

종합실습_배경

분석 배경

최근들어 넷플릭스, 왓챠, 리디북스 등과 같은 콘텐츠 기업에서는 특정 사용자가 관심을 가질 만한 정보 (영화, 책) 등을 추천하는 취향 저격 엔진을 만들어 소비자를 공략하고 있습니다. 본 분석은 웹에 있는 책 데이터를 바탕으로 목차, 평점 및 리뷰 데이터를 바탕으로 책을 추천하는 분석을 진행합니다.

NETFLIX

RIDIBOOKS

분석 방향

책 키워드를 검색하면 yes24, 네이버 서적에서 검색 후
키워드가 비슷하고 (목차 스크래핑) 평점, 리뷰가 좋은 책을 추천한다

종합실습_분석

분석 방향

- : 실제 웹에서 원하는 데이터를 가져오고, 분석이 가능하게 확인하고 정리함
- : 가져온 데이터를 텍스트 전처리를 통해 분석이 가능하게 만듦
- : 시각화와 분석을 통해 비즈니스에 활용할 수 있는 인사이트를 도출함

분석 기법

- : 교육 과정에서 학습한 스크래핑 방법과 텍스트 전처리를 활용하여 분석함



스크래핑을 활용한 데이터 수집



KoNLPy를 활용한 데이터 전처리



워드 클라우드



TF-IDF



코사인 유사도



기타 관심 있는 기법

종합실습_프로젝트 개요

■ 도서 추천 시스템 만들기

- : 네이버 도서 API를 사용해서 Data 수집
- : 수집한 데이터를 분석하기 위한 전처리 (불용어/어간추출)
- : TF-IDF와 Word2Vec을 사용해서 각각 모델링



네이버 도서 정보 스크래핑

NAVER 책에서 도서 정보를 스크래핑

<https://book.naver.com/>

NAVER 책 eBook 오디오북 파이썬 검색 통합검색 상세검색 > 우중은 > 99+

분야별 찾기 > 책 홈 베스트셀러 지서재 출판사공간

'파이썬'에 대한 검색결과

☒ 책 검색 ☐ 본문 검색

전체 (962건)

- 화제의 책

· 신간 (24)

- 분야별

컴퓨터/IT (900)

과학/공학 (27)

어린이 (14)

잡지 (14)

경제/경영 (13)

더보기

책 검색 (1-20 / 962건) | 전체

☒ 관련도순 ☒ 출간일순 ☒ 판매량순

☐ eBook ☐ 오디오북 ☐ 구입가능도서

파이썬 (비전공자를위한 파이썬 기초 입문서,한 권으로 끝내는)
김명호 저 | 에듀웨이 | 2019.05.25
★★★★★ 10 | 네티즌리뷰 4건 | 도서구매 48,000원 → 16,200원(-10%)
소개 막막했던 파이썬, 이제 한 권으로 기초를 배우는 파이썬 사용 설명서!전 세계적으로 유명한 구글(GOOGLE)이 자신의 많은 서비스를 파이썬으로 제공하고 있습니다. 이것은 구글의 뛰어난 서비스를 개인의 비즈니스에 쉽게 이용할 수 있다는 의미입니다.이러한 발전에 맞춰 파이썬 문법의 기초를...

모두의 파이썬 (20일 만에 배우는 프로그래밍 기초)
이승찬 저 | 길벗 | 2018.10.09
★★★★★ 0 | 네티즌리뷰 1건 | 도서구매 42,000원 → 10,800원(-10%) eBook구매 7,560원
소개 파이썬 분야 & 어린이 코딩 교육 분야 2년 연속 베스트셀러! 개정판 출간! 즐겁게 시작하는 나의 첫 프로그래밍! 프로그래밍을 한 번도 해본 적이 없어도 괜찮다. 파이썬이 무엇인지 몰라도... 어느새 파이썬 프로그램으로 멋진 그림을 그리고, 계산을 하고, 간단한 게임을 만들고, 수학 문제를...

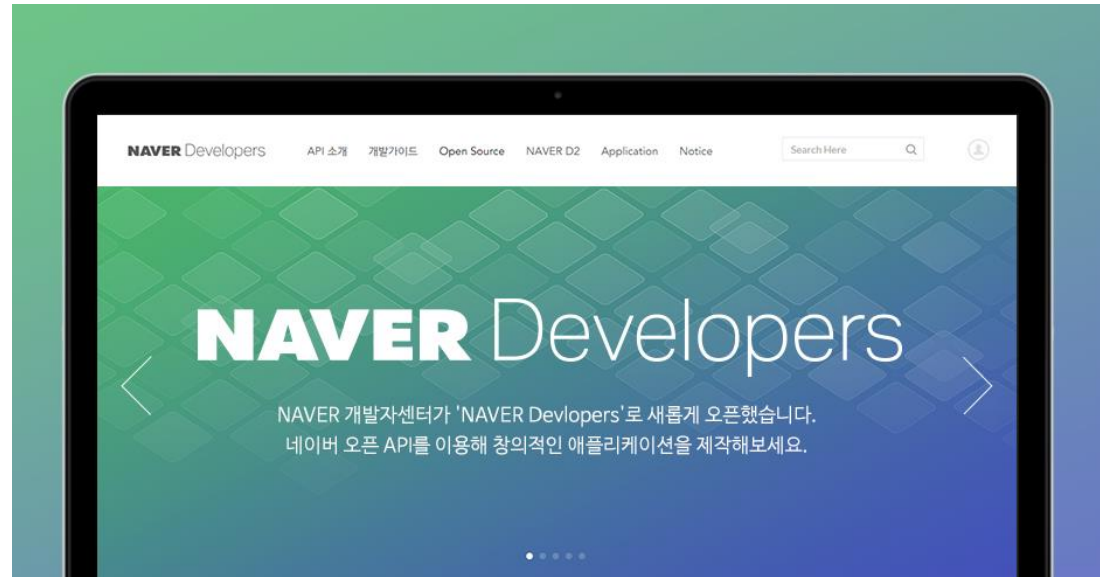
파이썬으로 데이터 주무르기 (독특한 예제를 통해 배우는 데이터 분석 입문)
민형기 저 | 비제이퍼블릭 | 2017.12.29
★★★★★ 0 | 네티즌리뷰 5건 | 도서구매 27,500원 → 24,750원(-10%) eBook구매 19,800원
소개 있을 파이썬 기초 문법책과 같은 내용이 아닌, 데이터 분석이라는 특별한 분야에서 초보를 위해 처음부터 끝까지 파이썬으로 진행되는 과정을 다룹니다. 서울시 범죄 현황 분석, 셀프 주유소 가격 정보 분석, 19대 대선 결과 분석 등 흥미 있는 목표를 이루기 위해서 파이썬의 기초를 익히고...

NAVER Open API 사용

네이버 개발자 센터

- : 다량의 정보를 스크래핑하기 위해 네이버가 제공하는 기능을 사용
- : 단순 스크래핑을 하면 Naver에서 IP를 차단
- : 기본적으로 하루 25,000 건의 데이터를 받을 수 있음

<https://developers.naver.com/docs/search/book/>



네이버 개발자 등록 (1/3)

NAVER Developers 가입

: 휴대폰 인증 필요

: <https://developers.naver.com/apps/#/wizard/register>

약관동의

> 계정 정보 등록

> 애플리케이션 등록

API 이용약관

가. 일반조건

1. 목적

NAVER API서비스 이용약관(이하 '본 약관'이라 합니다)은 네이버주식회사(이하 '회사'라 합니다)와 이용자 간에 회사가 제공하는 API 서비스의 제공 및 이용과 관련한 회사와 이용자의 권리, 의무 및 책임에 관한 제반 사항과 기타 필요한 사항을 구체적으로 규정함을 목적으로 합니다.

계정 설정

① 휴대폰

휴대폰 인증

초기 등록을 위해서 개인 인증이 필요합니다.
인증은 10분간 유효 합니다.

이메일

mm*****@gmail.com

이메일 정보 변경

회사 이름

프리랜서

☐ 소속 회사 이름은 제휴 접수 및 검토 등을 위해 사용되며, 서비스 이용 기간 동안 보관됩니다.

&"'() 문자는 허용하지 않습니다.

네이버 개발자 등록 (2/3)

애플리케이션 이름 / 사용 API / 로그인 환경 입력

애플리케이션 이름 ⇄	<div>도서정보</div> <ul style="list-style-type: none">• 네이버 아이디로 로그인할 때 사용자에게 표시되는 이름이므로 가급적 10자 이내의 간결한 이름을 사용해주세요.• 40자 이내의 영문, 한글, 숫자, 공백문자, "-", "_" 만 입력 가능합니다.
카테고리	<div>기타 ▼</div>
사용 API ⇄	<div>선택하세요. ▼</div> <div>검색</div> <div>×</div>
비로그인 오픈 API 서비스 환경	<div>환경 추가 ▼</div> <div><div>WEB 설정</div><div>×</div><div>^</div><div>웹 서비스 URL (최대 10개)</div><div>http://localhost:8080</div><div>+</div><ul style="list-style-type: none">• 텍스트 폼 우측 끝의 '+' 버튼을 누르면 행이 추가되며, '-' 버튼을 누르면 행이 삭제됩니다.• http와 https는 구분하지 않습니다.• www는 빼고 입력해 주세요. 예) http://naver.com• 서브 도메인이 있으면 대표 도메인명만 입력해 주세요. (예: http://naver.com)• 하이브리드 앱은 location.href 객체 출력 값을 입력하면 됩니다. (예: file://로컬 URI)</div>

네이버 개발자 등록 (3/3)

클라이언트 ID와 Secret ID를 확인

: API 호출 시 필요

도서정보

개요	API 설정	멤버관리	로그인 통계
----	--------	------	--------

애플리케이션 정보

Client ID	jowFygReKFKuBEtm2Ss8
Client Secret
	<div>보기</div>

API 동작 확인

본인의 ID를 사용해서 데이터 요청

: JSON 형태 (Dictionary)로 데이터를 출력

```
import requests

client_id = "YOUR_CLIENT_ID"
Client_secret = "YOUR_SECRET_ID"

url = "https://openapi.naver.com/v1/search/book.json?query=파이썬"
headers = {
    "X-Naver-Client-Id": client_id,
    "X-Naver-Client-Secret": client_secret,
}

resp = requests.get(url, headers=headers)
print(resp.json())
```

API 결과 확인 (1/2)

Return 값의 의미

요청 변수	Type	상세 설명
lastBuildDate	datetime	검색 결과를 생성한 시간
total	integer	검색 결과 문서의 총 개수
start	integer	검색 결과 문서 중, 문서의 시작점
display	integer	검색된 검색 결과의 개수
items	-	개별 검색 결과
title	string	검색 결과 문서의 제목
link	string	검색 결과 문서의 하이퍼텍스트 link
image	string	썸네일 이미지의 URL
author	string	저자 정보
price	integer	정가 정보

API 결과 확인 (2/2)

requests 결과 예제

```
{'lastBuildDate': 'Mon, 27 Jan 2020 17:10:12 +0900',  
'total': 3713,  
'start': 1,  
'display': 10,  
  
'items': [{'title': '<b>Python</b>으로 시작하는 빅데이터 분석 및 인공지능',  
'link': 'http://book.naver.com/bookdb/book_detail.php?bid=15407061',  
'image': 'https://bookthumb-phinf.pstatic.net/cover/154/070/15407061.jpg?type=m1&udate=20191023',  
'author': '조준모',  
'price': '24000',  
'discount': '21600',  
'publisher': '인피니티북스',  
'pubdate': '20190830',  
'isbn': '1185578463 9791185578460',  
'description': '이를 위해 현재 각광을 받고 있는 구글의 주피터 노트북(JUPITER NOTEBOOK)환경에서 넘파이(NUMPY)와 판다스(PANDAS)를 사용할 수 있는 파이썬(<b>PYTHON</b>)을 실습을 통해 익힌다.\n\n쉽고 간단한 예제를 통한 인공지능 세계로의 입문!\n\n인공지능은 전통적인 인공지능에서 출발하여 현재는 크게 기계학습과 딥러닝으로...'},  
  
{'title': '누구나 <b>PYTHON</b> 너도 데이터 가지고 놀 수 있어!',  
'link': 'http://book.naver.com/bookdb/book_detail.php?bid=16109062',  
'image': 'https://bookthumb-phinf.pstatic.net/cover/161/090/16109062.jpg?type=m1&udate=20200107',  
'author': '민형기',  
'price': '14500',  
'discount': '',  
'publisher': '잇플ITPLE',  
'pubdate': '20200108',  
'isbn': '1190283719 9791190283717',  
'description': '이 책은 한 권으로 모든 것을 설명하지 않습니다. 상세한 문법, 언어적 기술은 다른 멋지고 잘 만들어진 책에 맡기고, 우리 책은 난이도별 응용 예제에 집중했다고 생각해 주시면 됩니다.\n\n먼저 설치부터 시작해서, 데이터를 다루는데 필요한 Pandas라는 모듈, 데이터를 시각화하는데 필요한 matplotlib라는... }]
```

DataFrame으로 변환

items 항목의 데이터를 DataFrame 객체로 저장

```
import requests
import pandas as pd
```

코드 생략

```
resp = requests.get(url, headers=headers)
resp = resp.json()
data = resp['items']
```

DataFrame

```
df = pd.DataFrame(data=data)
print(df)
```

Index	title	link	image	author	price	discount	publisher	pubdate	isbn	description
0	Python...	http://boo...	https://bo...	조준모	24000	21600	인피니티북...	20190830	1185578463...	이를 위해 현재 각광을 받고 있는 구글의 주피터 노트...
0	누구나 ...	http://boo...	https://bo...	민형기	14500		잇플ITPLE	20200108	1190283719...	이 책은 한 권으로 모든 것을 설명하지 않습니다. 상세...
0	python...	http://boo...	https://bo...	김동근	26000	23400	가메	20180828	8980782993...	PYTHON으로 OPENCV 프로그래밍을 구현하였습니...
0	스타트 파...	http://boo...	https://bo...	황재호	19000	17100	코딩스쿨	20180901	1196440905...	학생과 성인 초보자를 위한 최상의 파이썬 학습서!온라...
0	Head First...	http://boo...	https://bo...	폴 배리	32000	28800	한빛미디어	20171204	1162240288...	또한 80개가 넘는 실습을 직접 해보면서 더 잘 배우고 ...
0	최적합 전...	http://boo...	https://bo...	장대경	22000	19800	성안당	20190424	8931555636...	본 책의 집필 의도는 답을 찾는 방법보다는 코드를 이...
0	작심 3일 ...	http://boo...	https://bo...	황덕창	18000	16200	스포츠라잇...	20190425	1187431168...	문과 전공 초보자를 본격 배려한 파이썬 기초 3일 특강...
0	Python...	http://boo...	https://bo...	요우시 리우	25000	22500	에이콘출판	20180907	116175203X...	▶ 파이썬 머신 러닝에 관한 내용을 담은 전문서적입니다...
0	사물인터넷...	http://boo...	https://bo...	박진현 이...	26000	23400	아이코스	20190823	1164260952...	물론, 이 과정에 필요한 모든 하드웨어 지식들도 포함...
0	Electromag...	http://boo...	https://bo...	Houle Jenn...	175540	168500	Wiley-IEEE...	20200122	1119565804...	

실습

■ 실습 (자연어 처리)

Title과 Description의 태그와 특수문자를 제거하라.
방법#1 문자열 치환 (Replace 메서드) 사용
방법#2 정규식 사용

	title	link	image	author	price	discount	publisher	pubdate
0	Python으로 시작하는 빅데이터 분석 및 인공지능	http://book.naver.com/bookdb/book_detail.php?b...	https://bookthumb-phinf.pstatic.net/cover/154/...	조준모	24000	21600	인피니티북스	20190830 1185597911855
1	누구나 PYTHON도 데이터 가지고 놀 수 있어!	http://book.naver.com/bookdb/book_detail.php?b...	https://bookthumb-phinf.pstatic.net/cover/161/...	민형기	14500		잇플ITPLE	20200108 1190297911902
2	python으로 배우는 OpenCV 프로그래밍	http://book.naver.com/bookdb/book_detail.php?b...	https://bookthumb-phinf.pstatic.net/cover/139/...	김동근	26000	23400	가메	20180828 8980797889807
	://book.naver.com/bookdb/book_detail.php?b...	https://bookthumb-phinf.pstatic.net/cover/139/...	요우시 리우	25000	22500	에이콘출판	20180907 116175203X9791161752037	▶ 파이썬 머신러닝에 관한 내용을 담은 전문서적입니다.
	://book.naver.com/bookdb/book_detail.php?b...	https://bookthumb-phinf.pstatic.net/cover/153/...	박진현 이동수	26000	23400	아이북스	20190823 11642609529791164260959	물론, 이 과정에 필요한 모든 하드웨어 지식들도 포함해서 말이죠. \n\n또한 소스 ...

실습 (정답)

```
import requests
import pandas as pd
from bs4 import BeautifulSoup
import re

# 코드 생략
# 태그 제거
def preprocessing(text):
    # HTML 태그 제거
    soup = BeautifulSoup(text, "html5lib")
    clean_text = soup.get_text()

    # 기호 제거
    p = re.compile("\W+")
    s = p.sub(" ", clean_text)
    return s

# DataFrame
df = pd.DataFrame(data=data)
df['title'] = df['title'].apply(preprocessing)
df['description'] = df['description'].apply(preprocessing)
print(df)
```

데이터 요청 형식

```
url = "https://openapi.naver.com/v1/search/book.json"
params = {
    "query": "파이썬",
    "start": 1,
    "display": 100
}
resp = requests.get(url, params=params, headers=headers)
print(resp.json())
```

API 요청 시 추가 옵션을 지정할 수 있음

요청 변수	Type	상세 설명
query	string	검색을 원하는 문자열
display	integer	검색 결과 출력 건수 지정 (기본 10 / 최대 100)
start	integer	검색 시작 위치 (기본 1 / 최대 1000)
sort	string	정렬 옵션 (기본 sim) - sim(유사도순), date(출간일순), count(판매량순)

실습

■ 실습 (자연어 처리)

검색 결과를 DataFrame으로 반환하는 함수를 정의하라.

최대 1000개 까지만 조회 가능

response의 total 키값에 전체 데이터 개수 저장

requests의 start 파라미터에 검색 시작위치 지정

```
>> df = request_book_by_query("파이썬")  
>> print( len(df) )  
1000
```


실습 (정답)

```
def request_book_by_query(query="파이썬"):
    client_id = "YOUR_CLIENT_ID"
    client_secret = "YOUR_SECRET_ID"

    url = "https://openapi.naver.com/v1/search/book.json"
    headers = {
        "X-Naver-Client-id": client_id,
        "X-Naver-Client-Secret": client_secret,
    }

    start = 1
    df_list = []
```

코드 계속

실습 (정답)

```
def request_book_by_query(query="파이썬"):
    # 코드 생략
    while start <= 1000:
        params = {
            "query": query,
            "start": start,
            "display": 100
        }

        resp = requests.get(url, params=params, headers=headers)
        print(resp.url)
        resp = resp.json()
        data = resp['items']
        start += len(data)

    # DataFrame
    df = pd.DataFrame(data=data)
    print(df)
    df['title'] = df['title'].apply(preprocessing)
    df['description'] = df['description'].apply(preprocessing)
    df_list.append(df)

    # break
    if len(data) != 100:
        break

    return pd.concat(df_list)
```

Word2Vec (1/4)

벡터화할 데이터를 하나의 Column으로 병합

: 스크래핑한 데이터의 제목과 설명을 Series로 가져옴

```
df = request_book_by_query("파이썬")  
  
target = df[ 'title' ] + ' ' + df[ 'description' ]
```

title	author	description	target
파이썬 비...	김명호	막막했던 ...	파이썬 비전공자를위한 파이썬 기초 입문서 한 권으로 끝내는 막막했
모두의 파...	이승찬	파이썬 분...	모두의 파이썬 20일 만에 배우는 프로그래밍 기초 파이썬 분야 어린C
파이썬으로...	민형기	있을 파이...	파이썬으로 데이터 주무르기 독특한 예제를 통해 배우는 데이터 분석
혼자 공부...	윤인성	1 1 과외하...	혼자 공부하는 파이썬 파이썬 최신 버전 반영 1 1 과외하듯 배우는 I
Do it 점프...	박응용	파이썬 4년...	Do it 점프 투 파이썬 파이썬 4년 연속 베스트셀러 1위 Do it 점프 투

Word2Vec (2/4)

명사만 추출 후 리스트로 변환

```
from konlpy.tag import Okt

def get_nouns(v) :
    okt = Okt()
    token = okt.pos(v, stem=True, norm=True)

    result = ""
    for x in token:
        if x[1] == "Noun" :
            result += x[0] + ' '
    return result
```

```
target = target.apply(get_nouns)
target = target.str.split()
print(target.head())
```

Index	target
0	['파이썬', '비', '전공자', '위', '파이썬', '기초', '입문', '권', '파이썬', '이제', '권', '기초...
1	['모두', '파이썬', '프로그래밍', '기초', '파이썬', '분야', '어린이', '코딩', '교육', '분야', '...
2	['파이썬', '데이터', '주무르기', '예제', '통해', '데이터', '분석', '입문', '파이썬', '기초', '...
3	['혼자', '공부', '파이썬', '파이썬', '최신', '버전', '반영', '과외', '듯', '파이썬', '프로그래...
4	['점프', '투', '파이썬', '파이썬', '연속', '베스트셀러', '위', '점프', '투', '파이썬', '전면', '...

Word2Vec (3/4)

트레이닝 및 연관 단어 조회

: Kernel Size 300 / Window 10

```
from gensim.models import Word2Vec

model = Word2Vec(target.to_list(), size=300, window=10, min_count=1)
model.init_sims(replace=True)

result = model.wv.most_similar("알고리즘", topn=10)
print("/".join([x[0] for x in result]))
```

구현/학습/신경망/핵심/이해/예제/수학/실습/코드/기초

✓ 추가: 불용어를 추가해서 검색의 질을 향상

Word2Vec (4/4)

저자 정보 추가 및 조회

: 구현의 편의를 위해 저자를 구분하기 위한 정보 추가

```
target = target.apply(get_nouns)
target = '(저)' + df['author'] + ' ' + target
target = target.str.split()

model = Word2Vec(target.to_list(), size=300, window=10, min_count=1)
model.init_sims(replace=True)

result = model.wv.most_similar("(저)조대표", topn=1000)

p = re.compile("\(저\)\\w+")
for x in result:
    m = p.findall(x[0])
    if len(m) > 0 :
        print(m[0])
```

```
(저)제이슨
(저)마이클
(저)마크
(저)존
(저)크리스
(저)마이크로소프트웨어
(저)이강성
(저)이호준
```

TF-IDF (1/4)

문서 단어 행렬(Document-Term Matrix, DTM)로 변환

```
from sklearn.feature_extraction.text import CountVectorizer
```

```
df = request_book_by_query("파이썬")
```

```
target = df['title'] + ' ' + df['description']
```

```
target = target.apply(get_nouns)
```

```
vector = CountVectorizer()
```

```
dtm = vector.fit_transform(target.to_list()).toarray()
```

```
print(dtm)
```

```
print(vector.vocabulary_)
```

```
print(len(vector.vocabulary_))
```

target에는 리스트에 문자열 데이터가 저장 됨

```
[  
    "파이썬 비 전공자 위 파이썬 기초 입문",  
    "모두 파이썬 프로그래밍 기초 파이썬 분야",  
    "파이썬 데이터 주무르기 예제",  
    ...  
]
```

```
[[0 0 0 ... 0 0 0]
```

```
[0 0 0 ... 0 0 0]
```

```
[0 1 0 ... 0 0 0]
```

```
...
```

```
[0 0 0 ... 0 0 0]
```

```
[0 0 0 ... 0 0 0]
```

```
[0 0 0 ... 0 0 0]]
```

```
{'파이썬': 2515, '전공자': 1941, '기초': 324, '입문': 1847, '이제': 1771, '사용': 1045, '설명': 1148, '세계': 1160, '구글':  
220, '자신': 1876, '서비스': 1122, '제공': 2029, '개인': 67, '비즈니스': 1020, '이용': 1766, '의미': 1733, '발전': 857, '문법':  
800,}
```

TF-IDF (2/4)

사이킷런의 TF-IDF를 자동 계산해주는 TfidfVectorizer

```
from sklearn.feature_extraction.text import TfidfVectorizer

df = request_book_by_query("파이썬")

target = df['title'] + ' ' + df['description']
target = target.apply(get_nouns)

tfidf = TfidfVectorizer()
tfidf = tfidf.fit_transform(target).toarray()
print(tfidf)
```

```
[[0. 0. 0. ... 0. 0. 0. ]
 [0. 0. 0. ... 0. 0. 0. ]
 [0. 0.18383122 0. ... 0. 0. 0. ]
 ...
 [0. 0. 0. ... 0. 0. 0. ]
 [0. 0. 0. ... 0. 0. 0. ]
 [0. 0. 0. ... 0. 0. 0. ]]
```


TF-IDF (3/4)

Cosine 유사도

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import linear_kernel

df = request_book_by_query("파이썬")

target = df['title'] + ' ' + df['description']
target = target.apply(get_nouns)

tfidf = TfidfVectorizer()
tfidf = tfidf.fit_transform(target).toarray()

cosine_similar = linear_kernel(tfidf, tfidf)
print(tfidf)
```

```
[[1.      0.06404185 0.11341927 ... 0.      0.      0.      ]
 [0.06404185 1.      0.08254748 ... 0.      0.      0.      ]
 [0.11341927 0.08254748 1.      ... 0.      0.      0.      ]
 ...
 [0.      0.      0.      ... 0.      0.      0.      ]
 [0.      0.      0.      ... 0.      0.      0.      ]
 [0.      0.      0.      ... 0.      0.      0.      ]]
```

TF-IDF (4/4)

조대표의 책과 가장 유사한 책은?

```
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics.pairwise import linear_kernel
```

```
df = request_book_by_query("파이썬")
```

```
target = df['title'] + ' ' + df['description']
target = target.apply(get_nouns)
```

```
tfidf = TfidfVectorizer()
tfidf = tfidf.fit_transform(target).toarray()
cosine_similar = linear_kernel(tfidf, tfidf)
```

```
cond = target['author'] == "조대표"
index = target[cond].index[0]
target['cosine'] = cosine_similar[index]
```

```
result = target.sort_values('cosine', ascending=False)
print(result.head(5))
```

title	author	description	cosine
파이썬으로 배우는 알고리즘 트레이...	조대표	파이썬 알고리즘 트레이딩 은...	1.000000
파이썬으로 배우는 알고리즘 트레이...	조대표	파이썬 알고리즘 트레이딩 은...	1.000000
파이썬으로 만드는 암호화폐 자동 ...	박재현	트레이딩 위 데이터 저장 가...	0.446243
알고리즘과 파이썬	채진석	이 책 알고리즘 파이썬 이론	0.346030
모두의 알고리즘 with 파이썬 컴퓨...	이승찬	모두 알고리즘 파이썬 은 주...	0.329432
파이썬으로 배우는 실전 알고리즘 ...	마시모 디 ...	이 책 알고리즘 분석 디자인 ...	0.279847
퀀트 전략 파이썬으로 세워라 누구...	박준규	파이썬 활용 퀀트 주식 투자 ...	0.229412

데이터에 중복이 존재 함
-> 중복 제거를 위한 디버깅

Description이 비어있는 경우 분석에서 제외