

junkyuhufs / 2024WorkshopForProfs

🔍

👤

▼

📁

🔗

<> Code

🕒 Issues

🔗 Pull requests

🔄 Actions

📁 Projects

📖 Wiki

🛡 Security

📈

📄

🔗 main ▼

2024WorkshopForProfs / README.md 📄

⋮

🔗

junkyuhufs Update README.md

5aa4a91 · 2 minutes ago

🕒

287 lines (202 loc) · 11.2 KB

Preview

Code

Blame

Raw

📄

📄

✏

⋮

🎓 멀티모달 도구와 Python을 활용한 교수 혁신 실습 🚀

📅 날짜: 2024-12-20

👤 발표자: 이준규 (한국외국어대학교 교육대학원)

📄

Overview of Today's Workshop

• Part I: 다양한 멀티모달 AI 도구들 소개

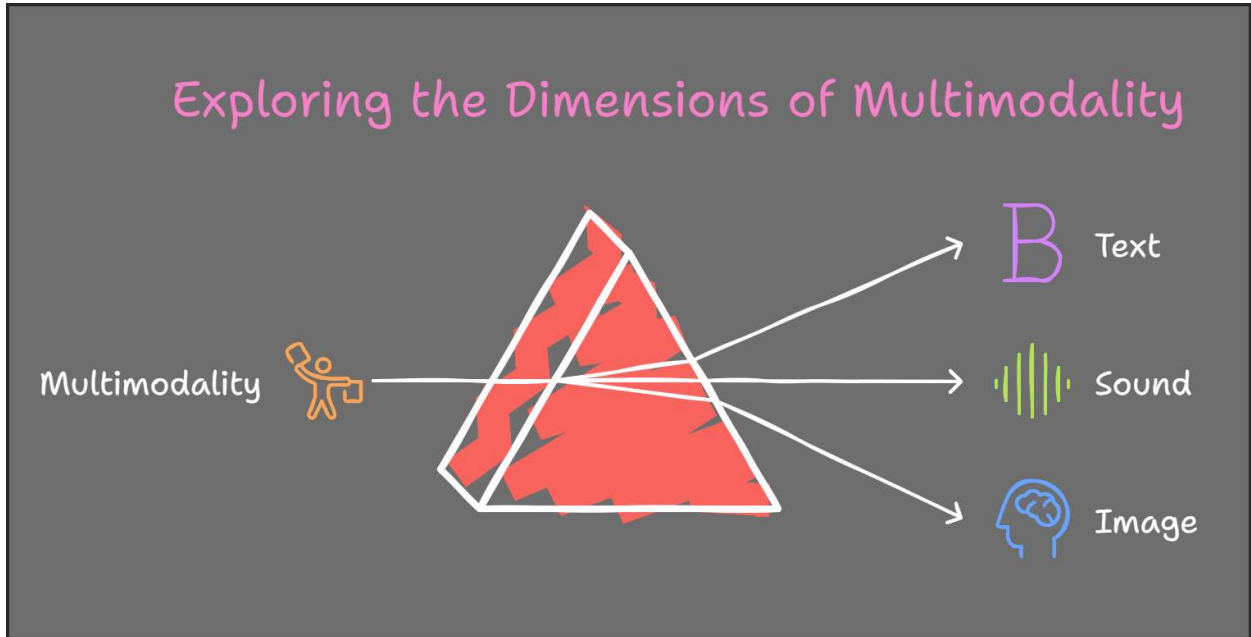
• Part II: ChatGPT를 활용한 Python 코딩으로 간단한 멀티모달 앱 만들기

🌐 Tools: Google Colab (통합 개발 환경 및 코딩 플랫폼)

https://github.com/junkyuhufs/2024WorkshopForProfs/blob/main/README.md

1/9

Multimodality



Multimodal Examples

멀티모달은 여러 입력 및 출력 모달리티(Text, Speech, Image 등)를 활용하여 데이터를 처리하는 기술입니다.

1. Text to Text (LM - Language Model)

- **설명:** 텍스트 입력을 통해 다른 텍스트를 생성하거나 변환하는 작업입니다.
- **예시:**
 - 문장 완성: "AI is..." → "AI is transforming industries worldwide."
 - 언어 번역: 영어 문장을 한국어로 번역.
 - 문체 변환: 캐주얼 문장을 학문적인 문장으로 변환.
- **도구:** [ChatGPT](#), [Google Translate](#), [DeepL](#).

2. Text to Speech (TTS)

- **설명:** 텍스트 데이터를 음성으로 변환하는 기술입니다.
- **예시:**
 - eBook 리더: 텍스트를 음성으로 읽어줌.
 - 스마트폰 음성 비서: Siri, Google Assistant.
- **도구:** [Google TTS](#), [Amazon Polly](#), [네이버 클로바](#), [카카오 음성 API](#).

3. Text to Image

- **설명:** 텍스트 설명을 기반으로 이미지를 생성하는 기술입니다.
- **예시:**
 - "A sunset over a mountain range with a river flowing through it." → 해당 이미지를 생성.
 - 그래픽 디자인: 브랜딩을 위한 비주얼 제작.
- **도구:** [DALL·E](#), [MidJourney](#), [Stable Diffusion](#).

4. Speech to Text (STT)

- **설명:** 음성을 텍스트로 변환하는 기술로, 회의록 작성이나 음성 명령 처리에 활용됩니다.
- **예시:**
 - 음성 메시지를 자동으로 텍스트화.
 - 비디오 자막 자동 생성.
- **도구:** [Google Speech-to-Text](#), [네이버 클로바 노트](#), [Otter.ai](#).

5. Image to Text

- **설명:** 이미지 속 텍스트를 추출하거나 이미지를 설명하는 텍스트를 생성하는 기술입니다.
- **예시:**
 - 스캔된 문서에서 텍스트 추출(OCR).
 - 이미지 설명 생성: "A cat sitting on a sofa."
- **도구:** [Google Vision API](#), [Tesseract OCR](#), [Microsoft Azure Vision](#).

6. Speech to Image

- **설명:** 음성을 기반으로 이미지를 생성하는 기술.
- **예시:**
 - "Draw a happy family under the moonlight"라는 음성 명령으로 이미지 생성.
- **도구:** [Stable Diffusion](#)과 음성 인식 API 연동.

7. Video to Text

- **설명:** 비디오 콘텐츠에서 텍스트 정보를 추출하는 기술입니다.

- **예시:**
 - 강의 영상에서 자막 자동 생성.
 - 영상 내 대화 내용 요약.
- **도구:** [YouTube 자동 자막](#), [Whisper API](#).

8. Image to Speech

- **설명:** 이미지 속 정보를 음성으로 설명하는 기술입니다. 주로 시각장애인을 위한 접근성 도구로 활용됩니다.
- **예시:**
 - 메뉴판 이미지를 읽어서 음성으로 설명.
 - 사진 속 텍스트를 음성으로 전달.
- **도구:** [Seeing AI](#), [Google Lookout](#).

9. Multimodal Fusion

- **설명:** 두 가지 이상의 모달리티를 결합하여 더욱 풍부한 경험을 제공하는 기술.
- **예시:**
 - 스마트 비서: 음성 명령(Speech) → 이미지 검색(Image) → 결과 설명(Text or Speech).
 - 언어 학습 앱: 단어(텍스트)와 발음(음성), 예문(이미지) 제공.
- **도구:** [GPT-4](#) (Multimodal), [DeepMind's Flamingo](#).

응용 및 실생활 사례

1. **교육:** AI 기반 튜터링 시스템은 텍스트 입력(T2T) → 음성 응답(TTS) → 시각 자료 생성(T2I)으로 학습 지원.
2. **의료:** 의사의 음성 진단(STT)을 텍스트화하여 의료 데이터베이스에 저장.
3. **고객 서비스:** 음성 명령(STT)을 통해 상품 이미지(T2I)를 검색하고 설명.

학습을 위한 팁

- 실습 중심으로 학습: TTS, STT, Text-to-Image를 체험하는 데모 제공.
- 한국어 데이터 활용: 한국어 기반 멀티모달 데모 시연.
- 도구 활용: [ChatGPT](#), [DALL·E](#) 등으로 기술을 실시간으로 시연.

💡 멀티모달 기술은 다양한 분야에서 창의적이고 실용적인 방식으로 활용될 수 있습니다. 이 문서를 통해 멀티모달의 가능성을 탐구해 보세요!

Part I. 🛠️ Diverse AI Multimodal Tools

🗣️ [1] Text to Speech

- 대표사이트: [ElevenLabs](#)

🎵 [2] Text to Song

- [Suno](#)

Sample Prompt:

너는 k-pop의 유명한 작사가야. 메가히트송을 여러 개 만들었어요. 오늘 노래가사를 하나 만들려고 해. 주제는 “멀티모달 도구와 Python을 활용한 교수 혁신 실습”에 참석한 한국외대 교수님들에게 희망의 메시지를 담은 노래를 만들고 싶어. 가사를 10줄정도로 만들어줘.



GPT 응답:

멀티모달의 날개 (가제)

- ☀️ 오늘도 배움의 문을 열어 새로운 길, 우리 함께 걸어
- 📚 Python과 도구들 손에 쥐고 교수의 꿈, 더 높이 빛나고
- 🌈 멀티모달 세상 속의 가능성 연결된 마음, 끝없는 반짝임
- 🚀 혁신의 날개로 날아올라 가르침 속 희망, 우리 곁에 살아
- 🎵 변화의 멜로디를 함께 불러 내일의 무대, 우리 빛날 거야

🖼️ [3] Text to Image

- [MidJourney](#)
- [DALL-E](#)
- [Canva](#)
- [Pika](#)

🎬 [4] Text to Video

- [Sora](#)

💡 실습 1: Text to Mind Map

Step 1: Prompt와 텍스트를 GPT에 입력

Sample Prompt:

Create a mind map for the below Critiques and suggestions for future research. Include relevant subtopics, tools, and methodologies to clearly show the key points. Provide a code box with Markdown language.



Step 2: Markdown 저장

- [Online markdown editor](#)

Step 3: Mind mapping software 사용

- [Xmind](#)

💡 실습 2: Text to Flowchart

1. Prompt 1: Flowchart로 만들고 싶은 내용 준비
2. Prompt 2: Mermaid 코드 생성
3. 사용 도구: [Draw.io](#)

🎧 Playing with Sound

- STT: [Whisper](#)
- Youtube to Text: [YouGlish](#)

💻 Part II. Welcome to the Coding World!

Sample Website

- HUFS GSE EE 학생들의 작품: [GitHub Example](#)

🐍 Python 코딩 첫걸음: Gradio로 간단한 앱 만들기 🚀

이 프로젝트는 Python 초보자들이 **Gradio** 라이브러리를 활용해 간단한 앱을 만들면서 코딩에 대한 자신감을 키울 수 있도록 설계되었습니다.

GPT를 활용해 코딩 과정을 더욱 쉽게 진행할 수 있다는 점도 함께 소개합니다! 😊

프로젝트 소개

이 프로젝트를 통해 배우게 될 내용:

1. Python과 라이브러리의 기본 개념 이해하기.
2. Gradio를 사용해 간단한 앱 만들기.
3. GPT를 활용해 코딩 문제 해결하기.

어떻게 작동하나요?

아래의 코드를 따라 하면 단어 수를 세어주는 간단한 웹 앱을 만들 수 있습니다!

1. 필요한 라이브러리 설치

Google Colab 또는 로컬 환경에서 아래 명령어로 Gradio를 설치하세요:

```
!pip install gradio
```



2. Python 코드 작성

아래 코드를 복사해 실행하세요:

```
# 1. 필요한 라이브러리 가져오기
import gradio as gr

# 2. 텍스트를 받아 단어 수를 세는 함수 정의
def count_words(text):
    return f"입력된 텍스트는 {len(text.split())}개의 단어로 이루어져 있습니다."

# 3. Gradio로 인터페이스 만들기
app = gr.Interface(fn=count_words, inputs="text", outputs="text")

# 4. 앱 실행
app.launch()
```



3. 결과 확인

1. 앱 실행 후 브라우저에서 열리는 링크를 클릭하세요.
2. 텍스트를 입력하면 단어 수가 출력됩니다.

GPT를 활용한 코딩 팁

GPT를 사용해 코딩 문제를 더 쉽게 해결할 수 있습니다!

질문 예시

- "Python으로 단어 수 세는 앱 만들고 싶어요. 코드를 알려주세요."
- "Gradio에서 입력창과 출력창을 어떻게 연결하나요?"

GPT의 도움으로 복잡한 문제도 해결 가능! 💡

확장 아이디어

이 프로젝트를 기반으로 다양한 기능을 추가할 수 있습니다:

- **문법 검사 앱**: 입력된 텍스트의 문법 오류를 검토.
- **텍스트 요약 앱**: 입력된 긴 문장을 간단히 요약.

마무리

Python과 Gradio는 초보자도 쉽게 앱을 만들 수 있도록 도와주는 강력한 도구입니다.

코딩은 어렵지 않다!는 자신감을 가지세요!

실습: Python & Gradio로 앱 만들기

1. TTS (Text-to-Speech)

[GitHub 링크](#)

2. Listening Comprehension Questions

[GitHub 링크](#)

3. Youtube 열기

[GitHub 링크](#)

4. Sequencing 앱

[GitHub 링크](#)

5. Writing Checker 앱

[GitHub 링크](#)

6. QR코드 생성 앱

[GitHub 링크](#)

참고자료

- [Hugging Face](#) 🌈