

SISTEMA WEB SUMARIZADOR AUTOMÁTICO COM INTELIGÊNCIA ARTIFICIAL PARA TRANSCRIÇÃO DE MULTIVÍDEOS E ENVIO ELETRÔNICO, INTEGRADO AO INTERPRETADOR DE LINGUAGEM DE SINAIS AVATAR DIGITAL DO TIPO VLBRAS

Paulo Gabriel Cayres ⁽¹⁾ (paulocayres@usp.br)

⁽¹⁾ Universidade de São Paulo (USP); Instituto de Ciências Matemáticas e de Computação

RESUMO: *O sistema apresentado neste artigo, foi desenvolvido com objetivo de gerar sumários e transcrições coerentes, por meio da identificação de eventos relevantes, de vídeos multimodal a partir de múltiplas fontes, mitigar o custo computacional e sobrecarga de informações, promover o acesso inclusivo à informação de forma assistida. Para o desenvolvimento deste sistema foi utilizado linguagem de programação Python e HTML, Interfaces de Programação de Aplicações - IDE Pycharm e bibliotecas como Moviepy, Flask e SpeechRecognition. Dos testes práticos realizados neste sistema web desenvolvido, foi possível observar a geração automática de sumário fidedignos, o respectivo vídeo extraído, por exemplo, de um Localizador Uniforme de Recursos – URL, bem com um avatar digital intérprete de linguagem de sinais implementado e operativo em uma mesma página web de resultados, possibilitando que pessoas com surdez possam consumir o conteúdo de um vídeo ou de uma coleção de vídeos. Este sistema também possibilitou a tradução textual, armazenamento local e envios eletrônicos automáticos de arquivos digitais oriundos da sumarização, sendo enviados um arquivo em formato mp4 e outro em formato txt, para múltiplos endereços de e-mail, contribuindo assim para produção e disseminação em massa de conteúdos úteis ao conhecimento e a promoção da inclusão social e democratização do conhecimento.*

PALAVRAS-CHAVE: *Sistema web, sumarização automática de multivídeos, inteligência artificial, avatar intérprete de linguagem de sinais, acesso inclusivo à informação.*

AUTOMATIC WEB SUMMARIZER SYSTEM WITH ARTIFICIAL INTELLIGENCE FOR MULTIMODAL MULTIVIDEO TRANSCRIPTION INTEGRATED WITH VLBRAS TYPE SIGN LANGUAGE INTERPRETER AVATAR

ABSTRACT: *The system presented in this article was developed with the aim of generating coherent summaries and transcripts by identifying relevant events in multimodal videos from multiple sources, mitigating computational costs and information overload, and promoting inclusive access to information in an assisted way. The system was developed using the Python and HTML programming languages, the Pycharm IDE and libraries such as Moviepy, Flask and SpeechRecognition. From the practical tests carried out on this web system, it was possible to observe the automatic generation of a reliable summary, the respective video extracted, for example, from a Uniform Resource Locator - URL, as well as a digital sign language interpreter avatar implemented and operating on the same results web page, enabling people with deafblindness to consume the content of a video or a collection of videos. This system also enabled textual translation, local storage and automatic electronic sending of digital files resulting from the summarization, with one file being sent in mp4 format and another in txt format to multiple email addresses, thus contributing to the mass production and dissemination of useful knowledge content and promoting social inclusion and the democratization of knowledge.*

KEYWORDS: *Web system, automatic multivideo summarization, artificial intelligence, sign language, interpreter avatar, inclusive information access.*

1. INTRODUÇÃO

No mundo contemporâneo inundado de informações, onde vídeos estão em toda parte digital, desde tutoriais online até notícias transmitidas ao vivo, a capacidade de processar e extrair insights valiosos deste conteúdo é fundamental. No entanto, com o volume crescente de vídeos disponíveis, encontrar o material relevante e consumi-lo de forma eficiente tornou-se um desafio.

Nesse contexto, a sumarização multivídeo com inteligência artificial emerge como uma solução promissora. Esta técnica combina os princípios da sumarização de texto com a análise de conteúdo visual para criar transcrições precisas e abrangentes de múltiplos vídeos. A sumarização multivídeo é o processo de condensar informações de vários vídeos em uma forma mais compacta e compreensível, sem perder os pontos-chaves ou a essência do conteúdo gerando transcrições.

Para Mayer e Moreno (2003), a apresentação de resumos visuais e verbais pode auxiliar na compreensão e na memorização de conteúdo e a aplicação desse conceito à sumarização de vídeos pode, portanto, melhorar a experiência de aprendizado dos usuários.

A aplicação de técnicas de processamento de linguagem natural (PLN) e aprendizado de máquina (ML) neste contexto contribui para a criação de sistemas mais eficientes (LI *et al.* 2020).

Ao contrário da sumarização de texto tradicional, que se baseia principalmente em análises linguísticas, a sumarização multivídeo com Inteligência Artificial - IA incorpora elementos visuais, como reconhecimento de objetos, reconhecimento facial e detecção de movimento, para obter uma compreensão mais abrangente do conteúdo.

A. Benefícios da sumarização de multivídeos

Pode-se citar que, de forma geral, a sumarização automática de multivídeo com IA oferece uma série de benefícios significativos, por exemplo, os abaixo relacionados:

- **Economia de tempo:** ao condensar informações de vários vídeos em sumários concisos, os usuários economizam tempo precioso que seria gasto assistindo a cada vídeo na íntegra. A sumarização de vídeos pode reduzir em até 70% o tempo necessário para encontrar informações específicas dentro de um conteúdo audiovisual (XU *et al.* 2020).
- **Acessibilidade Rápida a Informações Relevantes:** os sumários permitem que os usuários identifiquem rapidamente os pontos-chaves e as informações mais relevantes de cada vídeo, facilitando a navegação e a busca de conteúdo específico. A sumarização de vídeos também contribui para a inclusão e a acessibilidade de conteúdos audiovisuais. A disponibilização de resumos em formatos textuais ou audiovisuais permite que pessoas com diferentes habilidades de aprendizado ou deficiências auditivas se beneficiem do

conteúdo. A sumarização de vídeos melhorou o acesso a informações educacionais para estudantes com dificuldades auditivas (WANG e JIANG, 2021).

- Compreensão abrangente: ao combinar análises de texto e conteúdo visual, a sumarização automática de multivídeo oferece uma compreensão mais abrangente do conteúdo, capturando tanto os aspectos linguísticos quanto visuais. A sumarização de vídeos também desempenha um papel importante no processo de aprendizado. A apresentação de conteúdos resumidos facilita a compreensão e a retenção das informações. Estudantes que utilizaram vídeos resumidos apresentaram um desempenho superior em testes de compreensão em comparação àqueles que visualizaram vídeos de forma integral (LIU *et al.* 2019).
- Facilitar a tomada de decisão: auxiliar na tomada de decisões ao fornecer insights valiosos e informações relevantes de vídeos.

B. Campo de aplicação da sumarização de multivídeos

As aplicações práticas da sumarização automática de multivídeos são variadas e abrangem diferentes setores e necessidades específicas. Abaixo é apresentado algumas das principais aplicações:

- Educação e treinamento: criação de materiais educacionais: professores e instrutores podem usar a sumarização automática para criar versões textuais e acessíveis de vídeos educacionais curtos ou longos, facilitando a revisão e o aprendizado dos alunos. Na educação, a sumarização de vídeos permite que os alunos absorvam materiais de forma mais eficiente. Vídeos educacionais longos podem ser resumidos para destacar os conceitos mais importantes, melhorando a retenção e a compreensão. A utilização de vídeos resumidos em ambientes de aprendizado online resultou em um aumento significativo no desempenho dos alunos (TIAN *et al.* 2020).
- Treinamento corporativo: empresas podem utilizar a transcrição automática para criar legendas ou scripts de treinamento, tornando o conteúdo mais acessível e permitindo que os funcionários revisem informações importantes de maneira eficiente. Também permite a elaboração de atas de reuniões e arquivamentos dos respectivos conteúdos e abordagens.
- Mídia e jornalismo: resumos e transcrição de notícias em que jornalistas podem usar a sumarização automática para rapidamente gerar resumos de vídeos de notícias e eventos, facilitando a produção de conteúdo para publicações online. Pode-se também trabalhar na análise de mídia social, pois empresas de mídia podem usar a transcrição automática

para analisar vídeos compartilhados nas redes sociais, identificando tendências e sentimentos do público. Ferramentas que resumem vídeos para destacar momentos chave são amplamente utilizadas por influenciadores e marcas para engajar o público. A estratégia de resumo atrai mais visualizações e interações em plataformas sociais (HE *et al.* 2019).

- Entretenimento e cultura: legendagem automática em plataformas de streaming pode usar a transcrição automática para gerar legendas automáticas em diferentes idiomas, melhorando a acessibilidade do conteúdo para um público global. No setor de entretenimento, a sumarização de vídeos é utilizada por serviços de streaming para criar trailers e teasers que sintetizam a essência de filmes e séries. Isso permite que os espectadores decidam rapidamente o que assistir. Geração de trailers a partir de filmes ou séries (DENG *et al.* 2020). Estudos mostram que a sumarização de vídeos aumenta a taxa de cliques em recomendações de conteúdo (WANG *et al.* 2021).
- Segurança e monitoramento: no monitoramento de segurança, organizações de segurança podem usar a transcrição automática para analisar vídeos de vigilância, identificando incidentes e comportamentos suspeitos de maneira mais eficiente. No contexto da vigilância, a sumarização de vídeos é empregada para filtrar grandes volumes de gravações e destacar eventos de interesse, como atividades suspeitas. A sumarização pode reduzir o tempo de revisão de gravações de câmeras de segurança em até 80%, aumentando a eficiência das operações de segurança (ALHUSSEINI *et al.* 2020). Em análise forense, por exemplo, a sumarização automática pode ser usada em investigações forenses para identificar evidências chave em vídeos relacionados a crimes ou incidentes ou até mesmo auxiliando os trabalhos dos escrivães.

Essas aplicações demonstram como a sumarização e a transcrição automáticas de multivídeos podem melhorar a eficiência, acessibilidade e utilidade de conteúdos audiovisuais em uma variedade de contextos práticos e profissionais. À medida que a tecnologia avança, é esperado que novos métodos e aplicações continuem a emergir, transformando fundamentalmente a maneira como interagimos e consumimos conteúdo audiovisual na era digital.

C. Paradigmas da sumarização de multivídeos

A seguir cita-se alguns dos principais desafios enfrentados pela sumarização de multivídeo e estratégias recomendadas e em desenvolvimento para superá-los, podendo ser:

- Complexidade e diversidade do conteúdo visual: um dos principais desafios da sumarização de multivídeo é lidar com a complexidade e a diversidade do conteúdo visual. Vídeos podem conter uma ampla gama de elementos, como cenas, objetos, pessoas e texto em movimento, o que torna a identificação de informações relevantes uma tarefa desafiadora. Além disso, a diversidade de estilos e formatos de vídeos, que vão desde palestras acadêmicas até vlogs de viagem, requer abordagens adaptativas para a sumarização eficaz.
- Integração de dados multimodais: a integração de dados multimodais, como áudio, texto e vídeo, apresenta outro desafio significativo para a sumarização de multivídeo. Embora abordagens baseadas apenas em análise visual possam ser eficazes em algumas situações, a combinação de diferentes modalidades de dados pode levar a uma compreensão mais completa do conteúdo. No entanto, integrar esses diferentes tipos de dados de forma coerente e eficaz é uma tarefa complexa que requer técnicas avançadas de processamento de multimídia. Ambiguidade na linguagem falada e na interpretação de imagens pode resultar em resumos imprecisos. A compreensão do contexto é fundamental para a eficácia da sumarização (YU *et al.* 2019). A integração de dados de diferentes modalidades (áudio, vídeo e texto) é complexa, exigindo algoritmos que sejam capazes de capturar informações relevantes de todas as fontes (WANG *et al.* 2020).
- Variabilidade de vídeos: embora a sumarização de vídeos tenha avançado significativamente, ainda existem desafios a serem superados. A variabilidade dos formatos de vídeo, a diversidade de conteúdos e as diferenças culturais podem dificultar a criação de sistemas de sumarização que funcionem de maneira eficaz em todos os contextos. A diversidade de tipos de vídeos (educacionais, de entretenimento, documentários, etc.) torna difícil o desenvolvimento de uma abordagem única que atenda a todos os formatos (ZHOU *et al.* 2021). Além disso, a necessidade de compreender o contexto e a intenção por trás do conteúdo é um desafio significativo para as tecnologias atuais (HASSANNIA *et al.* 2021).

Para que a sumarização de vídeos continue a evoluir, é crucial investir em pesquisa que aborda os desafios mencionados. Inovações em inteligência artificial, especialmente em técnicas de aprendizado profundo, podem melhorar a precisão e a relevância da sumarização (ZHANG *et al.* 2020).

D. Front-end na sumarização de multivídeos

No contexto atual da tecnologia da informação, o front-end refere-se à parte da aplicação web com a qual os usuários interagem diretamente. Engloba tudo, desde o layout e design até a interação com o usuário. O avanço constante das tecnologias front-end tem transformado a maneira como os desenvolvedores pensam sobre a construção de aplicações web (MARCOTTE, 2011).

Desenvolver o front-end de um sistema de sumarização multivídeo envolve várias considerações importantes para garantir uma experiência de usuário eficiente, intuitiva e acessível.

O Design Responsivo, com o aumento do uso de dispositivos móveis, o design responsivo se tornou uma norma. Utilizar unidades relativas e médias queries garante que as aplicações sejam funcionalmente adequadas em diferentes tamanhos de tela (MARCOTTE, 2011).

Sugere-se a seguir alguns aspectos a serem considerados no desenvolvimento de um front-end.

- A compatibilidade com dispositivos móveis para garantia que o design seja responsivo, funcionando bem em smartphones, tablets e desktops, a experiência do usuário (UX) pois a integração de inteligência artificial pode personalizar experiências do usuário, automatizar a criação de interfaces e até mesmo ajudar no desenvolvimento por meio de assistentes inteligentes (GONZALEZ, 2020).
- A acessibilidade deve ser uma prioridade no desenvolvimento front-end. A aplicação de técnicas que permitam que pessoas com deficiências utilizem a interface é essencial para a inclusão (MASON, 2018).
- O desempenho da aplicação deve ser otimizado para oferecer uma melhor experiência ao usuário, as tecnologias e ferramentas das arquiteturas de micro front-ends permitem dividir uma aplicação em partes menores e independentes, facilitando a escalabilidade e a manutenção (SCHILLING, 2021) e, os testes e depuração para prova de conceito de operação.

Cabe ressaltar que o desenvolvimento web tem evoluído rapidamente nas últimas décadas, com a crescente demanda por aplicações dinâmicas e interativas. A linguagem de programação Python, por exemplo, é versátil juntamente com o microframework Flask e, tem se tornado uma escolha popular para o desenvolvimento de aplicações web sendo estes recursos adotados no desenvolvimento do sistema web deste artigo.

A linguagem Python, também é conhecido por sua simplicidade e legibilidade, tem ganhado espaço no desenvolvimento web, especialmente no back-end. Flask, um microframework leve, permite que os desenvolvedores construam aplicações web robustas rapidamente e o Flask é leve, fácil de entender e escalável (GRINBERG, 2018).

Python é uma linguagem de programação de alto nível, amplamente utilizada por sua facilidade de uso e extensa biblioteca de módulos. Sendo uma linguagem interpretada, o desenvolvimento e a depuração se tornam mais ágeis, tornando-a ideal para protótipos e aplicações em produção (LUTZ, 2013).

Um projeto típico em Flask como o sistema web deste artigo, inclui uma estrutura de diretórios que facilita a organização do código. As pastas comuns incluem a pasta *templates* onde estão os arquivos HTML das páginas, a pasta *static* para arquivos estáticos como CSS, JavaScript e imagens e a pasta *app.py* do arquivo principal onde a aplicação é definida.

Algumas vantagens do uso de Flask para front-end podem ser percebidas como, por exemplo, a flexibilidade pois o referido microframework permite que os desenvolvedores usem bibliotecas e ferramentas que possibilitam um desenvolvimento mais customizado.

Comunidade de desenvolvedores e de suporte ativa, também é uma das vantagens do ecossistema Python e Flask pois comumente fornecem extensões e suportes, tornando mais fácil encontrar soluções para problemas comuns (GRINBERG, 2018).

Integração com bibliotecas também é salutar no contexto de aplicação da Python e do Flask, pois a linguagem Python é amplamente utilizada em ciência de dados e aprendizado de máquina e o Flask permite integrar modelos e algoritmos diretamente nas aplicações web, oferecendo uma maneira eficiente de criar aplicações baseadas em dados (RASCHKA e MIRJALILI, 2019).

Neste sentido, o sistema web aqui proposto visa possibilitar maior alcance e disponibilidade das informações de vídeos multimodais da web, promovendo a inclusão social, acessibilidade e integração ao conhecimento das pessoas com surdez e assim, democratizando o acesso conhecimento.

Pode-se citar como exemplos que se justifica o uso deste sistema, os casos de pessoas que queiram consumir conteúdo de vídeos que não sejam de seu idioma ou uma pessoa com surdez queria assistir vídeo aulas que não eram disponibilizadas com tradução em linguagem de sinais para surdos.

2. MATERIAIS E MÉTODO

No desenvolvimento do sistema web sumarizador automático, apresentado neste artigo, foi considerado princípios da computação exulta resultando num sistema de fácil operação.

A. Sequência de operação do referido sistema

Primeiro o usuário deverá copiar no campo Vídeos do front-end o caminho em que está o vídeo a ser transcrito. Caso o usuário deseje sumarizar mais de um vídeo em uma única operação, deverá clicar no botão que está escrito “ADICIONAR OUTRO VÍDEO +” que um novo campo abrirá para cada inserção de mais link.

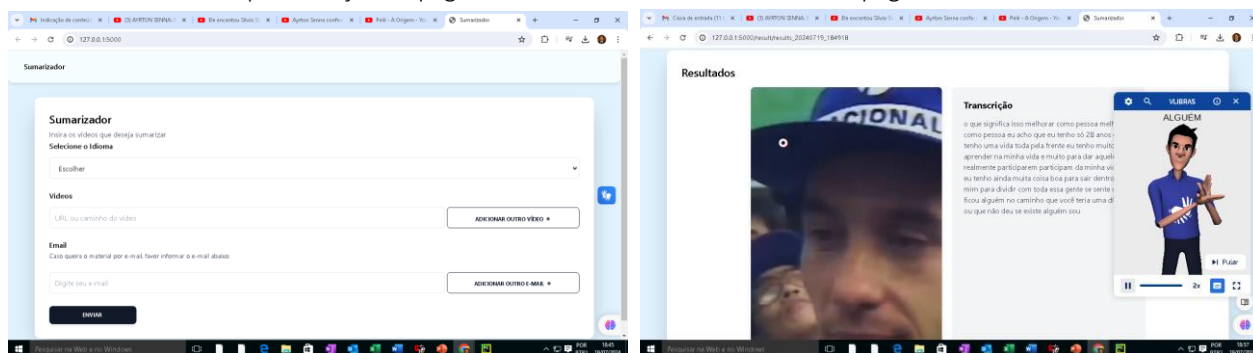
Na sequência o usuário poderá selecionar o idioma que deseja a transcrição. Por exemplo, se a língua nativa do(s) vídeo(s) está no idioma inglês, bastará selecionar a opção para língua portuguesa que a tradução sairá traduzida, possibilitando o uso do avatar intérprete de linguagem de sinais que no momento é específico apenas para língua portuguesa.

Caso o usuário desejar enviar ou receber os arquivos do vídeo, concatenado ou não, em formato mp4 e a transcrição em formato txt, bastará incluir o e-mail destino que o sistema enviar os respectivos arquivos automaticamente para o endereço eletrônico preenchido no formulário do front-end. Após os preenchimentos é necessário clicar no botão do formulário do front-end que está escrito “ENVIAR” e isto iniciará os processamentos e abrirá automaticamente uma página web de

RESULTADOS onde será possível assistir e ouvir o vídeo ou vídeos concatenados, ler na mesma tela a transcrição e acionar o avatar digital intérprete de linguagem de sinais, no caso o VLibras.

A Figura 1 apresenta a página web do formulário de entrada e página web dos resultados que disponibiliza o vídeo extraído do URL, a transcrição correspondente e o intérprete de linguagem de sinais avatar digital realizando a interpretação da transcrição possibilitando acessibilidade ao conteúdo de pessoas com surdez.

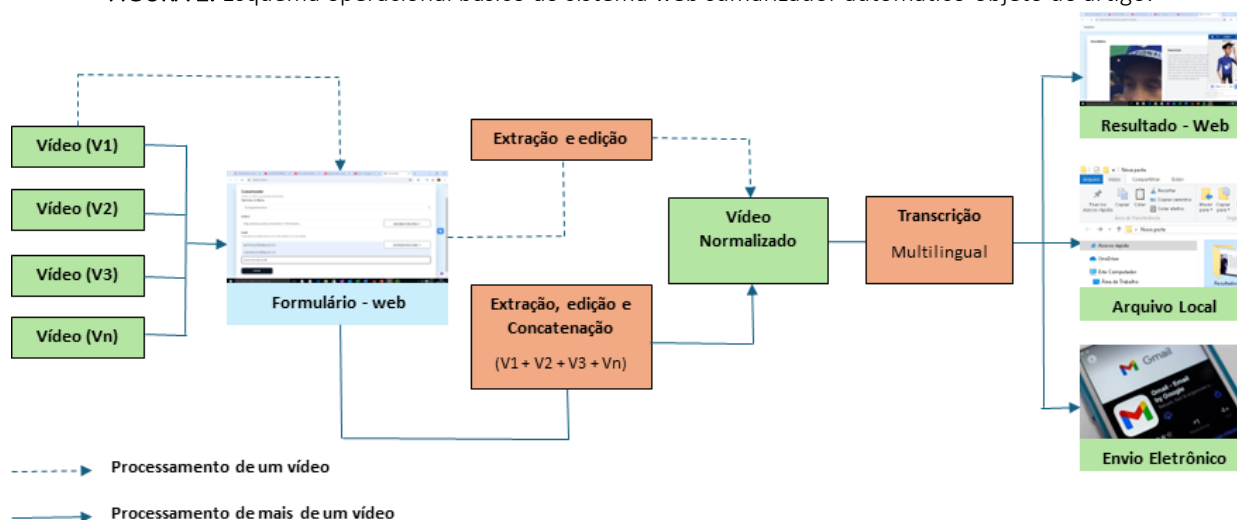
FIGURA 1. Apresentação da página web do formulário de entrada e a página web dos resultados.



Fonte: Próprio autor (2024).

Ressalta-se que é possível acionar legenda do avatar, modificar o personagem e a velocidade dos gestos de comunicação dentro do padrão oferecido. O esquema operacional básico do sistema web sumarizador automático com inteligência artificial para transcrição de multivídeos e envio eletrônico, integrado ao interpretador de linguagem de sinais avatar digital do tipo VLibras é apresentado na Figura 2, onde é possível observar o respectivo formulário de entradas (front-end), bem com o processamento e as multisaídas possíveis do sistema.

FIGURA 2. Esquema operacional básico do sistema web sumarizador automático objeto do artigo.



Fonte: Próprio autor (2024).

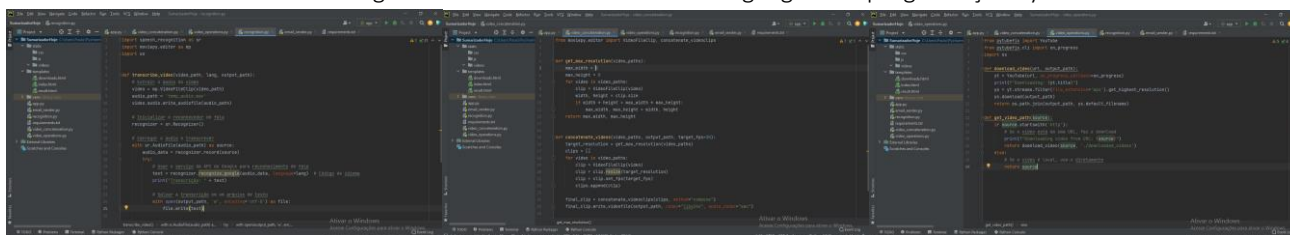
B. Back-end

O back-end foi desenvolvido em nove programas separados em microserviços e, em mais de uma linguagem de programação, sendo um destes o programa middleware que é responsável por iterar todos de forma eficaz, contrastando assim com a arquitetura monolítica, onde todas as funcionalidades estão integradas em um único bloco de código.

A arquitetura de microserviços permite a utilização de diferentes tecnologias e linguagens de programação para diferentes serviços, proporcionando flexibilidade para escolher a melhor ferramenta para cada tarefa específica (GAJANAN, J. 2019).

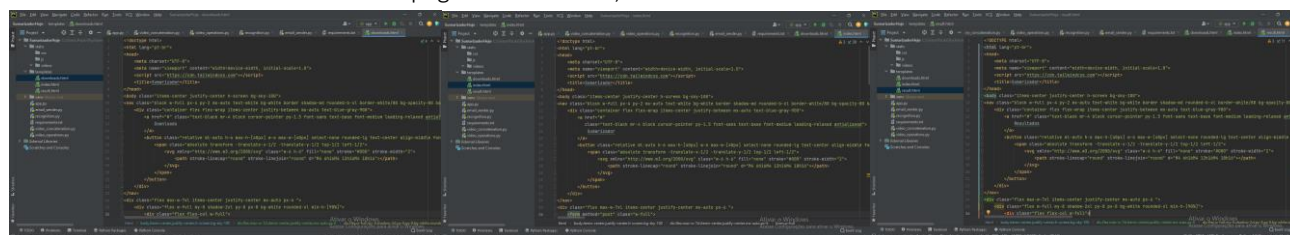
A Figura 3 abaixo, demonstra os arranjos dos programas do back-end que foram desenvolvidos em linguagem de programação Python. A Figura 4 demonstra os arranjos dos programas do back-end que foram desenvolvidos em linguagem de programação HTML das páginas downloads, index (formulário de entradas) result. Os códigos dos programas em Python e em HTML foram implementados no IDE PyCharm.

FIGURA 3. Códigos do back-end do sistema em linguagem de programação Python.



Fonte: Próprio autor (2024).

FIGURA 4. Códigos do back-end do sistema desenvolvidos na linguagem de programação HTML - para geração das páginas downloads, index e result do sistema.



Fonte: Próprio autor (2024).

C. Middleware do sistema

Com o crescimento exponencial da complexidade nos sistemas computacionais, a integração de diferentes aplicações e plataformas tornou-se um desafio significativo.

O middleware surgiu como uma solução eficaz para permitir a interoperabilidade, tratando das questões de comunicação, escalabilidade e gerenciamento de dados entre sistemas heterogêneos (POPEK, 2010).

Middleware é definido como uma camada de software que reside entre o sistema operacional e as aplicações em uma rede (KOZLOV, 2015).

Para Sharma *et al.* (2018), as funções principais de um middleware visam facilitar a comunicação entre diferentes componentes de software, abstrair detalhes de implementação da comunicação em rede e fornecer serviços de gerenciamento de transações, segurança e persistência de dados.

Existem vários tipos de middleware, entre os quais se destacam:

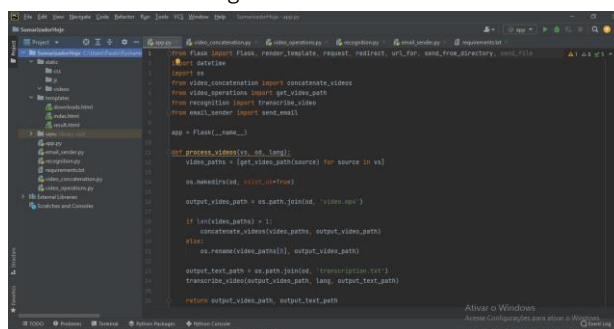
- Middleware de Mensagens sendo os que facilitam a comunicação assíncrona entre aplicativos por meio de sistemas de mensageria (HOHPE e WOOLF, 2009).
- Middleware de Objetos Remotos sendo os que permitam que objetos em diferentes locais se comuniquem, como CORBA (Common Object Request Broker Architecture) e RMI (Remote Method Invocation) (BICHLER e MACKIE-MASON, 2008),
- Middleware para Web Services que suportam a criação e uso de serviços web, permitindo a interoperabilidade via protocolos como SOAP e REST (BREWSTER et al., 2016).
- Middleware em Nuvem que são os facilitam a integração e o gerenciamento de recursos em ambientes de computação em nuvem (KUMAR *et al.* 2019).

Cabe mencionar algumas vantagens do uso de middleware, tais como: a interoperabilidade possibilitando que sistemas diferentes se comuniquem, independentemente das plataformas subjacentes, a escalabilidade para facilitar a adição de novos serviços e aplicativos sem a necessidade de reconfiguração extensiva e, o desacoplamento possibilitam os desenvolvedores focarem na lógica de negócios sem se preocupar com as complexidades de comunicação e integração de sistemas (ZHANG *et al.* 2020).

Conforme Agboola *et al.* (2023), a evolução do middleware está intimamente ligada às novas tecnologias emergentes, como inteligência artificial e machine learning, que podem automatizar a integração e a comunicação em sistemas complexos.

A Figura 5 apresenta o programa middleware do sistema sumarizador automático do artigo em linguagem de programação Python.

FIGURA 5. Programa middleware do sistema.



Fonte: Próprio autor (2024).

D. Laboratório de Sistemas Distribuídos e Programação Concorrente – LASDPC: ambiente de desenvolvimento e testes parciais do sistema

O Laboratório de Sistemas Distribuídos e Programação Concorrente (LaSDPC) do ICMC/USP, ambiente da pesquisa, nasceu em 1990, quando alguns professores do ICMC voltaram ao Brasil após o doutorado.

As linhas de pesquisa nas quais o LaSDPC vem trabalhando recentemente incluem: simulação, prototipagem e benchmark, índices de carga para web services, computação móvel, computação verde, computação de alto desempenho e desenvolvimento de aplicações de alta performance, testes de aplicações distribuídas e paralelas, agendamento de processos, sistemas distribuídos grid e cloud computing, arquitetura orientada a serviços, QoS, software livre / open source, técnicas e metodologias para desenvolvimento aberto de software, recursos educacionais abertos, redes de computadores de segurança em redes de computadores, protocolos, middlewares para redes ad-hoc voadoras, computação ubíqua, sistemas computacionais adaptativos modelagem dinâmica de sistemas computacionais, simulação computacional de sistemas naturais, autogestão e otimização em tempo real, sistemas web, processamento de dados estruturados e não-estruturados gerais e inteligência artificial.

O LaSDPC, Figura 6, possui três cluster integrados para computação paralela e concorrente, sendo estes o Cluster Halley, o Cluster Cosmos e o Cluster Andrômeda, porém, a seguir é apresentada as características técnicas apenas do Cluster Halley que foi utilizado no desenvolvimento e testes do sistema objeto deste artigo.

FIGURA 6. LaSDPC/USP - Área de desenvolvimento prático da pesquisa.



Fonte: Próprio autor (2024).

Às características técnicas de hardware do Halley, empregados na pesquisa são: Nodes (13 hosts / 104 virtuais) – 12 slaves nodes / 1 master node, Intel® Core™ I7 Processor – LGA -1150 – 4790 3.60GHZ DMI 5GT/S 8MB, 32 GB RAM DDR3 Corsair Vegeance, Motherboard Gigabyte GA-Z97X-SLI ATX, Placa Gráfica Nvidia GTX 650 – 1GB, HD 2TB Seagate SATA III 7200RPM e Fonte ATX 650W Real. A Figura 6 acima, apresenta parte da área do ambiente do desenvolvimento prático do sistema web sumarizador automático onde é possível observar ao lado direito os Clusters Halley, Cosmos e o Cluster Andrômeda.

Foram utilizadas as seguintes bibliotecas operacionais no desenvolvimento do back-end e do front-end foram: blinker 1.8.2, certifi 2024.6.2, charset-normalizer 3.3.2, click 8.1.7, colorama 0.4.6, decorator 4.4.2, ffmpeg-python 0.2.0, flask 3.0.3, future 1.0.0, idna 3.7, imageio 2.34.2, imageio-

ffmpeg 0.5.1, importlib-metadata 8.0.0, itsdangerous 2.2.0, jinja2 3.1.4, MarkupSafe 2.1.5, moviepy 1.0.3, numpy 2.0.0, opencv-python 4.10.0.84, pillow 10.3.0, proglog 0.1.10, pytube 15.0.0, requests 2.32.3, secure-smtp lib 0.1.1, SpeechRecognition 3.10.4, tqdm 4.66.4, typing-extensions 4.12.2, urllib3 2.2.2, werkzeug 3.0.3 e zipp 3.19.2.

O sistema web foi desenvolvido utilizando o sistema operacional Windows V10 da empresa Microsoft em ambiente de desenvolvimento integrado multiplataforma PyCharm Community Edition da empresa JetBrains, essencialmente em linguagem de programação Python e HTML.

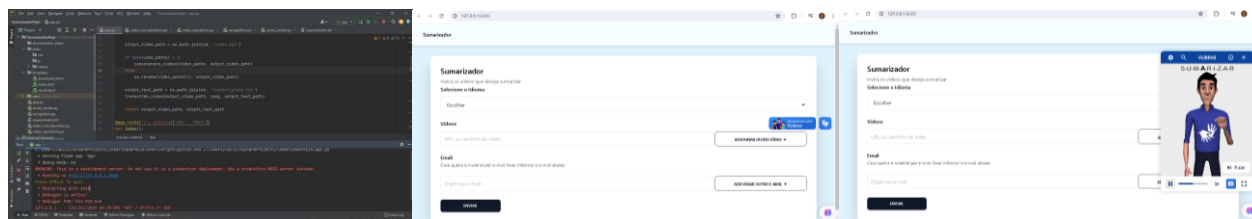
3. RESULTADOS

Os testes práticos visaram a apuração funcional das funções do sistema web, sendo estas: operacionalidade dos códigos do back-end e middleware cominando a disponibilização de uma página web de formulário para entradas, funcionalidade do formulário com as entradas dos URL dos vídeos e dos e-mails que receberão os arquivos, a obtenção de uma página web de resultados com juntos os vídeos concatenados ou não, o texto da respectiva transcrição geral e o avatar gráfico digital interprete de linguagem de sinais integrado. Também foi testado as funções extração dos vídeos do URL, armazenamento local juntamente com o arquivo da transcrição e o envio automático destes arquivos para e-mails cadastrado no formulário da página web inicial. A seguir apresenta-se os resultados destes testes.

A. Inicialização do formulário web de entrada

Primeiramente, na Figura 7, apresenta-se a abertura do formulário para inclusão do(s) vídeo(s) e do(s) endereço(s) eletrônicos para envio e recebimento dos arquivos do vídeo em formato mp4 e do arquivo do sumário no formato txt. Nota-se a presença do avatar intérprete de linguagem de sinais para auxílio quanto o sistema for operado por pessoa com surdez melhorando a acessibilidade ao conteúdo e a inclusão social.

FIGURA 7. Abertura do formulário web do sistema para inclusão dos caminhos dos vídeos a serem sumarizados.

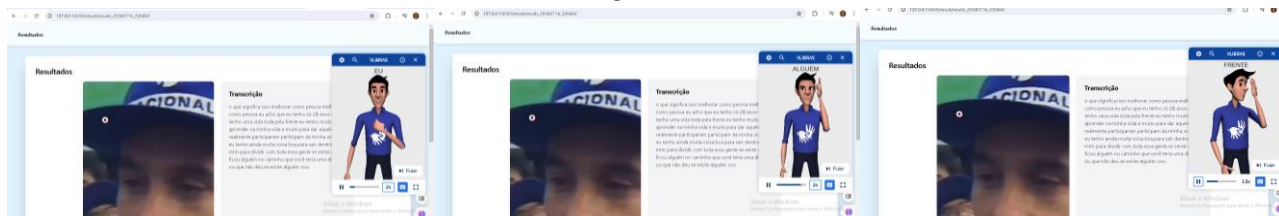


Fonte: Próprio autor (2024).

B. Resultados obtidos com a sumarização automática de um único vídeo

A Figura 8 demonstra a página web gerada automaticamente com o vídeo, sumário e o intérprete de sinais avatar digital referente a entrada de um vídeo.

FIGURA 8. Página web gerada automaticamente com o vídeo, sumário e intérprete de linguagem de sinais avatar digital - um vídeo.



Fonte: Próprio autor (2024).

A Figura 9 demonstra o resultado do envio automático para um endereço eletrônico cadastrado na página do formulário web. Nessa figura, é possível observar que o endereço eletrônico recebeu um e-mail com um arquivo de vídeo em formato mp4, ou seja, o vídeo extraído do URL original, e um arquivo textual em formato txt, referente a transcrição do vídeo escolhido.

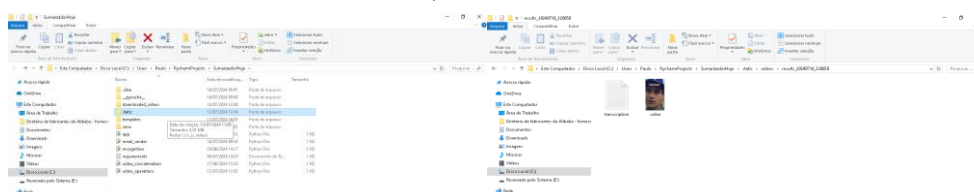
FIGURA 9. Arquivo de um vídeo em formato mp4 e da transcrição em formato txt que foi enviado automaticamente pelo sistema (e-mail, download e visualização dos vídeos e do sumário).



Fonte: Próprio autor (2024).

A Figura 10 a seguir, demonstra o arquivamento automático na máquina local do vídeo em formato mp4 extraído do URL, e o arquivo textual referente a transcrição deste vídeo em formato txt.

FIGURA 10. Arquivos gerados e armazenados automaticamente na máquina local - um vídeo em formato mp4 e um arquivo em formato txt.



Fonte: Próprio autor (2024).

Cabe mencionar que os resultados foram os mesmos obtidos com a sumarização de um vídeo em idioma estrangeiro, no caso foi testado um vídeo em inglês.

C. Resultados obtidos com multivídeos

A Figura 11 demonstra o resultado web da sumarização de mais de um vídeo onde é possível observar que os vídeos de entrada foram concatenados automaticamente e que foi gerado um único texto da respectiva transcrição dos vídeos que foram unificados.

FIGURA 11. Página web gerada automaticamente com o vídeo, sumário e intérprete de linguagem de sinais avatar digital - dois vídeos concatenados.



Fonte: Próprio autor (2024).

A Figura 12 demonstra o resultado web da sumarização de mais de um vídeo onde é possível observar que os vídeos de entrada foram concatenados automaticamente e que foi gerado um único texto geral da respectiva transcrição dos vídeos que foram unificados.

FIGURA 12. Arquivo de dois vídeos concatenados em formato enviados por e-mail.



Fonte: Próprio autor (2024).

A Figura 13 demonstra o resultado do arquivamento local e automático dos respectivos arquivos, sendo um do vídeo resultante do processo de concatenação em formato mp4 e o outro a transcrição/sumário geral em formato txt.

D. Resultados com envios para multiendereço eletrônicos

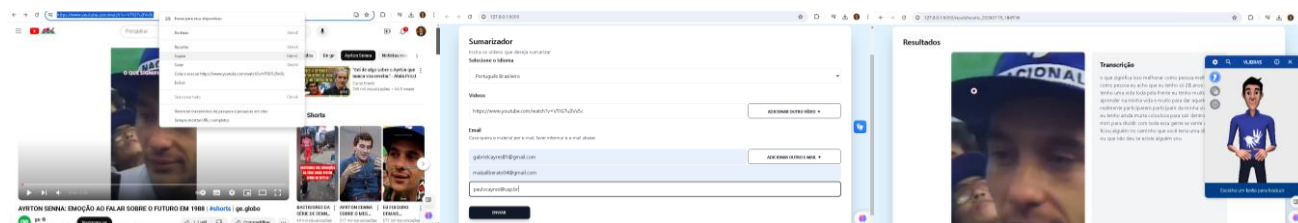
A Figura 14 demonstra o cadastramento do vídeo a ser sumarizado, mais de um endereço eletrônico sendo cadastrado para receber o arquivo do vídeo em formato mp4 e o arquivo textual em formato txt e o resultado na página web.

FIGURA 13. Arquivos gerados e armazenados automaticamente na máquina local - dois vídeos concatenados vídeo em formato mp4 e um arquivo do sumário geral em formato txt.



Fonte: Próprio autor (2024).

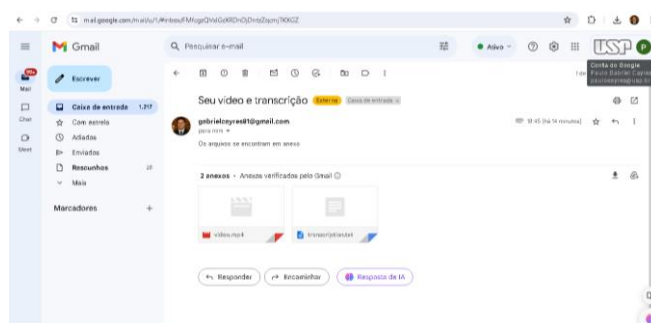
FIGURA 14. Resultado de envio automático do vídeo em formato mp4 e da sumarização automática em formato txt para múltiplos e-mails.



Fonte: Próprio autor (2024).

A Figura 15 demonstra que o primeiro e-mail paulocayres@usp.br, cadastrado no formulário web recebeu o arquivo do vídeo estratificado do caminho URL em formato mp4 e o arquivo do sumário em formato txt a partir de uma única sumarização pelo sistema.

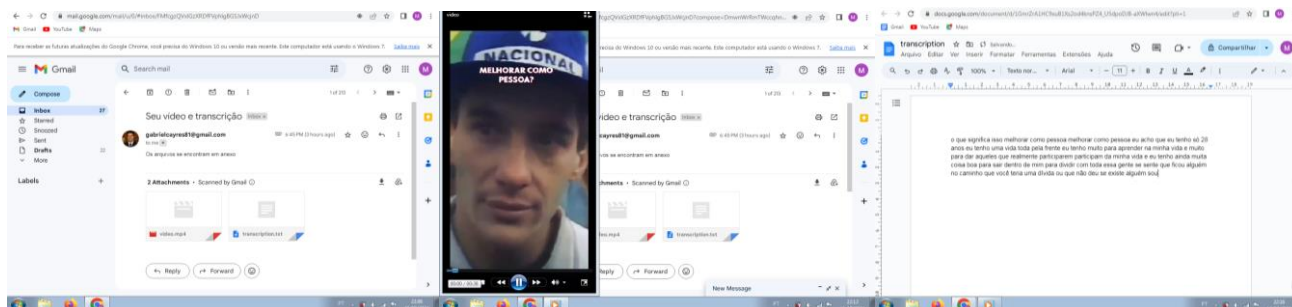
FIGURA 15. Arquivos recebido pelo e-mail paulocayres@usp.br com o arquivo mp4 e txt anexos da sumarização enviados automaticamente pelo sistema web objeto deste artigo.



Fonte: Próprio autor (2024).

A Figura 16 demonstra que o primeiro e-mail maizaliberato04@gmail.br, cadastrado no formulário web recebeu o arquivo do vídeo estratificado do caminho URL em formato mp4 e o arquivo do sumário em formato txt a partir de uma única sumarização pelo sistema.

FIGURA 16. E-mail maizaliberato04@gmail.br com arquivo mp4 e txt anexos oriundos da sumarização recebidos do sistema web sumarizador automático desenvolvido.



Fonte: Próprio autor (2024).

5. CONCLUSÕES

Dada todas simulações e respectivos resultados apresentados neste trabalho, pode-se concluir que o sistema objeto deste artigo, é capaz de produzir resultados positivos quanto a geração de conhecimentos, insights, acesso à informação e para educação inclusiva assistida por máquinas.

Apesar de já existir sumarizadores de vídeos com inteligência artificial, fica claro o grau de inovação e aplicabilidade do sistema aqui proposto, por possibilitar a sumarização de multivídeos, ou seja, gerar a respectiva transcrição geral dos vídeos concatenados e disponibilizar pelo menos um avatar digital intérprete de linguagem de sinais lado a lado, ao mesmo tempo que o sistema gera armazenamento em máquina local e envios eletrônicos automáticos para uma quantidade ilimitada de endereços de e-mails, permitindo-se receber os arquivos correspondentes em qualquer local que tenha sinal de internet, quer seja em computadores ou em aparelhos telefônicos celulares, sendo estes arquivos enviados a partir de uma mesma plataforma otimizando assim o tempo, os recursos e mão-de-obra na produção de conteúdos por meio da sumarização de multivídeo multimodal.

Pode-se também apontar que, até o momento da escrita deste artigo, os intérpretes de linguagem de sinais do tipo avatar digital, não realizam comunicação eficaz para leitura labial pois esta tecnologia encontra-se em desenvolvimento. Mas, no caso deste sistema web proposto, o vídeo fica disponível próximo ao avatar possibilitando a pessoa surda observar também os movimentos labiais do comunicador no vídeo ao falar, ao mesmo tempo que o usuário assiste à interpretação em linguagem de sinais feita pelo avatar digital, sem a perda de informações visuais representativas que possam estar no ambiente do vídeo em visualização, tudo isto em um mesmo lay-out ergonomicamente favorável, permitindo, portanto, capturar tanto aspectos linguísticos como visuais.

Contudo, comprova-se que é possível em um único sistema web gerar sumarização automática de multivídeo multimodal, gerar arquivos e envios eletrônicos em massa dos resultados deste e promover acessos ao conhecimento de forma assistida por meio da tecnologia digital.

DECLARAÇÃO DE RESPONSABILIDADE

O autor é o único responsável por este trabalho.

REFERÊNCIAS

- AGBOOLA, O. et al. (2023). AI-Driven Middleware Solutions for Distributed Systems. *Journal of Computer Networks*, 15(2), 123-135.
- ALHUSSEINI, M., DIAB, A., & ALSHEHRI, M. (2020). Video summarization techniques: A review. *Future Generation Computer Systems*, 108, 345-367.
- BICHLER, M., & MACKIE-MASON, J. K. (2008). Architectures for the Global Information Infrastructure. *Communications of the ACM*, 51(10), 112-120.
- DENG, R., NIU, Q., & ZHANG, Y. (2020). Video Summarization Using Hierarchical Deep Reinforcement Learning. *IEEE Transactions on Multimedia*, 22(9), 2302-2313.
- GRINBERG, M. (2018). *Flask Web Development: Developing Web Applications with Python*. O'Reilly Media.
- HASSANNIA, S., RAGHUNATHAN, A., & ANANTHARAM, P. (2021). A study on video summarization challenges. *Journal of Visual Communication and Image Representation*, 75, 103058.
- HE, Z., ZHANG, H., & LI, Z. (2019). Social media video summarization: An overview and future directions. *Journal of Network and Computer Applications*, 135, 1-12.
- HOHPE, G., & WOOLF, B. (2009). *Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions*. Addison-Wesley.
- KOZLOV, S. (2015). Middleware: The Glue of Distributed Systems. *Distributed Computing Journal*, 28(4), 305-318.
- KUMAR, M. et al. (2019). Cloud Middleware: A Review of Services and Features. *International Journal of Cloud Computing and Services Science*, 8(1), 12-29.
- LI, Y., YU, H., & LUO, Z. (2020). Deep learning for video summarization: A review. *Computer Science Review*, 38, 100276.
- LIU, H., WANG, Y., & ZHANG, J. (2019). The impact of video summarization on learning outcomes. *Educational Technology Research and Development*, 67(3), 601-620.
- LUTZ, M. (2013). *Learning Python*. O'Reilly Media.
- MARCOTTE, E. (2011). *Responsive Web Design*. A Book Apart.
- MASON, J. (2018). *Accessibility for Web Design: A Guide to Making Your Site Inclusive*. O'Reilly Media.
- MAYER, R. E., & MORENO, R. (2003). Nine ways to reduce cognitive overload in multimedia learning. *Educational Psychologist*, 38(1), 43-52.

- POPEK, G. (2010). Middleware for Distributed Systems. IEEE Software, 27(6), 23-29.
- RASCHKA, S., & MIRJALILI, V. (2019). Python Machine Learning. Packt Publishing.
- SCHILLING, C. (2021). Micro Frontends in Action. Manning Publications.
- SHARMA, G. et al. (2018). Middleware Technologies for Cloud Computing. Journal of Cloud Computing: Advances, Systems and Applications, 7(1), 3-20.
- TIAN, Y., LIU, S., & ZHANG, D. (2020). Enhancing learning outcomes through video summarization in online education. Computers & Education, 149, 103798.
- WANG, X., & JIANG, H. (2021). Video summarization for trailers and teasers: A comprehensive analysis. IEEE Transactions on Multimedia, 23, 2345-2354.
- XU, Y., ZHANG, L., & YU, L. (2020). Time-efficient methods for video content retrieval: A comparative study. Computers in Human Behavior, 104, 106189.
- YU, Y., et al. (2019). Learning to Summarize with Human Attention. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (ACL), 116-126.
- WANG, X., & JIANG, H. (2021). Accessibility of educational video content through summarization techniques. International Journal of Educational Technology in Higher Education, 18(1), 1-16.
- ZHANG, Y., MOHAMMED, F., & CHEN, L. (2020). Video summarization via neural networks: Insights and applications. IEEE Transactions on Multimedia, 22(12), 3167-3181.
- ZHAU, WAYNE XIN; et al. (2023). Uma pesquisa de grandes modelos de linguagem. arXiv: 2303.18223 [cs.CL].

AGRADECIMENTOS

O autor agradece ao CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico e à USP - Universidade de São Paulo pelo apoio fornecido ao longo deste trabalho.