

Part1. 프로그램 언어의 이해



Point

- 각 소프트웨어별 특성화된 개발 영역에 대한 이해
- 파이썬 언어의 특성
- 파이썬에서 모듈(내장, 외장)이란_개념만

<https://imasoftwareengineer.tistory.com/56>

Part1. 파이썬의 유래 및 모듈 이해

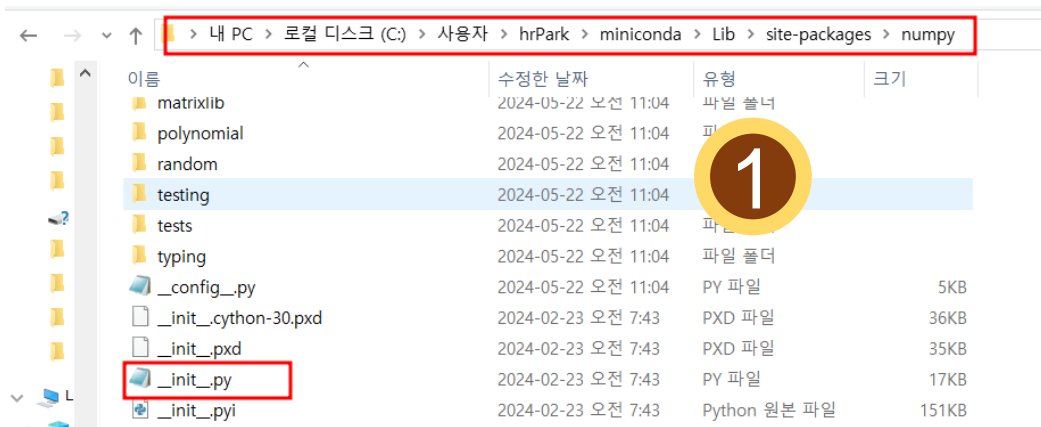
파이썬의 유래

<https://joone.net/tag/python/>

- 파이썬은 어떤 결과물을 원하는가에 따라 필요한 모듈(라이브러리, 패키지)을 선택하여야함. -> 즉 그 **모듈에 관련한 문법을 학습**해야함
 - 각각의 모듈은 최소한의 파이썬프로그램규칙(문법)을 사용하게 하며 이미 프로그램을 작성하여서 제공함.
 - 작성되어 있는 프로그램은 def(함수)와 def의 모음인 class(클래스)로 구성된 000.py 로 제공됨 (추후 모듈에서 학습함)
- 모듈 설치의 pip 또는 conda **install 모듈명** 이며 상황에 따라서는 모듈을 제공하는 사이트에 가서 000.whl 인 wheel 파일을 다운로드 받고 직접 wheel파일을 내컴퓨터에 설치해야 하기도 함. <https://blog.naver.com/PostView.nhn?blogId=youndok&logNo=222149853952>

import 모듈의 개념

- [1] import numpy 는 numpy폴더의 __init__.py를 실행함 (이 파일에는 수많은 import가 있음)
- [2] vscode에서 import numpy 코드에서 마우스 우측 클릭하고 [Go to Definition]을 선택하면
- [3] 폴더 경로 확인됨



(참고) from PIL import Image
→ PILLOW 폴더(모듈 버전 변경으로 인해 PILLOW에서 PIL로 모듈명변경)
에서 Image.py 파일을 실행

Part1. 구글검색

소프트웨어개발자 분야

소프트웨어 개발자 분야

✕

🗨️ 🎤 🖼️ 🔍

전체 이미지 뉴스 쇼핑 동영상 더보기 도구

브런치스토리
https://brunch.co.kr > ...
SW 개발 분야의 종류

2018. 6. 21. — SW 개발 분야의 종류 · 1. 프론트엔드 · 2. 모바일 · 3. 백엔드/서버 · 4. 게임 · 5. 임베디드 · 6. 데스크탑 네이티브 · 7. 연구 및 알고리즘 모델링 · 8 ...

티스토리
https://die-romantische-schule.tistory.com > ...

컴퓨터(CS/CE) 관련 소프트웨어 개발 분야 정리 - 낭만주의 학교

2021. 8. 5. — 컴퓨터(CS/CE) 관련 소프트웨어 개발 분야 정리 · 1. 한 눈에 보는 개발 분야 · 2. 세부 분야 정리 1) 프론트엔드 2) 백엔드 3) 데이터 4) 임베디드 5) 보안

Naver Blog
https://blog.naver.com > ...

[IT분야 SW개발 직무] 소프트웨어개발자의 종류, 어떻게 나눌까?

2018. 2. 19. — 소프트웨어의 개발과 공급, 유지 보수 등과 관련된 업무를 기획하고 지휘, 조정하는 직무입니다. 어떤 소프트웨어를 개발할 것인지를 기획하는 것부터 C, ...

컴파일러, 인터프리터

파이썬 프론트엔드

파이썬 백엔드

파이썬 앱 개발

파이썬 네트워크 제어

Part2. 파이썬 코드 작성



Point

- 변수, list, index, for, if의 단순한 문법만 알아도
사용하는 모듈 상황에 따라 원하는 결과물이 나올수 있음에 대하여 알기
- 파이썬에 특화된 검색용어 포함하여 검색하는 연습
(gpt 에서의 프롬프트 / 구글에서의 검색용어)

gpt에게 물어보고 사전학습하기

- 'print의 f포맷'
- '파이썬 for enumerate 설명 및 예제'

- [1] 자료준비 'c:/data/pdf실습/' 폴더를 작성한뒤 '0000.pdf' 파일을 준비함.
- [2] 파이썬 코드를 실행하기 위해 구글검색 '파이썬에서 pdf를 이미지로'
- [2] 구글에서 제공하는 코드 보면서 vscode에서 실행함



[주의사항1] 모듈이 설치되어 있지 않다면 conda 환경 또는 주피터노트북에서 설치해야함.
주피터노트북에서는 !pip로 설치

[구글자료]

```
pip install PyMuPDF
```

[수정코드]

```
1 !pip install PyMuPDF
```

[주의사항2] pdf가 있는 파일의 경로 및, 각 pdf를 이미지 파일로 저장할 경로는 본인이 알아서 지정
이 작업에서는 'c:/data/pdf실습/GPT활용_파이썬기초 및_ 데이터분석 및 시각화.pdf' 를 c:/data/pdf실습 폴더에 각각 이미지 파일로 저장하고자 함.

[구글자료]

```
# 예제 PDF 파일 경로 .은 현재폴더
PDF_FILE_PATH = "./data/sample.pdf"
```

```
import fitz
doc = fitz.open(PDF_FILE_PATH)
for i, page in enumerate(doc):
    img = page.get_pixmap()
    img.save(f"./data/{i}.png")
```

[수정코드]

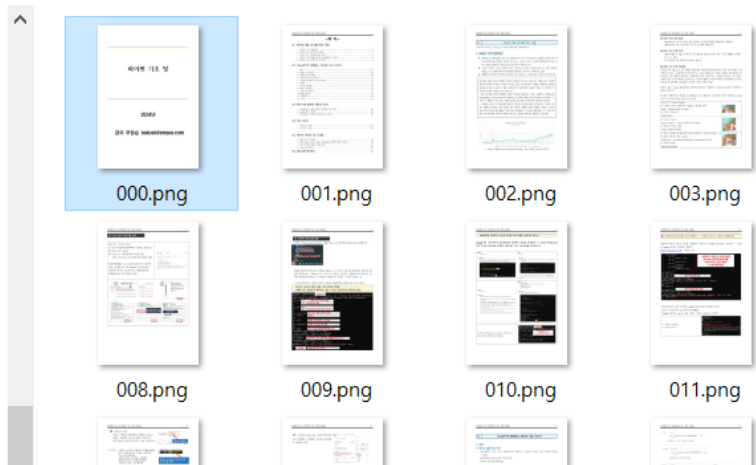
```
1 # 예제 PDF 파일 경로
2 PDF_FILE_PATH = "c:/data/pdf실습/GPT활용_파이썬기초 및_ 데이터분석 및 시각화.pdf"
✓ 0.0s
```

```
1 import fitz
2 doc = fitz.open(PDF_FILE_PATH)
3 ✓ for i, page in enumerate(doc):
4     | img = page.get_pixmap()
5     | img.save(f"c:/data/pdf실습/{i}.png")
```



5페이지 코드 결과자료

data > pdf실습



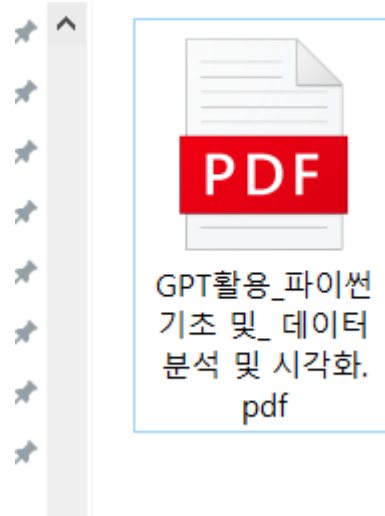
나

파이썬 폴더에서 png 확장자만 모두 삭제

[미션]
GPT에게 질문후
코드 작성하여 실행

결과자료

> data > pdf실습



Part2. [실습] 코드실행하고 이해하기 – pdf 에서 페이지 단위로 텍스트를 읽어내는 기능임.

for의 enumerate / print f 옵션의 쓰임새 알기

```
import fitz
doc = fitz.open(PDF_FILE_PATH)
for i, page in enumerate(doc):
    txt=page.get_text()
    print( f'▶Page {i+1:03d}-----')
    print(txt)○
```

포함하는 글자에 해당하는 if 조건 in 데이터: 에 대한 쓰임새 알기

```
import fitz
doc = fitz.open(PDF_FILE_PATH)
findWord='리스트'
for i, page in enumerate(doc):
    txt=page.get_text()
    if findWord in txt:
        print( f'▶Page {i+1:03d}-----')
```

'리스트'글자가 있다면

```
▶Page 000-----
파이썬 기초 및
2024년
강사 부성순 bakpak@empas.com
```

결과자료

```
▶Page 001-----
파이썬기초 및 데이터분석 및 시각화 (2024)
<목 차>
01. 파이썬 개
...
*GPT에 업로드한 이미지
- GPT분석 결과 -
*재미나 실행결과
```

클릭하면 모든 내용
다 display됨

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#)

결과자료

```
▶Page 002-----
▶Page 034-----
▶Page 035-----
▶Page 036-----
▶Page 037-----
▶Page 038-----
▶Page 039-----
▶Page 046-----
```

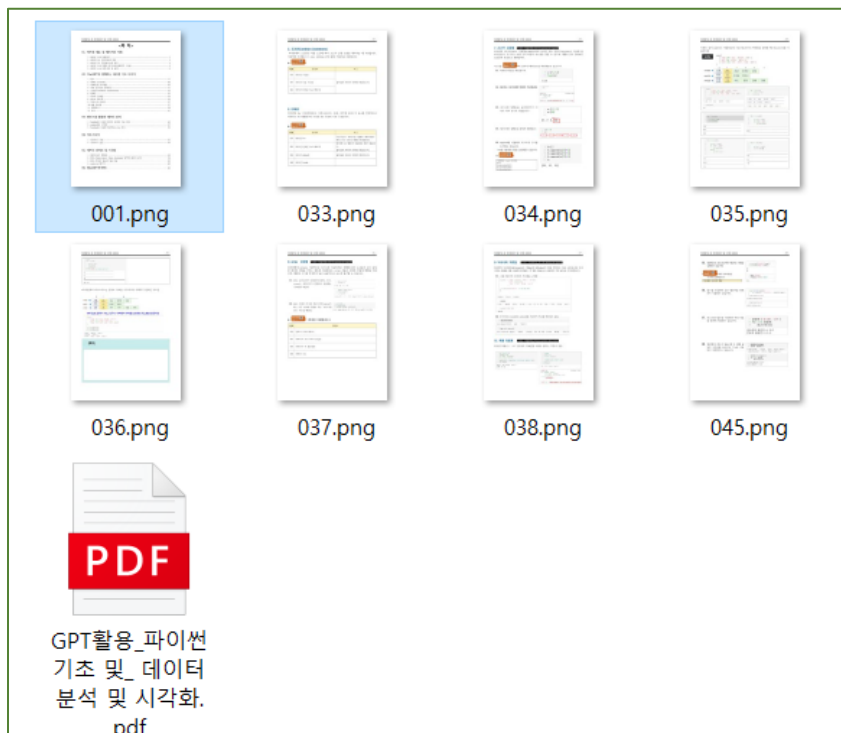


Part2. [미션] gpt또는 구글 검색하여 작업하기

[미션1]

'pdf' 페이지내의 '리스트'글자가
있는 페이지만 이미지로 저장

결과물



[미션2]

png 자료 모두 삭제

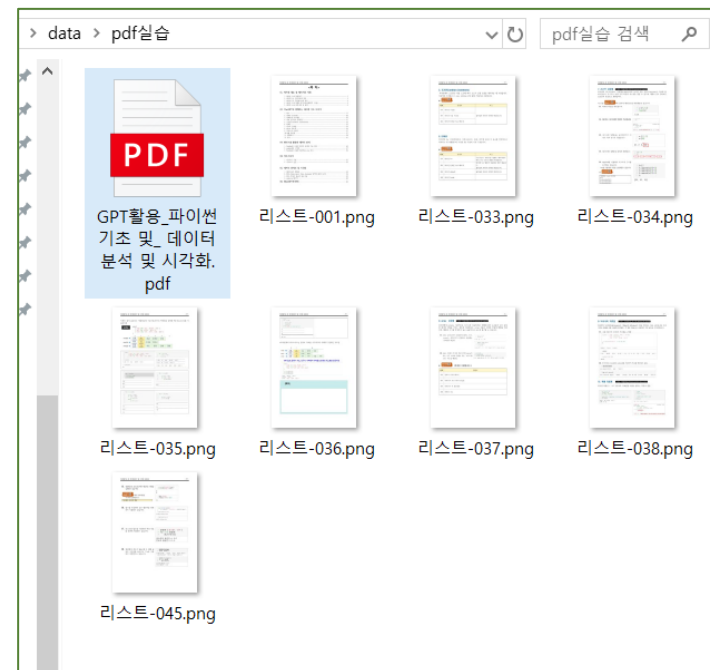
결과물



[미션3]

출력결과물이 '리스트-001.png'

결과물





Part2. [미션: 모듈 doc를 이용한 문제해결]

오픈 소스 Python PDF 파서 라이브러리 - gpt나 구글에서 검색이 원활치 않은 경우에는 doc문서를 이용해야함.

<https://products.documentprocessing.com/ko/parser/python/pymupdf/>

타 작업은 코드 복잡도가 있어, 위의 링크사이트에서 하단에 있는 [PDF에 텍스트 삽입] 코드만 doc 자료를 보면서 실습하여 봅니다.

PDF에 텍스트 삽입

아래 Python 코드 조각은 PDF 파일에 텍스트를 삽입하고 수정된 PDF를 te

```
1 # Import PyMuPDF
2 import fitz
3
4 # Open a PDF file
5 doc = fitz.open("documentprocessing.pdf")
6 page = doc[0] # Access the first page
7
8 # Define the starting point for the text
9 p = fitz.Point(75, 150)
10
11 # Define the text
12 text = "Some text"
13
14 # Insert the text on the page
15 rc = page.insert_text(p, text)
16
17 # Print the number of lines printed on the page
18 print("%i lines printed on page %i." % (rc, page.number))
19
20 # Save the modified PDF to a new file
21 doc.save("text.pdf")
```

코드 살펴보기



나

파이썬 fitz.Point(75, 150) 쉽게 해석



ChatGPT

파이썬 라이브러리 중 하나인 fitz.Point(75, 150)는 x좌표가 75이고 y좌표가 150인 점을 나타냅니다. 주로 PDF 문서에서 텍스트를 삽입하거나 수정하는 작업에서 사용됩니다.



나

x,y는 무엇



ChatGPT

x와 y는 일반적으로 2차원 공간에서 좌표를 나타냅니다.

코드 학습

----- Python PDF Test by 2024.05.31-----

결과: 코드 실행

3

pdf파일의 `p = fitz.Point(30,20)` 위치에 텍스트 내용 추가
단 한글은 별도 글꼴명령 추가해야함.

출력파일을 열어보면 첫페이지만 되어 있음.

Part2. [이해: for 를 이용한 각 페이지 마다 텍스트 추가하는 작업 반복 처리]



```
1 #####3
2 ## 맨 첫장만 추가하기
3 ## https://products.documentprocessing.com/ko/parser/python/pymupdf/
4 #####
5 import fitz
6
7 # Open some document, for example a PDF (could also be EPUB,
8 PDF_FILE_PATH = "c:/data/pdf실습/GPT활용_파이썬기초 및_ 데이터분
9 text = " -- pdf python test -- 2024.05.31 ---"
10 doc = fitz.open(PDF_FILE_PATH)
11 page=doc[0] #첫장
12
13 p = fitz.Point(30, 20)
14 rc = page.insert_text(p, text)
15 print("%i lines printed on page %i." % (rc, page.number))
16 doc.save("c:/data/pdf실습/text.pdf")
✓ 0.0s
```

9page 정답

결과: 모든 페이지 위에 글자 추가됨

모든 페이지
마다 txt
추가

-- pdf python test -- 2024.05.31 ---

파이썬기초 및 데이터분석 및 시각화 (2024)



구글에서 검색해 봅니다.

-- pdf python test -- 2024.05.31 ---

파이썬기초 및 데이터분석 및 시각화 (2024)

5. 파이썬 프로그램 설치 및 실행

▶ 방법1: 파이썬 프로그램을 설치 (<http://>)

```
1 # import package PyMuPDF
2 import fitz
3
4 # Open some document, for example a PDF (could also be EPUB, XPS, etc.)
5 PDF_FILE_PATH = "c:/data/pdf실습/GPT활용_파이썬기초 및_ 데이터분석 및 시각화.pdf"
6 text = " -- pdf python test -- 2024.05.31 ---"
7 doc = fitz.open(PDF_FILE_PATH)
8
9 for page in doc:
10     p = fitz.Point(30, 20)
11     rc = page.insert_text(p, text)
12     print("%i lines printed on page %i." % (rc, page.number))
13 doc.save("c:/data/pdf실습/text.pdf")
```



Part2. [필수작업 아님]

[작업]

png 자료 모두 삭제

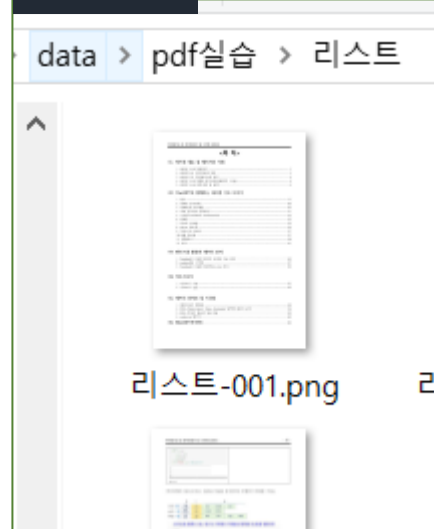
결과물



[작업]

검색하는 자료
에 해당하는 **폴더명**
제작하고 그 **폴더명**에 save

결과물



```
#####
```

```
## [작업] 텍스트 글자중에 포함,
```

```
## 폴더생성
```

```
## 파일명은 'c:/data/pdf샘플/리스트-001.png'
```

```
#####
```

```
import fitz
```

```
import os
```

```
PDF_FILE_PATH = "c:/data/pdf실습/GPT활용_파이썬기초 및_ 데이터분석 및 시각화.pdf"
```

```
doc = fitz.open(PDF_FILE_PATH)
```

```
findWord='리스트'
```

```
# 새로운 폴더를 생성할 경로 지정
```

```
folder_path = "c:/data/pdf실습/" + findWord
```

```
# 폴더가 존재하지 않는 경우에만 폴더를 생성
```

```
if not os.path.exists(folder_path):
```

```
    os.mkdir(folder_path)
```

```
    print(f"폴더 '{folder_path}'가 생성되었습니다.")
```

```
else:
```

```
    print(f"폴더 '{folder_path}'는 이미 존재합니다.")
```

```
for i, page in enumerate(doc):
```

```
    txt=page.get_text()
```

```
    if findWord in txt:
```

```
        print( f'▶ Page {i+1:03d}-----')
```

```
        img = page.get_pixmap()
```

```
        img.save(f"{folder_path}/{findWord}-{i:03d}.png")
```



Part2. [필수작업 아님: 함수에 대해 미리 이해하기]

폴더삭제 명령으로 변경

```
import fitz
import os
import shutil
```

def delete_folder(folder_path):

```
if os.path.exists(folder_path):
    shutil.rmtree(folder_path)
    print(f"폴더 '{folder_path}'가 삭제되었습니다.")
else:
    print(f"폴더 '{folder_path}'는 이미 존재하지 않습니다.")
```

def create_folder(folder_path):

```
if not os.path.exists(folder_path):
    os.mkdir(folder_path)
    print(f"폴더 '{folder_path}'가 생성되었습니다.")
else:
    print(f"폴더 '{folder_path}'는 이미 존재합니다.")
```

def extract_images_from_pdf(pdf_file_path, find_word, folder_path):

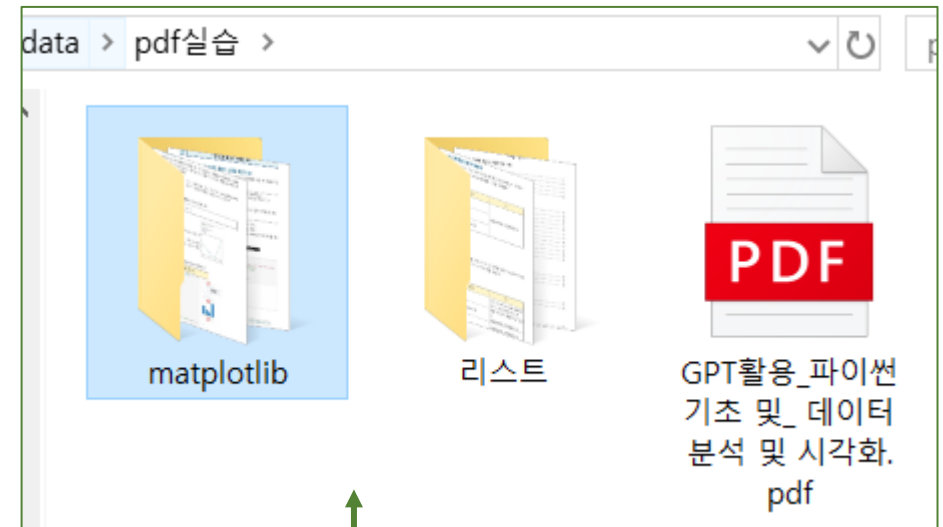
```
doc = fitz.open(pdf_file_path)
delete_folder(folder_path)
create_folder(folder_path)

# PDF 페이지를 순회하며 특정 단어가 있는 페이지를 이미지로 저장
for i, page in enumerate(doc):
    txt = page.get_text()
    if find_word in txt:
        print(f'▶ Page {i+1:03d}-----')
        img = page.get_pixmap()
        img.save(f"{folder_path}/{find_word}-{i:03d}.png")
```

함수 호출

```
PDF_FILE_PATH = "c:/data/pdf실습/GPT활용_파이썬기초 및 데이터분석 및 시각화.pdf"
findWord = 'matplotlib'
FOLDER_PATH = "c:/data/pdf실습/" + findWord
extract_images_from_pdf(PDF_FILE_PATH, findWord, FOLDER_PATH)
```

이 함수만 모아서
별도의 파일로
만들어서 000.py
로 저장하면
외부 모듈이 됨



findWord 별도 폴더가 작성됨



구글검색 ' GPT 할루시네이션 '

GPT 3.5

나
대한민국 대통령

ChatGPT
2022년 1월 현재, 대한민국 대통령은 문재인 대
시점과 달라질 수 있습니다. 최신 정보를 확인하

GPT 4,

2024년 기준으로 대한민국의 대통
대 대통령으로 재임 중입니다. 윤석
검찰총장을 역임한 바 있습니다. 그

GPT 3.5

나
부성순간온수

ChatGPT
"부성순간온수"는 '부성(浮聲)'이란 용어와 '순간(瞬間) 온수'가 합쳐진 표현입니다. "부성(浮聲)"은 물속
에서 소리가 떠다니는 것을 의미하며, "순간 온수"는 잠시 동안 뜨거운 물을 의미합니다. 이런 표현을
사용한다면, 뜨거운 물을 사용하다가도 순간적으로 차가운 물이 나오는 현상을 묘사하는 것으로 이해
될 수 있습니다.

나
문재인간온수

ChatGPT
"문재인간온수"라는 표현은 제가 알고 있는 지식 범위 내에서는 의미가 없습니다. "문재인"은 대한민국
의 전 대통령인 문재인의 이름입니다. 그러나 "간온수"라는 용어는 저에게는 익숙하지 않습니다. 이 표
현이 특정한 맥락이나 의미를 갖고 있다면 그것을 알려주시면 더 정확한 설명을 드릴 수 있을 것입니
다



Part2. [GPT 3,4 의 결과물이 다름에 대한 이해만 필요함, 강사 실습 내용만 보아도 됨]

- pdf에서 검색한 단어 리스트를 이용해서 몇 페이지에 있는 지를 딕셔너리 형태로 구성 하고자 함
- import fitz 모듈은 GPT 3.5와 GPT4버전의 code 가 다르게 제공됨, 이 모듈은 gpt4 이상 권장함.

제약조건

```
import fitz
PDF_FILE_PATH = "c:/data/pdf실습/GPT활용_파이썬기초 및_ 데이터분석 및 시각화.pdf"
findWordList=['matplotlib', 'gpt', 'index']
for i, page in enumerate(doc):
    txt=page.get_text()
    if findWord in txt:
```

출력

```
[{'matplotlib': [2, 7, 15, 44, 49, 69]},
{'gpt': [7, 11, 70, 85, 87, 88]},
{'index': [3, 55]}]
```

Part3. 파이썬 기본기 익히기



Point

- 파이썬이 다양한 작업에서 모두 기본이 되는 문법
- 특히 코딩테스트(코테)와 같은 특수분야에서는 아주 중요함

문법공부만 원한다면: 프로그래머스, 백준

→ 문법을 심도 있게 학습할때는 보통 알고리즘에 대한 이해도가 있어야 가능

- 하지만 파이썬은 모듈중심임
- (프로그램 상황에 따라 다르며, 코드는 정답이 없음, 결과와 최적화만 확인함)
 - 예를 들어서 Numpy에서는 최대한 for를 사용하지 않고 작업하는것을 권장함
<https://hyeonchan523.tistory.com/13> https://rphabet.github.io/posts/python_numpy_loop/
 - 판다스에서도 for를 사용하지 않고도 다양한 반복문이 가능함
<https://aldente0630.github.io/data-science/2018/08/05/a-beginners-guide-to-optimizing-pandas-code-for-speed.html>

GPT
에게
질문

파이썬에서 한글은 왜 깨지는가

인코딩이란

ASCII 코드는

'a'가 ascii코드로 얼마
인지 출력하는 파이
썬 코드

주피터노트북 추천 마크다운

html언어로

파이썬 자료형 설명과 샘플

각 자료형의 쓰임새 샘플

정수 계산 샘플

문자열 처리 샘플

문자열길이 len을 날코딩으로

대소문자 변환함수를 날코딩으로

리스트 처리 샘플

딕셔너리 처리 샘플

List Comprehension 예제를 날코딩과 함께 설명

리스트와 딕셔너리 for 샘플을 다양하게