Dr. NGP INSTITUTE OF TECHNOLOGY

MINI PROJECT

VIDEO SYNTHESIS

Presented by,

KAPILAN SD	(22AD021)
LOGESH M	(22AD028)
SANJAY S	(22AD044)
VAIGAI VENDHAN K	(22AD056)

Under the Guidance of
Ms. M. Sarmila AP/AI&DS

Abstract

The "Video Synthesis" project transforms text into captivating video presentations through advanced NLP, image retrieval, and TTS technologies. Key features include automated keyword extraction, image database integration, and customizable audio synthesis. With a user-friendly interface and browser extension, it streamlines content creation. "Video Synthesis" empowers users across education, marketing, and entertainment sectors to effortlessly produce engaging videos from textual content.

Problem Statement

The traditional process of converting text to video is cumbersome and lacks efficiency, relying on manual editing and synchronization. Existing tools often lack advanced features like automated keyword extraction and customizable audio synthesis, resulting in subpar output quality. To address these challenges, there's a pressing need for a comprehensive solution that leverages advanced technologies to automate video synthesis from text, enhancing efficiency and output quality

Literature survey

Tune-A-Video: One-Shot Tuning of Image Diffusion Models for Text-to-Video Generation

ICCV 2023 · Jay Zhangjie Wu, Yixiao Ge, Xintao Wang, Weixian Lei, YuChao Gu, Yufei Shi, Wynne Hsu, Ying Shan, XiaoHu Qie, Mike Zheng Shou

We make two key observations:

1. T2I models can generate still images that represent verb terms.
2. Extending T2I models to generate multiple images concurrently exhibits surprisingly good content consistency.
3. To further learn continuous motion, we introduce Tune-A-Video

Literature survey

Make-A-Video: Text-to-Video Generation without Text-Video Data

29 Sep 2022 · Uriel Singer, Adam Polyak, Thomas Hayes, Xi Yin, Jie An, Songyang Zhang, Qiyuan Hu, Harry Yang, Oron Ashual, Oran Gafni, Devi Parikh, Sonal Gupta, Yaniv Taigman ·

1) An approach for directly translating the tremendous recent progress in Text-to-Image (T2I) generation to Text-to-Video (T2V).
2) We design a spatial temporal pipeline to generate high resolution and frame rate videos with a video decoder.

Literature survey

Sync-DRAW: Automatic Video Generation using Deep Recurrent Attentive Architectures30 Nov 2016 Gaurav Mittal, Tanya Marwah , Vineeth N. Balasubramanian

1. This paper introduces a novel approach for generating videos called Synchronized Deep Recurrent Attentive Writer (Sync-DRAW).
2. Generating GIF using Sync - Deep Recurrent Attentive Architectures.

Dr. NGP INSTITUTE OF TECHNOLOGY

Literature survey

CogVideo : Large-scale Pretraining for Text-to-Video Generation via Transformers

29 May 2022 · Wenyi Hong, Ming Ding, Wendi Zheng, Xinghan Liu, Jie Tang ·

1) Large-scale pretrained transformers have created milestones in text (GPT-3) and text-to-image (DALL-E and CogView) generation.
2) Currently it only supports ***simplified Chinese input.***

Existing System

Our system, inspired by existing platforms like Synthesia, InVideo, D-ID, and Hugging Face, offers an automated approach to video creation from text-based content. Similar to these platforms, our system harnesses advanced technologies to transform text into audio narration, identify key points, and source relevant images. By leveraging state-of-the-art algorithms, our system streamlines the process of generating engaging videos, catering to a wide range of users and applications. With a focus on simplicity and efficiency, our system aims to revolutionize content creation, offering an intuitive solution for converting textual content into compelling video presentations.

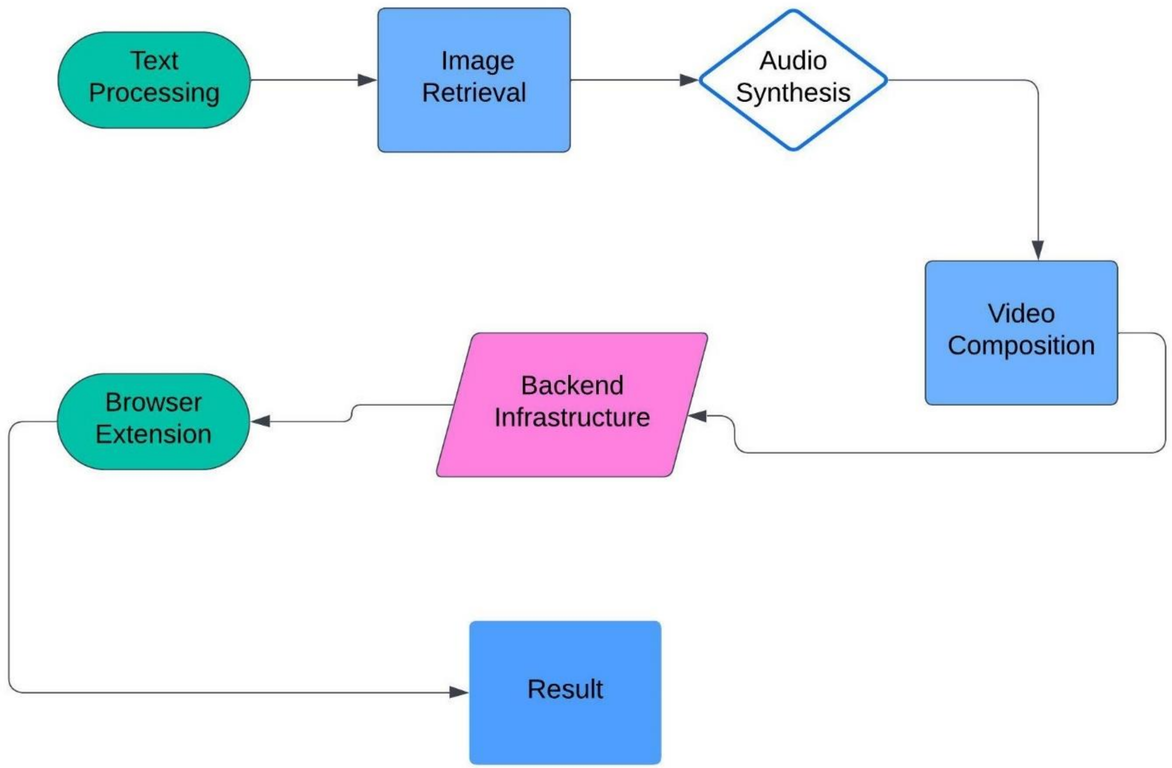
Existing System

1. Synthesiya - https://www.synthesia.io/
2. Invideo - https://invideo.io/
3. D-ID - https://www.d-id.com/
4. Hugging Face - https://huggingface.co/
5. Sora - https://openai.com/sora

Proposed System

The proposed system employs advanced NLP, image retrieval, and audio synthesis technologies to revolutionize video content creation from textual inputs. It automates text-to-video conversion by extracting keywords, retrieving images, and synthesizing audio narration. Integrated with a user-friendly interface and browser extension, it enables seamless generation of engaging videos directly from web pages. The scalable backend architecture ensures optimal performance, while comprehensive documentation and support resources facilitate user adoption and maximize platform benefits.

Proposed System Architecture



List of Hardware

- Processor: Intel Core i5 or equivalent AMD (or higher)
- RAM: 8 GB or higher
- Graphics Card: NVIDIA GTX 1060 or AMD
- Radeon RX 580 (or equivalent)
- Storage: SSD with at least 256 GB
- Internet Connection: Broadband

Dr. NGP INSTITUTE OF TECHNOLOGY

List of Software

- Operating System: Windows 10, macOS, or Linux
- Python: Latest version of Python 3.x
- Development Environment: Anaconda or Miniconda
- Libraries: moviepy, spacy, requests, pyttsx3, tkinter, Flask, Flask-CORS
- Video Editing Software: Optional
- Web Browser: Latest version of Chrome, Firefox, or Safari
- IDE: PyCharm, VS Code, or Jupyter8. API Keys: Unsplash, etc.

Dr. NGP INSTITUTE OF TECHNOLOGY

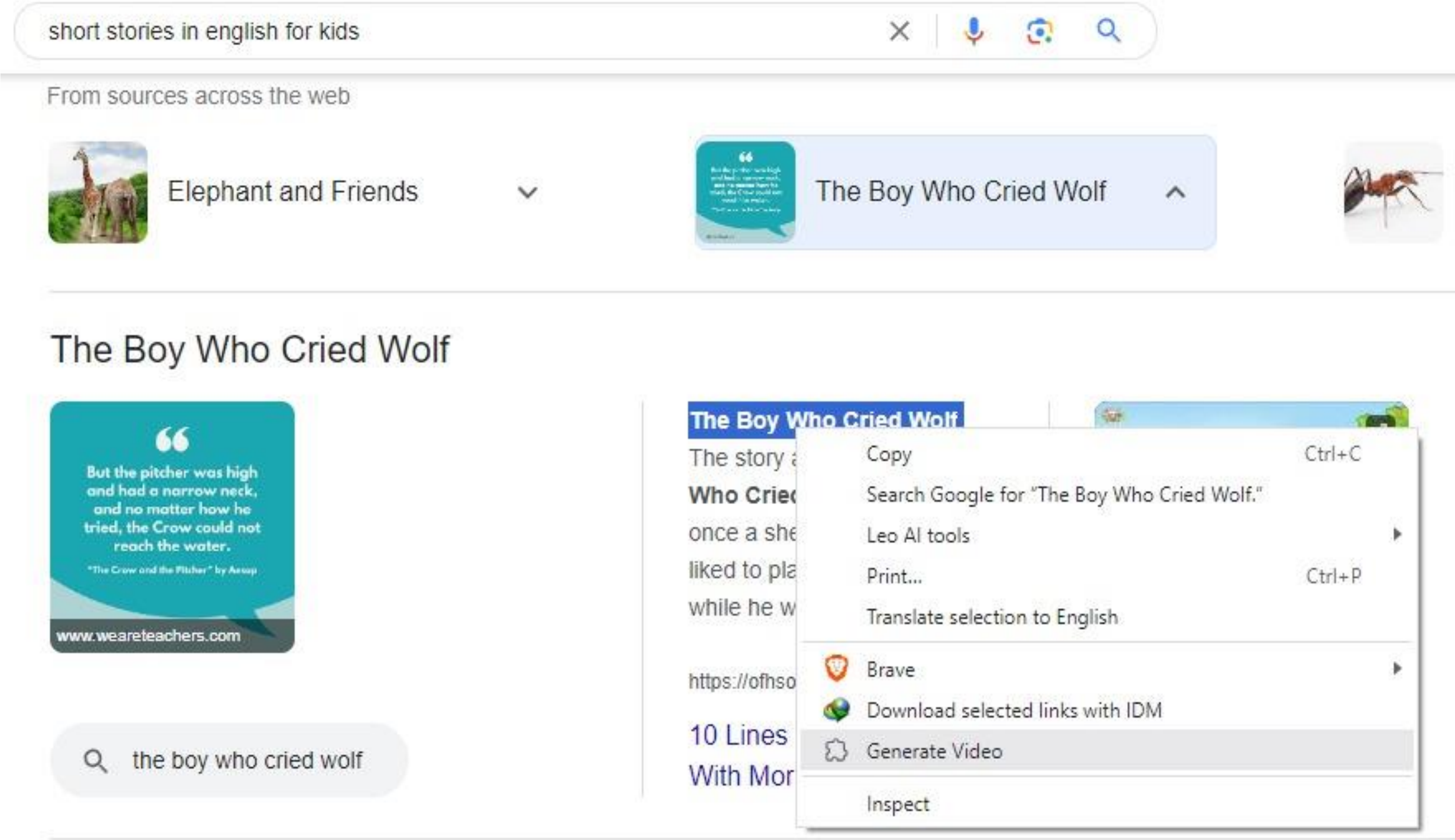
List of Modules

- Text Processing
- Image Retrieval
- Audio Synthesis
- Video Composition
- Backend Infrastructure
- User Interface
- Browser Extension

Module Description

1. Text Processing: Analyzes text for video content extraction.
2. Image Retrieval: Fetches images relevant to text keywords.
3. Audio Synthesis: Converts text to audio for video narration.
4. Video Composition: Combines images and audio for final video creation.
5. Backend Infrastructure: Manages system resources for processing.
6. User Interface: Provides intuitive interaction for users.
7. Browser Extension: Enables web-based video synthesis.

Module output screenshot



short stories in english for kids

From sources across the web

Elephant and Friends

The Boy Who Cried Wolf

The Boy Who Cried Wolf

But the pitcher was high and had a narrow neck, and no matter how he tried, the Crow could not reach the water.

"The Crow and the Pitcher" by Aesop

www.weareteachers.com

the boy who cried wolf

The Boy Who Cried Wolf

The story a

Who Cried

once a she

liked to pla

while he w

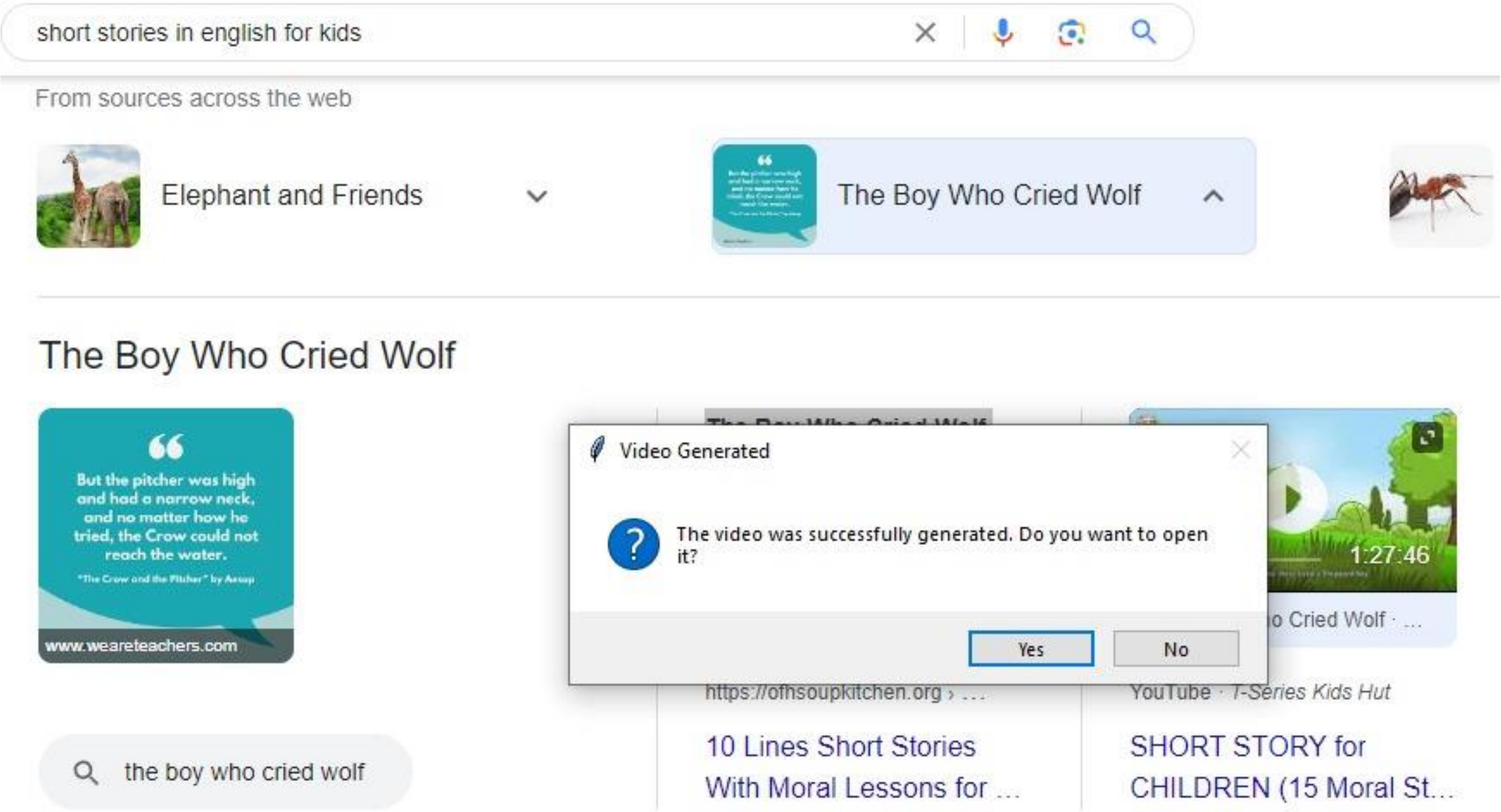
https://ofhso

10 Lines

With Mor

Copy Ctrl+C
Search Google for "The Boy Who Cried Wolf."
Leo AI tools
Print... Ctrl+P
Translate selection to English
Brave
Download selected links with IDM
Generate Video
Inspect

Module output screenshot



short stories in english for kids

From sources across the web

Elephant and Friends

The Boy Who Cried Wolf

The Boy Who Cried Wolf

But the pitcher was high and had a narrow neck, and no matter how he tried, the Crow could not reach the water.

"The Crow and the Pitcher" by Aesop

www.weareteachers.com

the boy who cried wolf

Video Generated

The video was successfully generated. Do you want to open it?

https://ofhsoupkitchen.org › ...

10 Lines Short Stories With Moral Lessons for ...

YouTube · T-Series Kids Hut

SHORT STORY for CHILDREN (15 Moral St...

Dr. NGP INSTITUTE OF TECHNOLOGY

Conclusion

In summary, the Video Synthesis project has revolutionized content creation by automating the process of converting text into compelling video presentations. With its user-friendly interface and seamless integration with web browsers, the system caters to diverse users, including educators, marketers, and creators. By leveraging advanced technologies like natural language processing and image retrieval, it streamlines workflow and enhances efficiency. Moving forward, the project's versatility and potential for further innovation hold promise for transforming communication and engagement across various domains.

References

1. Hinton, G., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2012). Improving neural networks by preventing co-adaptation of feature detectors.
2. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need.
3. Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate.
4. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory.
5. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition.