

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Kaique Vandersar Ribeiro Lopes

Recuperação da informação em vídeos do YouTube

Uberlândia, Brasil

2022

UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Kaique Vandesar Ribeiro Lopes

Recuperação da informação em vídeos do YouTube

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Sistemas de Informação.

Orientador: Prof. Dr. Wendel Alexandre Xavier de Melo

Universidade Federal de Uberlândia – UFU

Faculdade de Ciência da Computação

Bacharelado em Sistemas de Informação

Uberlândia, Brasil

2022

Kaique Vandesar Ribeiro Lopes

Recuperação da informação em vídeos do YouTube

Trabalho de conclusão de curso apresentado à Faculdade de Computação da Universidade Federal de Uberlândia, Minas Gerais, como requisito exigido parcial à obtenção do grau de Bacharel em Sistemas de Informação.

Trabalho aprovado. Uberlândia, Brasil, 26 de maio de 2022:

**Prof. Dr. Wendel Alexandre Xavier de
Melo**
Orientador

Professor

Professor

Uberlândia, Brasil
2022

Resumo

Este trabalho auxilia os usuários fornecendo uma ferramenta simples, visual, interativa e dinâmica para facilitar o processo de busca de informações em vídeos. O desenvolvimento foi dividido em levantamento de requisitos, estudo das tecnologias necessárias para implantação do sistema e desenvolvimento do sistema. O projeto resultou em uma aplicação web feita com JavaScript e Python com auxílio das ferramentas Bootstrap, Flask e IBM Watson, sendo um sistema para realizar busca de informações em vídeos. Desta maneira, o sistema oferece a possibilidade de adicionar novos vídeos à base de dados e realizar consultas a fim de recuperar documentos relevantes. A ferramenta poderá ser utilizada por qualquer pessoa que queira realizar busca de informação em vídeos.

Palavras-chave: Ferramenta Web, Buscador de vídeos, Busca.

Lista de ilustrações

Figura 1 – Processo de Indexação	11
Figura 2 – Modelo Básico do Processo de Recuperação	12
Figura 3 – Funcionamento da ferramenta	22
Figura 4 – Tela de cadastro de novo canal	23
Figura 5 – Tela de busca	24
Figura 6 – Tela de Inclusão de canal preenchida	25
Figura 7 – Tela de processamento de inclusão de canal	26
Figura 8 – Tela de Inclusão de canal finalizada com sucesso	26
Figura 9 – Tela de Pesquisa	27
Figura 10 – Tela de Pesquisa em processamento	27
Figura 11 – Tela de Pesquisa com exibição de resultado	28
Figura 12 – Tela de Pesquisa com exibição do texto do video	29
Figura 13 – Tela de Pesquisa com exibição de múltiplos vídeos como resposta	30

Lista de abreviaturas e siglas

HTML Hyper Text Markup Language

CSS Cascading Style Sheets

Sumário

1	INTRODUÇÃO	7
1.1	Objetivos	8
1.1.1	Objetivos Específicos	8
1.2	Metodologia	8
2	FUNDAMENTAÇÃO TEÓRICA	9
2.1	Recuperação de informação	9
2.1.1	Processo de Recuperação da Informação	9
2.2	Modelos de RI	10
2.2.1	Modelo Vetorial	12
2.2.2	TF-IDF	13
2.3	Processamento de Língua Natural	14
2.3.1	Normalização	14
2.3.2	Remoção de <i>Stopwords</i>	14
2.3.3	Correção Ortográfica	14
2.3.4	<i>Stemização</i> e Lematização	15
2.4	Computação Cognitiva	15
2.4.1	Speech to text	16
2.5	Trabalhos Correlatos	17
3	DESENVOLVIMENTO	19
3.1	Levantamento de Requisitos	19
3.1.1	Requisitos Funcionais	19
3.1.2	Requisitos Não Funcionais	20
3.2	Tecnologias necessárias para implementação do sistema	20
3.3	Funcionamento	20
4	RESULTADOS	25
5	CONCLUSÃO	31
5.1	Trabalhos futuros	32
	REFERÊNCIAS	33

1 Introdução

A internet vem se popularizando cada vez mais, e, com isso, o acesso e a obtenção de informação por meio digital também se tornam acessíveis à maior parte da população. Há mais de 4,95 bilhões de pessoas que fazem o uso da internet (DATAREPORTAL, 2022) buscando vários conteúdos, desde lazer à busca por conhecimento em assuntos diversos. Os livros e enciclopédias foram trocados pelos artigos e wikis, as discussões físicas foram trocadas pelo espaço virtual com o uso de redes sociais e fóruns, os cursos e aulas já estão migrando para serviços de streaming. Além disso, a geração de conteúdo está mais acessível. Qualquer pessoa com acesso a internet consegue consumir e realizar a criação de conteúdo em diversos formatos, como áudio, vídeo, imagens, textos, entre outros.

Existem diversas plataformas onde é possível a geração e o consumo de conteúdo, alguns mais conhecidos como Twitter¹, Facebook², Instagram³, Twitch⁴ e YouTube⁵. O Youtube por exemplo, é o segundo maior mecanismo de pesquisa da internet, e é acessado por mais de 2,3 bilhões de pessoas em todo o mundo ao menos uma vez por mês (INSIGHT, 2022).

Todos os dias, as pessoas assistem a mais de um bilhão de horas de vídeo e geram bilhões de visualizações (INSIGHT, 2022). O YouTube conta com cerca de 38 milhões de canais ativos, e além disso, é a plataforma social mais utilizada no mundo para fins de pesquisa entre os tomadores de decisão *Business-to-Business* com 50,9% de usuários (INSIGHT, 2022).

Com um número tão grande de mídias, faz-se necessário que ferramentas automatizadas auxiliem os usuários na busca de informações que são relevantes aos seus interesses. Desse modo surgiram os buscadores como o Google⁶ e o Bing⁷. Além destes, surgem cada vez mais buscadores específicos para determinado conteúdo, como Google Shopping⁸ e Buscapé⁹ que são buscadores voltados para compras. Em muitos casos, há dificuldades marcantes em encontrar vídeos que atendam uma demanda por informação, especialmente quando o título, descrição e palavras-chave não estão adequadamente preenchidos. Não é incomum usuários precisarem assistir longos vídeos para, ao final, se darem conta de que o conteúdo buscado não se encontra naquele material, o que pode lhes tomar demasiado

¹ <<https://twitter.com/>>

² <<https://www.facebook.com/>>

³ <<https://www.instagram.com/>>

⁴ <<https://www.twitch.tv/>>

⁵ <<https://www.youtube.com/>>

⁶ <<https://www.google.com/>>

⁷ <<https://www.bing.com/>>

⁸ <<https://shopping.google.com.br/>>

⁹ <<https://www.buscape.com.br/>>

tempo. Dessa forma, para auxiliar os usuários na busca de informações em vídeos, buscadores de vídeos que levem em consideração o conteúdo do vídeo se fazem necessários. Visando preencher essa lacuna, desenvolvemos neste trabalho uma ferramenta que indexa o conteúdo de vídeos de modo a fornecer suporte para que usuários possam encontrar bons materiais que atendam sua requisição por informação.

1.1 Objetivos

Este trabalho tem por objetivo apresentar uma ferramenta que possibilita a recuperação de informação em vídeos de acordo com seu conteúdo. Espera-se que a ferramenta forneça suporte ao usuário para que o mesmo localize vídeos de interesse de forma assertiva, evitando que seja necessário assistir conteúdo que não atenda suas requisições por informação.

1.1.1 Objetivos Específicos

- Indexar vídeos pelo seu conteúdo através da extração e transcrição de áudio;
- Permitir que usuários localizem vídeos que efetivamente tratem suas requisições de informação de forma mais assertiva através de todo o seu conteúdo, e não apenas do título e da descrição fornecidos pelo cadastrador do vídeo.
- Possibilitar ao usuário visualizar a transcrição completa do vídeo, visto que muitos preferem ler conteúdo textual a ouvir o áudio do vídeo, já que, em geral, este último é mais demorado.

1.2 Metodologia

Este trabalho utilizou as linguagens de programação Python e Javascript em conjunto com o *framework* Flask para a exposição das APIs e o GitHub para versionamento do código fonte. Também foi usado o IBM Watson para realizar a transcrição do áudio. Para a implementação do site foram usadas as tecnologias CSS3, HTML5 e a biblioteca Bootstrap 4. O sistema possibilita o cadastro de novos canais que terão os vídeos baixados, e seu áudio extraído. Após a extração, ocorre a transcrição do áudio, onde ele é transformado em um texto que é usado para realizar a indexação utilizando um modelo de recuperação de informação. Uma interface fica responsável por receber as consultas e respondê-las de acordo com um modelo de Recuperação de Informação.

2 Fundamentação Teórica

2.1 Recuperação de informação

A Recuperação da Informação (RI) é uma área da Ciência da Computação que lida com o armazenamento de documentos e a recuperação automática de informação contida neles provendo aos usuários o acesso fácil às informações de seu interesse. Atualmente a pesquisa em RI inclui modelagem, classificação de textos, arquitetura de sistemas, interfaces de usuário, visualização de dados, filtragem e linguagens (BAEZA-YATES, 2013).

A área pode ser estudada sob dois pontos de vista distintos e complementares: um centrado no computador e outro centrado no usuário. Na visão centrada no computador, a RI consiste principalmente na construção de índices eficientes, no processamento de consultas e no desenvolvimento de algoritmos de ranqueamento a fim de obter o melhor resultado. Na visão centrada no usuário, a RI consiste principalmente no estudo do comportamento do usuário, o entendimento de suas principais necessidades e como estas afetam a organização e a operação do sistema de recuperação (BAEZA-YATES, 2013).

A recuperação da informação engloba o processo de encontrar uma informação desejada em uma base de dados, ou seja, os sistemas de recuperação da informação devem oferecer recursos para facilitar a busca por essas informações. O usuário de um sistema de recuperação de informação está mais interessado em recuperar informação sobre um determinado assunto do que dados ou documentos que atendam sua expressão de busca.

A eficácia de um sistema de recuperação da informação está diretamente ligada, dentre outros fatores, a qualidade da análise dos documentos e do modelo de recuperação adotado.

2.1.1 Processo de Recuperação da Informação

O processo de Recuperação da Informação em um sistema de buscas consiste em basicamente três passos: indexar, armazenar e recuperar. O propósito principal da indexação é a representação dos conteúdos contidos nos documentos, para fins de recuperação da informação (LANCASTER, 2004). Já o armazenamento e a recuperação são etapas que devem estar sincronizadas com qualquer modelo de busca e recuperação de informação para satisfazer a necessidade de busca de informação por parte do usuário. Geralmente o texto não é armazenado por completo, para cada texto, são criadas estruturas de dados com o objetivo de acelerar a sua recuperação (BAEZA-YATES, 1999).

É importante distinguir a recuperação baseada em conteúdo do baseado em meta-

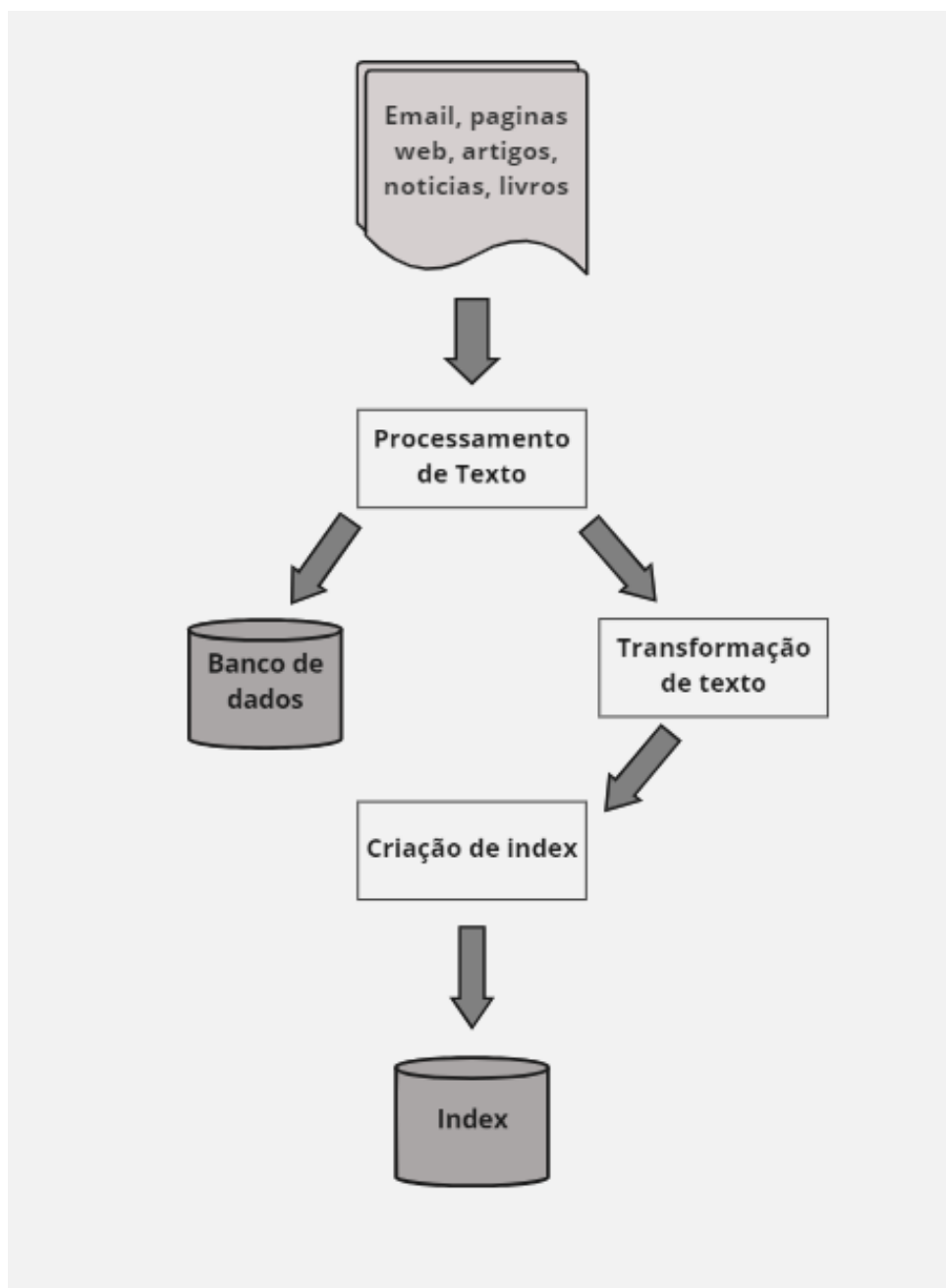
dados, como o título de um vídeo, música ou áudio. A recuperação de metadados é viável para arquivos estruturados com bancos de dados, como catálogos de músicas ou filmes. No entanto, a recuperação baseada em conteúdo é mais importante em dados gerados pelo usuário, que muitas vezes tem pouca estrutura ou metadados (ONCESCU et al., 2021).

Com o avanço da tecnologia, ferramentas e algoritmos na área da computação, armazenar e recuperar informações se torna uma tarefa mais eficiente, por haver recursos automatizados que permitem maior rapidez para indexar, armazenar e recuperar informações. Todavia, novos desafios são gerados a medida em que se aumenta o volume de dados sendo gerados e tratados.

2.2 Modelos de RI

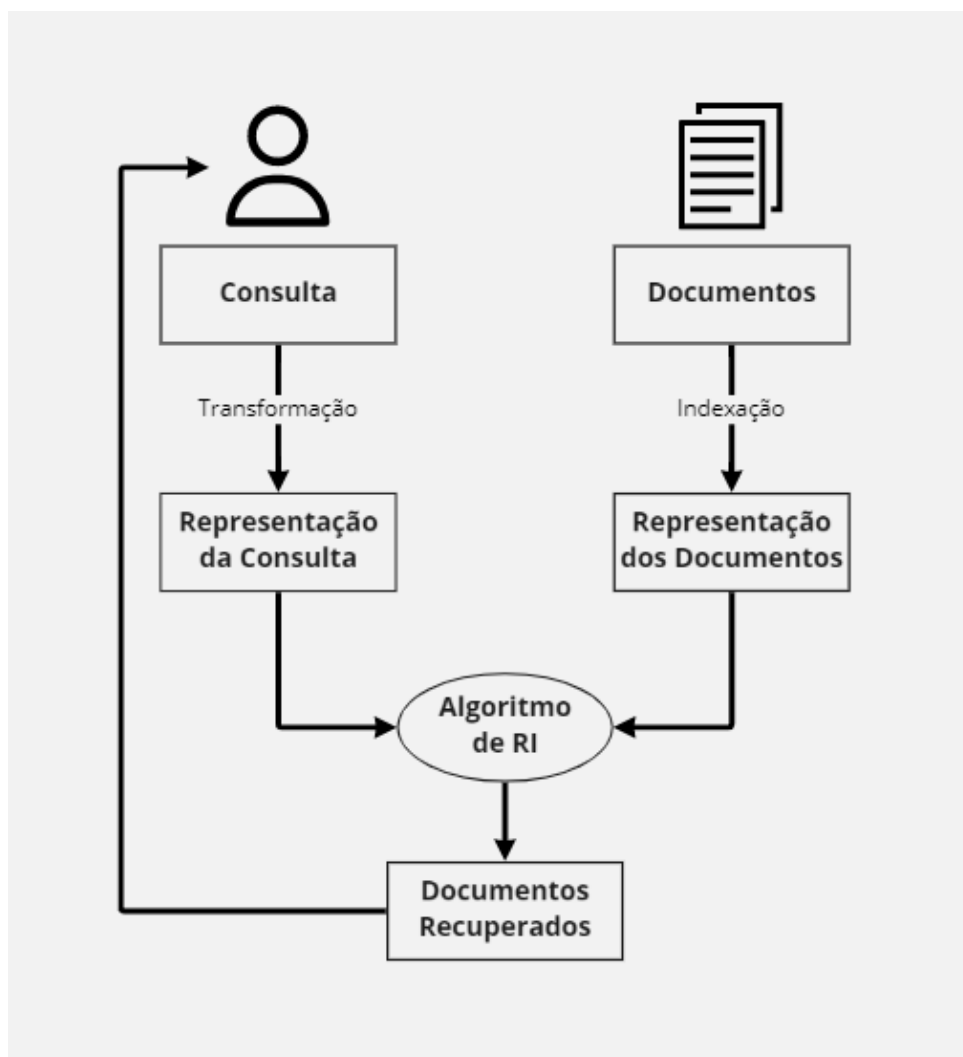
A modelagem em RI é um processo que busca produzir uma função de ranqueamento para atribuir escores a documentos em relação a uma consulta. Esse ranqueamento possivelmente é o processo mais importante em um sistema de RI, pois é através dele que serão escolhidos os documentos que satisfazem à consulta do usuário. Em sistemas de RI, geralmente são usados termos de indexação para indexar e recuperar documentos. Termos de indexação são palavras-chave (ou um grupo de palavras-chaves relacionadas) que tem significado próprio, ou seja, que normalmente tem o mesmo papel de um substantivo. Palavras em outras classificações como verbos e adjetivos, entre outros, também podem ser usadas como termo de indexação. De modo mais geral, qualquer palavra que apareça no texto do documento pode ser um termo de indexação (BAEZA-YATES, 2013). A Figura 1 mostra um exemplo de como ocorre o processo de indexação, enquanto a Figura 2 exemplifica um modelo básico de recuperação da informação.

Figura 1 – Processo de Indexação



Adaptado de: (CROFT DONALD METZLER, 2010)

Figura 2 – Modelo Básico do Processo de Recuperação



Adaptado de: (CROFT DONALD METZLER, 2010)

A eficiência do ranqueamento depende dos índices e a eficácia depende do modelo de recuperação (CROFT DONALD METZLER, 2010), ou seja, a execução em conjunto da indexação e do ranqueamento que ditam a qualidade da recuperação da informação.

2.2.1 Modelo Vetorial

O modelo vetorial (*Vector Space Model*) propõe um ambiente no qual é possível obter documentos que respondem parcialmente a uma expressão de consulta, usada para realizar a busca. Isto é feito associando pesos aos termos de indexação dos documentos e aos termos de indexação utilizados na expressão de consulta. Esses pesos são utilizados para formar vetores de pesos que então passam a representar documentos e consultas. O grau de similaridade entre um documento e uma consulta é calculado de acordo com a similaridade entre seus respectivos vetores (SALTON; WONG; YANG, 1975).

2.2.2 TF-IDF

TF-IDF (abreviação do inglês *Term Frequency–Inverse Document Frequency*, que significa frequência do termo–inverso da frequência nos documentos), é uma medida estatística que avalia a relevância de uma palavra para um documento em uma coleção de documentos (RAMOS et al., 2003).

Isso é feito multiplicando-se duas métricas: uma delas leva em conta a frequência de um termo de indexação em um documento (TF), e a outra leva em conta a frequência inversa do número de documentos que contém um determinado termo de indexação (IDF) (RAMOS et al., 2003).

- O TF (*Term Frequency*) de uma palavra em um documento pode ser calculado de varias maneiras, sendo uma das mais simples uma contagem bruta de ocorrências em que uma palavra aparece em um documento (QAISER; ALI, 2018). Assim, a ponderação TF mede a importância de um termo k em um documento d pelo número de aparições deste k em d .
- O IDF (*Inverse Document Frequency*) indica o quão comum ou rara é uma palavra em todo o conjunto de documentos. Quanto mais próximo de 0, mais comum é uma palavra. O IDF de um termo k pode ser calculado através do logaritmo da divisão do numero total de documentos pelo numero de documentos que contém k (QAISER; ALI, 2018):

$$idf(k_i) = \log\left(\frac{N}{n_i}\right)$$

onde:

- N : número total de documentos
- n_i : número de documentos com o termo k_i

Portanto, se a palavra for muito comum e aparecer em muitos documentos, esse número se aproximará de 0. Caso contrário, se aproximará de 1.

A multiplicação dos resultados resulta na pontuação TF-IDF de uma palavra em um documento. Quanto maior a pontuação, mais relevante é a respectiva palavra naquele documento específico. Assim, a ponderação TF-IDF, ao mesclar as duas métricas, mede a importância de um termo k em um documento d pela frequência de k em d pela raridade de k na base. A ponderação TF-IDF tem muitos usos, principalmente na análise automatizada de texto, e é muito útil para pontuar palavras em algoritmos de aprendizado de máquina para Processamento de Língua Natural (PLN).

2.3 Processamento de Língua Natural

O Processamento da Linguagem Natural (PLN) trata os aspectos da comunicação humana como sons, discursos, palavras e sentenças com o objetivo de encontrar métodos de interpretar o dialeto de maneira computacional. Em resumo, o PLN propõe-se a compreender a linguagem humana através do processamento em níveis diferentes de entendimento: fonético e fonológico, morfológico, sintático, semântico e pragmático (GONZALEZ, 2003).

Em PLN, há alguns artifícios que são importantes para a manipulação de textos tais como Normalização, Remoção de *Stopwords* e Lematização. Esse pré-processamento faz com que o vocabulário seja reduzido, tornando os dados menos esparsos, característica conveniente para o processamento computacional.

2.3.1 Normalização

A normalização engloba tratativas como a tokenização, remoção de caracteres especiais, transformação de letras maiúsculas para minúsculas, remoção de tags HTML, Javascript, CSS, dentre outras. O processo de tokenização tem como objetivo separar palavras ou sentenças em unidades. A tokenização lexical marca cada palavra como um token no texto, identificando-a mesmo se estiver encostada em alguma pontuação.

2.3.2 Remoção de *Stopwords*

Esse método muito usado no pré-processamento de textos consiste em remover palavras vazias de significado, tais como “a”, “de”, “o”, “da”, “que”, “e”, “do” entre outras, pois em sua maioria são palavras que não possuem significância quando isoladas, portanto, frequentemente não fornecem informações relevantes para a construção do texto. As *Stopwords* devem ser removidas somente quando não forem relevantes para a tarefa almejada. Em um contexto de análise de sentimentos, por exemplo, não poderíamos remover a stopword “não”, pois traz uma conotação de negatividade para a sentença, indicando justamente o sentimento transmitido. Listas de *Stopwords* podem ser facilmente encontradas na internet.

2.3.3 Correção Ortográfica

Os corretores ortográficos (*Spell Checkers*) são muito usuais para tratar um *dataset* que contém erros de digitação, abreviações e vocabulário informal. Esses erros são prejudiciais por gerarem novos *tokens*, aumentando a esparsidade dos dados.

2.3.4 Stemização e Lematização

O processo de *stemização* consiste em reduzir as palavras ao seu radical ou forma-raiz. A palavra “meninas” se reduziria a “menin”, assim como “meninos” e “meninhos”. As palavras “gato”, “gata”, “gatos” e “gatas” reduziriam-se para “gat”. Já a lematização reduz a palavra à forma no masculino singular. No caso de verbos, o lema é o infinitivo. Por exemplo, as palavras “gato”, “gata”, “gatos” e “gatas” são todas variações do mesmo lema: “gato”. A vantagem de se aplicar a *Stemização* e Lematização é a redução de vocabulário e abstração de significado de termos que, em essência, possuem o mesmo significado com pouca variação entre eles.

2.4 Computação Cognitiva

As tecnologias cognitivas são uma evolução na computação que imita alguns aspectos dos processos de pensamento humano em uma escala maior. Nesse caso, escala refere-se à capacidade de processar os volumes de dados e informações disponíveis no domínio científico. Os desenvolvedores de tecnologia perceberam que o raciocínio humano, o aprendizado e a inferência constituem um dos sistemas de pensamento mais sofisticados que existem (DHARMENDRA et al., 2011). Ainda assim, a cognição humana tem limitações, das quais duas incluem escalabilidade e viés (LEHNE et al., 2015). Os sistemas cognitivos tentam imitar aspectos do pensamento humano enquanto adicionam a capacidade de lidar com grandes quantidades de informações e avaliá-las sem viés.

Na comunidade de computação, a definição de computação cognitiva é frequentemente associada à Inteligência Artificial (IA), um campo da tecnologia que abrange amplos aspectos da inteligência humana (VELDE, 2010). A IA inclui não só as habilidades relacionadas ao raciocínio e resolução de problemas, mas também à percepção (reconhecimento facial e visão) e à capacidade de manipular objetos (robótica) (VELDE, 2010).

Para entender como a computação cognitiva funciona, é útil comparar e contrastar como os seres humanos e as tecnologias cognitivas se envolvem na descoberta e em várias formas de processos de tomada de decisão. Uma maneira de descrever isto é através de processos como a observação, interpretação, avaliação e decisão (CHEN; ARGENTINIS; WEBER, 2016).

A observação de dados refere-se à agregação, integração e exame de dados como base para avaliação e descoberta. Os humanos observam através de diferentes canais sensoriais, como ler publicações relevantes ou ouvir outras pessoas. Os humanos também costumam ter uma base pré-existente de informações obtidas por meio de sua própria observação, educação e experiências de vida. Essas observações são retidas na memória como parte de uma base de conhecimento mais ampla (CHEN; ARGENTINIS; WEBER,

2016).

Para fazer observações, uma solução cognitiva requer acesso a volumes de dados. A identificação, compra, licenciamento e normalização de dados devem ser coordenados. Com um sistema de computação cognitiva, centenas de fontes de conteúdo externas, públicas, licenciadas e privadas, que podem conter dados relevantes são agregadas. No caso do *Watson* (plataforma de serviços cognitivos da IBM para negócios), a IBM agrega esses dados em um único repositório chamado *Watson corpus*. Um corpus *Watson* exclusivo é estabelecido para cada domínio ao qual o *Watson* é aplicado. Portanto, em direito, medicina, engenharia e finanças, um corpus *Watson* personalizado pode ser criado com conjuntos de dados e conteúdo relevantes para cada um desses domínios. O conteúdo é normalizado e limpo em um conjunto de dados formatado que pode ser usado para análise (HIGH, 2017).

A interpretação implica na capacidade de compreender os dados, neste caso, a linguagem, além das definições de termos individuais, para deduzir o significado de frases e parágrafos. Como seres humanos, absorvemos conteúdo e informação. Lemos e reconhecemos palavras, traduzindo-as em significado abstrato. Por exemplo, um químico reconhecerá compostos de artigos publicados ou patentes e criará uma representação mental de compostos relacionados e os recursos que os definem (CHEN; ARGENTINIS; WEBER, 2016).

Da mesma forma, um componente-chave de um sistema cognitivo envolve aprender a linguagem de uma indústria ou domínio específico. Para permitir a compreensão da linguagem, um sistema deve ser fornecido com dicionários e tesouros relevantes. Estes podem incluir uma lista conhecida de nomes de genes humanos ou nomes químicos, mas também incluem os verbos que criam relações entre eles, como “expressar ou inibir”. Compreender os verbos, substantivos e preposições em cada frase torna os sistemas cognitivos diferentes da pesquisa de palavras-chave e da análise de texto que pode identificar apenas os substantivos de interesse ou confiar na correspondência de palavras individuais para encontrar informações relevantes. A capacidade de entender verbos, adjetivos e preposições permite a compreensão do que a linguagem significa em oposição a apenas o que ela diz (VELDE, 2010).

2.4.1 Speech to text

O serviço *Speech to Text* converte a fala em texto legível de acordo com o idioma especificado pelo usuário. Existem várias plataformas que oferecem esse tipo de serviço, e uma delas é o IBM Watson. O serviço é capaz de transcrever a fala de vários idiomas e formatos de áudio para texto com baixa latência (SANTIAGO PALLAVI, 2012). Este serviço pode ser usado em aplicações práticas como:

- Transcrever chamadas em um call center para identificar o que está sendo discutido quando encaminhar chamadas e entender o conteúdo de vários oradores.
- Transcrever uma explicação técnica fornecida por um especialista sobre algum assunto para ajudar a criar a documentação sobre o mesmo.
- Transcrever a fala de um filme para criar legendas.
- Transcrever mensagens de correio de voz para obter um registro escrito das mensagens deixadas, por exemplo, por clientes ou pacientes.
- Criar aplicativos de controle de voz, por exemplo, um sistema automatizado de controle doméstico baseado em fala. As funções podem incluir ligar e desligar a luz, controlar dispositivos eletrônicos ou até mesmo abrir a porta remotamente.

2.5 Trabalhos Correlatos

Este trabalho se relaciona com temas da literatura como legendagem de áudio, recuperação com base em áudio, recuperação com base em texto, entre outros. A legendagem de áudio consiste em gerar uma descrição de texto para um som (DROSSOS, 2017). Isso requer uma análise mais detalhada e compreendida do som do que simplesmente mapear o som para um conjunto de rótulos. Vários conjuntos de dados de legendas já foram introduzidos em áudios usando ferramentas de geração automática, como o CLOTHO (DROSSOS, 2020) que foi usado na legendagem de áudio automatizada na DCASE challenge 2020 (DCASE, 2020), Audio Caption (WU, 2019) e AUDIOCAPS (KIM et al., 2019).

A recuperação com base em áudio aproveita a semelhança de características de som que representam diferentes aspectos de sons (por exemplo, tom ou volume) (FOOTE, 1997). Ultimamente tem sido cada vez mais comum o uso de estruturas de rede neural siamesa para aprender a identificar áudio semanticamente semelhante. O trabalho (JIN et al., 2012) aborda a detecção de arquivos multimídia usando apenas dados de áudio, enquanto o trabalho (AVGOUSTINAKIS et al., 2021) aborda a recuperação de vídeos duplicados por recuperação de áudio. Estes são métodos puramente baseados em áudio que são aplicados a conjuntos de dados de vídeo, mas sem usar informações visuais. (HOU; ZHOU, 2013) propõem uma abordagem de duas etapas para recuperação de vídeo que usa informações de áudio (grossas) e visuais (finas) juntas.

O trabalho (RILEY; HEINEN; GHOSH, 2008) apresenta uma abordagem para recuperação baseada no conteúdo de músicas digitais, que consiste em representar as músicas como *"Bagof-Audio-Words"* (Resultado de um processamento do áudio que consiste em etapas de processamento, segmentação, extração do áudio, quantização vetorial

e construção de histograma.) e realizar os cálculos de similaridade diretamente no modelo vetorial de termos (SALTON; WONG; YANG, 1975).

Já (ONCESCU et al., 2021) propõe um *framework* que permite a consulta de um banco de dados de áudio de forma livre, realizando a descrição dos áudios em linguagem natural (por exemplo, “Um homem falando enquanto a música está tocando seguido por um coaxar de um sapo.”), possibilitando a recuperação de dados de áudio que correspondem a sequência de eventos na consulta. Além disso, as consultas em linguagem natural são familiares em interfaces amplamente utilizadas nos motores de busca atuais.

3 Desenvolvimento

Este capítulo apresenta as etapas realizadas para a construção deste trabalho, divididas em levantamentos de requisitos, estudo de tecnologias necessárias e desenvolvimento da ferramenta.

3.1 Levantamento de Requisitos

Existem dois tipos principais de requisitos de sistema que devem ser coletados por aqueles que trabalham em projetos de *software*. Os requisitos do sistema podem ser classificados como requisitos funcionais ou não funcionais (SOMERVILLE, 2019). Compreender a diferença entre os dois ajuda a garantir que os desenvolvedores entregarão um produto com o desempenho esperado.

Os requisitos funcionais podem ser definidos de forma mais simples como algo que o sistema deve fazer (SOMERVILLE, 2019). Se o sistema não atender a um requisito funcional, ele falhará. Isso ocorre porque ele não será capaz de alcançar algo que deve fazer para operar corretamente. O conceito de requisito funcional também pode ser entendido através da revisão do sistema em termos de entradas e saídas, já que os requisitos funcionais especificam qual resposta o sistema deve produzir à diferentes entradas (SOMERVILLE, 2019).

Já os requisitos não funcionais em engenharia de *software* podem ser explicados como requisitos que descrevem como o sistema funciona (SOMERVILLE, 2019). Os requisitos não funcionais são focados em como o sistema entrega uma função específica. À primeira vista, eles podem ser vistos como menos importantes do que os requisitos funcionais, mas ambos têm um papel fundamental na construção de um bom sistema. Os requisitos não funcionais não afetam a funcionalidade do sistema, mas sim o desempenho dele (SOMERVILLE, 2019). Em suma, os requisitos não funcionais estão ligados à usabilidade do sistema. Se os requisitos não funcionais não forem atendidos, os usuários podem ficar frustrados com o seu funcionamento e adotarem um sistema concorrente.

3.1.1 Requisitos Funcionais

- RF-1: Possibilitar a inclusão de novos canais de vídeo do Youtube na base de dados

Nos dias atuais, temos observado uma crescente disponibilização de conteúdo de vídeo na internet, fazendo com que seja necessário uma maneira para que um usuário com a devida permissão possa apontar canais onde será realizada indexação e onde a busca será feita.

- RF-2: Possibilitar a realização de consultas à base de dados

Para possibilitar a realização de consultas à base de dados, é necessário que o sistema apresente uma interface onde o usuário possa inserir a expressão de consulta que deseja fazer na base de dados.

- RF-3: Exibir os resultados ranqueados

Quando realizada a consulta, o sistema deve mostrar, de maneira ordenada, os itens que possuem maior similaridade com a consulta.

3.1.2 Requisitos Não Funcionais

- RNF-1: Plataforma Web

Com o intuito de evitar a dependência de um sistema operacional específico e possibilitar um maior alcance de usuários, foi definido que a plataforma a ser utilizada deverá ser a plataforma Web.

- RNF-2: Uso de Design responsivo nas interfaces gráficas

A interface do sistema deverá se comportar adequadamente independente do front-end que será utilizado para acesso – Browser, Smartphone ou Tablet.

3.2 Tecnologias necessárias para implementação do sistema

Foi utilizado a linguagem de programação Python para realizar a criação do *Backend*, junto com o *framework* Flask. Além destas, também foi usado HTML5, Javascript e Bootstrap 4, um dos mais famosos *frameworks* para web criado para facilitar o desenvolvimento de uma interface responsiva e amigável para o usuário (MORETO, 2016).

A partir disso, nesta etapa foram estudadas tecnologias para realizar a obtenção dos links do canal do Youtube e para realizar o *download* e extração do áudio do vídeo.

3.3 Funcionamento

O sistema possui dois perfis de usuários, administradores do sistema, que poderão cadastrar os canais, e usuários ordinários, que apenas realizarão consultas. Para realizar a adição de novos documentos à base de dados e realizar as consultas para responder as buscas dos usuários, a ferramenta realiza algumas ações. A primeira é realizar a obtenção

dos links dos vídeos do canal do YouTube. Para realizar tal tarefa foi usado a Youtube Data API¹. Esta API oferece várias funcionalidades como enviar vídeos, gerenciar lista de reprodução e inscrições, atualizar as configurações do canal e muito mais, Entre os recursos oferecidos pela API, está também a possibilidade de listar os últimos vídeos de um canal, e esta funcionalidade é usada para realizar a coleta das URLs dos vídeos. Após coletado as URLs, o próximo passo é realizar o *download* dos vídeos, e, para isso, foi usado o *youtube-dl*² que é um gerenciador de *download* de código aberto para vídeo e áudio do YouTube e de outros sites de hospedagem de vídeo. Após realizar o download do vídeo, o próximo passo é realizar a extração do seu áudio, e para isto foi usado a ferramenta *FFmpeg*³ que é um *software* multiplataforma usada para gravar, converter e transmitir áudio e vídeo. Após ser realizado a extração do áudio, ele é enviado para a ferramenta *Speech To Text*⁴ do Watson, que é uma ferramenta usada para transcrever áudio para texto em diferentes idiomas com o reconhecimento e a transcrição de fala desenvolvidos com IA. Após o processo de transcrição ser finalizado, o resultado é salvo em um arquivo de texto que passa a compor a base de dados da aplicação de busca, como é mostrado na Figura 3.

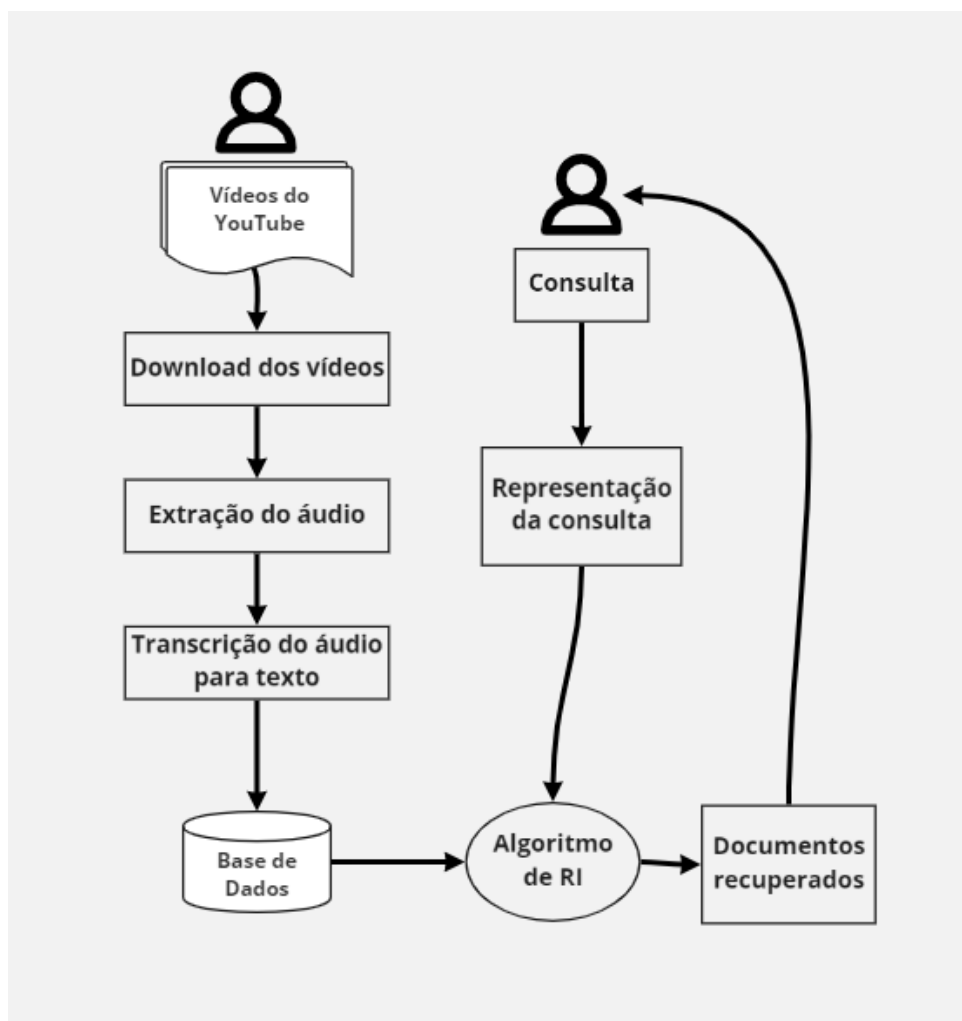
¹ <<https://developers.google.com/youtube/v3>>

² <<http://yt-dl-org.github.io/youtube-dl/about.html>>

³ <<https://ffmpeg.org/>>

⁴ <<https://www.ibm.com/br-pt/cloud/watson-speech-to-text>>

Figura 3 – Funcionamento da ferramenta



Fonte: próprio autor

A entrada de dados é realizado pelo usuário através de uma página *Web* dividida em duas seções. A Figura 4 mostra a primeira tela, onde o usuário faz a inclusão de novos canais e vídeos à base de dados. Para realizar a inclusão e usuário precisa fornecer no campo “Id do canal” o Id (identificador) do canal que deseja incluir e logo abaixo informar o número de vídeos que deseja incluir. Após preencher estas informações, o usuário pode realizar a inclusão do canal à base de dados.

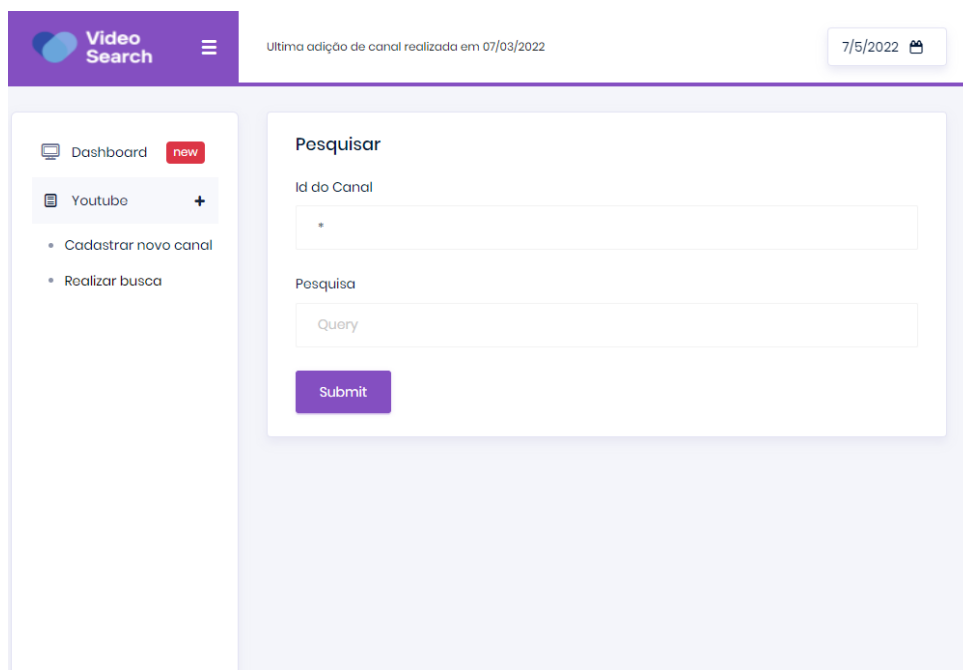
Figura 4 – Tela de cadastro de novo canal

The screenshot shows a web application interface for 'Video Search'. The top navigation bar is purple and contains the logo, a menu icon, the text 'Última adição de canal realizada em 07/05/2022', and a date '7/5/2022' with a calendar icon. On the left, a sidebar menu lists 'Dashboard' (with a 'new' badge), 'Youtube' (with a '+' icon), and two sub-items: 'Cadastrar novo canal' and 'Realizar busca'. The main content area is titled 'Adicionar Canal' and contains two input fields: 'Id do Canal' with the placeholder 'id' and 'Numero maximo de videos' with the value '1'. A purple 'Submit' button is located below the second input field.

Fonte: próprio autor

A Figura 5 mostra a segunda parte, que é onde os usuários conseguem realizar as buscas. Nesta etapa, o usuário pode inserir a pesquisa que deseja realizar na base de dados, e, após clicar no botão de realizar busca, uma requisição é enviada ao servidor *Back-End* que realiza o processamento da requisição e consulta a base de dados calculado à similaridade com base na ponderação TF-IDF e o método Similaridade de Cosseno (Modelo Vetorial). Após a requisição ser processada, é retornado para o *Front-End* o resultado contendo os documentos que tem maior grau de similaridade com a consulta recebida, e, a partir desta resposta, é apresentado na página *Web* os vídeos correspondentes.

Figura 5 – Tela de busca



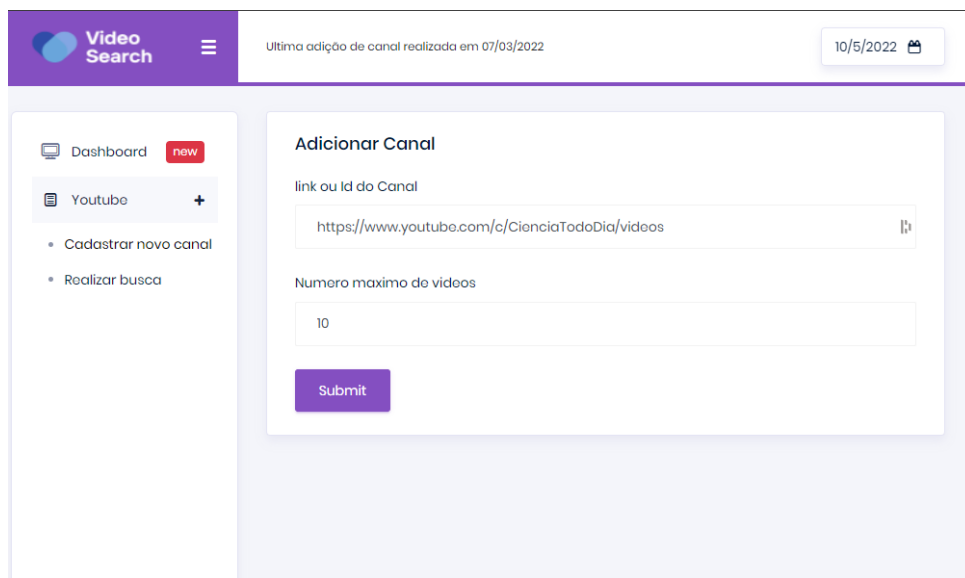
Fonte: próprio autor

4 Resultados

Neste capítulo serão apresentados os resultados do desenvolvimento do software através de capturas de tela do mesmo durante o seu funcionamento, realizando a inclusão de novos vídeos a base de dados e realizando consultas.

Na tela de inclusão de canais há dois campos. O primeiro campo é destinado para informar o endereço URL do canal do Youtube ou o id do canal. O segundo campo serve para informar a quantidade de vídeos do canal que devem ser adicionados à base de busca. Neste ensaio, foi realizada a inclusão de um canal de ciências à base de dados. A Figura 6 mostra os campos “link do canal” e “número de vídeos” preenchidos, enquanto a Figura 7 mostra a tela de carregamento que é apresentada enquanto o sistema realiza as ações necessárias para a inclusão dos vídeos à base de dados. Já a figura 8 mostra a tela que é apresentada caso todo o processo de inclusão ocorra com sucesso

Figura 6 – Tela de Inclusão de canal preenchida



A captura de tela mostra a interface de usuário 'Video Search'. No topo, há um cabeçalho com o logo 'Video Search' e um menu. À direita do cabeçalho, há uma barra de status que indica 'Ultima adição de canal realizada em 07/03/2022' e uma data '10/5/2022' com um ícone de calendário. O conteúdo principal é dividido em duas seções. À esquerda, há um menu lateral com opções: 'Dashboard' (com um botão 'new' em vermelho), 'Youtube' (com um botão '+'), 'Cadastrar novo canal' e 'Realizar busca'. À direita, há uma seção intitulada 'Adicionar Canal'. Esta seção contém dois campos de entrada: 'link ou Id do Canal' com o valor 'https://www.youtube.com/c/CienciaTodoDia/videos' e 'Numero maximo de videos' com o valor '10'. Abaixo dos campos, há um botão 'Submit' em um fundo roxo.

Fonte: próprio autor

Figura 7 – Tela de processamento de inclusão de canal

The screenshot shows the 'Adicionar Canal' (Add Channel) page in the Video Search application. The page has a purple header with the 'Video Search' logo and a navigation menu. The main content area is white and contains the following elements:

- Header:** 'Video Search' logo, a hamburger menu icon, and the text 'Ultima adição de canal realizada em 07/03/2022'. A date and notification icon are in the top right corner, showing '10/5/2022'.
- Left Sidebar:** A navigation menu with 'Dashboard' (marked 'new') and 'Youtube' (with a plus sign). Below it are two options: 'Cadastrar novo canal' and 'Realizar busca'.
- Main Content Area:**
 - Title:** 'Adicionar Canal'
 - Form:** A form with two input fields. The first is labeled 'link ou Id do Canal' and contains the URL 'https://www.youtube.com/c/CienciaTodoDia/videos'. The second is labeled 'Numero maximo de videos' and contains the number '1'.
 - Submit Button:** A purple button labeled 'Submit'.
 - Feedback:** Below the form, there is a message: 'Adicionando videos! Isto leva cerca de 3 minutos para cada video. Você pode continuar navegando normalmente enquanto os videos são adicionados'. Below this message is a green progress bar.

Fonte: próprio autor

Figura 8 – Tela de Inclusão de canal finalizada com sucesso

The screenshot shows the 'Adicionar Canal' (Add Channel) page in the Video Search application, indicating successful completion. The layout is identical to Figure 7, but with the following changes:

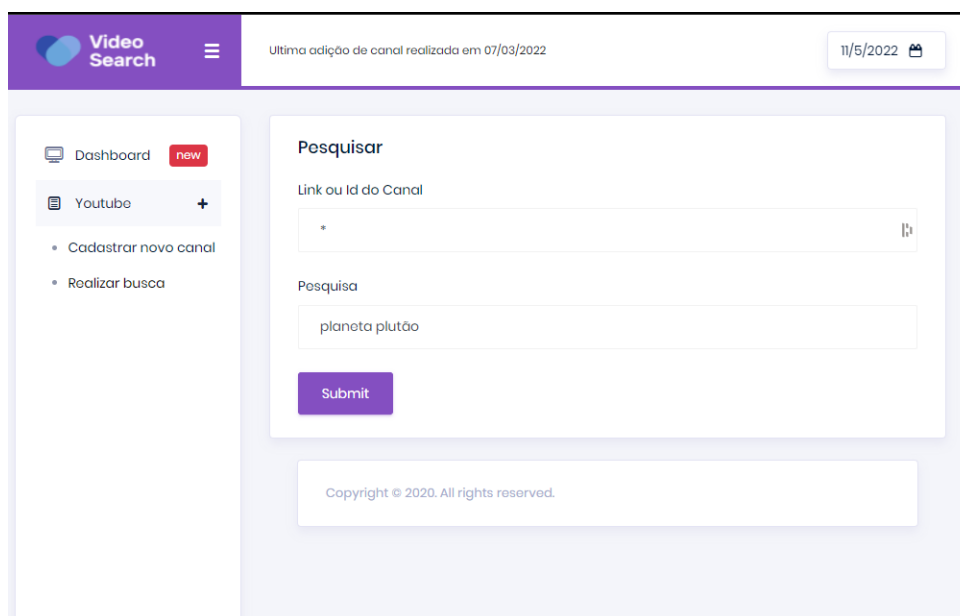
- Form:** The input fields are now empty. The 'link ou Id do Canal' field has a placeholder 'Link' and a copy icon. The 'Numero maximo de videos' field has a placeholder '1'.
- Submit Button:** The purple 'Submit' button is still present.
- Feedback:** A green message 'Canal Adicionado com sucesso' (Channel Added successfully) is displayed below the form.

Fonte: próprio autor

No próximo ensaio, foi realizado uma consulta usando o termo “planeta plutão”. Para realizar a busca, é necessário preencher dois campos, o primeiro campo “Link ou Id do Canal” deve ser usado para informar em qual canal do Youtube a busca deve ocorrer. Ao usar o caractere “*” neste campo, a busca ocorre em todos os canais que estão cadastrados na base de dados. O segundo campo é o campo “Pesquisa” que deve ser usado para informar a query a ser respondida. A Figura 9 mostra os campos “Link ou Id do Canal” e “Pesquisa” preenchidos. Já a Figura 10 mostra a tela que é apresentada

enquanto a busca está sendo processada e a Figura 11 mostra a tela que apresenta os resultados que respondem à pesquisa realizada. O modelo de recuperação de informação adotado casa a consulta com a transcrição dos áudios dos vídeos para obter os resultados.

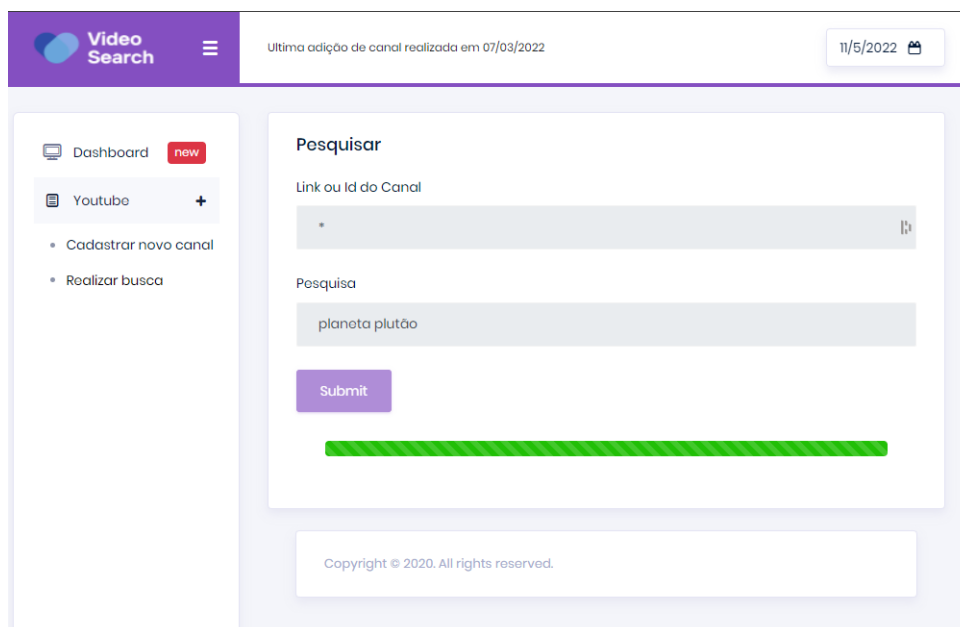
Figura 9 – Tela de Pesquisa



The screenshot shows the 'Video Search' application interface. At the top, there is a purple header with the logo and a navigation menu. Below the header, the main content area is divided into a left sidebar and a central search form. The sidebar contains a 'Dashboard' link with a 'new' badge and a 'Youtube' link with a '+' icon. Below these are two menu items: 'Cadastrar novo canal' and 'Realizar busca'. The central search form is titled 'Pesquisar' and contains a 'Link ou Id do Canal' input field with an asterisk, a 'Pesquisa' input field containing the text 'planeta plutão', and a purple 'Submit' button. At the bottom of the form, there is a copyright notice: 'Copyright © 2020. All rights reserved.'

Fonte: próprio autor

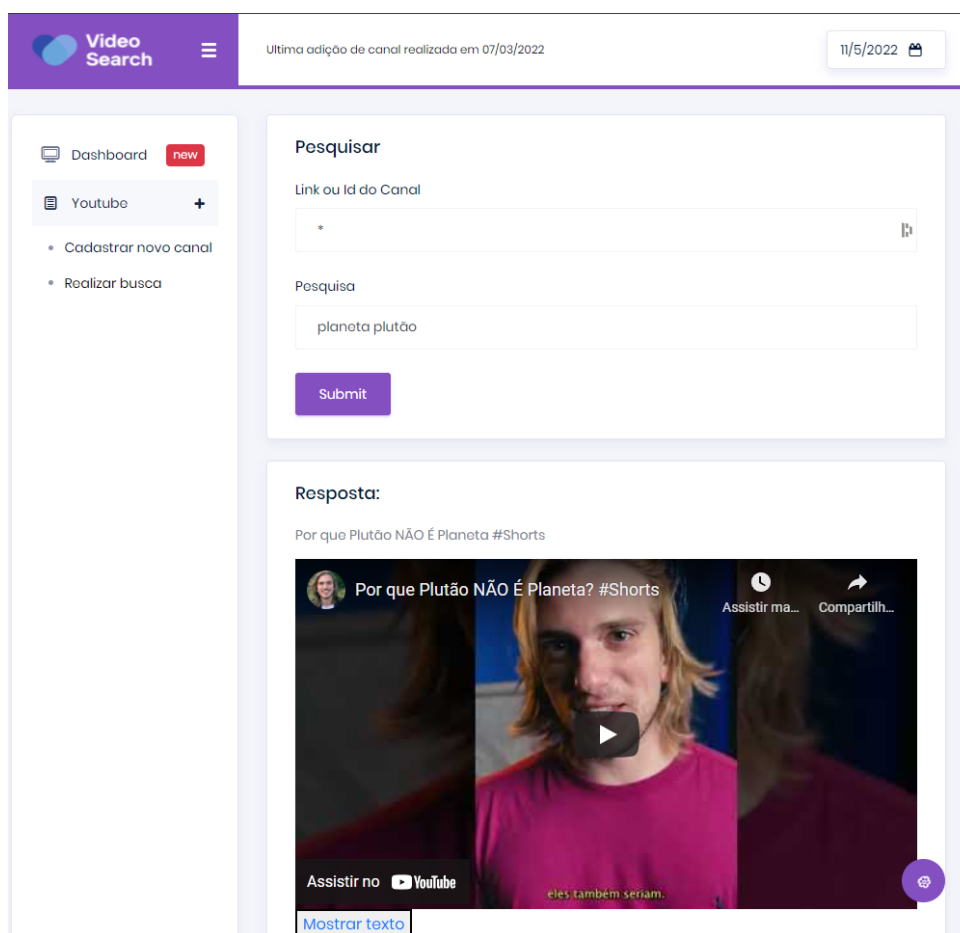
Figura 10 – Tela de Pesquisa em processamento



This screenshot is identical to Figure 9, showing the search page with the text 'planeta plutão' entered. The key difference is the presence of a green progress bar at the bottom of the search form, indicating that the search is currently in progress. The rest of the interface, including the sidebar and header, remains the same.

Fonte: próprio autor

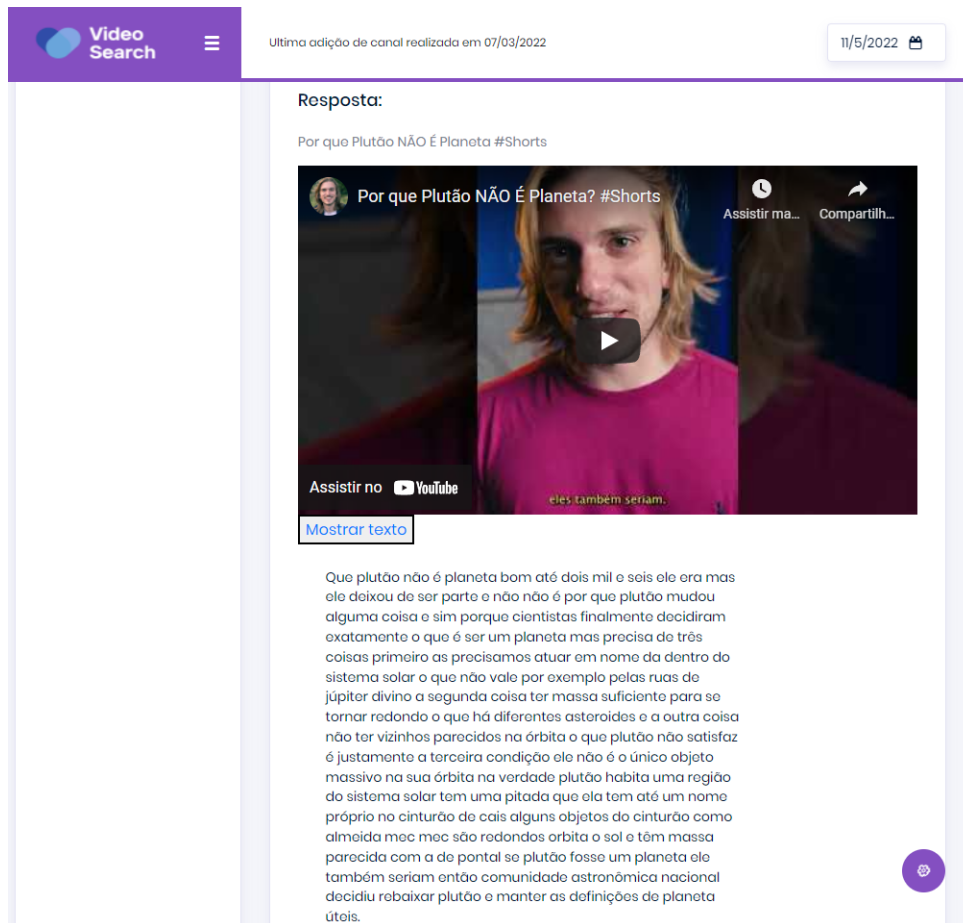
Figura 11 – Tela de Pesquisa com exibição de resultado



Fonte: próprio autor

Além de mostrar os vídeos que respondem a pesquisa realizada, a ferramenta ainda possibilita a opção de mostrar o conteúdo do vídeo em formato de texto. Para realizar tal ação há o botão “Mostrar texto” que aparece logo abaixo dos vídeos de resposta apresentados, conforme exemplificado na Figura 12.

Figura 12 – Tela de Pesquisa com exibição do texto do video



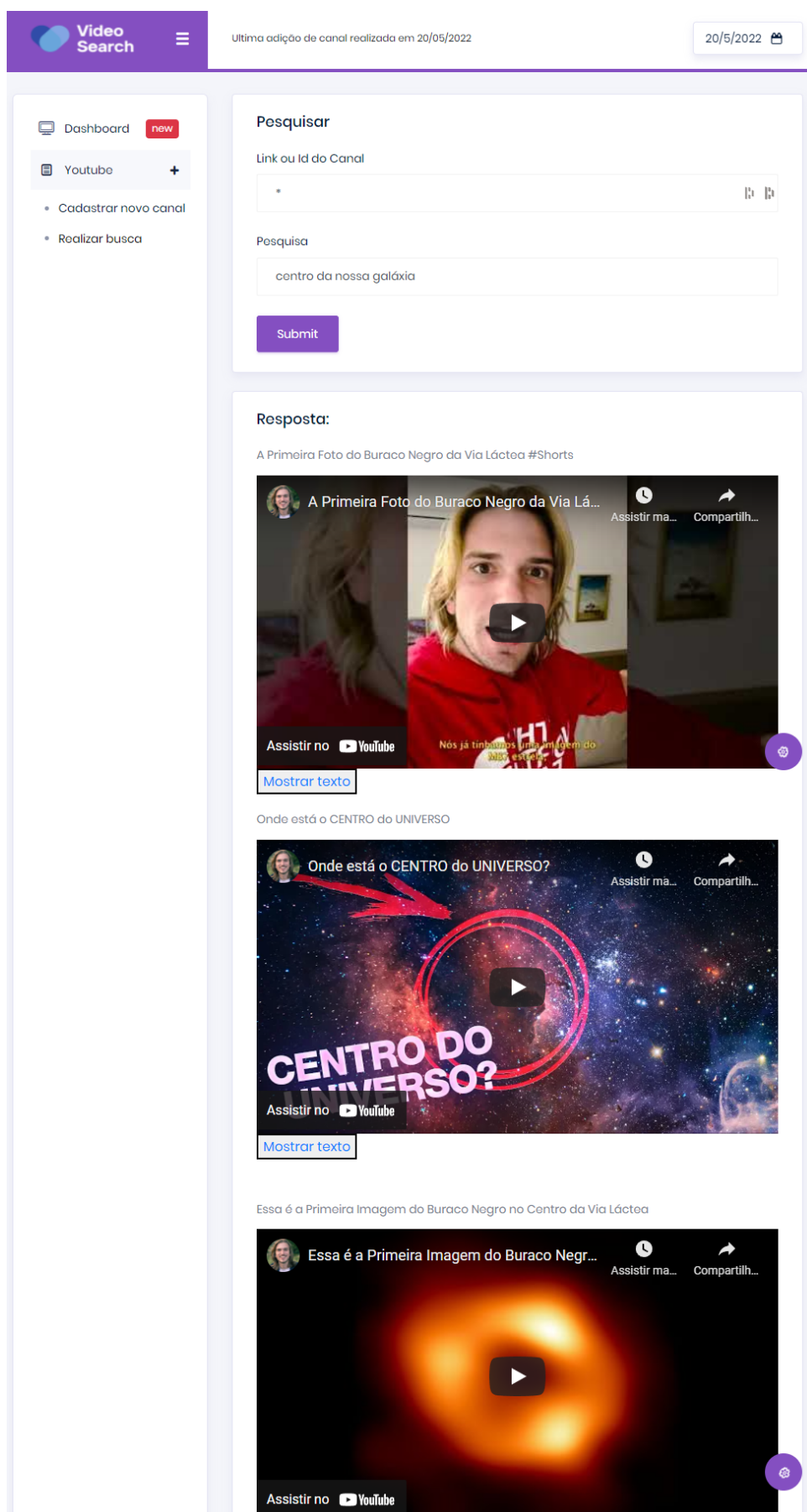
The screenshot shows a search interface for 'Video Search'. At the top, there is a purple header with the 'Video Search' logo and a menu icon. To the right, it says 'Última adição de canal realizada em 07/03/2022' and a date '11/5/2022' with a calendar icon. The main content area is titled 'Resposta:' and features a video thumbnail for 'Por que Plutão NÃO É Planeta? #Shorts'. The thumbnail shows a man with long hair in a pink shirt. Below the thumbnail is a 'Mostrar texto' button. The transcript text is as follows:

Que plutão não é planeta bom até dois mil e seis ele era mas ele deixou de ser parte e não não é por que plutão mudou alguma coisa e sim porque cientistas finalmente decidiram exatamente o que é ser um planeta mas precisa de três coisas primeiro as precisamos atuar em nome da dentro do sistema solar o que não vale por exemplo pelas ruas de júpiter divino a segunda coisa ter massa suficiente para se tornar redondo o que há diferentes asteroides o a outra coisa não ter vizinhos parecidos na órbita o que plutão não satisfaz é justamente a terceira condição ele não é o único objeto massivo na sua órbita na verdade plutão habita uma região do sistema solar tem uma pitada que ela tem até um nome próprio no cinturão de cois alguns objetos do cinturão como almeida mec mec são redondos orbita o sol e têm massa parecida com a de pontal se plutão fosse um planeta ele também seriam então comunidade astronômica nacional decidiu rebaixar plutão e manter as definições de planeta úteis.

Fonte: próprio autor

A figura 13 ilustra a realização de uma consulta usando o termo “centro da nossa galáxia”. Como esse termo aparece em vários vídeos no canal buscado, é exibido para o usuário de maneira ranqueada os vídeos que mais se assemelham a consulta.

Figura 13 – Tela de Pesquisa com exibição de múltiplos vídeos como resposta



Fonte: próprio autor

5 Conclusão

Os avanços dos recursos tecnológicos presentes hoje em dia possibilitaram a criação de muitos mecanismos de recuperação de informação. Este trabalho de conclusão de curso teve como objetivo o desenvolvimento de um sistema visando auxiliar usuários que possuem uma necessidade de realizarem busca de informações em vídeos, levando em consideração o seu conteúdo em vez de somente o título e metadados como a maioria dos buscadores fazem.

Para isto foi, criada uma ferramenta que é executada em plataforma WEB, que pode ser acessada de qualquer dispositivo através de um browser e acesso à internet, o que facilita o acesso dos usuários. A ferramenta foi construída usando HTML5, CSS3, JavaScript, Bootstrap4, Python e Flask. Inicialmente foi realizado o levantamento de requisitos do sistema e, a partir, daí, a construção da ferramenta foi iniciada, começando pela a lógica de conexão com o Youtube e download dos vídeos, que posteriormente são transformados em texto e adicionados à base de dados para serem usados para responder às consultas. As consultas são respondidas usando o modelo vetorial com base na ponderação TF-IDF e cálculo de similaridade pelo cosseno do ângulo. Para auxiliar no desenvolvimento foram usadas também algumas bibliotecas, como a biblioteca Spacy, que foi usada para realizar a lematização do texto, e a biblioteca Sklearn que foi usada para auxiliar no cálculo do TF-IDF.

Os resultados do desenvolvimento do software foram apresentados por meio de capturas de tela do mesmo durante a demonstração do fluxo de adição de canais à base de dados, e durante o fluxo de consultas à base de dados (ver Figuras 6-12). Em cada um dos fluxos foram apresentados os parâmetros de entrada usados para preencher cada um dos campos necessários para conclusão do fluxo, além do resultado obtido.

O software fornece uma interface simples e amigável ao usuário, permitindo que ele faça a inclusão de qualquer canal do Youtube na base de dados e realize consultas de forma livre. Além disso, o usuário pode assistir aos vídeos que respondem a sua consulta diretamente na página da ferramenta apresentada, ou até mesmo, caso não queira assistir o vídeo, pode somente ler o texto do vídeo que a ferramenta oferece.

O processo de elaboração do software foi desafiador em alguns momentos. Além da dificuldade de implementar todo o sistema em uma parte que faz todo o processamento das requisições e outra que faz a coleta e exibição das afirmações para o usuário, enfrentou-se também o desafio de implementar uma maneira de obter os links dos últimos vídeos de um canal do YouTube e realizar o armazenamento e gerenciamento dos dados de texto dos vídeos. Apesar desses desafios, o *software* desenvolvido no presente trabalho, no que

tange recuperação de informação em vídeos, apresenta o potencial de auxiliar os usuários a buscarem vídeos que satisfaçam as suas necessidades de busca de informação, já que através desta ferramenta o usuário consegue por uma interface amigável realizar as suas buscas e visualizar os vídeos que respondem a sua consulta, e, além disso, conseguem visualizar o conteúdo do vídeo em formato de texto, otimizando a busca do usuário já que ele não mais fica preso a necessidade de assistir todo o vídeo para que no final verifique se realmente o vídeo respondeu à sua necessidade de informação.

5.1 Trabalhos futuros

Ainda há espaço para melhorias no software, o que poderia ser o foco de trabalhos futuros. Estes poderiam trabalhar temas diversos, como: Melhorar o sistema para ser capaz de indicar o momento do vídeo em que responde a busca do usuário; melhorar o processamento dos dados para que ocorra em paralelo e o tempo de inclusão de novos vídeos a base de dados seja reduzido; realizar testes usando outras ferramentas de transcrição buscando uma melhor assertividade; expandir o suporte para outras plataformas de vídeo para que seja possível realizar a busca em diferentes plataformas; fazer implementações com o intuito de melhorar a performance das consultas, como, utilizar cache e persistir em memória dicionários usados nos cálculos de similaridade; implementar uma parte de login para usuários administradores que possuem permissão para adicionar novos vídeos à base de dados; implementar funcionalidade de indexação automática de novos canais, podendo esta até receber parâmetros para indicar o tipo de canal indexado.

Referências

- AVGOUSTINAKIS, P. et al. Audio-based near-duplicate video retrieval with audio similarity learning. In: IEEE. *2020 25th International Conference on Pattern Recognition (ICPR)*. 2021. p. 5828–5835. Disponível em: <<https://doi.org/10.1109/ICPR48806.2021.9413056>>. Citado na página 17.
- BAEZA-YATES, B. R.-N. R. *Modern Information Retrieval*. ACM Press, 1999. Disponível em: <<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.27.7690&rep=rep1&type=pdf>>. Citado na página 9.
- BAEZA-YATES, B. R.-N. R. *Recuperação De Informação*. [S.l.]: bookman, 2013. Citado 2 vezes nas páginas 9 e 10.
- CHEN, Y.; ARGENTINIS, J. E.; WEBER, G. Ibm watson: how cognitive computing can be applied to big data challenges in life sciences research. *Clinical therapeutics*, Elsevier, v. 38, n. 4, p. 688–701, 2016. Disponível em: <<https://doi.org/10.1016/j.clinthera.2015.12.001>>. Citado 2 vezes nas páginas 15 e 16.
- CROFT DONALD METZLER, T. S. B. *Search Engines: Information Retrieval in Practice*. Pearson, 2010. Disponível em: <http://library.mpib-berlin.mpg.de/toc/z2009_2465.pdf>. Citado 2 vezes nas páginas 11 e 12.
- DATAREPORTAL. *Global Overview Report*. 2022. Disponível em: <<https://datareportal.com/reports/digital-2022-global-overview-report>>. Citado na página 7.
- DCASE. *DCASE2020 challenge task 6: Automated audio captioning*. 2020. Disponível em: <<http://dcase.community/challenge2020/task-automatic-audio-captioning>>. Citado na página 17.
- DHARMENDRA, S. et al. *Unite neuroscience, supercomputing, and nanotechnology to discover, demonstrate, and deliver the brain's core algorithms*. San Jose. CA. USA: IBM Research-Almaden, 2011. Disponível em: <<https://dl.acm.org/doi/abs/10.1145/1978542.1978559>>. Citado na página 15.
- DROSSOS, K. Automated audio captioning with recurrent neural networks. IEEE WASPAA, 2017. Disponível em: <<https://doi.org/10.1109/WASPAA.2017.8170058>>. Citado na página 17.
- DROSSOS, K. *Clotho: An audio captioning dataset*. ICASSP, 2020. Disponível em: <<https://doi.org/10.1109/ICASSP40776.2020.9052990>>. Citado na página 17.
- FOOTE, J. T. *Content-based retrieval of music and audio*. SPIE, 1997. Disponível em: <<https://doi.org/10.1117/12.290336>>. Citado na página 17.
- GONZALEZ, V. L. S. d. L. M. Recuperação de informação e processamento da linguagem natural. PUCRS, 2003. Disponível em: <<https://www.marilia.unesp.br/Home/Instituicao/Docentes/EdbertoFerneda/mri-06---gonzales-e-lima-2003.pdf>>. Citado na página 14.

- HIGH, R. *The Era of Cognitive Systems: An Inside Look at IBM Watson and How it Works*. [S.l.]: IBM Redbooks, 2017. Citado na página 16.
- HOU, S.; ZHOU, S. Audio-visual-based query by example video retrieval. *Mathematical Problems in Engineering*, Hindawi, v. 2013, 2013. Disponível em: <<https://doi.org/10.1155/2013/972438>>. Citado na página 17.
- INSIGHT, G. M. *YouTube User Statistics 2022*. 2022. Disponível em: <<https://www.globalmediainsight.com/blog/youtube-users-statistics/>>. Citado na página 7.
- JIN, Q. et al. Event-based Video Retrieval Using Audio. 9 2012. Disponível em: <https://kilthub.cmu.edu/articles/journal_contribution/Event-based_Video_Retrieval_Using_Audio/6473363>. Citado na página 17.
- KIM, C. D. et al. Audiocaps: Generating captions for audios in the wild. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. [s.n.], 2019. p. 119–132. Disponível em: <<http://dx.doi.org/10.18653/v1/N19-1011>>. Citado na página 17.
- LANCASTER, F. W. *Indexação e resumos: teoria e prática*. Brinquet De Lemos Livros, 2004. Disponível em: <<https://revista.ibict.br/ciinf/article/download/509/509>>. Citado na página 9.
- LEHNE, M. et al. Reading a suspenseful literary text activates brain areas related to social cognition and predictive inference. *PloS one*, Public Library of Science San Francisco, CA USA, v. 10, n. 5, p. e0124550, 2015. Disponível em: <<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0124550>>. Citado na página 15.
- MORETO, S. *Bootstrap 4 By Example*. Packt Publishing Ltd, 2016. Disponível em: <<https://dl.acm.org/doi/abs/10.5555/3006353>>. Citado na página 20.
- ONCESCU, A.-M. et al. Audio retrieval with natural language queries. *arXiv preprint arXiv:2105.02192*, 2021. Disponível em: <<https://doi.org/10.5120/ijca2018917395>>. Citado 2 vezes nas páginas 10 e 18.
- QAISER, S.; ALI, R. Text mining: use of tf-idf to examine the relevance of words to documents. *International Journal of Computer Applications*, v. 181, n. 1, p. 25–29, 2018. Disponível em: <https://www.researchgate.net/profile/Shahzad-Qaiser/publication/326425709_Text_Mining_Use_of_TF-IDF_to_Examine_the_Relevance_of_Words_to_Documents/links/5b4cd57fa6fdcc8dae245aa3/Text-Mining-Use-of-TF-IDF-to-Examine-the-Relevance-of-Words-to-Documents.pdf>. Citado na página 13.
- RAMOS, J. et al. Using tf-idf to determine word relevance in document queries. In: CITESEER. *Proceedings of the first instructional conference on machine learning*. 2003. v. 242, n. 1, p. 29–48. Disponível em: <<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.121.1424&rep=rep1&type=pdf>>. Citado na página 13.
- RILEY, M.; HEINEN, E.; GHOSH, J. A text retrieval approach to content-based audio retrieval. In: *Int. Symp. on Music Information Retrieval (ISMIR)*. [S.l.: s.n.], 2008. p. 295–300. Citado na página 17.

SALTON, G.; WONG, A.; YANG, C.-S. A vector space model for automatic indexing. *Communications of the ACM*, ACM New York, NY, USA, v. 18, n. 11, p. 613–620, 1975. Disponível em: <<https://doi.org/10.1145/361219.361220>>. Citado 2 vezes nas páginas 12 e 18.

SANTIAGO PALLAVI, S. S. F. *Building Cognitive Applications with IBM Watson Services: Volume 6 Speech to Text and Text to Speech*. [S.l.]: IBM Redbooks, 2012. Citado na página 16.

SOMERVILLE, I. *Engenharia de Software*. [S.l.]: Pearson Universidades, 2019. Citado na página 19.

VELDE, F. Van der. Artificial intelligence and neuroscience meet: The search for grounded architectures of cognition. *Adv Artif Intell*, 2010. Disponível em: <<https://www.proquest.com/openview/2e1d2b7071ab9a5e435b477202d3b4b6/1?pq-origsite=gscholar&cbl=237361>>. Citado 2 vezes nas páginas 15 e 16.

WU, M. *Audio caption: Listen and tell*. ICASSP, 2019. Disponível em: <<https://doi.org/10.1109/ICASSP.2019.8682377>>. Citado na página 17.