

## Software de transcrição de videoaulas em PDFs compactos

Autora: Ligia Keiko Carvalho | Orientadora: Patrícia Gagliardo de Campos | Coorientador: Rafael Eiki Matheus Imamura

### Introdução

**4,3 milhões de brasileiros** que têm conexão limitada à internet ou aparelhos eletrônicos com baixa capacidade de processamento não tiveram acesso adequado ao ensino remoto (IBGE, 2019). No Brasil a **educação remota é limitada**, e com a chegada da pandemia da Covid-19 essas limitações ficaram ainda mais evidentes.

**47,5%** das secretarias municipais e estaduais brasileiras disponibilizaram **videoaulas**.

Fonte: CIEB, 2020

São gastos **517MB/10min** em videochamadas. Consumindo **TUDO PACOTE** de um plano **3G** comum.

Fonte: SILVA e JUNIOR, 2013

### Objetivo

Desenvolver uma sistema que consiga transcrever e processar as imagens do conteúdo de uma videoaula para um PDF compacto, para alunos com conexão à internet limitada.

### Metodologia

1

#### Levantamento de requisitos

Processo de Engenharia de Software, avaliação de características técnicas e pedagógicas

#### Requisitos técnicos

- Sumarizar o conteúdo de forma concisa
- Gerar um PDF que sirva de material de estudo
- Capturar os keyframes que possuem conteúdo central da videoaula gravada
- Transcrever o conteúdo de forma precisa

3

#### Validação e testes

#### 1ª Etapa da validação

- Avaliação dos parâmetros técnicos
- Metrificação qualitativa

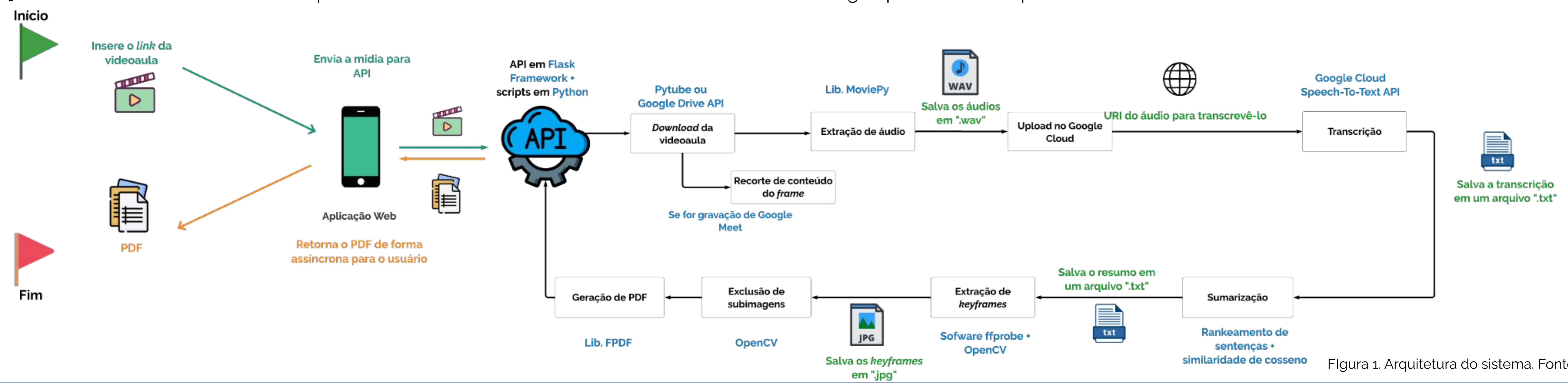
#### 2ª Etapa da validação

- Validação do PDF gerado c/ 5 docentes
- Metrificação com Framework ROUGE, Word Error Rate, Word Information Lost e Distância de Levenshtein

2

#### Desenvolvimento da aplicação

O sistema conta um cliente *Web* que envia o link da videoaula a ser processada. Através de uma API *RESTful* é produzido o PDF com o conteúdo da videoaula enviada e logo após retornado para o usuário.



### Resultados

Parâmetros	Algor. testados:	Resultados obtidos:
<b>Agorit. Transcrição:</b>  Precisão da transcrição quando comparada ao áudio.	<ul style="list-style-type: none"> <li>✓ Reconhecimento da fala e transcrição c/ <i>Google Speech API</i> + lib. <i>SpeechRecognition</i> (GS API)</li> <li>✓ <i>Google Cloud Speech-To-Text API</i> (GCSTT API)</li> </ul>	<ol style="list-style-type: none"> <li><b>Google Cloud Speech-To-Text API:</b> <ol style="list-style-type: none"> <li>48,56%/51,10% de <i>Word Error Rate (WER)</i></li> <li>60,85% <i>Word Information Lost (WIL)</i></li> <li>Distância de Levenshtein: 302</li> </ol> </li> </ol>
<b>Agorit. Sumarização:</b>  Correspondência do resumo gerado com o tema central do texto analisado.	<ul style="list-style-type: none"> <li>✗ Ranqueamento de sentenças:</li> <li>✓ Ranqueamento de sentenças + método de similaridade de cosseno de matrizes.</li> </ul>	<ol style="list-style-type: none"> <li><b>Tamanho do texto adequado</b></li> <li><b>Baixa precisão e recall</b></li> <li><b>Concordância ruim</b> do texto</li> <li><b>Priorização incorreta</b> do conteúdo</li> </ol>
<b>Agorit. Extração de keyframes:</b>  Tempo de processamento e quantidade de quadros repetidos.	<ul style="list-style-type: none"> <li>✗ 2 algorit. de diferença de histograma entre 2 frames consecutivos;</li> <li>✗ Diferenças de cor e posição de elementos presentes no quadro;</li> <li>✓ Utilização de dados provenientes de softwares analisadores de fluxo multimídia simples (<b>ffprobe</b>).</li> </ul>	<ol style="list-style-type: none"> <li>Aula de slides (<b>sem efeitos funciona bem</b>)</li> <li>Mais movimentação → mais frames → <b>PDF menos compacto</b></li> </ol>

#### Desafios

- GS API (TESTADO):**
- Limitações de 10MB por requisição e máximo de 60 min./mês do *Google Speech API*;
- GCSTT API (IMPLEMENTADO):**
- Não suporta formatos de áudios compactos;
  - Não transcreve coerentemente duas línguas diferentes.
1. Dificuldades para processar textos com muitos caracteres especiais.
2. Exclui partes importantes da videoaula;
- Captura muitas subimagens;
  - Dependendo da videoaula não é eficiente.
  - Reduzir excedente de keyframes extraídos sem perder conteúdo;

#### Soluções

- Desenvolveu-se algoritmos que cortam o vídeo em partes de 3 min. e extraem seu áudio para mídia ser transcrita;
1. Modificou-se o algoritmo de forma que uma quantidade maior de caracteres especiais fossem reconhecidos
- Possíveis soluções:**
- Clusterização de sentenças;
  - Análise semântica latente;
  - Uso de grafo-semântico;
  - Uso de recursos linguísticos;
  - Machine learning*.
1. Remoção de frames c/ faces por:
- Cálculo do tamanho da face;**
  - Recortar face do frame deixando só o conteúdo.**
- ✗ Perda de um frame c/ conteúdo

O algoritmo de transcrição retornou taxas de *WER*, *WIL* e *word distance* relativamente altas. Presença de mais de um idioma e fórmulas decaem a precisão.

**Textos em português e sem fórmulas retornam alta precisão**

O algoritmo de sumarização não retornou bons resultados quando foi avaliado pelo *framework* ROUGE, deixando o resumo desconexo.

**O algorit. de sumarização não está coletando todas as partes chaves do texto.**

A extração de *keyframes* auxiliada aos algoritmos de exclusão de subimagens e frames com faces se demonstra eficiente, no entanto ainda necessita de melhorias para reduzir a quantidade de subimagens.

**Ótimos resultados se auxiliado aos outros dois algoritmos**

### Conclusões

Ao fim da pesquisa, foi possível desenvolver um sistema capaz de gerar PDFs com economias de acima de 90%. A meta era de 95% em economia, mas quanto maior a videoaula, maior o PDF para não haver perda de conteúdo textual.

No quesito de avaliação, a extração de *keyframes* foi bem avaliado pois conseguiu capturar todo o conteúdo e reduzir a quantidade de *subimagens*, contribuindo para deixar o PDF compacto. A parte textual apresentou bons resultados em cenários comuns, mas em situações onde há mistura de idiomas e presença de fórmulas a transcrição demonstrou-se pouco precisa. E o algoritmo de sumarização não conseguiu extrair todo o conteúdo principal do texto.

Através do desenvolvimento da pesquisa, foi possível cumprir os objetivos relacionados à economia de dados, redução de custo de impressão e extração de *keyframes*. No entanto, o sistema ainda requer melhorias para ter melhor desempenho técnico e melhor avaliação pedagógica.

### Referências

DUBEY, Praveen. Understand Text Summarization and create your own summarizer in python. Disponível em: <https://towardsdatascience.com/understand-text-summarization-and-create-your-own-summarizer-in-python-b26a9f09c70x>. Publicado em 23 de Dez. de 2018.

CIEB. Planejamento das Secretarias de Educação do Brasil para Ensino Remoto, 2020. Disponível em: <http://cieb.net.br/pesquisa-analise-estrategias-de-ensino-remoto-de-secretarias-de-educacao-durante-a-crise-da-covid-19/>. Acesso em 5 de out. de 2021>. Acesso em 22 de jun. de 2021

PNAD Contínua TIC 2019. internet chega a 82,7% dos domicílios do país. Agência de Notícias IBGE. Publicado em 14 Abr. de 2021. Disponível em: <https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/30521-pnad-continua-tic-2019-inter-net- chega-a-82-7-dos-domicilios-do-pais>. Acesso em 02 Set. de 2021.

Ícones disponibilizados de forma gratuita em: www.vecteezy.com, www.freevector.com, flyclipart.com e www.pngwing.com

SILVA, Ricardo; JUNIOR, ELI. ORIGEM E UTILIZAÇÃO DA TECNOLOGIA 3G NO BRASIL. ETIC-ENCONTRO DE INICIAÇÃO CIENTÍFICA-ISSN 21-78-8498, v. 9, n. 9, 2013.

WIDYASSARI, A. P.; RUSTAD, S.; SHIDIK, G. F.; NOERSASONGKO, E.; SYUKUR, A.; AFFANDY, A. Review of automatic text summarization techniques & methods. Journal of King Saud University-Computer and Information Sciences, 2020.