dificuldades para interagir com a sociedade no âmbito sociocultural, visto que, nestes ambientes, raramente são disponibilizados recursos de acessibilidade que viabilizem sua participação.

A deficiência visual é um problema que atinge pessoas do mundo inteiro. A *World Health Organization* (WHO) [27] estima que existem cerca de 253 milhões de pessoas com algum nível da deficiência, onde, dentre esses, 36 milhões são completamente cegas e 217 milhões possuem perda visual moderada ou severa. No Brasil, de acordo com o censo demográfico 2010 realizado pelo Instituto Brasileiro de Geografia e Estatística (IBGE), existem aproximadamente 35,7 milhões de pessoas com algum grau de deficiência visual, o que representa cerca de 18,8% da população brasileira [12]. Dentre as deficiências investigadas, a deficiência visual é apontada como a de maior incidência na população, ocasionando, dessa forma, limitações tanto no acesso, quanto no consumo de informações dependentes da visão.

Diante disso, a audiodescrição (AD) se apresenta como um meio de contornar essa situação, fornecendo informações adicionais a este público. A audiodescrição é um recurso de acessibilidade desenvolvido para beneficiar as pessoas com deficiência visual no contexto de compreensão de informações visuais, através da descrição de imagens, cenários, detalhes visuais, ações, entre outros. No entanto, esse tipo de conteúdo é geralmente produzido de forma manual, o que acaba requerendo bastante tempo [15]. Segundo Lakritz e Salway [15], para um profissional da área descrever um filme de 2 horas, são necessárias, em média, cerca de 60 horas de trabalho. De acordo com Szarkowska [23], o processo de produção demorado e os altos custos envolvidos são alguns dos principais obstáculos que prejudicam uma maior disponibilização de audiodescrição.

Na literatura, alguns trabalhos vêm sendo desenvolvidos utilizando estratégias computacionais para reduzir os problemas relacionados ao tempo, custo e escala na produção de conteúdos de AD. Esses trabalhos sugerem a geração automática das descrições usando como base informações extraídas do roteiro [1, 15] e do vídeo [5, 6, 20, 22, 24, 28].

Contudo, de maneira geral, esses trabalhos são focados em etapas específicas do processo de geração de audiodescrição (*e.g.* sumarização de texto e classificação de imagens). Ou seja, eles não apresentam uma solução computacional integrada que contenha todas as etapas necessárias para a geração automática de trilhas de AD a partir do roteiro ou do vídeo. Desta forma, será apresentada na seção seguinte, a definição do problema

do presente trabalho, que trata diretamente da necessidade de um serviço de geração automática de trilhas de audiodescrição.

1.1 Definição do Problema

Diante do exposto, torna-se perceptível a necessidade do desenvolvimento de tecnologias capazes de automatizar todo o processo de geração de trilhas de audiodescrição para vídeos digitais.

Assim sendo, o seguinte problema de pesquisa foi elaborado: como disponibilizar uma abordagem capaz de produzir automaticamente trilhas de AD e contribuir com o aumento da disponibilização de vídeos digitais acessíveis para cegos?

1.2 Objetivo geral

O objetivo geral deste trabalho é projetar e desenvolver um serviço que viabilize a geração automática de audiodescrição de vídeos digitais através da web. Desta forma, aplicações multiplataforma poderão ser construídas para contribuir com a distribuição de recursos de AD em ambientes que não dispõem de acessibilidade para pessoas cegas.

1.3 Objetivos específicos

Para alcançar o objetivo geral deste trabalho, foram definidos os seguintes objetivos específicos:

1. Definir uma abordagem baseada em *Web Services* para a geração automática de audiodescrição;
2. Especificar uma arquitetura baseada no estilo REST para prover acesso aos serviços Web;
3. Utilizar estratégias de geração automática de audiodescrição e síntese de voz para desenvolver um componente gerador de trilhas de AD.

1.4 Estrutura da monografia

Esta monografia encontra-se organizada por seis capítulos, como descrito a seguir. A sequência dos capítulos procura refletir o processo adotado durante o desenvolvimento deste trabalho.

Este capítulo introduziu a monografia e a sua finalidade. Ainda neste capítulo, foi apresentada a definição o problema, bem como os objetivos gerais e específicos para

responder o problema definido. Em resumo, ele identifica o problema, o raciocínio por trás do mesmo e a estrutura do trabalho.

O restante do trabalho está organizado da seguinte forma: no Capítulo 2, são apresentados os conceitos fundamentais para o entendimento deste trabalho. O Capítulo 3 apresenta os estudos relacionados ao problema definido no Capítulo 1 e que estão mais próximos a este trabalho. No Capítulo 4, é descrita a solução proposta para a disponibilizar, através de um *Web Service*, a geração automática de audiodescrição. Os resultados obtidos são apresentados e discutidos no Capítulo 5. Por fim, o Capítulo 6 apresenta as considerações finais sobre o desenvolvimento da proposta e cenário estudado.

2 FUNDAMENTAÇÃO TEÓRICA

A seguir, serão abordados, com suporte da literatura, os principais temas que formam o referencial teórico deste trabalho, a saber: Audiodescrição e *Web Services*.

2.1 Audiodescrição

A audiodescrição (AD) é um recurso que possibilita às pessoas com algum tipo de deficiência visual usufruírem dos mais variados tipos de produções audiovisuais, evitando que informações-chave transmitidas visualmente e contribuem para a contextualização da produção sejam perdidas. De acordo com Motta [16], audiodescrição é um recurso de acessibilidade que permite que as pessoas com deficiência visual possam entender melhor obras audiovisuais como filmes, peças teatrais, programas de TV, exposições, mostras, musicais, entre outros. É um processo de tradução que converte toda a informação visual em verbal, ou seja, transforma aquilo que é visto no que é ouvido.

A audiodescrição surgiu nos Estados Unidos na década de 1970 a partir de ideias desenvolvidas por Gregory Frazier em sua dissertação de mestrado. Entretanto, a AD teve seu início apenas na década seguinte graças ao trabalho do casal Margaret Pfanstiehl e Cody Pfanstiehl [10].

Levando em consideração o tipo de produção audiovisual, a audiodescrição pode ser simultânea ou pré-gravada [17]. O recurso de AD simultânea geralmente está presente em eventos ao vivo, no qual podem ocorrer situações inesperadas, como peças teatrais, palestras, eventos esportivos, entre outros. A audiodescrição pré-gravada é usada em conjunto com conteúdos estáticos ou pré-gravados como filmes, programas de TV, obras de arte etc.

De maneira geral, a audiodescrição deve descrever elementos visuais essenciais que permitem o ouvinte entender a obra audiovisual. De acordo com Casado [2], os elementos a serem descritos são classificados como elementos visuais não verbais e elementos visuais verbais. Entre os elementos visuais não verbais estão as características de personagens (figurino, físico, expressões, etnia, idade etc), a ambientação (localização dos personagens, data e hora) e as ações. Os elementos visuais verbais são legendas, títulos, subtítulos e outros elementos textuais como nomes de estabelecimentos, títulos de livros, cartas lidas por personagens, dentre outros. As Figuras 1 e 2 apresentam exemplos de descrições¹.

¹Feitas por Patrícia Braille, coordenadora da educação especial no estado da Bahia e autora do projeto Pra Cego Ver. Imagens obtidas na internet.



Cartum desenhado em preto e branco. Em uma escura sala de cinema, vários tubarões sorridentes estão sentados, com seus copos de refrigerante e sacos de pipoca. Na tela aparece um navio naufragando e, na parte que ainda não está submersa, está escrito: "Titanic". O tubarão da primeira fileira diz: "Adoro finais Felizes!" [fala do personagem está dentro de um balão].

Figura 1: Descrição de imagem (Exemplo 1).



Clássica fotografia em tons de cinza da atriz Marilyn Monroe tentando segurar seu vestido que está sendo levantado pelo vento.

Figura 2: Descrição de imagem (Exemplo 2).

2.2 Web Services

Na literatura não existe um consenso para uma definição precisa do termo “*Web Service*”. Pode-se entender por *Web Service* um sistema de software que torna possível que aplicações diferentes interajam entre si e sistemas de plataformas diferentes tornem-se compatíveis. Além disso, os *Web Services* permitem que aplicações enviem e recebam dados em formatos diversificados. Cada aplicação pode ter sua própria linguagem, que é traduzida para uma linguagem universal, como é o caso do formato XML.

Para o W3C [26], a definição de “*Web Service*” dá-se como um sistema projetado para suportar a interoperabilidade entre máquinas sobre uma rede. Ele possui uma interface descrita em um formato processável por máquina (especificamente *Web Services Description Language* - WSDL). Outros sistemas interagem com o *Web Service* por meio de mensagens SOAP (*Simple Object Access Protocol*), geralmente transmitidas usando HTTP com serialização XML em conjunto com outros padrões relacionados à web.

Uma característica fundamental da arquitetura de *Web Services*, diz respeito à

capacidade de permitir a utilização de diferentes formas de transmissão de dados pela rede. Logo, as mensagens de comunicação de *Web Services* podem ser transportadas por uma variedade de protocolos de rede, tais como HTTP, SMTP, FTP, RMI-IIOP ou protocolos de mensagens proprietários [25].

Os *Web Services* (WS) especificados pelo W3C podem ser categorizados em duas gerações. A primeira geração possuía várias limitações em seu funcionamento, desde limitações na linguagem usada para a descrição dos serviços a seu funcionamento propriamente dito. A segunda geração, conhecida como “*WS-**”, representou um complemento, contribuindo com novos recursos e funcionalidades para suprir as limitações da geração anterior. A arquitetura “*WS-**” é composta por mais de 20 especificações. Dentre essas, destacam-se *WS-Coordination*, *WS-Transaction*, *BPEL4WS*, *WS-Security*, *WS-ReliableMessaging*, *WS-Policy* e *WS-Attachments* [8]. Entretanto, o grande número de especificações torna esse formato complexo e burocrático, além de difícil de ser dominado por uma só pessoa.

De acordo com Pautasso, Zimmerman e Leymann [18], recentes tendências tecnológicas no domínio de *Web Services* indicam que a solução capaz de eliminar a complexidade presente nos padrões “*WS-**” é vista no REST. Segundo Fielding [9], o REST fornece um conjunto de restrições arquiteturais que, quando aplicadas como um todo, enfatizam a escalabilidade de interações de componentes, generalidade de interfaces, implantação independente de componentes e a redução de latência de interação através de componentes intermediários, além de reforçar a segurança e encapsular sistemas legados. Este trabalho adotou o REST para a especificação da solução proposta. A seguir, serão apresentados mais detalhes sobre o REST.

2.2.1 REST

De acordo com Fielding [9], a *World Wide Web* teve sucesso em grande parte pelo fato de sua arquitetura ter sido projetada para atender às necessidades de um sistema de hipermídia distribuído em escala de internet. A *web* teve seu desenvolvimento de forma iterativa, passando por várias modificações nos padrões que definem sua arquitetura.

Segundo Fielding [9], o REST pode ser compreendido como um estilo arquitetônico para ser usado como guia para o desenvolvimento de sistemas de hipermídia distribuídos.

Os sistemas que seguem os princípios REST são comumente referenciados como “*RESTful*”. Para Richardson e Ruby [19], o termo “*RESTful*” tem o significado semelhante ao utilizado para caracterizar “orientação a objetos”. Um *framework* ou uma aplicação podem ser projetados em uma linguagem orientada a objetos (e.g. C++ e Ruby), mas isso não implica que sua arquitetura seja orientada a objetos. O mesmo serve para “*RESTful*”.

A Figura 3 ilustra uma arquitetura *RESTful* por duas perspectivas, uma abstrata e outra concreta. Para fazer o mapeamento da forma abstrata para a forma concreta utilizou-se o HTTP. No centro de uma Arquitetura REST está o conjunto de recursos. Esses recursos são representados por URIs e uma representação interna (e.g. dados do tipo JSON). Por fim, há um conjunto de operações através das quais é possível manipular os recursos. Conforme a ilustração, as operações podem ser mapeadas para os métodos HTTP GET, POST, DELETE, entre outros.

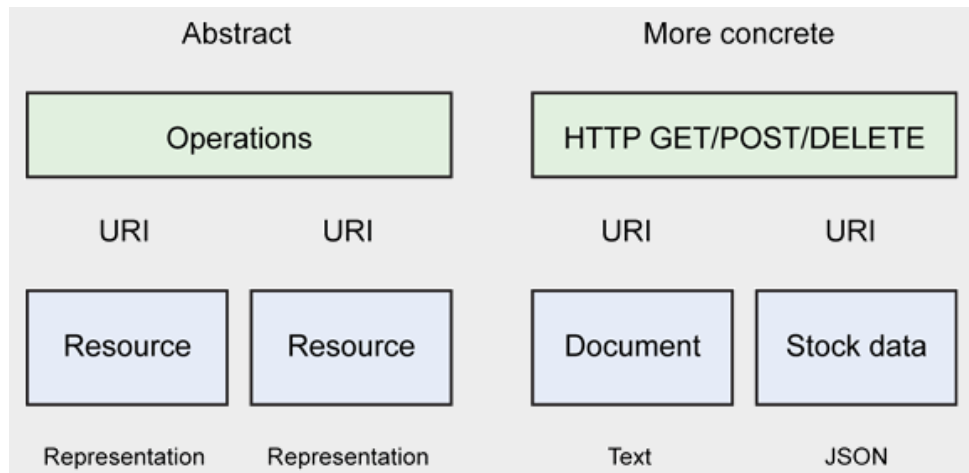


Figura 3: Visão alto nível de uma arquitetura *RESTful* [13].

2.2.2 Princípios do REST

Conforme Fielding [9], para que os princípios do REST sejam respeitados, deve-se obedecer um conjunto de restrições estabelecidas pelo estilo, as quais serão detalhadas a seguir.

- **Cliente-Servidor** - esta característica é frequentemente encontrada em aplicações que fazem uso do estilo arquitetônico baseado em rede (Aplicações Web). Um servidor, oferecendo um conjunto de serviços, atende requisições a esses serviços. Um cliente, desejando que um serviço do servidor seja executado, envia uma requisição para o servidor. O servidor então pode tanto rejeitar, quanto executar o serviço solicitado e devolve uma resposta ao cliente. A separação de responsabilidades é o princípio por trás dessa restrição.
- **Stateless (Sem Estado)** - outra restrição imposta pelo estilo REST refere-se a interação entre cliente e servidor. Cada requisição feita do cliente ao servidor deve conter todas as informações necessárias para atender a solicitação e nenhuma informação deve ser armazenada no servidor. A desvantagem é que isso pode diminuir o desempenho da rede aumentando a quantidade de dados redundantes enviados em uma série de requisições.

- **Cache** - uma maneira de diminuir o impacto da desvantagem trazida pela redução de desempenho da rede é a utilização de *cache*. Esta restrição exige que os dados de resposta de uma requisição sejam rotulados como *cacheable* ou *non-cacheable* (passíveis ou não de utilização da *cache*). Se uma resposta for marcada como *cacheable* então o cache do cliente poderá reutilizá-la para futuras requisições equivalentes. A vantagem de usar as restrições de *cache* é que elas têm o potencial de eliminar parcialmente ou completamente algumas interações, aumentando a eficiência, a escalabilidade, e o desempenho percebido pelo usuário, reduzindo a latência média de uma série de requisições.
- **Interface Uniforme** - a principal característica que diferencia o estilo arquitetural REST de outros estilos baseados em rede é sua ênfase em uma interface uniforme entre componentes. Para obter essa uniformidade, o REST define quatro restrições interface: (1) identificação de recursos; (2) manipulação de recursos através de representações; (3) mensagens auto-descritivas; e (4) hipermídia como mecanismo de estado da aplicação. O HTTP é o protocolo mais indicado para se trabalhar com o REST. O mesmo fornece uma interface uniforme com um conjunto fixo de quatro operações básicas. Os métodos GET, POST, PUT e DELETE compõem estas operações. O GET é utilizado para recuperar um recurso. Com o POST é possível criar um novo recurso. O PUT atualiza um recurso. Por fim, o DELETE permite remover um recurso.
- **Multicamada** - com a intenção de aprimorar os requisitos de escalabilidade da internet, foi incorporado ao REST a restrição de divisão em camadas. O estilo de sistema em camadas permite que uma arquitetura seja composta de camadas hierárquicas restringindo o comportamento de cada componente de acordo com a funcionalidade, de forma que cada componente seja capaz de interagir apenas com sua camada imediata.
- **Código por Demanda** - a última característica do estilo REST é uma restrição opcional. O REST permite que funcionalidades do cliente sejam estendidas através do download e execução de códigos na forma de *applets* e *scripts*. Isso torna a implementação do cliente mais simples, reduzindo a quantidade de recursos necessários para ser pré-implementado.

3 TRABALHOS RELACIONADOS

Atualmente, a criação de conteúdos de audiodescrição é um processo que demanda muito tempo, tendo em vista que a maior parte do trabalho de produção é feito manualmente [15]. Por essa razão, alguns trabalhos vêm sendo desenvolvidos com o auxílio de tecnologias computacionais para dar suporte às etapas envolvidas no processo de criação de audiodescrição. Neste capítulo serão apresentados alguns trabalhos relacionados a tecnologias para geração de recursos de AD.

3.1 Descrevendo vídeos *on-line* com síntese de fala para narração

Em Kobayashi [14] é apresentada uma plataforma que usa síntese de voz para adicionar AD a vídeos *on-line* usando tecnologia *text-to-speech* (TTS) que permite que roteiros de AD sejam utilizados no formato de texto sem a necessidade de um audiodescritor-narrador.

A arquitetura da plataforma consiste de um editor de roteiro, um *player* de vídeo, um repositório de metadados e um formato baseado em texto utilizado para a troca de roteiros entre os componentes. A Figura 4 apresenta esta arquitetura.

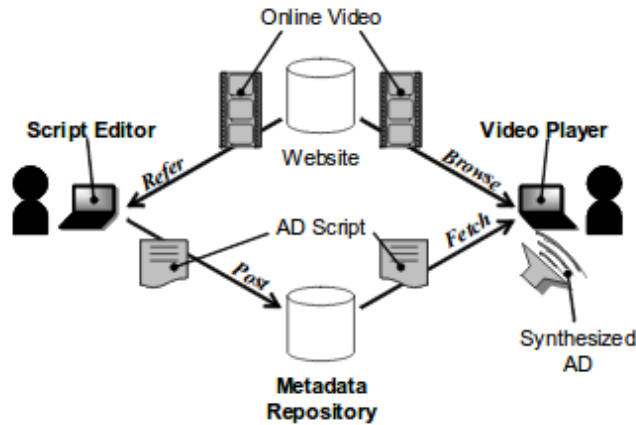


Figura 4: Arquitetura proposta em [14].

O editor de roteiro é uma ferramenta que fornece uma interface visual que permite a criação e edição de roteiros de AD. Nela é possível editar as sentenças de AD e especificar, na linha do tempo do vídeo, os instantes em que cada sentença deve ser narrada. Essa ferramenta também permite a modificação dos parâmetros de voz, tais como gênero do narrador e a velocidade.

Para facilitar a identificação da localização de espaços onde narrações de audiodescrição podem ser inseridas, a interface também exibe a forma de onda da trilha de áudio original. Ao inserir uma sentença de AD, a duração da narração é exibida na

linha do tempo. Esses recursos de visualização permitem ao usuário criar e revisar a AD. Finalmente, os roteiros criados são armazenados e compartilhados em um repositório de metadados. Uma vez no repositório, o tocador de vídeo busca a AD do vídeo acessando o repositório de metadados e apresenta as narrações assim que o usuário reproduz o vídeo associado. A plataforma proposta permite sintetizar as narrações tanto no servidor, quanto no cliente.

Experimentos foram feitos com pessoas com deficiência visual para analisar a aceitabilidade dos sintetizadores de AD, bem como o impacto da qualidade das vozes na inteligibilidade dos vídeos. Para isso, foram exibidos vídeos de diversas categorias com AD em diferentes qualidades da voz sintetizada e também com a versão gravada por um narrador profissional humano para servir de apoio na análise dos resultados. Os resultados mostraram que a voz sintetizada era aceitável e poderia melhorar consideravelmente a experiência do usuário em relação a vídeos que não apresentam AD. O trabalho, contudo, não apresenta testes feitos com a ferramenta de edição de roteiro.

3.2 Anotação colaborativa para acessibilidade de vídeo

No trabalho de Champin [3], é apresentando um estudo preliminar conduzido com usuários cegos a fim de abordar problemas de multimodalidade e, em particular, como renderizar vídeos que foram enriquecidos com anotações. A partir disso, foram extraídos requisitos para o desenvolvimento de um sistema com o objetivo de melhorar a acessibilidade em vídeos.

O sistema apresenta uma abordagem que fornece descrições detalhadas de conteúdos de vídeo para personalizar a renderização do conteúdo de acordo com as deficiências sensoriais do usuário. A arquitetura geral do sistema e o fluxo de trabalho são apresentados na Figura 5, onde os blocos pretos representam os componentes a serem desenvolvidos.

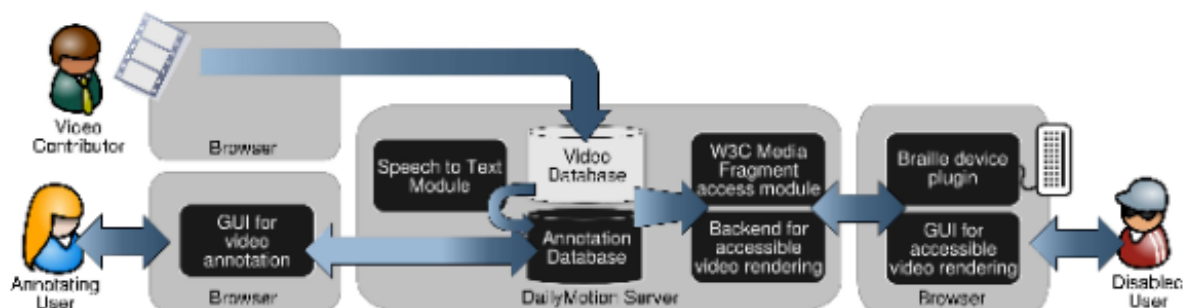


Figura 5: Sistema proposto em [3].

De acordo com a Figura 5, o servidor Dailymotion tem um banco de dados contendo vídeos enviados por uma variedade de colaboradores. Esse banco de dados é complementado com uma base de anotações contendo todas as informações adicionais necessárias

para tornar os vídeos acessíveis. Essas informações devem ser produzidas por uma comunidade de usuários responsáveis por criar anotações através de uma GUI específica para ajudá-los na tarefa. Como as transcrições de vídeos serão, obviamente, uma parte importante das anotações, os usuários também são auxiliados por um módulo integrado de fala para texto. Por fim, o vídeo e suas anotações são combinados para fornecer aos usuários com deficiência visualizações adaptadas.

O acesso no lado do cliente deverá ser feito através de interfaces desenvolvidas especificamente para pessoas cegas, usando tecnologias disponíveis em navegadores modernos. Um *plugin* de navegador de código aberto para a condução de dispositivos braille também foi desenvolvido.

Por ser de caráter exploratório, e ter como objetivo apenas definir um sistema que contribua com o enriquecimento da acessibilidade em vídeos, o trabalho não apresenta testes e tem como trabalho futuros o desenvolvimento do sistema.

3.3 Descrição de vídeo acessível sob demanda

Fornecer às pessoas cegas e com deficiência visual as descrições dos principais elementos visuais pode melhorar muito a acessibilidade de vídeos, cinema e televisão. Este projeto apresenta uma plataforma *web* para a descrição de vídeo (DV) usando um tocador adaptado [4].

O objetivo deste trabalho é testar a usabilidade de um *player* acessível que fornece aos usuários finais vários níveis de descrição de vídeo sob demanda. Para isso, foi desenvolvido um site acessível que apresentou cinco curtas que puderam ser exibidos com o VDPlayer. Também foram produzidas descrições DV com a renderização de uma voz sintética e fornecidas em dois níveis de DV: 1) o modo padrão que deu a DV nos segmentos sem voz e 2) o modo estendido que oferecia todas as DVs produzidas mesmo que excedendo os segmentos sem fala. Com relação a geração das descrições de vídeo, o trabalho não menciona se é feita manualmente ou de forma automática.

O VDPlayer foi projetado para 1) oferecer controles padrão de vídeo, como reproduzir, pausar, retroceder, avançar e mudar o volume, e 2) fornecer controles específicos para a DV como, por exemplo, permitir que os usuários selecionem o nível de DV.

A validação deste trabalho foi dividida em duas partes: avaliação do VDPlayer e avaliação da descrição de vídeo. A primeira parte teve como objetivo estabelecer a relevância e a facilidade de uso dos controles do *player* de vídeo relacionados a DV. O resultados mostraram que 90% dos participantes classificaram os controles do VDPlayer como relevantes e fáceis de usar. A segunda parte foi feita através de uma série de nove afirmações que, segundo os autores do trabalho, tiveram tom negativo e positivo, porém não são apresentados exemplos. Os participantes tiveram que escolher o nível de

concordância ou discordância para cada uma delas. Os resultados mostraram que 92% concordaram com as afirmações positivas. Enquanto que apenas 15% concordaram com as afirmações negativas.

3.4 Troca de Vídeo Descritivo

A *Descriptive Video Exchange* (DVX) é uma nova tecnologia que vem sendo desenvolvida pelo Miele Lab [7]. Essa tecnologia permite aos espectadores de vídeo descrever verbalmente o que veem em um filme e compartilhar esses cliques de descrição e suas informações de sincronização pela Internet, aumentando assim a acessibilidade generalizada para espectadores cegos. A plataforma DVX é baseada no modelo Cliente/Servidor que armazena todos os cliques de descrição, as informações de sincronização e os identificadores exclusivos do conteúdo de vídeo original em um servidor dedicado. O cliente DVX é um software tocador de vídeo que reproduz vídeos e sincroniza automaticamente os cliques de descrição com o vídeo original, enquanto o reproduz a partir de sua fonte original. O cliente DVX também é usado para gravar os cliques descritivos e carregá-los automaticamente no servidor DVX. A Figura 6 apresenta uma visão geral da plataforma que está em desenvolvimento.



Figura 6: Visão geral da plataforma em desenvolvimento.

Os testes com essa plataforma estão seguindo um programa de pesquisa rigorosa para avaliar 1) a eficácia do DVX na gravação e distribuição da descrição de áudio amador, 2) a eficácia das ferramentas digitais automatizadas para melhorar a apresentação de descrição amadora de áudio e 3) a eficácia do uso de redes sociais e comunidades online para o recrutamento e treinamento de audiodescritores voluntários.

3.5 Descrição remota em tempo real

Este trabalho descreve uma técnica que está sendo desenvolvida e permitirá que um descritor em qualquer lugar do mundo forneça uma descrição em tempo real de um fluxo de vídeo visualizado por um aluno com deficiência visual em casa, na sala de aula ou em qualquer lugar [21]. No *Remote real-time description* (RRTD), um vídeo é transmitido para o descritor que passa o áudio (e opcionalmente o vídeo) para o aluno, juntamente com a descrição ao vivo adicionada. Isso garante a sincronização do fluxo de áudio/vídeo original com o comentário descritivo. Se o *webcast* ou *webinar* for interativo, qualquer áudio e/ou vídeo do aluno será transmitido diretamente de volta à fonte original para permitir que o aluno faça perguntas ou faça comentários.

Um dos pontos fortes do RRTD é que ele exigirá apenas um descritor disposto e uma conexão com a Internet de ambas as partes. Esta técnica possibilitará um parente em casa fornecer serviços de descrição gratuita para um aluno em sala de aula. Também permitirá que um descritor profissional em um local remoto forneça serviços descritivos para um ou mais alunos com deficiências visuais.

3.6 Características gerais dos trabalhos

De modo a facilitar o entendimento dos trabalhos relacionados em comparação com o trabalho apresentado nesta monografia, a Tabela 1 apresenta um resumo esquemático comparativo. Neste resumo são apresentadas as principais características dos trabalhos relacionados no que se refere ao tipo de geração, entradas, saídas e a forma como a AD é distribuída.

Conforme pode ser observado na Tabela 1, os trabalhos relacionados apresentam a forma de geração de AD manual ou semi-automática, isso significa que nessas abordagens sempre há a interferência humana no processo.

Em relação aos dados de entrada, a maioria dos trabalhos usam o vídeo como a principal mídia para a geração da audiodescrição. Em alguns trabalhos, outros arquivos também são necessários para dar suporte na geração e distribuição da AD. As mídias resultantes do processo de criação de AD desses trabalhos são arquivos de áudio contendo voz sintetizada ou gravada que são distribuídos em conjunto com os vídeos de entrada através de reprodutores de vídeo e navegadores de internet projetados para essa finalidade.

Apesar de oferecerem soluções que contribuem para facilitar e aumentar a distribuição de recursos de audiodescrição em vídeos, os trabalhos apresentados neste capítulo ainda dependem da intermediação humana para a criação de AD. Por outro lado, este trabalho propõe a geração de AD totalmente automática, baseando-se apenas nos conteúdos enviados pelos usuários. Outra característica importante é a forma de distribuição das

Tabela 1: Características gerais dos trabalhos relacionados.

Trabalho	Tipo de Geração	Tipo de Entrada	Tipo de Saída	Forma de distribuição
Descrevendo vídeos <i>on-line</i> com síntese de fala para narração	semi-automática	Vídeo e Arquivo de texto temporizado	Vídeo e Voz Sintetizada	Reprodutor de vídeo
Anotação colaborativa para acessibilidade de vídeo	semi-automática	Vídeo e Anotações	Vídeo, Braille e Voz Sintetizada	Navegador de internet / Plugin de braille
Descrição de vídeo acessível sob demanda	Não identificada	Não identificada	Vídeo e Voz Sintetizada	Reprodutor de vídeo em um Website
Troca de Vídeo Descritivo	Manual	Vídeo e Voz Humana Gravada	Vídeo e Voz Gravada	Reprodutor de vídeo
Descrição remota em tempo real	Manual	Streaming de Vídeo	Streaming de Áudio	Webcast ou Webinar

mídias geradas, que nos trabalhos relacionados é feita através de tecnologias desenvolvidas especificamente para reprodução de AD em vídeos, enquanto que neste trabalho o recurso de AD já vem integrado no vídeo, o que permite que o conteúdo seja consumido através de reprodutores de vídeos tradicionais. O capítulo seguinte apresentará a solução proposta do presente trabalho.

4 SOLUÇÃO PROPOSTA

De acordo com o que foi exposto no Capítulo 1, a solução tem como objetivo contribuir para a redução dos problemas de acessibilidade enfrentados por pessoas cegas em diferentes ambientes que oferecem experiências audiovisuais através de vídeos digitais. Nesse contexto, a proposta deste trabalho é permitir que usuários, por meio de aplicações multiplataforma, possam submeter vídeos a um serviço *web* capaz de gerar trilhas de AD automaticamente, tomando como base informações extraídas do próprio vídeo, da legenda SRT associada e, quando disponível, do roteiro cinematográfico. Para isso, foram desenvolvidos um *Web Service* e um gerador de audiodescrição responsáveis por atender requisições de aplicações e gerar de forma automática trilhas de AD, respectivamente.

O funcionamento geral do sistema é apresentado na Seção 4.1. Nas Seções 4.2 e 4.3 são descritos os componentes que integram a solução proposta deste trabalho.

4.1 Serviço Web de Geração de AD

O serviço de geração de AD tem como objetivo contribuir com a ampliação da acessibilidade para pessoas com deficiência visual. Para isso, a abordagem consiste na disponibilização de um serviço *web* de geração automática de trilhas de audiodescrição que tem como dados de entrada três mídias digitais: vídeo, legenda no formato SRT, e roteiro cinematográfico, sendo esta última, opcional.

O serviço é composto por um conjunto de componentes responsáveis por: atender as requisições de aplicações; gerar um roteiro de AD a partir de informações extraídas dos dados de entrada; transformar o roteiro de AD em uma trilha de áudio utilizando técnicas de síntese de voz; sincronizar a trilha de AD com vídeo enviado pela aplicação e disponibilizá-los para o requerente. A visão esquemática da solução proposta é ilustrada na Figura 7.

De acordo com a Figura 7, o serviço funciona da seguinte maneira: uma aplicação que precise tornar um vídeo digital acessível envia por meio da API de Comunicação os dados de entrada ao Servidor de Aplicação que, através de uma fila de mensagens, e em conjunto com alguns parâmetros de configuração, os encaminha para o processo de Geração Automática de AD, que por sua vez gera a trilha de AD e a retorna para o servidor de aplicação usando uma segunda fila de mensagens. Por fim, a aplicação pode realizar o download da mídia gerada.

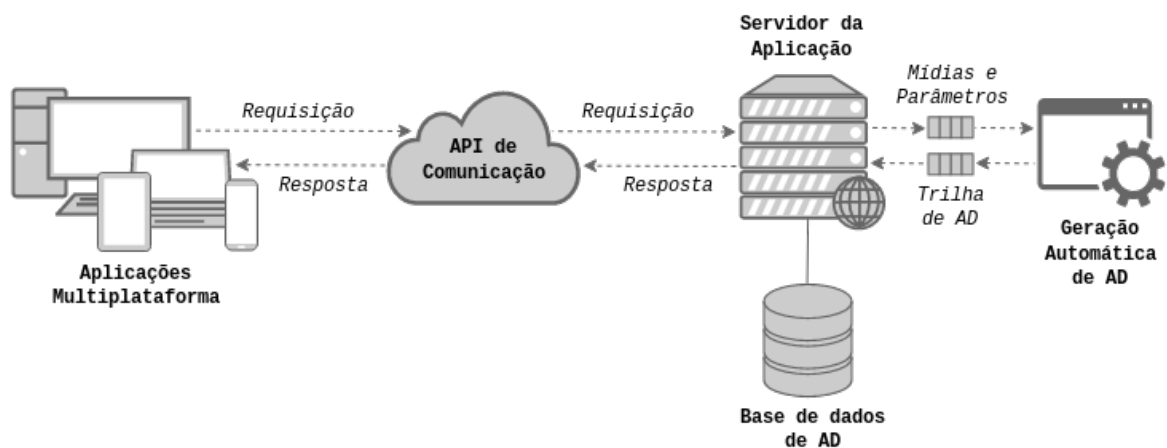


Figura 7: Visão Esquemática do Serviço Proposto.

4.2 Web Service

Para tornar a geração de trilhas de AD disponível para ser acessada por diferentes aplicações na *web*, foi desenvolvido um *Web Service* contendo um servidor de aplicação responsável pela manipulação das solicitações de geração de audiodescrição. O serviço também define uma API REST, por onde as aplicações podem fazer requisições ao servidor de aplicação. Além disso, a API conta com um conjunto de *endpoints* que expõem como o serviço pode ser acessado para a geração de conteúdos de audiodescrição. Esses *endpoints* são apresentados a seguir.

- **Endpoint de geração de trilha de AD integrada (com roteiro):** recebe os arquivos de vídeo, legenda e roteiro cinematográfico associados, gera automaticamente a trilha de audiodescrição e retorna para a aplicação o vídeo com o recurso de AD integrado;
- **Endpoint de geração de trilha de AD integrada (sem roteiro):** semelhante ao *endpoint* de geração de trilha de AD integrada com roteiro, esse *endpoint* gera automaticamente a trilha de audiodescrição, porém, utiliza apenas as informações extraídas do vídeo e da legenda. O retorno desse *endpoint* também devolve o vídeo com a trilha de AD embutida;
- **Endpoint de geração de trilha de AD (com roteiro):** recebe os arquivos de vídeo, legenda e roteiro cinematográfico e retorna apenas a trilha de AD gerada automaticamente (i.e., sem mixá-la com a trilha de áudio original do vídeo);
- **Endpoint de geração de trilha de AD (sem roteiro):** assim como o *endpoint* de geração de trilha de AD com roteiro e seguindo o raciocínio do *endpoint* de geração de trilha de AD integrada sem roteiro, esse *endpoint* recebe apenas o vídeo e a legenda e retorna a trilha de audiodescrição;

- **Endpoint de geração de roteiro de AD (com roteiro):** diferente dos demais *endpoints* que geram arquivos de áudio e vídeo, esse *endpoint* é responsável por produzir automaticamente o roteiro de AD a partir do vídeo, da legenda e do roteiro original (de cinema). Através desse *endpoint*, especialistas em audiodescrição podem revisar o roteiro de AD antes de submetê-lo ao processo de síntese de fala ou à locução humana;
- **Endpoint de geração de roteiro de AD (sem roteiro):** recebe os arquivos de vídeo e legenda para gerar o roteiro de AD. Esse *endpoint* é usado quando o roteiro original não está disponível;
- **Endpoint de geração de trilha de AD integrada (com roteiro de AD):** recebe os arquivos de vídeo e roteiro AD, gera automaticamente a trilha de AD usando apenas a síntese de fala e retorna o vídeo com a audiodescrição mixada. Esse *endpoint* é um complemento que permite que o roteiro de AD revisado possa ser usado como mídia de entrada para a geração da trilha de AD;
- **Endpoint de geração de trilha de AD (com roteiro de AD):** esse *endpoint* recebe apenas o arquivo do roteiro de AD e retorna a trilha de áudio contendo a audiodescrição.

Adicionalmente, para facilitar o acesso e o monitoramento da geração dos conteúdos de AD, foram desenvolvidos os *endpoints* para **verificação de status da geração de mídia** e **download da mídia gerada**. Para mais detalhes sobre as especificações dos endpoints, a documentação da API pode ser consultada no Apêndice A.

Com relação às tecnologias utilizadas para o desenvolvimento desse componente do sistema, foram usados o ambiente Node.js², um interpretador de códigos JavaScript focado no lado do servidor, em conjunto com o *framework* para construção de aplicativos web Express³ e o software de banco de dados NoSQL MongoDB⁴.

4.3 Geração Automática de AD

Conforme mencionado na Seção 4.1, os vídeos enviados por aplicações clientes são encaminhados para o processo de geração automática de audiodescrição, que é responsável pela geração dos conteúdos de AD. Essa parte do sistema é composta por um conjunto de componentes onde cada um deles é encarregado por uma etapa específica do processo de geração das mídias de AD.

²<<https://nodejs.org/en/>>

³<<https://expressjs.com/pt-br/>>

⁴<<https://www.mongodb.com/>>

De acordo com a Figura 8, o componente Controlador é responsável por receber os arquivos e parâmetros enviados pelo servidor de aplicação. Esse componente conduz todo o processo de geração automática de AD e é responsável por invocar os demais componentes de acordo com o tipo de requisição recebida e dos conteúdos disponibilizados.

Após receber o pedido de geração de AD, o Controlador decide então por encaminhar o vídeo para a Análise de Vídeo, que processa os quadros do vídeo, identifica os objetos, o cenário e as ações, ou, caso o roteiro esteja disponível para a Análise de Roteiro que seleciona as sentenças mais relevantes para serem descritas. Em ambas as análises o resultado é um arquivo de texto temporizado (Roteiro de AD). Esses componentes são propostos por Campos (Analisador de roteiro) [1] e pelo Grupo de Trabalho *Accessibility as a Service* (Analisador de vídeo) [11].

Em seguida, após a criação do roteiro de AD, o controlador invoca o componente Sintetizador de Voz, que é responsável pelos processos de síntese de voz e mixagem dos arquivos de áudio do vídeo e da audiodescrição. Esse componente faz a leitura do roteiro de AD em busca das sentenças de descrição e, para cada sentença encontrada, executa uma ferramenta *Text-To-Speech* (TTS) que gera um arquivo de áudio no formato *WAVEform* contendo o texto falado. Continuando o processo, as mídias de áudio são concatenadas levando-se em conta as marcações de tempo do roteiro de AD, resultando num arquivo único. Por fim, de acordo com a requisição feita pela aplicação cliente, a trilha de AD pode ainda ser mixada com o áudio original do vídeo.

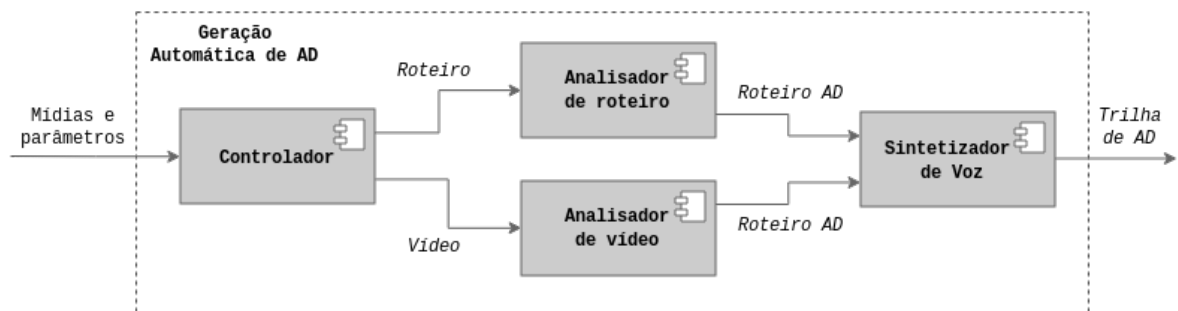


Figura 8: Visão esquemática do processo de geração automática de AD.

Para o desenvolvimento dos componentes Controlador e Sintetizador de Voz, é utilizada a linguagem de programação C++. No processo de síntese de voz, são usadas duas ferramentas de TTS: Amazon Polly⁵, do *framework* Amazon Web Service (AWS), e o Espeak⁶, um software de código aberto para Linux. A mixagem da trilha original do vídeo com a trilha de audiodescrição é feita com o suporte do FFmpeg⁷. A comunicação entre o processo de Geração Automática de AD e o Servidor de Aplicação é feita através

⁵<<https://aws.amazon.com/pt/polly/>>

⁶<<http://espeak.sourceforge.net/>>

⁷<<https://www.ffmpeg.org/>>

do software de mensageria RabbitMQ⁸ e de um *Wrapper* na linguagem de programação Python no lado do Controlador.

⁸<<https://www.rabbitmq.com/>>

5 RESULTADOS E DISCUSSÕES

Os resultados obtidos a partir de experimentos com a solução proposta serão apresentados e discutidos neste capítulo. O objetivo é validar a solução proposta em relação à qualidade dos conteúdos de acessibilidade gerados e ao desempenho computacional do sistema, através de um conjunto de testes que foi planejado e executado em duas partes.

A primeira parte consistiu na avaliação do nível de compreensão de usuários cegos sobre o contexto geral e alguns aspectos chaves da história de vídeos com a audiodescrição gerada pela solução proposta. Na segunda parte dos testes, a proposta foi medir o tempo que o serviço leva para processar por completo as solicitações de geração de conteúdos de AD. A primeira e a segunda parte dos testes serão apresentadas nas Seções 5.1 e 5.2, respectivamente.

5.1 Compreensão Geral dos Vídeos

Conforme mencionado no início do Capítulo 5, a primeira parte dos testes teve como objetivo avaliar o nível de compreensão de usuários cegos sobre o contexto geral da história ao assistirem vídeos com o recurso de AD gerado pelo serviço proposto. Nessa etapa, foram utilizados dois vídeos, sem considerar o respectivos roteiros, isto é, a AD foi gerada pelo componente de análise de vídeo. Com isso, foi possível avaliar o nível de compreensão geral do vídeo pelos usuários cegos em cenários onde o roteiro do vídeo não está disponível. A Tabela 2 apresenta as informações gerais dos vídeos utilizados.

Tabela 2: Características dos Vídeos da Primeira Parte dos Testes.

Informações Gerais	Vídeo A	Vídeo B
Duração (HH:MM:SS)	00:13:52	00:14:34
Número de cenas	7	26
Número de diálogos	39	35

De acordo com a Tabela 2, o Vídeo A contém 7 cenas, 39 diálogos e a duração de 13 minutos e 52 segundos. Já o Vídeo B contém uma duração de 14 minutos e 34 segundos, 26 cenas e 55 diálogos.

Os testes foram realizados com um grupo de usuários composto por onze pessoas com deficiência visual da Universidade Federal da Paraíba, Brasil. Inicialmente, os usuários preencheram um questionário que coletou alguns dados pessoais. O grupo de usuários era formado por sete homens e quatro mulheres na faixa etária de 12 à 60 anos, com idade média de 34,4 anos. Quanto ao grau de deficiência visual, todos os usuários eram completamente cegos, sendo sete usuários com cegueira congênita e quatro com cegueira adquirida.

Após completar a primeira parte do questionário, os usuários foram convidados a assistir os dois vídeos e a responder às seguintes perguntas:

1. O que acontece na história?
2. Onde você acha que a história acontece? Consegue identificar o ambiente?
3. Quantos personagens participam da história? Qual o gênero dos personagens (mulher ou homem)?
4. A história se passa de dia, de noite, de madrugada?

As perguntas tiveram como objetivo avaliar se, ao assistirem os vídeos com a AD gerada pela solução proposta, os usuários com deficiência visual eram capazes de compreender o contexto geral do vídeo, o ambiente onde a história acontecia, além de outras informações importantes no contexto audiovisual como, por exemplo, informações relacionadas aos personagens e ao horário em que a história acontecia.

Para o Vídeo A, todos os usuários responderam a Pergunta 1 com uma descrição geral adequada sobre a história. Cinco usuários apresentaram uma descrição mais detalhada da história, incluindo informações específicas sobre os personagens, enquanto cinco usuários apresentaram descrições mais resumidas. Em relação à Pergunta 2, seis usuários descreveram corretamente o ambiente, dois descreveram de forma incorreta, e dois usuários não responderam por não conseguirem identificar o ambiente. Além disso, dos seis usuários que descreveram o ambiente corretamente, dois deles reportaram que não tinham certeza se a descrição do ambiente era realmente aquela. Com relação à Pergunta 3, todos os usuários responderam corretamente, identificando também o gênero dos personagens. Finalmente, com relação à Questão 4, nove usuários identificaram corretamente que a história se passava durante o dia, enquanto dois usuários reportaram que não foram capazes de identificar o horário em que a história se passava.

Para o Vídeo B, com relação à Pergunta 1, dez usuários apresentaram uma descrição detalhada e correta sobre a história, enquanto que um usuário reportou que conseguiu compreender a história apenas pela narrativa, pois não gostou da audiodescrição gerada. Sobre a Pergunta 2, dez usuários descreveram o ambiente com exatidão, enquanto um usuário reportou que não conseguiu identificar onde aconteceu a história pois a audiodescrição não era clara. Em relação à Pergunta 3, todos os usuários conseguiram apresentar uma boa descrição dos personagens. Com relação à Pergunta 4, a história continha cenas que ocorriam durante o dia intercaladas com cenas que ocorriam durante a noite. Nesta pergunta, sete usuários descreveram o horário corretamente, dois responderam que a história acontecia apenas durante o dia e dois responderam que não conseguiu

identificar o horário. Dentre os usuários que responderam corretamente, um deles também reportou que não conseguiu identificar o horário com precisão.

Baseado nesses resultados, é possível observar que a maioria dos usuários conseguiu compreender corretamente os aspectos gerais da história, entretanto, alguns usuários tiveram dificuldades para identificar os ambientes e os horários. Por conta disso, continuar investigando estratégias que utilizam técnicas para melhorar a descrição de vídeos quando não há roteiro disponível é uma das propostas de trabalhos futuros.

Finalmente, para conferir o nível de satisfação dos participantes em relação ao serviço proposto, eles foram convidados a responder a mais duas questões subjetivas. Na primeira pergunta, utilizando uma escala de 1 à 6, eles responderam se a AD gerada pelo serviço era fácil de entender, em que 6 corresponde a “muito fácil de entender” e 1 corresponde a “muito confuso”. Na segunda pergunta, eles responderam se essa audiodescrição ajudava na compreensão do vídeo e também utilizando uma escala de 1 à 6, em que 6 representa “ajuda muito” e 1 representa “muito confusa”. O resultados obtidos para essas perguntas são apresentados nas Figuras 9 e 10.

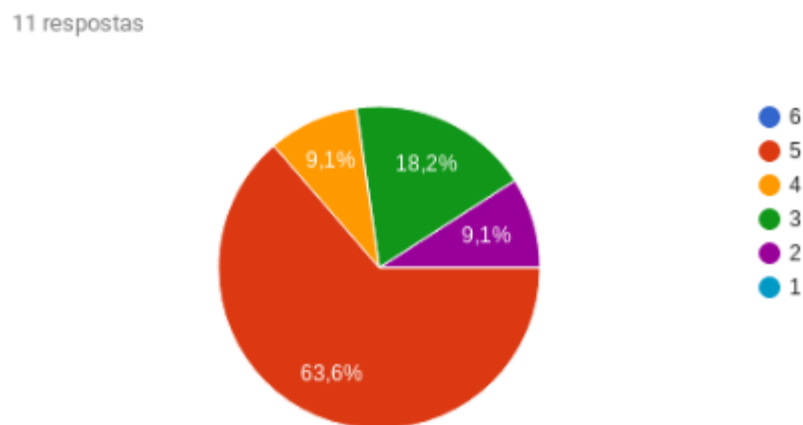


Figura 9: Avaliação dos usuários sobre a facilidade de entendimento da AD gerada pelo protótipo.

De acordo com a Figura 9, sete usuários (63,6%) que participaram do teste, deram nota 5 para a facilidade de compreender a audiodescrição gerada pelo serviço, um usuário (9,1%) deu nota 4, dois (18,2%) deram nota 3 e um usuário (9,1%) deu nota 2. Analisando esses dados, é possível observar que aproximadamente 73% dos usuários atribuíram um conceito maior ou igual a 4 sobre a facilidade de entender a AD gerada pela solução proposta, enquanto que aproximadamente 27% atribuíram um conceito inferior ou menor do que 3.

Com relação ao questionamento se a AD gerada pelo serviço ajuda na compreensão dos conteúdos (ver Figura 10), nove usuários (81,9%) atribuíram um conceito igual ou superior a 4, enquanto que apenas dois usuários (18,1%) atribuíram um conceito menor

ou inferior a 3, o que é um indício de que a AD gerada pelo serviço pode auxiliar na compreensão dos conteúdos.

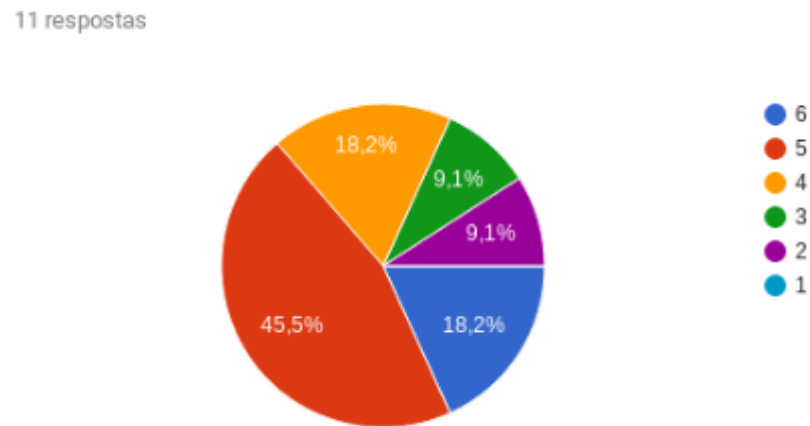


Figura 10: Avaliação dos usuários sobre contribuição da AD gerada pelo protótipo na compreensão dos vídeos.

5.2 Tempo de Geração das Trilhas de AD

Na segunda parte dos testes, o objetivo foi avaliar o tempo que a solução proposta demorava para atender as solicitações de geração de trilhas de AD. Para isso, foram selecionados dois vídeos que continham os seus respectivos roteiros: um curta-metragem (Vídeo C) e um longa-metragem (Vídeo D). A Tabela 3 apresenta as informações gerais dos vídeos utilizados.

Tabela 3: Características dos Vídeos da Segunda Parte dos Testes.

Informações Gerais	Vídeo C	Vídeo D
Duração (HH:MM:SS)	00:14:33	02:04:45
Resolução	640x360	528x336

De acordo com a Tabela 3, o Vídeo C tem uma resolução de 640x360 pixels e uma duração de 14 minutos e 33 segundos. Já o Vídeo D tem uma resolução de 528x336 pixels e a sua duração é de 2 horas, 4 minutos e 45 segundos.

Para a medição do tempo de geração das mídias de AD, os *endpoints* de Geração de trilha de AD integrada (com roteiro) e Geração de trilha de AD integrada (sem roteiro) foram selecionados, uma vez que esses *endpoints* demandam todas as etapas do processo de geração automática de audiodescrição. A máquina utilizada na execução dos testes possuía as seguintes configurações: processador Intel Xeon E5-2640 2.50 GHz, placa de vídeo Nvidia GeForce GTX 1080 Ti, 16 Gigabytes de memória RAM e Sistema Opera-

cional Linux Ubuntu 16.04 LTS Xenial Xerus. Quanto ao tempo para realizar os testes, foram necessárias cerca de 65,6 horas de processamento, distribuídas em uma semana.

O conjunto de testes que mediu o tempo necessário para o serviço proposto gerar as trilhas de AD foi projetado e executado em três etapas:

1. Mensuração do tempo de geração variando a duração dos vídeos;
2. Mensuração do tempo de geração variando a resolução dos vídeos;
3. Mensuração do tempo de geração para um conjunto de cem requisições.

A divisão do conjunto de testes nessas três etapas teve como objetivo avaliar o desempenho da solução proposta em cenários reais onde os vídeos recebidos variam de duração e resolução. Além de observar o comportamento do serviço ao processar uma quantidade crescente de requisições.

5.2.1 Mensuração do Tempo de Geração Variando a Duração dos Vídeos

Como disposto na Seção 5.2, a primeira etapa teve como foco a medição do tempo de geração da trilha de AD quando a duração dos vídeos variam. As Figuras 11 e 12 apresentam os resultados obtidos para os dois casos avaliados onde a solução proposta dispõe e não dispõe do roteiro, respectivamente.

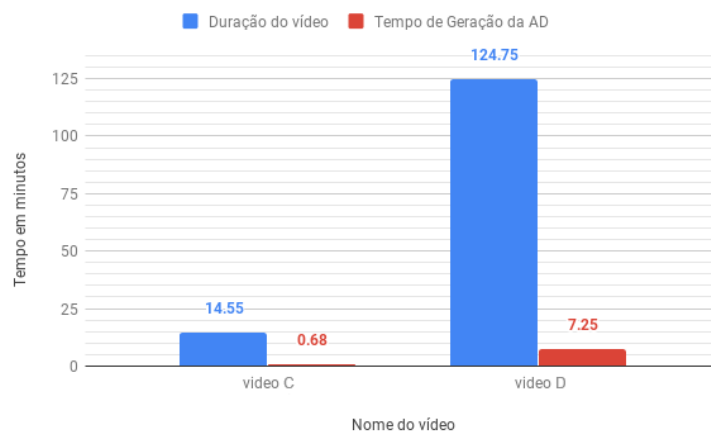


Figura 11: Tempo de geração da trilha de AD com roteiro.

Conforme a Figura 11, para o vídeo com duração de 14,55 minutos (Vídeo C) o tempo gasto na geração da audiodescrição foi de 0,68 minutos, isto é, aproximadamente 41 segundos. Para o Vídeo D, com duração de 124,75 minutos, o tempo da geração da trilha de AD foi de 7,25 minutos. Contrastando com os resultados obtidos no primeiro

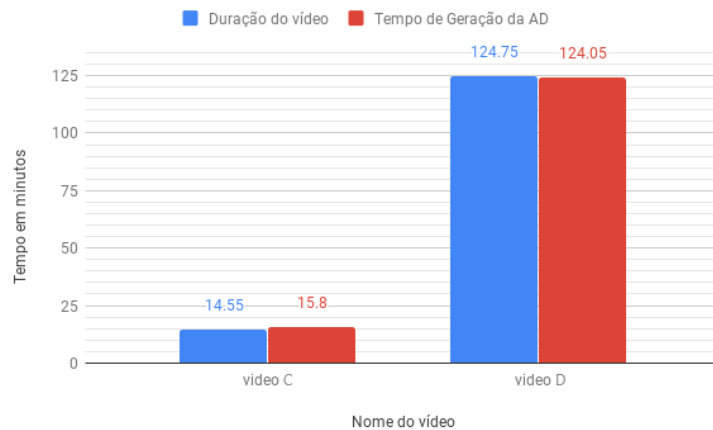


Figura 12: Tempo de geração da trilha de AD sem roteiro.

caso, o tempo de geração da trilha de AD quando o roteiro não está presente (ver Figura 12), passa a ser 15,8 minutos para o vídeo C, e 124,05 minutos para o Vídeo D.

Esses resultados apresentaram uma grande diferença entre as duas abordagens de geração de trilha de AD usadas na solução proposta. Enquanto o processo que dispõe dos roteiros manteve um tempo de geração relativamente baixo comparado com a duração dos vídeos, a geração da trilha de AD usando informações extraídas apenas dos vídeos manteve o tempo de processamento próximo da duração dos vídeos. Dessa maneira, é preferível a disponibilização do roteiro para gerar a mídia de audiodescrição, no entanto, a maioria dos vídeos disponíveis atualmente não possuem seus respectivos roteiros, o que acaba gerando a necessidade de aprimoramento das técnicas de geração de trilhas de AD baseada apenas em vídeo. Ainda assim, comparando-se com a produção manual feita por profissionais, ambas as estratégias de geração automática de AD são mais vantajosas em relação ao tempo gasto na criação de audiodescrição.

5.2.2 Mensuração do Tempo de Geração Variando a Resolução dos Vídeos

Para a segunda etapa, os testes tiveram como proposta mensurar o tempo de geração das trilhas de AD quando a resolução dos vídeos varia. Nesse caso, a utilização do roteiro não se fez necessária, visto que a mudança da resolução dos vídeos somente causa impacto na análise de vídeo sem roteiro.

A Figura 13 apresenta os resultados obtidos após a execução da solução proposta usando o Vídeo C com a resolução original (640x360) e com a resolução reduzida em uma escala de 2 (320x180) como entrada. Com a duração de 14,55 minutos, o Vídeo C na resolução 640x360 teve sua trilha de AD gerada pelo serviço em 15,8 minutos, enquanto que com a resolução reduzida para 320x180 o tempo de geração reduziu para 11,85 minutos. Da mesma forma, como mostra a Figura 14, para o Vídeo D com 124,75 minutos de

duração, os resultados foram 124,05 minutos com a resolução original (528x336) e 99,83 minutos com a resolução reduzida (264x168).

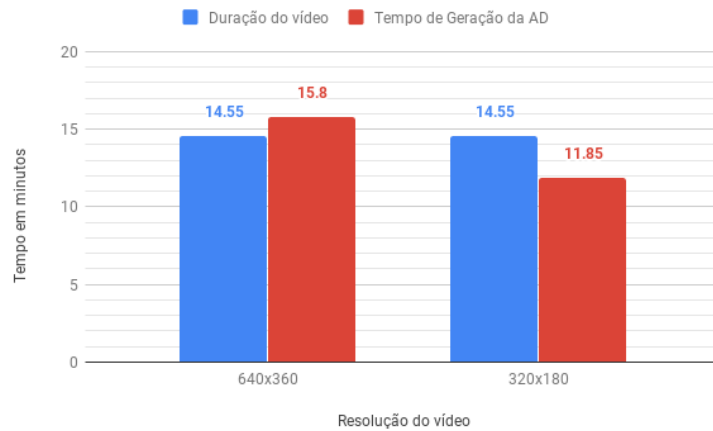


Figura 13: Tempo de geração da trilha de AD variando a resolução do vídeo C.

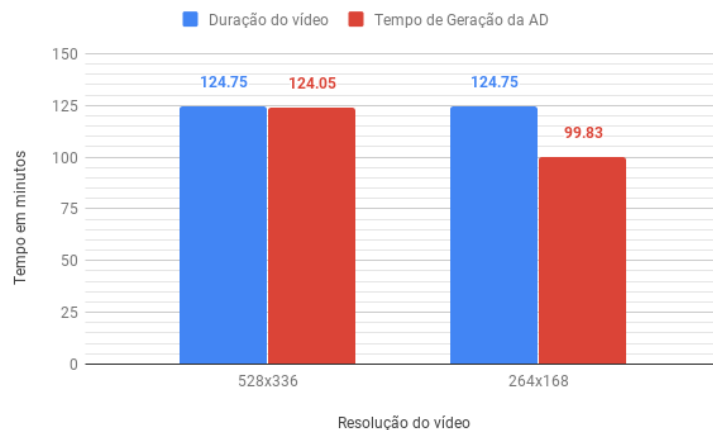


Figura 14: Tempo de geração da trilha de AD variando a resolução do vídeo D.

Os resultados desta etapa mostraram que a diminuição da resolução dos vídeos contribui com a redução da geração das trilhas de AD, contudo, fica claro que a mudança de resolução não reduz o tempo de geração da audiodescrição de forma proporcional. No primeiro experimento com o Vídeo C, a mudança da resolução de 640x360 para 320x180 garantiu uma redução no tempo de geração de 25%, enquanto que no segundo experimento que usou o Vídeo D como entrada, a diminuição do tempo de geração alterando a resolução de 528x336 para 264x168 foi de aproximadamente 19,5%.

5.2.3 Mensuração do tempo de geração para um conjunto de cem requisições

A terceira e última etapa dos testes de geração teve como objetivo medir o tempo que a solução proposta demorava para atender um conjunto de requisições de geração de trilhas de audiodescrição, com um número crescente de instâncias do componente de Geração Automática de AD.

O experimento realizado nesta etapa contou com o Vídeo C como entrada, além de um conjunto de cem requisições enviadas inicialmente para o *endpoint* de geração de AD integrada (com roteiro) e de um número de instâncias do componente de Geração Automática de AD definido em uma, duas e quatro instâncias (*workers*). Posteriormente, o mesmo experimento foi repetido com o *endpoint* de geração de AD integrada (sem roteiro). As Tabelas 4 e 5 em conjunto com as Figuras 15 e 16 apresentam um resumo dos resultados obtidos nesta etapa de testes. As tabelas contendo os dados completos podem ser consultadas no Apêndice B.

Tabela 4: Sumarização do tempo de processamento de 100 requisições com roteiro disponível.

Número de Requisições	Tempo para Geração de AD com 1 <i>worker</i>	Tempo para Geração de AD com 2 <i>workers</i>	Tempo para Geração de AD com 4 <i>workers</i>
1	0,7 min	0,8 min	0,8 min
2	1 min	0,8 min	0,8 min
5	3 min	2 min	1 min
10	6 min	3 min	2 min
50	33 min	17 min	9 min
100	67 min	34 min	18 min

De acordo com a Tabela 4, as primeiras requisições recebidas pelo servidor da solução proposta tiveram seus tempos de processamento bastante próximos a 1 minuto, notando-se uma pequena vantagem no tempo de geração de AD com apenas um *worker*. No entanto, os benefícios de se usar mais de um *worker* começam a se tornar perceptíveis a partir da quinta sequência de requisições, onde a espera para a geração de AD com um *worker* durou 3 minutos, enquanto que com dois e quatro *workers* durou 2 minutos e 1 minuto, respectivamente. Essa diferença pode ser melhor visualizada através dos gráficos da Figura 15, que mostram com precisão o momento em que o acréscimo de mais *workers* ao sistema começa a fazer a diferença no tempo de processamento de cem requisições.

Na Figura 15, o gráfico da esquerda apresenta o tempo gasto no processamento do conjunto de 100 requisições, enquanto que no gráfico da direita são apresentados o tempos de geração das trilhas de AD das 10 primeiras requisições desse conjunto.

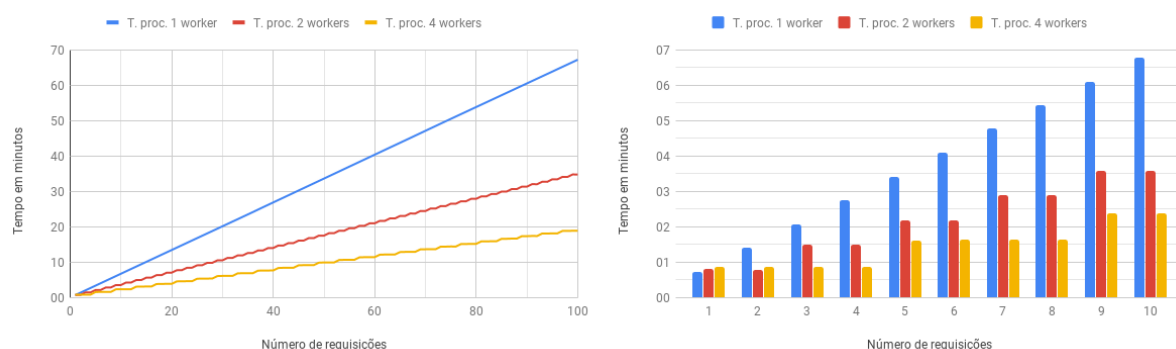


Figura 15: Tempo de processamento de 100 requisições com 1, 2 e 4 *workers* e com roteiro.

De maneira idêntica aos testes da geração automática de AD usando o roteiro, os testes com a geração automática de AD baseada apenas em vídeo mediram o tempo necessário para o processamento de um conjunto de 100 requisições com uma, duas e quatro instâncias do gerador automático de audiodescrição.

Tabela 5: Sumarização do tempo de processamento de 100 requisições com ausência do roteiro.

Número de Requisições	Tempo para Geração de AD com 1 <i>worker</i>	Tempo para Geração de AD com 2 <i>workers</i>	Tempo para Geração de AD com 4 <i>workers</i>
1	16 min	23 min	34 min
2	31 min	23 min	34 min
5	79 min	69 min	68 min
10	157 min	114 min	102 min
50	788 min	573 min	439 min
100	1574 min	1147 min	838 min

Como pode ser observado na Tabela 5, as primeiras requisições feitas à solução proposta foram processadas em 16 minutos com um *worker*, 23 minutos com dois *workers* e 34 minutos com quatro *workers*. Isso deixa clara a desvantagem inicial de usar mais de uma instância do gerador automático de AD quando a geração é baseada em informações apenas do vídeo, entretanto, da mesma forma que aconteceu com a geração de audiodescrição usando o roteiro, os benefícios de usar mais de um *worker* começam a ser percebidos a partir da quinta sequência de requisições, onde o tempo de processamento passa a ser 79 minutos com um *worker*, 69 minutos com dois *workers* e 68 minutos com quatro *workers*.

Essa vantagem torna-se ainda maior na centésima sequência de requisições que apresenta uma redução no tempo de processamento em 427 minutos de um para dois *workers* e 736 minutos de um para quatro *workers*. Essas observações também podem

ser feitas através dos gráficos da Figura 16, que fazem o comparativo, entre o número de *workers*, o tempo gasto para processar um conjunto de 100 requisições (gráfico da esquerda) e as 10 primeiras requisições desse conjunto (gráfico da direita).

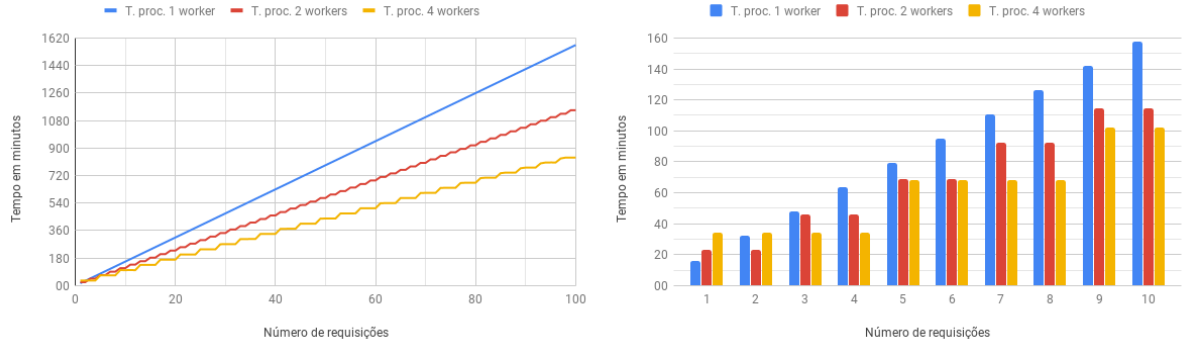


Figura 16: Tempo de processamento de 100 requisições com 1, 2 e 4 workers e sem roteiro.

Nesta etapa de testes, os resultados obtidos aumentando o número de *workers* disponíveis apresentaram uma redução considerável no tempo de processamento de um conjunto maior de requisições, em contrapartida, ao aumentar o número de instâncias do componente gerador de AD, o tempo necessário para gerar uma trilha de AD também aumentou (ver primeira linha das Tabelas 4 e 5), ocasionando um tempo de espera extra para cada requisição a ser atendida.

Outro ponto importante a ser discutido é se a redução do tempo necessário para atender uma certa quantidade de requisições é proporcional ao aumento da quantidade de *workers* de geração de AD, ou seja, se o tempo para processar n requisições diminui em duas vezes com 2 *workers* e em quatro vezes com 4 *workers*. Para 100 requisições feitas a geração de AD usando o roteiro, o fator de redução se aproximou de 1,97 de um para dois *workers* e 3,72 de um para quatro *workers*. Com relação às 100 requisições feitas a geração de AD baseada apenas em vídeo, o fator de redução obtido de um para dois *workers* foi de 1.37 e de um para quatro *workers* chegou a aproximadamente 1,87.

6 CONSIDERAÇÕES FINAIS

Neste trabalho, foi apresentado um Web Service Restful para a geração automática de trilhas de audiodescrição de vídeos digitais. Com essa solução é possível que novas aplicações sejam construídas, favorecendo o aumento da disponibilização de recursos de acessibilidade direcionados à pessoas cegas ou com baixa visão.

Para a validação da solução proposta, foi conduzido um experimento através de um conjunto de testes com o objetivo de avaliar a qualidade das trilhas de AD geradas e aferir o desempenho computacional do sistema.

A avaliação da qualidade realizada com usuários que assistiram os filmes com o recurso de AD gerado pelo sistema mostrou que a maioria deles foi capaz de compreender aspectos gerais das histórias dos filmes, porém, alguns usuários sentiram dificuldades em identificar os ambientes e horários em que as histórias aconteciam. Adicionalmente, os usuários foram convidados a responder numa escala de 1 a 6, o quão satisfeitos ficaram com a AD gerada pelo serviço, usando como parâmetros a facilidade de entendimento e a contribuição na compreensão do vídeo. Os resultados mostraram que 73% dos participantes atribuíram um conceito maior ou igual a 4 sobre a facilidade de entender a AD gerada pelo serviço, enquanto que aproximadamente 27% atribuíram um conceito inferior ou menor do que 3. Para a contribuição na compreensão do vídeo, 81,9% atribuíram um conceito igual ou superior a 4, enquanto que apenas 18,1% atribuíram um conceito menor ou inferior a 3.

Os resultados obtidos ao aferir o desempenho computacional do sistema mostraram que existe uma grande diferença entre as duas abordagens de geração de audiodescrição, sendo a geração de trilhas de AD a partir do vídeo e roteiro mais vantajosa que a geração de trilhas de AD que usa somente o vídeo. Também foi conferido que a resolução do vídeo tem pouca influência no tempo de geração de AD sem roteiro, visto que o vídeo D com resoluções menores obteve 19,5% de redução no tempo de geração da audiodescrição, enquanto o vídeo C, com resoluções maiores, obteve 25% de diminuição no tempo de geração da AD. Com relação a capacidade do serviço tornar-se escalável, os resultados mostraram que o aumento do número de instâncias do componente Gerador de AD reduziu consideravelmente o tempo de processamento de um conjunto relativamente grande de requisições, no entanto, essa redução não foi proporcional ao aumento do número de instâncias quando o processo de geração de AD é baseado apenas em vídeo, visto que o fator de redução de uma para duas instâncias foi de 1,37 e de uma para quatro instâncias chegou a aproximadamente 1,87.

Entre as propostas de trabalhos futuros, pretende-se investigar novas estratégias que visem melhorar a escalabilidade da geração de audiodescrição de vídeos digitais quando não há roteiro disponível. Além disso, também é pretendido desenvolver uma

versão distribuída do sistema de geração automática de AD. Outra proposta é o desenvolvimento de aplicações multiplataforma que permitam aos usuários terem acesso ao serviço apresentado neste trabalho.

REFERÊNCIAS

- [1] CAMPOS, V. P. **Um sistema de geração automática de roteiros de audio-descrição**. 2015. Dissertação (Mestrado em Informática) – Universidade Federal da Paraíba, João Pessoa, 2015.
- [2] CASADO, A. La audiodescripción: apuntes sobre el estado de la cuestión y las perspectivas de investigación. **Tradterm**, v. 13, p. 151-169, 18 dez. 2007. Article No. 17.
- [3] CHAMPIN, P. A. *et al.* Towards collaborative annotation for video accessibility. In: International Web for All Conference: Addressing Information Barriers, 16, Raleigh. **Anais eletrônicos...** ACM: ACM New York, NY, USA—, 2010.
- [4] CHAPDELAINE, C.; GAGNON, L. Accessible videodescription On-Demand. In: International ACM SIGACCESS Conference on Computers and Accessibility, 11, Pittsburgh. **Anais eletrônicos...** ACM: ACM New York, NY, USA—, 2009.
- [5] CHEN, M. *et al.* Reference Based LSTM for Image Captioning. In: AAAI Conference on Artificial Intelligence. 2017. p. 3981-3987.
- [6] DONAHUE, J. *et al.* Long-Term Recurrent Convolutional Networks for Visual Recognition and Description. **IEEE Transactions on Pattern Analysis and Machine Intelligence**. v. 39, n. 4, p. 677-691, 1 Abr. 2017.
- [7] DVX, Descriptive Video Exchange. Disponível em: <<http://www.mielelab.org/projects/dvx>>. Acesso em: 25 out. 2018.
- [8] ERL, T. **Service-Oriented Architecture: A Field Guide to Integrating XML and Web Services**. Upper Saddle River: Prentice Hall PTR, 2004.
- [9] FIELDING, R. T. **Architectural Styles and the Design of Network-based Software Architectures**. 2000. Tese (Doctoral dissertation) – University of California, Irvine, 2000.
- [10] FRANCO, E. P. C.; SILVA, M. C. C. C. **Audiodescrição: breve passeio histórico**. In: Livia Maria Villela de Mello Motta, Paulo Romeu Filho. Transformando Imagens em Palavras. São Paulo: Secretaria de Estado dos Direitos da Pessoa com Deficiência, 2010.
- [11] GTAAaaS 2.0, Grupo de Trabalho: Accessibility as a Service. Disponível em: <<http://gtad.lavid.ufpb.br/>>. Acesso em: 24 out. 2018.

- [12] IBGE, Instituto Brasileiro de Geografia e Estatística. Censo Demográfico 2010: características gerais da população, religião e pessoas com deficiência. Rio de Janeiro: IBGE, 2010. Disponível em: <https://ww2.ibge.gov.br/home/estatistica/populacao/censo2010/caracteristicas_religiao_deficiencia/default_caracteristicas_religiao_deficiencia.shtm>. Acesso em: 1 out. 2018.
- [13] JONES, M. T. Understand Representational State Transfer (REST) in Ruby. 2012. Disponível em: <<https://www.ibm.com/developerworks/library/os-understand-rest-ruby/index.html>>. Acesso em: 6 out. 2018.
- [14] KOBAYASHI, M. *et al.* Describing online videos with text-to-speech narration. In: International Web for All Conference: Addressing Information Barriers, 16, Raleigh. **Anais eletrônicos...** ACM: ACM New York, NY, USA—, 2010. Article No. 19.
- [15] LAKRITZ, J.; SALWAY, A. **The Semi Automatic Generation of Audio Description from Screenplays**. Guildford: School of Electronics and Physical Sciences, Department of Computing, University of Surrey, 2005. (Technical Report CS-06-05).
- [16] MOTTA, L. M. V. M. Audiodescrição - recurso de acessibilidade para inclusão cultural das pessoas com deficiência visual. 2008. Disponível em: <<http://www.bengalalegal.com/livia>>. Acesso em: 6 out. 2018.
- [17] NUNES, E.V.; MACHADO, F.O.; VANZIN T. **Audiodescrição como Tecnologia Assistiva para o Acesso ao Conhecimento Por Pessoas Cegas**. In: Vania Ribas Ulbricht, Tarcísio Vanzin e Vilma Villarouco. Ambiente virtual de aprendizagem inclusivo. Florianópolis: Pandion, 2011.
- [18] PAUTASSO, C.; ZIMMERMANN, O.; LEYMANN, F. RESTful Web Services vs. "Big" Web Services: making the right architectural decision. In: International Conference on World Wide Web, 17, Beijing. **Anais eletrônicos...** ACM: ACM New York, NY, USA—, 2008. CD-ROM.
- [19] RICHARDSON, L.; RUBY, S. **RESTful Web Services**. Sebastopol: O'Reilly, 2007.
- [20] ROHRBACH, A.; ROHRBACH M.; SCHIELE B. The Long-Short Story of Movie Description. In: German Conference on Pattern Recognition, 37, Aachen. **Anais eletrônicos...** Springer: Springer, Cham—, 2015. p. 209-221.
- [21] RRTD, Remote real-time description. Disponível em: <<http://www.vdrdc.org/research/rrtd>>. Acesso em: 25 out. 2018.

- [22] SHETTY, R.; LAAKSONEN J. Video captioning with recurrent networks based on frame- and video-level features and visual content classification. In: ICCV Large Scale Movie Description Challenge workshop. 2015.
- [23] SZARKOWSKA, Agnieszka. Text-to-speech audio description: towards wider availability of AD. **Journal of Specialised Translation**, v. 15, p. 142-163, jan. 2011.
- [24] VENUGOPALAN, S. *et al.* Translating Videos to Natural Language Using Deep Recurrent Neural Networks. In: Human Language Technologies: Annual Conference of the North American Chapter of the ACL, 2015, Denver. 2015. p. 1494-1504.
- [25] W3C, World Wide Web Consortium. Architecture of the World Wide Web, Volume One. Cambridge: W3C, 2004. Disponível em: <<https://www.w3.org/TR/webarch/>>. Acesso em: 1 out. 2018.
- [26] W3C, World Wide Web Consortium. Web Services Glossary. Cambridge: W3C, 2004. Disponível em: <<https://www.w3.org/TR/ws-gloss/>>. Acesso em: 4 out. 2018.
- [27] WHO, World Health Organization. Blindness and vision impairment. Geneva: WHO, 2018. Disponível em: <<http://www.who.int/en/news-room/fact-sheets/detail/blindness-and-visual-impairment#>>. Acesso em: 30 set. 2018.
- [28] WU, Z. *et al.* **Deep Learning for Video Classification and Captioning**. In: Association for Computing Machinery and Morgan & Claypool. Frontiers of Multimedia Research. 2017.

APÊNDICE A – DOCUMENTAÇÃO DA API DO SERVIÇO

apiDoc: adgen-api - 0.0.0

file:///home/wesnydy/Workspace/Lavid/adgen-ap...

adgen-api

Midia

Midia - Download da AD gerada

GET

/media/download/:requestTag

Parameter

Field	Type	Description
requestTag	String	ID da requisição

404

Name	Description
EtiquetaNaoReconhecida	Não existe requisição associada a etiqueta informada
MidiaNaoDisponivel	A mídia não está disponível para download

500

Name	Description
ErroInterno	Ocorreu um erro interno no servidor

- Etiqueta Nao Reconhecida [#error-examples-Midia-GetMediaDownloadRequesttag-0_0_0-0]
- Midia Nao Disponivel [#error-examples-Midia-GetMediaDownloadRequesttag-0_0_0-1]
- Erro Interno [#error-examples-Midia-GetMediaDownloadRequesttag-0_0_0-2]

HTTP/1.1 404 There is no request associated with this mediaTag

HTTP/1.1 404 File not available, check the processing status

HTTP/1.1 500 Internal Server Error

Midia - Status da geração da AD

GET

/media/status/:requestTag

Parameter

Field	Type	Description
requestTag	String	ID da requisição

Success 200

Field	Type	Description
-------	------	-------------

Field	Type	Description
status	String	Status da geração da AD
media	Object	Informações da mídia gerada
requestDate	Date	Data da requisição
finishDate	Date	Data do fim da geração da AD
processingTimeInSec	Number	Tempo gasto na geração da AD

- Success [#success-examples-Midia-GetMediaStatusRequesttag-0_0_0-0]

```
HTTP/1.1 200 OK
{
  "status": "done",
  "media": {
    {
      "name": "50102560-b51b-11e8-9a58-b5b0a18e05db.srt",
      "location": "/tmp/adgen-media/50102560-b51b-11e8-9a58-b5b0a18e05db"
    },
    "requestDate": "2018-09-10T17:02:34.683Z",
    "finishDate": "2018-09-10T17:02:43.569Z",
    "processingTimeInSec": 8.886
  }
}
```

404

Name	Description
EtiquetaNaoReconhecida	Não existe requisição associada a etiqueta informada

500

Name	Description
ErroInterno	Ocorreu um erro interno no servidor

- Etiqueta Nao Reconhecida [#error-examples-Midia-GetMediaStatusRequesttag-0_0_0-0]
- Erro Interno [#error-examples-Midia-GetMediaStatusRequesttag-0_0_0-1]

```
HTTP/1.1 404 There is no request associated with this mediaTag
```

```
HTTP/1.1 500 Internal Server Error
```

Roteiro

Roteiro - Geração de roteiro de AD

```
POST
/script/ad
```

Parameter

Field	Type	Description
script	File	Roteiro (.celtx)
subtitle	File	Legenda (.srt)
video	File	Vídeo (.mp4)

Success 200

Field	Type	Description
requestTag	String	ID da requisição

- Success [#success-examples-Roteiro-PostScriptAd-0_0_0-0]

```
HTTP/1.1 200 OK
{
  "requestTag": "609a22c0-bb1d-11e8-bf13-513879813097"
}
```

422

Name	Description
ArquivosInsuficientes	Arquivos isuficientes para a geração da AD
ArquivoVazio	Não é possível carregar arquivos vazios

500

Name	Description
ErroInterno	Ocorreu um erro interno no servidor

- Arquivos Insuficientes [#error-examples-Roteiro-PostScriptAd-0_0_0-0]
- Arquivo Vazio [#error-examples-Roteiro-PostScriptAd-0_0_0-1]
- Erro Interno [#error-examples-Roteiro-PostScriptAd-0_0_0-2]

```
HTTP/1.1 422 Not enough files for AD generation from script
```

```
HTTP/1.1 422 Cannot upload empty files
```

```
HTTP/1.1 500 Internal Server Error
```

Roteiro - Geração de trilha de AD integrada

POST

```
/script/mixed/ad
```

Parameter

Field	Type	Description
script	File	Roteiro (.celtx)

Field	Type	Description
subtitle	File	Legenda (.srt)
video	File	Vídeo (.mp4)

Success 200

Field	Type	Description
requestTag	String	ID da requisição

- Success [#success-examples-Roteiro-PostScriptMixedAd-0_0_0-0]

```
HTTP/1.1 200 OK
{
  "requestTag": "609a22c0-bb1d-11e8-bf13-513879813097"
}
```

422

Name	Description
ArquivosInsuficientes	Arquivos isuficientes para a mixagem da trilha de AD com o vídeo
ArquivoVazio	Não é possível carregar arquivos vazios

500

Name	Description
ErroInterno	Ocorreu um erro interno no servidor

- Arquivos Insuficientes [#error-examples-Roteiro-PostScriptMixedAd-0_0_0-0]
- Arquivo Vazio [#error-examples-Roteiro-PostScriptMixedAd-0_0_0-1]
- Erro Interno [#error-examples-Roteiro-PostScriptMixedAd-0_0_0-2]

```
HTTP/1.1 422 Not enough files for mix AD and Video
```

```
HTTP/1.1 422 Cannot upload empty files
```

```
HTTP/1.1 500 Internal Server Error
```

Roteiro - Geração de trilha de AD

POST

```
/script/track/ad
```

Parameter

Field	Type	Description
script	File	Roteiro (.celtx)
subtitle	File	Legenda (.srt)

Field	Type	Description
video	File	Vídeo (.mp4)

Success 200

Field	Type	Description
requestTag	String	ID da requisição

- Success [#success-examples-Roteiro-PostScriptTrackAd-0_0_0-0]

```
HTTP/1.1 200 OK
{
  "requestTag": "609a22c0-bb1d-11e8-bf13-513879813097"
}
```

422

Name	Description
ArquivosInsuficientes	Arquivos isuficientes para a geração da trilha de AD
ArquivoVazio	Não é possível carregar arquivos vazios

500

Name	Description
ErroInterno	Ocorreu um erro interno no servidor

- Arquivos Insuficientes [#error-examples-Roteiro-PostScriptTrackAd-0_0_0-0]
- Arquivo Vazio [#error-examples-Roteiro-PostScriptTrackAd-0_0_0-1]
- Erro Interno [#error-examples-Roteiro-PostScriptTrackAd-0_0_0-2]

```
HTTP/1.1 422 Not enough files for synthesise speech
```

```
HTTP/1.1 422 Cannot upload empty files
```

```
HTTP/1.1 500 Internal Server Error
```

RoteiroAD

RoteiroAD - Geração de trilha de AD integrada

POST

```
/subtitle/mixed/ad
```

Parameter

Field	Type	Description
subtitle	File	Roteiro de AD (.srt)

Field	Type	Description
video	File	Vídeo (.mp4)

Success 200

Field	Type	Description
requestTag	String	ID da requisição

- Success [#success-examples-RoteiroAD-PostSubtitleMixedAd-0_0_0-0]

```
HTTP/1.1 200 OK
{
  "requestTag": "609a22c0-bb1d-11e8-bf13-513879813097"
}
```

422

Name	Description
ArquivosInsuficientes	Arquivos isuficientes para a mixagem da trilha de AD com o vídeo
ArquivoVazio	Não é possível carregar arquivos vazios

500

Name	Description
ErroInterno	Ocorreu um erro interno no servidor

- Arquivos Insuficientes [#error-examples-RoteiroAD-PostSubtitleMixedAd-0_0_0-0]
- Arquivo Vazio [#error-examples-RoteiroAD-PostSubtitleMixedAd-0_0_0-1]
- Erro Interno [#error-examples-RoteiroAD-PostSubtitleMixedAd-0_0_0-2]

```
HTTP/1.1 422 Not enough files for mix AD and Video
```

```
HTTP/1.1 422 Cannot upload empty files
```

```
HTTP/1.1 500 Internal Server Error
```

RoteiroAD - Geração de trilha de AD

POST

```
/subtitle/track/ad
```

Parameter

Field	Type	Description
subtitle	File	Roteiro de AD (.srt)

Success 200

Field	Type	Description
requestTag	String	ID da requisição

- Success [#success-examples-RoteiroAD-PostSubtitleTrackAd-0_0_0]

```
HTTP/1.1 200 OK
{
  "requestTag": "609a22c0-bb1d-11e8-bf13-513879813097"
}
```

422

Name	Description
ArquivosInsuficientes	Arquivos isuficientes para a geração da trilha de AD
ArquivoVazio	Não é possível carregar arquivos vazios

500

Name	Description
ErroInterno	Ocorreu um erro interno no servidor

- Arquivos Insuficientes [#error-examples-RoteiroAD-PostSubtitleTrackAd-0_0_0]
- Arquivo Vazio [#error-examples-RoteiroAD-PostSubtitleTrackAd-0_0_0-1]
- Erro Interno [#error-examples-RoteiroAD-PostSubtitleTrackAd-0_0_0-2]

```
HTTP/1.1 422 Not enough files for synthesise speech
```

```
HTTP/1.1 422 Cannot upload empty files
```

```
HTTP/1.1 500 Internal Server Error
```

Video

Video - Geração de roteiro de AD

POST

```
/video/ad
```

Parameter

Field	Type	Description
subtitle	File	Legenda (.srt)
video	File	Vídeo (.mp4)

Success 200

Field	Type	Description
requestTag	String	ID da requisição

- Success [#success-examples-Video-PostVideoAd-0_0_0]

```
HTTP/1.1 200 OK
```

```
{
  "requestTag": "609a22c0-bb1d-11e8-bf13-513879813097"
}
```

422

Name	Description
ArquivosInsuficientes	Arquivos insuficientes para a geração da AD
ArquivoVazio	Não é possível carregar arquivos vazios

500

Name	Description
ErroInterno	Ocorreu um erro interno no servidor

- Arquivos Insuficientes [#error-examples-Video-PostVideoAd-0_0_0-0]
- Arquivo Vazio [#error-examples-Video-PostVideoAd-0_0_0-1]
- Erro Interno [#error-examples-Video-PostVideoAd-0_0_0-2]

HTTP/1.1 422 Not enough files for AD generation from script

HTTP/1.1 422 Cannot upload empty files

HTTP/1.1 500 Internal Server Error

Video - Geração de trilha de AD integrada

POST

/video/mixed/ad

Parameter

Field	Type	Description
subtitle	File	Legenda (.srt)
video	File	Vídeo (.mp4)

Success 200

Field	Type	Description
requestTag	String	ID da requisição

- Success [#success-examples-Video-PostVideoMixedAd-0_0_0-0]

```
HTTP/1.1 200 OK
{
  "requestTag": "609a22c0-bb1d-11e8-bf13-513879813097"
}
```

422

Name	Description
------	-------------

Name	Description
ArquivosInsuficientes	Arquivos isuficientes para a mixagem da trilha de AD com o vídeo
ArquivoVazio	Não é possível carregar arquivos vazios

500

Name	Description
ErroInterno	Ocorreu um erro interno no servidor

- Arquivos Insuficientes [#error-examples-Video-PostVideoMixedAd-0_0_0-0]
- Arquivo Vazio [#error-examples-Video-PostVideoMixedAd-0_0_0-1]
- Erro Interno [#error-examples-Video-PostVideoMixedAd-0_0_0-2]

HTTP/1.1 422 Not enough files for mix AD and Video

HTTP/1.1 422 Cannot upload empty files

HTTP/1.1 500 Internal Server Error

Video - Geração de trilha de AD

POST

/video/track/ad

Parameter

Field	Type	Description
subtitle	File	Legenda (.srt)
video	File	Vídeo (.mp4)

Success 200

Field	Type	Description
requestTag	String	ID da requisição

- Success [#success-examples-Video-PostVideoTrackAd-0_0_0-0]

HTTP/1.1 200 OK
{
 "requestTag": "609a22c0-bb1d-11e8-bf13-513879813097"
}

422

Name	Description
ArquivosInsuficientes	Arquivos isuficientes para a geração da trilha de AD
ArquivoVazio	Não é possível carregar arquivos vazios

500

Name	Description
ErroInterno	Ocorreu um erro interno no servidor

- Arquivos Insuficientes [#error-examples-Video-PostVideoTrackAd-0_0_0-0]
- Arquivo Vazio [#error-examples-Video-PostVideoTrackAd-0_0_0-1]
- Erro Interno [#error-examples-Video-PostVideoTrackAd-0_0_0-2]

HTTP/1.1 422 Not enough files for synthesise speech

HTTP/1.1 422 Cannot upload empty files

HTTP/1.1 500 Internal Server Error

APÊNDICE B – TABELAS COMPLETAS DOS RESULTADOS

Tabela 6: Tempo de processamento de 100 requisições com roteiro disponível.

Vídeo A 640x360 com duração 00:14:33					
1 Worker		2 Workers		4 Workers	
Nº Requisição	T. processamento (HH:MM:SS)	Nº Requisição	T. processamento (HH:MM:SS)	Nº Requisição	T. processamento (HH:MM:SS)
1	00:00:44	1	00:00:48	1	00:00:51
2	00:01:25	2	00:00:47	2	00:00:53
3	00:02:05	3	00:01:29	3	00:00:51
4	00:02:45	4	00:01:30	4	00:00:52
5	00:03:25	5	00:02:11	5	00:01:37
6	00:04:06	6	00:02:12	6	00:01:38
7	00:04:46	7	00:02:53	7	00:01:38
8	00:05:26	8	00:02:54	8	00:01:38
9	00:06:07	9	00:03:35	9	00:02:23
10	00:06:47	10	00:03:35	10	00:02:24
11	00:07:27	11	00:04:17	11	00:02:25
12	00:08:08	12	00:04:17	12	00:02:24
13	00:08:48	13	00:04:59	13	00:03:07
14	00:09:28	14	00:04:59	14	00:03:09
15	00:10:08	15	00:05:41	15	00:03:10
16	00:10:49	16	00:05:41	16	00:03:10
17	00:11:29	17	00:06:23	17	00:03:52
18	00:12:09	18	00:06:23	18	00:03:54
19	00:12:49	19	00:07:05	19	00:03:55
20	00:13:30	20	00:07:05	20	00:03:55
21	00:14:10	21	00:07:46	21	00:04:37
22	00:14:50	22	00:07:47	22	00:04:39
23	00:15:30	23	00:08:28	23	00:04:40
24	00:16:10	24	00:08:29	24	00:04:40
25	00:16:51	25	00:09:10	25	00:05:22
26	00:17:31	26	00:09:11	26	00:05:24
27	00:18:11	27	00:09:51	27	00:05:25
28	00:18:51	28	00:09:52	28	00:05:25
29	00:19:31	29	00:10:34	29	00:06:07
30	00:20:11	30	00:10:34	30	00:06:10
31	00:20:52	31	00:11:15	31	00:06:10
32	00:21:32	32	00:11:16	32	00:06:11
33	00:22:12	33	00:11:57	33	00:06:53
34	00:22:52	34	00:11:58	34	00:06:56
35	00:23:32	35	00:12:39	35	00:06:56
36	00:24:13	36	00:12:40	36	00:06:57
37	00:24:54	37	00:13:21	37	00:07:39
38	00:25:35	38	00:13:22	38	00:07:42
39	00:26:15	39	00:14:03	39	00:07:42
40	00:26:56	40	00:14:03	40	00:07:42
41	00:27:36	41	00:14:45	41	00:08:24
42	00:28:17	42	00:14:45	42	00:08:28
43	00:28:57	43	00:15:27	43	00:08:28
44	00:29:38	44	00:15:27	44	00:08:28
45	00:30:18	45	00:16:09	45	00:09:10
46	00:30:59	46	00:16:09	46	00:09:13
47	00:31:39	47	00:16:51	47	00:09:13
48	00:32:20	48	00:16:51	48	00:09:13
49	00:33:00	49	00:17:33	49	00:09:55
50	00:33:40	50	00:17:33	50	00:09:58
51	00:34:21	51	00:18:15	51	00:09:58
52	00:35:02	52	00:18:15	52	00:09:58

Nº Requisição	T. processamento (HH:MM:SS)	Nº Requisição	T. processamento (HH:MM:SS)	Nº Requisição	T. processamento (HH:MM:SS)
53	00:35:42	53	00:18:57	53	00:10:40
54	00:36:23	54	00:18:56	54	00:10:43
55	00:37:04	55	00:19:39	55	00:10:43
56	00:37:44	56	00:19:38	56	00:10:43
57	00:38:24	57	00:20:20	57	00:11:25
58	00:39:05	58	00:20:20	58	00:11:27
59	00:39:45	59	00:21:02	59	00:11:28
60	00:40:26	60	00:21:02	60	00:11:29
61	00:41:06	61	00:21:43	61	00:12:10
62	00:41:47	62	00:21:43	62	00:12:12
63	00:42:27	63	00:22:25	63	00:12:14
64	00:43:08	64	00:22:25	64	00:12:14
65	00:43:48	65	00:23:07	65	00:12:54
66	00:44:28	66	00:23:06	66	00:12:57
67	00:45:09	67	00:23:48	67	00:12:59
68	00:45:50	68	00:23:48	68	00:12:59
69	00:46:30	69	00:24:30	69	00:13:40
70	00:47:11	70	00:24:30	70	00:13:42
71	00:47:52	71	00:25:12	71	00:13:43
72	00:48:32	72	00:25:12	72	00:13:43
73	00:49:13	73	00:25:53	73	00:14:25
74	00:49:53	74	00:25:53	74	00:14:26
75	00:50:33	75	00:26:35	75	00:14:29
76	00:51:13	76	00:26:34	76	00:14:28
77	00:51:53	77	00:27:16	77	00:15:11
78	00:52:33	78	00:27:16	78	00:15:11
79	00:53:14	79	00:27:57	79	00:15:14
80	00:53:54	80	00:27:57	80	00:15:14
81	00:54:34	81	00:28:38	81	00:15:55
82	00:55:14	82	00:28:38	82	00:15:55
83	00:55:54	83	00:29:20	83	00:15:58
84	00:56:34	84	00:29:20	84	00:15:58
85	00:57:14	85	00:30:01	85	00:16:40
86	00:57:54	86	00:30:01	86	00:16:39
87	00:58:34	87	00:30:43	87	00:16:43
88	00:59:14	88	00:30:42	88	00:16:42
89	00:59:55	89	00:31:24	89	00:17:24
90	01:00:35	90	00:31:24	90	00:17:24
91	01:01:15	91	00:32:06	91	00:17:27
92	01:01:55	92	00:32:05	92	00:17:27
93	01:02:35	93	00:32:46	93	00:18:09
94	01:03:15	94	00:32:46	94	00:18:08
95	01:03:55	95	00:33:28	95	00:18:11
96	01:04:36	96	00:33:28	96	00:18:11
97	01:05:16	97	00:34:10	97	00:18:53
98	01:05:56	98	00:34:09	98	00:18:53
99	01:06:36	99	00:34:50	99	00:18:56
100	01:07:16	100	00:34:50	100	00:18:55

Tabela 7: Tempo de processamento de 100 requisições com ausência do roteiro.

Vídeo A 640x360 com duração 00:14:33					
1 Worker		2 Workers		4 Workers	
Nº Requisição	T. processamento (HH:MM:SS)	Nº Requisição	T. processamento (HH:MM:SS)	Nº Requisição	T. processamento (HH:MM:SS)
1	00:16:14	1	00:23:23	1	00:34:04
2	00:31:58	2	00:23:21	2	00:34:04
3	00:47:42	3	00:46:15	3	00:34:12
4	01:03:25	4	00:46:14	4	00:34:04
5	01:19:09	5	01:09:08	5	01:08:02
6	01:34:56	6	01:09:08	6	01:08:04
7	01:50:39	7	01:32:02	7	01:08:06
8	02:06:25	8	01:32:03	8	01:08:07
9	02:22:10	9	01:54:49	9	01:42:10
10	02:37:55	10	01:54:50	10	01:42:07
11	02:53:41	11	02:17:46	11	01:42:11
12	03:09:26	12	02:17:46	12	01:42:07
13	03:25:12	13	02:40:39	13	02:16:18
14	03:40:57	14	02:40:40	14	02:16:16
15	03:56:43	15	03:03:38	15	02:16:18
16	04:12:27	16	03:03:38	16	02:16:19
17	04:28:13	17	03:26:35	17	02:50:08
18	04:44:00	18	03:26:35	18	02:50:15
19	04:59:46	19	03:49:33	19	02:50:07
20	05:15:30	20	03:49:33	20	02:50:10
21	05:31:14	21	04:12:32	21	03:24:14
22	05:46:59	22	04:12:32	22	03:24:07
23	06:02:43	23	04:35:27	23	03:24:13
24	06:18:28	24	04:35:26	24	03:24:08
25	06:34:13	25	04:58:28	25	03:57:57
26	06:49:57	26	04:58:28	26	03:57:58
27	07:05:42	27	05:21:29	27	03:58:02
28	07:21:28	28	05:21:28	28	03:57:47
29	07:37:13	29	05:44:29	29	04:31:29
30	07:52:59	30	05:44:29	30	04:31:48
31	08:08:44	31	06:07:26	31	04:31:51
32	08:24:30	32	06:07:26	32	04:31:50
33	08:40:16	33	06:30:18	33	05:05:02
34	08:55:58	34	06:30:18	34	05:05:22
35	09:11:44	35	06:53:16	35	05:05:30
36	09:27:29	36	06:53:16	36	05:05:27
37	09:43:10	37	07:16:15	37	05:38:38
38	09:58:53	38	07:16:15	38	05:38:51
39	10:14:36	39	07:39:11	39	05:39:01
40	10:30:21	40	07:39:11	40	05:39:01
41	10:46:02	41	08:02:07	41	06:12:04
42	11:01:46	42	08:02:07	42	06:12:21
43	11:17:29	43	08:24:53	43	06:12:29
44	11:33:29	44	08:24:52	44	06:12:33
45	11:49:13	45	08:47:45	45	06:45:37
46	12:04:56	46	08:47:45	46	06:45:50
47	12:20:41	47	09:10:39	47	06:46:01
48	12:36:26	48	09:10:40	48	06:45:56
49	12:52:24	49	09:33:36	49	07:19:13
50	13:08:08	50	09:33:36	50	07:19:23
51	13:23:51	51	09:56:35	51	07:19:37
52	13:39:34	52	09:56:34	52	07:19:26
53	13:55:19	53	10:19:28	53	07:52:50
54	14:11:04	54	10:19:27	54	07:53:01
55	14:26:52	55	10:42:27	55	07:53:11
56	14:42:35	56	10:42:27	56	07:53:13
57	14:58:16	57	11:05:23	57	08:26:28
58	15:14:01	58	11:05:23	58	08:26:40
59	15:29:47	59	11:28:41	59	08:26:47
60	15:45:32	60	11:28:41	60	08:26:44
61	16:01:15	61	11:51:39	61	08:59:53
62	16:17:00	62	11:51:38	62	09:00:19
63	16:32:44	63	12:14:37	63	09:00:18
64	16:48:29	64	12:14:37	64	09:00:19
65	17:04:13	65	12:37:40	65	09:33:22
66	17:19:56	66	12:37:40	66	09:33:49
67	17:35:45	67	13:00:32	67	09:33:54
68	17:51:27	68	13:00:32	68	09:33:49
69	18:07:10	69	13:23:34	69	10:06:36
70	18:22:53	70	13:23:33	70	10:07:19
71	18:38:37	71	13:46:32	71	10:07:16
72	18:54:21	72	13:46:31	72	10:07:22
73	19:10:02	73	14:09:32	73	10:39:33
74	19:25:47	74	14:09:32	74	10:40:36

Nº Requisição	T. processamento (HH:MM:SS)	Nº Requisição	T. processamento (HH:MM:SS)	Nº Requisição	T. processamento (HH:MM:SS)
75	19:41:31	75	14:32:29	75	10:40:36
76	19:57:11	76	14:32:30	76	10:40:39
77	20:12:53	77	14:55:27	77	11:12:01
78	20:28:37	78	14:55:27	78	11:13:37
79	20:44:21	79	15:18:14	79	11:13:48
80	21:00:02	80	15:18:14	80	11:13:48
81	21:15:44	81	15:41:11	81	11:44:07
82	21:31:28	82	15:41:11	82	11:46:44
83	21:47:12	83	16:04:05	83	11:46:54
84	22:02:56	84	16:04:05	84	11:46:57
85	22:18:41	85	16:27:06	85	12:16:13
86	22:34:25	86	16:27:05	86	12:19:53
87	22:50:09	87	16:50:06	87	12:20:06
88	23:05:51	88	16:50:06	88	12:20:03
89	23:21:35	89	17:13:05	89	12:48:18
90	23:37:16	90	17:13:05	90	12:52:50
91	23:52:56	91	17:36:04	91	12:53:09
92	24:08:41	92	17:36:03	92	12:53:05
93	24:24:24	93	17:59:05	93	13:20:20
94	24:40:07	94	17:59:05	94	13:26:07
95	24:55:49	95	18:22:08	95	13:26:16
96	25:11:31	96	18:22:08	96	13:26:14
97	25:27:15	97	18:45:01	97	13:52:23
98	25:42:59	98	18:45:02	98	13:58:18
99	25:58:44	99	19:07:50	99	13:58:21
100	26:14:29	100	19:07:52	100	13:58:17