Video Retrival

Guilherme G. Silverio[1], **Marcelo A. Mendonça**[2], **Matheus A. de Aguiar**[3]

[1]Instituto de Informática – Universidade Federal de Goiás (UFG)

1. Introdução

Este trabalho tem como objetivo o desenvolvimento de um sistema de busca multimodal capaz de localizar segmentos de vídeos que correspondam a descrições textuais fornecidas pelo usuário. O sistema utilizará modelos de aprendizado de máquina, como o CLIP, para mapear texto e vídeo em um espaço compartilhado, permitindo a análise de similaridade semântica. O foco principal será em vídeos de diversas categorias, desde conteúdos simples aos mais complexos.

2. Trabalhos Relacionados

FAISS é uma biblioteca desenvolvida pelo Facebook AI Research (FAIR) para busca eficiente de vetores em grandes conjuntos de dados. Ele é otimizado para operações de busca por similaridade, como encontrar os vetores mais próximos (k-nearest neighbors) em um espaço de alta dimensionalidade.

2.1. Vantagens de uso

Eficiência: Projetado para lidar com milhões ou até bilhões de vetores.

Flexibilidade: Suporta diferentes métricas de similaridade, como cosseno, distância euclidiana, etc.

Escalabilidade: Oferece opções de busca aproximada (ANN - Approximate Nearest Neighbors) para melhorar o desempenho em grandes conjuntos de dados.

2.2. Funcionamento

Indexação: Os vetores (embeddings) são armazenados em uma estrutura de dados otimizada para busca rápida.

O FAISS oferece diferentes tipos de índices, como Índice Flat (busca exata) ou Índice IVF (busca aproximada com partições).

Busca: Dado um vetor de consulta (por exemplo, o embedding de uma descrição textual), o FAISS encontra os vetores mais próximos no índice.

A busca pode ser exata (retorna os vizinhos mais próximos com 100% de precisão) ou aproximada (mais rápida, mas com uma pequena perda de precisão).

Ranqueamento: Os resultados são retornados em ordem de relevância, com base na métrica de similaridade escolhida (por exemplo, similaridade de cosseno). O FAISS é uma ferramenta poderosa para ranquear resultados por relevância com base em embeddings. É útil ao projeto, pois permite buscar vídeos relevantes com base em descrições textuais de forma rápida e eficiente. Combinado com modelos multimodais como o X-CLIP, ele pode ser a peça para construir um sistema de busca preciso.

2.3. Como CLIP e FAISS trabalham juntos?

Geração de Embeddings: O CLIP gera embeddings para a descrição textual digitada pelo usuário e para frames ou segmentos de vídeo no repositório.

Indexação no FAISS: Os embeddings dos vídeos são indexados no FAISS para permitir buscas rápidas.

Busca e Ranqueamento: Quando o usuário digita uma descrição, o X-CLIP gera o embedding do texto. O FAISS usa esse embedding para buscar e ranquear os vídeos mais relevantes no repositório.

Retorno dos Resultados: Os vídeos mais relevantes são retornados ao usuário, ordenados por relevância.

Em um repositório com 1 milhão de vídeos, sem FAISS, o X-CLIP teria que calcular a similaridade entre o embedding do texto e os embeddings de todos os 1 milhão de vídeos.Isso seria muito lento e inviável para uma aplicação em tempo real. Com FAISS, os embeddings dos vídeos são pré-indexados. Quando o usuário faz uma consulta, o FAISS encontra os top 5 vídeos mais relevantes em milissegundos.

2.4. Ferramenta de Avaliação de Busca

A biblioteca ir_metrics é uma ferramenta em Python desenvolvida para avaliar a qualidade de sistemas de recuperação de informações, como o FAISS, que são usados para buscar itens semelhantes em grandes conjuntos de dados. Ela facilita a medição de desempenho de modelos de busca com base em métricas padrões, como Recall@K, Precision@K, MRR (Mean Reciprocal Rank), MAP (Mean Average Precision) e NDCG (Normalized Discounted Cumulative Gain), entre outras. Sua principal função é ajudar a analisar se os resultados retornados por um sistema de busca estão atendendo às expectativas em termos de relevância e precisão.

Ao utilizar ir_metrics, serão fornecidos os resultados da busca (os vídeos retornados pelo FAISS) e um conjunto de respostas relevantes esperadas, ou seja, os itens que, teoricamente, deveriam ser retornados para determinada consulta. A ferramenta, então, compara esses dois conjuntos de dados — os resultados da busca e as respostas relevantes — para calcular várias métricas de avaliação. Por exemplo, ao calcular o Recall@K, será possível verificar qual proporção de itens relevantes foi recuperada dentro dos primeiros K resultados. Já a Precision@K avalia quantos dos K primeiros itens retornados são realmente relevantes, enquanto o MRR foca na posição do primeiro item relevante dentro da lista, com a média do ”rank” inverso.

Enquanto o FAISS é utilizado para realizar a busca em si, ou seja, para encontrar os vídeos mais semelhantes a partir de uma consulta, ir_metrics não realiza a busca. Ele apenas recebe os resultados da busca realizada (por exemplo, uma lista de vídeos ou itens com suas respectivas pontuações) e compara esses resultados com o conjunto de dados relevantes para calcular as métricas mencionadas. Dessa forma, será possível avaliar como o sistema de busca está se comportando e se ele está retornando os resultados que são mais relevantes ou úteis para os usuários.

No contexto de um projeto, o FAISS será utilizado para indexar os vídeos e buscar os mais semelhantes com base em uma consulta fornecida. Depois, ir_metrics será usado para avaliar se o conjunto de vídeos retornado contém a maioria dos vídeos realmente

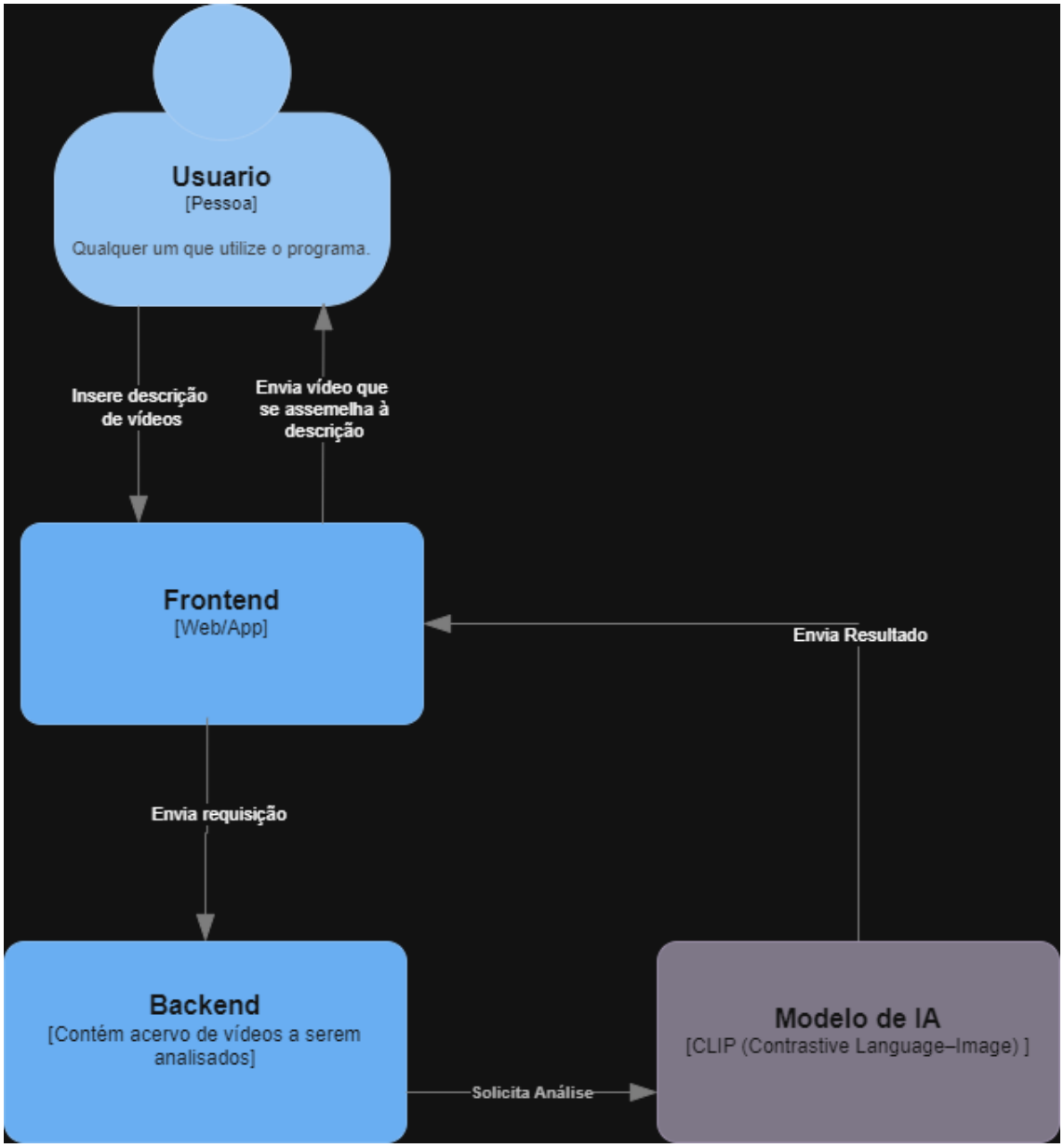
relevantes, verificando se os vídeos mais relevantes aparecem nas primeiras posições, se o sistema está cobrindo uma boa quantidade dos itens relevantes e qual a precisão da busca ao longo dos K primeiros resultados. Essa avaliação será essencial para entender o desempenho do sistema de busca e identificar possíveis ajustes para melhorar a precisão e a relevância dos resultados entregues ao usuário.

3. Dataset Escolhido

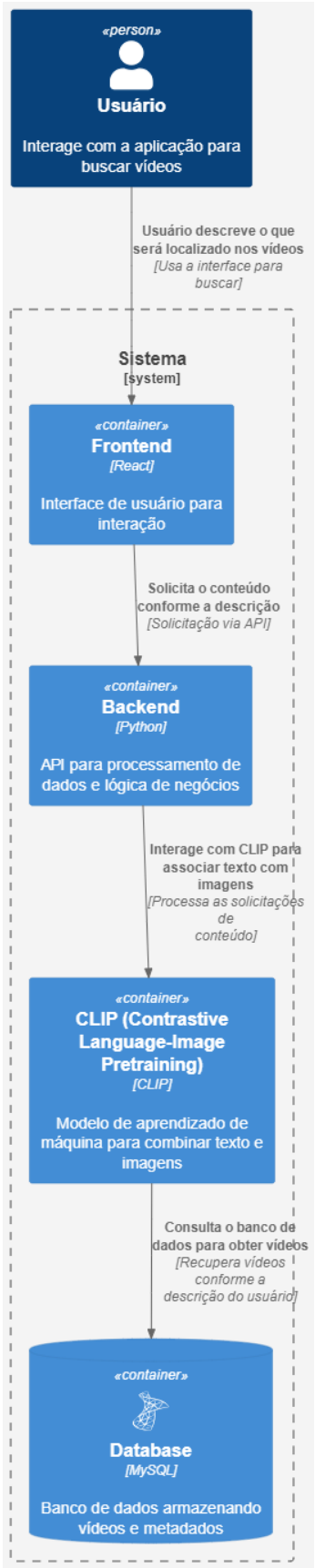
MSR-VTT (Microsoft Research Video to Text)

4. Proposta Arquitetural

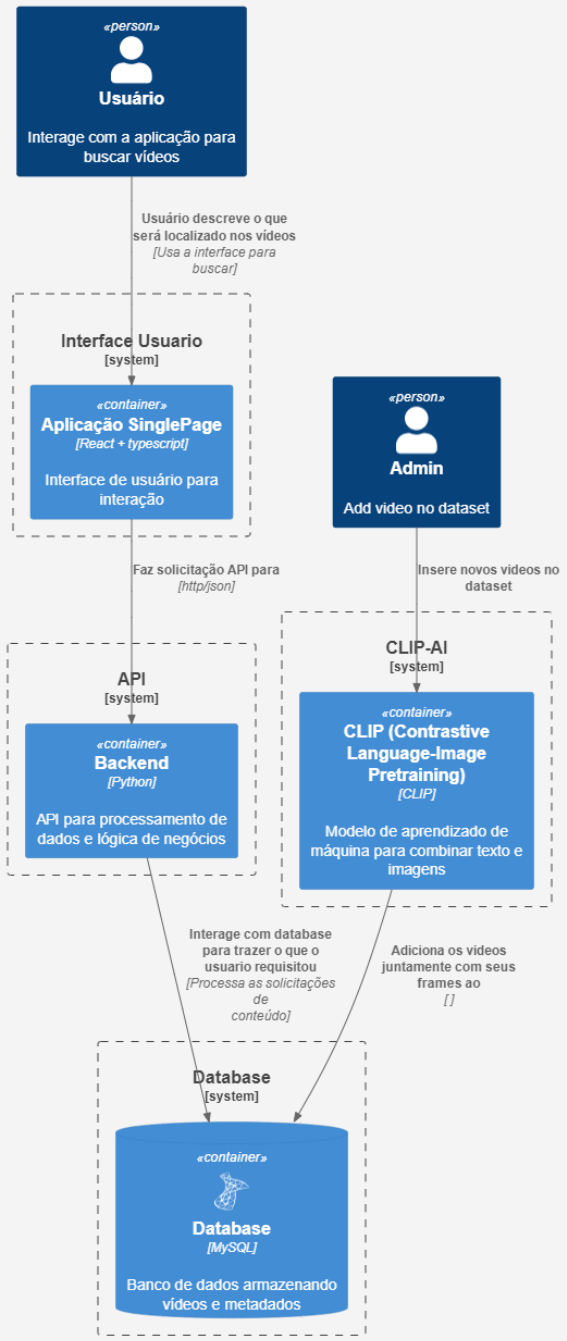
4.1. Diagrama de Contexto



4.2. Diagrama de Container



4.3. Diagrama de Componente



5. References

References

[1] Facebook Research, *FAISS*, GitHub, disponível em: https://github.com/facebookresearch/faiss.

[2] FAISS, *Official FAISS website*, disponível em: https://faiss.ai/.

[3] Zilliz Learn, *Information Retrieval Metrics*, Medium, disponível em: https://medium.com/@zilliz_learn/information-retrieval-metrics-0b50ffc5873b#:~:text=Information%20Retrieval%20(IR)%20systems%20are,relevant%20information%20using%20ranking%20metrics, acesso em: Fev. 2025.