

Sumarização Automática de Vídeos: Geração de Descrições Narrativas

Ester Adaianne O. Ferreira¹, João Pedro de Brito Tomé², Leonardo C. Filho³

¹Instituto de Informática – Universidade Federal de Goiás (UFG)
Goiânia – GO – Brazil

{esteradaianne, joaopedrodebritotome, leonardo.cortes}@discente.ufg.br

Abstract. *With the exponential growth of videos on digital platforms, it has become unfeasible to consume and describe all content organically. This project proposes an automated system for generating descriptive and coherent video summaries by combining techniques from computer vision and natural language processing (NLP). To support the proposal, related studies on video summarization were reviewed, and the C4 model was used as the basis for the system's architecture. Preliminary results indicate that the integration of these technologies can enable the automatic generation of accurate and contextualized descriptions, optimizing the consumption and indexing of content on a large scale.*

Resumo. *Com o aumento exponencial de vídeos em plataformas digitais, tornou-se inviável consumir e descrever todo o conteúdo de forma orgânica. Este projeto propõe um sistema automatizado para a geração de sumários descritivos e coerentes de vídeos, combinando técnicas de visão computacional e processamento de linguagem natural (PLN). Para embasar a proposta, foram revisados estudos relacionados à sumarização de vídeos, e o modelo C4 foi utilizado como base para a arquitetura do sistema. Os resultados preliminares indicam que a integração dessas tecnologias pode viabilizar a geração automática de descrições precisas e contextualizadas, otimizando o consumo e a indexação de conteúdo em larga escala.*

1. Introdução

Nos últimos anos, o volume de vídeos publicados e consumidos em plataformas digitais cresceu, impulsionado pela popularização de redes sociais, algoritmos que beneficiam vídeos curtos e interessantes e ferramentas de criação de conteúdo remuneradas [DOMO, 2024]. Com a quantidade de vídeos publicados por minuto em plataformas como o TikTok podendo chegar a mais de 15 mil e mais de 500 horas de conteúdo sendo adicionadas ao Youtube nesse mesmo período de tempo, torna-se inviável consumir e indexar ou sumarizar todo o conteúdo de maneira orgânica. [DOMO 2024, Youtube 2025]. Diante disto, a sumarização automatizada surge como um possível facilitador para o processo de analisar, indexar, resumir e sumarizar vídeos em plataformas digitais. [Tiwari and Bhatnagar 2021]

2. Trabalhos relacionados

A fim de compreender o atual estado da literatura e pesquisa sobre sumarização automática de vídeos e sumarização utilizando técnicas como processamento de linguagem natural (PLN) foram realizadas pesquisas na plataforma Google Scholar utilizando as seguintes strings de busca: "Sumarização Automática de Vídeos: Geração

de Descrições Narrativas”e multi-video summarization”OR multi-video synopsis”OR ”video-summarization”OR ”video-synopsis.

A pesquisa resultou em muitos artigos e trabalhos acadêmicos, assim como alguns capítulos de livros abordando o assunto de sumarização de vídeos. Após a leitura dos artigos e trabalhos encontrados que tinham as palavras chaves mais relevantes ao tema, chegou-se à conclusão de que as pesquisas sobre sumarização de vídeos foram bastante utilizadas no contexto de sumarização de vídeos de sistemas de segurança com muitas horas de imagens ociosas ou sem grandes acontecimentos. Há também menção à pesquisa voltada para a sumarização eficiente de emoções em, por exemplo, telejornais. Encontrou-se ainda alguns artigos sobre deep learning na sumarização.

Tamires Barbieri propôs uma abordagem que leva em consideração estratégias humanas para gerar sumários multivídeos relevantes semanticamente para algoritmos de usuários de plataformas digitais. [BARBIERI 2021] Hari K.C. e Manish Pokharel fizeram uma revisão aprofundada sobre sumarização de vídeos utilizando deep learning. [Hari and Pokharel 2024]

Os trabalhos considerados mais relevantes foram devidamente documentados no repositório do github e estarão também mencionados nas referências bibliográficas do relatório.

3. Proposta arquitetural

Para a documentação da proposta arquitetural para a solução foram utilizados os diagramas do Modelo C4. Este modelo foi escolhido pois permite uma boa visão hierárquica do sistema, facilitando a compreensão da arquitetura proposta.

3.1. Diagrama de contexto

O diagrama de contexto permite uma visão ampla do software, demonstrando com que personas e outros sistemas ele interage.



Figura 1. Diagrama de contexto

3.2. Diagrama de container

O diagrama de container mostra a distribuição de responsabilidades entre os principais módulos do sistema e também como os containers se comunicam entre si.

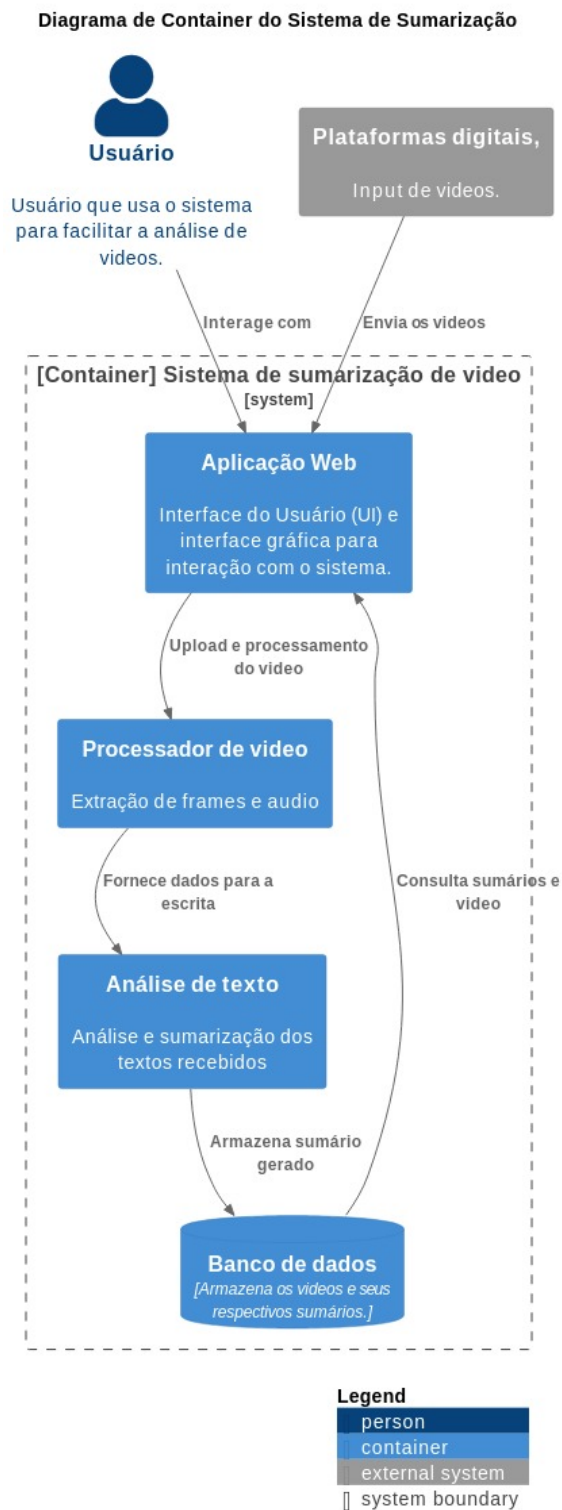


Figura 2. Diagrama de container

3.3. Diagrama de componentes

Mostra de que componentes é feito um container e alguns detalhes relevantes da implementação.

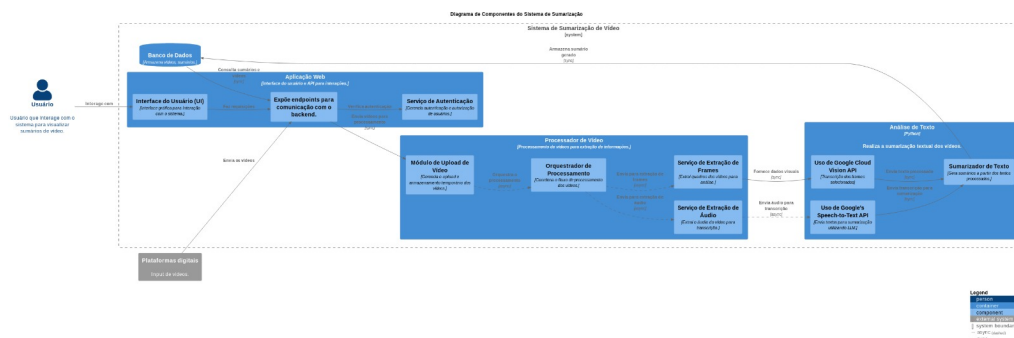


Figura 3. Diagrama de componentes

Referências

- BARBIERI, T. T. d. S. (2021). *Sumarização automática multivídeo baseada em estratégias humanas*. PhD thesis, Instituto de Ciências Matemáticas e de Computação, University of São Paulo.
- DOMO (2024). Data never sleeps 12.0. Technical report, DOMO.
- Hari, K. and Pokharel, M. (2024). A review on video summarization using deep learning: Approaches, challenges and future direction. *SSRN*.
- Tiwari, V. and Bhatnagar, C. (2021). A survey of recent work on video summarization: approaches and techniques. *Springer Nature*.
- Youtube (2025). Youtube in numbers. Technical report, YouTube.