

/* elice */

파이썬 크롤링

여러 페이지 크롤링하기



김경민 선생님

Query

Query

🔒 sports.donga.com/ent?p=1

🔒 sports.donga.com/ent?p=21

🔒 sports.donga.com/ent?p=41

이 뉴스 웹사이트는 각 페이지의 URL에서 **p=(숫자)** 부분이

20씩 증가하고 있는 규칙이 있습니다.

이 사이트에서 여러 페이지를 크롤링하려면 어떻게 해야 할까요?

Query

```
for i in range(0, 5) :  
    url = "http://sports.donga.com/Enter?p="+str((i*20+1))  
    ...
```

쉬운 방법으로는 URL을 문자열 연산으로 처리하여
새로운 URL을 얻는 것입니다.

Query

하지만, URL의 **query(쿼리)**를 이용하면
이 작업을 더 효과적으로 할 수 있습니다.

Query

웹 서버에 GET 요청을 보낼 때

조건에 맞는 정보를 표현하기 위한 변수

예) 번호가 1번인 학생을 보여줘라

전체 기사 중 페이지가 21인 기사들을 보여줘라

Query

🔒 google.com/search?q=elice

google에 'elice'를 검색한 결과입니다.

q라는 변수에 elice라는 값이 담겨,

전체 데이터 중 elice라는 키워드로 검색한 결과만을 보여줍니다.

Query

 movie.naver.com/movie/bi/mi/basic.nhn?code=168058

네이버 영화 서비스에서 특정 영화를 클릭하면,

code라는 변수에 영화 코드가 담겨

해당 영화에 대한 정보를 보여줍니다.

requests 라이브러리

```
url = "https://www.google.com/search"  
result = requests.get(url, params = {'q': 'elice'})
```

requests의 get 메소드로 GET 요청을 보낼 때

params 매개변수에 **딕셔너리**를 전달함으로써

쿼리를 지정할 수 있습니다.

requests 라이브러리

```
code = ... # 영화 코드에 대한 정보를 얻는다.  
result = requests.get(url, params = {'movie':code})
```

전체 영화 데이터에서 영화 코드에 대한 정보를 찾고,
다시 requests를 이용하여 특정 영화에 대한 정보를
얻는 요청을 할 수 있습니다.

Tag Attribute

태그와 속성

```
<div class="elice" id="title">제목</div>
```

태그 속성 속성

HTML에는 여러 종류의 태그와,
태그에 특정 기능이나 유형을 적용하는 속성이 있습니다.

태그와 속성

```
div = soup.find("div")  
print(div.attrs)
```

어떤 태그의 속성이 무엇이 있는지 확인할 때는
attrs 멤버변수를 출력합니다.

태그와 속성

```
print(div['class'])
```

attrs 딕셔너리의 키로 인덱싱하여,
태그의 속성에 접근할 수 있습니다.

href 속성

```
<a href="https...">기사 제목</a>
```

a 태그는 하이퍼링크를 걸어주는 태그로써
이동할 URL을 href 속성에 담고 있습니다.

href 속성

```
a = soup.find("a")  
href_url = a["href"]
```

위와 같이 href 속성을 이용하여
웹 페이지에 존재하는 하이퍼링크의 URL을 얻을 수 있습니다.

Children, Name

Children, Name

웹 사이트의 구조가 복잡한 경우

다양한 옵션을 적용해야 할 수도 있습니다.

children은 어떤 태그가 포함하고 있는 태그를,

name은 어떤 태그의 이름을 의미하는 속성입니다.

Children

```
<div>  
  <span>span1</span>  
  <span>span2</span>  
  <p>p tag</p>  
  <img ... />  
</div>
```

옆의 div 태그는
여러 태그들을 갖고 있습니다.

Children

beautifulsoup의 **children** 속성으로

어떤 태그가 **포함하고 있는 태그**들도 조회할 수 있습니다.

Children

```
soup.find("div").children
```

#span, p, img 태그를 갖는 리스트를 얻습니다.

위의 코드는 어떤 div 태그를 찾고,

그 div 태그에 **포함된 태그들의 리스트**를 얻는 코드입니다.

Name

```
children = soup.find("div").children  
for child in children :  
    print(child.name)  
  
# span, span, p, img가 각각 출력됩니다.
```

어떤 태그의 이름을 알고 싶다면 name 속성을 이용할 수 있습니다.

태그가 존재하지 않는 경우 None 값을 얻습니다.

실전 크롤링

실전 크롤링

배운 내용으로 크롤링 코드를 직접 작성해보도록 하겠습니다.

여러 페이지의 기사 제목 수집하기



'소녀의 세계' 권현빈, 강렬한 첫 등장...아린과 훈훈 케미

'소녀의 세계' 권현빈, 강렬한 첫 등장...아린과 훈훈 케미 '소녀의 세계' 권현빈이 오마이걸 아린과 귀여운 케미를 선보였다. 권현빈은 지난 22일 공개된 tVN D 웹드라마 '소녀의 세계' 1화 에필로그에 ...

2020-04-23 09:43:00



[TV북마크] '기막힌 유산' 강세정X신정윤, 심상치 않은 재회...박...

[TV북마크] '기막힌 유산' 강세정X신정윤, 심상치 않은 재회...박인환 재산정리 어제(22일) KBS1 새 저녁 일일드라마 '기막힌 유산'(연출 김형일 극본 김경희) 3회에서는 자식들에게 증여를 준비하는 박...

2020-04-23 09:43:00

동아스포츠의 연예부 기사의
제목 부분을 크롤링해보려고 합니다.

여러 페이지의 기사 제목 수집하기



[DA:차트] 조정석 '아로하', 16주차 가온차트서 2관왕 영예

배우 조정석의 '아로하'가 16주차 가온차트에서 2관왕을 차지했다. 가온차트를 운영하는 사단법인 한국음악콘텐츠협회는, "16주차(2020.04.12~2020.04.18) 디지털차트, 스트리밍차트에서 슬기로운 의...

2020-04-23 09:28:00

1 2 3 4 5 6 7 8 9 10 >

사이트의 하단에서 버튼을 눌러 페이지를 이동할 수 있습니다.

1페이지부터 5페이지까지의 기사 제목을 크롤링하려고 합니다.

여러 페이지의 기사 제목 수집하기

<https://sports.donga.com/ent?p=1>

이 사이트는 URL의 쿼리 부분에서 p의 값에 따라 페이지가 결정됩니다.

한 페이지에 기사가 20개씩 존재하므로

p=1이면 1페이지, p=21이면 2페이지와 같은 식입니다.

여러 페이지의 기사 제목 수집하기

URL을 문자열의 덧셈 연산으로 만드실 수도 있지만,

requests.get 함수의 **params 매개변수**로

쿼리 변수를 추가할 수 있습니다.

여러 페이지의 기사 제목 수집하기

['나띠, 데뷔 스케줄러 공개...5월7일 쇼케이스 'M2 방송', '[DA:차트] '너를 만나'→'우리 만남이'...폴킴, 연속 음원차트 1위', '컴백' 공원소녀, 4色 오버뷰 영상 공개...전곡 하이라이트', '원어스, '쉽게 쓰여진 노래' 성공적 마무리 "'로드 투 킹덤'에서 봐요"', '소녀의 세계' 권현빈, 강렬한 첫 등장...아린과 훈훈 케미', '[TV북마크] '기막힌 유산' 강세정x신정윤, 심상치 않은 재회...박인환 재산정...', '신원호 PD "'슬의생' 2막...5인방 중심 주변 서사 깊어질 것"', '1TEAM(원팀), '덕분에 챌리지' 동참...코로나19 의료진 응원', '블랙핑크, 레이디가가와 콜라보...피처링 참여 [공식]', '본 어게인' 장기용·진세연·이수혁, 새 문명 시작...전생 악연 풀릴까', '가족입니다' 메인 포스터...때론 타인보다 낯선 가족 이야기', '리밋, '날말' MV 티저 공개...썬엔딩' 주인공 재회', '모던 패밀리' 김영록 "친구·이순재 활동 보면서 연기 자극 받아"', '[DA:할리우드] 샤를리즈테론 기부, 코로나19 관련 100만달러 쾌척', '끼리끼리' 인교진→하승진, 24일 '굿모닝FM' 출격 (ft. 장성규) [공식]', '뭉쳐야 찬다' 윤성빈 vs 모태범, 허벅지 힘겨루기 한 판', '사냥의 시간', 오늘 공개...안재홍, 타투+탈색 '파격 변신', '컴백' 파나티스, 6色 '바비걸' 콘셉트컷 공개', '[DA:차트] 조정석 '아로하', 16주차 가온차트서 2관왕 영예', '엠카' 측 "갯세븐·솔라·에이프릴, 오늘 컴백 무대 최초 공개" [공식]', '엠카' 측 "갯세븐·솔라·에이프릴, 오늘 컴백 무대 최초 공개" [공식]', '넷플릭스x연상호 '지옥' 제작 확정...송곳' 작가 공동집필 [공식]', '나 혼자 산다' 송승헌, 무지개 회원 입성...10년차 자취일상 최초 공개', '오인택 결혼, 9월 송무원 출신 예비 신부와 백년가약 [공식]', '쌍갑포차' 이준혁, 머리부터 발끝까지 올 화이트...新 저승사자 탄생', '29일 컴백' NCT DREAM, 오늘(23일) 트랙 비디오 공개', '미스터트롯' 이대

크롤링한 결과는 다음과 같습니다.

각 기사의 href 수집



'소녀의 세계' 권현빈, 강렬한 첫 등장...아린과 훈훈 케미

'소녀의 세계' 권현빈, 강렬한 첫 등장...아린과 훈훈 케미 '소녀의 세계' 권현빈이 오마이걸 아린과 귀여운 케미를 선보였다. 권현빈은 지난 22일 공개된 tV N D 웹드라마 '소녀의 세계' 1화 에필로그에 ...

2020-04-23 09:43:00



[TV북마크] '기막힌 유산' 강세정X신정윤, 심상치 않은 재회...박...

[TV북마크] '기막힌 유산' 강세정X신정윤, 심상치 않은 재회...박인환 재산정리 어제(22일) KBS1 새 저녁 일일드라마 '기막힌 유산'(연출 김형일 극본 김경희) 3회에서는 자식들에게 증여를 준비하는 박...

2020-04-23 09:43:00

이번에는 각 기사로 이동할 수 있는 href를 수집해봅시다.

각 기사의 href 수집

href는 a 태그의 속성으로 존재하며
크롤링된 a 태그에 접근하여 얻을 수 있습니다.

각 기사의 href 수집

```
['https://sports.donga.com/ent/article/all/20200423/100773540/1', 'https://sports.donga.com/ent/article/all/20200423/100773507/1',  
'https://sports.donga.com/ent/article/all/20200423/100773479/1', 'https://sports.donga.com/ent/article/all/20200423/100773425/1',  
'https://sports.donga.com/ent/article/all/20200423/100773430/1', 'https://sports.donga.com/ent/article/all/20200423/100773413/1',  
'https://sports.donga.com/ent/article/all/20200423/100773388/1', 'https://sports.donga.com/ent/article/all/20200423/100773334/1',  
'https://sports.donga.com/ent/article/all/20200423/100773314/1', 'https://sports.donga.com/ent/article/all/20200423/100773308/1',  
'https://sports.donga.com/ent/article/all/20200423/100773312/1', 'https://sports.donga.com/ent/article/all/20200423/100773280/1',  
'https://sports.donga.com/ent/article/all/20200423/100773265/1', 'https://sports.donga.com/ent/article/all/20200423/100773258/1',  
'https://sports.donga.com/ent/article/all/20200423/100773216/1', 'https://sports.donga.com/ent/article/all/20200423/100773200/1',  
'https://sports.donga.com/ent/article/all/20200423/100773088/1', 'https://sports.donga.com/ent/article/all/20200423/100773129/1',  
'https://sports.donga.com/ent/article/all/20200423/100773094/1', 'https://sports.donga.com/ent/article/all/20200423/100773059/1']
```

코드 실행이 완료되었습니다.

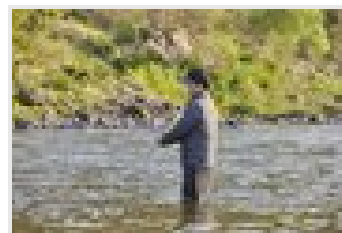
크롤링한 결과는 다음과 같습니다.

네이트 최신뉴스 href 수집하기

최신뉴스

전체 | 정치 | 경제 | 사회 | 세계 | IT/과학 | 스포츠 | 연예 | 칼럼

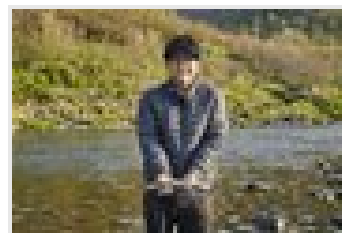
제목 | 제목+내용 | 포토



[ESC] 회 먹기 전에 낚시?

향이네 식탁 지금도 선합니다. 낚시가 취미였던 부친은 어린 저를 데리고 강에 자주 가셨습니다. 전 꼬물꼬물 몸을 비트는 지렁이를 뽐족한 바늘에 무자비하게 꽂았더랬지...

한겨레 04-23 10:06



[ESC] 피라미, 네가 거기서 왜 나와

섬진강에 발 담그고 나 홀로 낚시 줄 묶고 풀고 미끼 달며 쾅쾅 아기자기한 손맛, 청정 자연은 덤. 낚시 초보, 꺾지 루어낚시에 도전하다 “낚시는 원래 혼자 하는 거다.” ...

한겨레 04-23 10:06

네이트의 최신뉴스 목록에서

각 기사의 href를 수집해보겠습니다.

네이트 최신뉴스 href 수집하기

```
[ 'http://news.nate.com/view/20200423n11258?mid=n0100', 'http://news.nate.com/view/20200423n11257?mid=n0100',  
'http://news.nate.com/view/20200423n11256?mid=n0100', 'http://news.nate.com/view/20200423n11255?mid=n0100',  
'http://news.nate.com/view/20200423n11249?mid=n0100', 'http://news.nate.com/view/20200423n11253?mid=n0100',  
'http://news.nate.com/view/20200423n11252?mid=n0100', 'http://news.nate.com/view/20200423n11251?mid=n0100',  
'http://news.nate.com/view/20200423n11250?mid=n0100', 'http://news.nate.com/view/20200423n09470?mid=n0100',  
'http://news.nate.com/view/20200423n11247?mid=n0100', 'http://news.nate.com/view/20200423n11221?mid=n0100',  
'http://news.nate.com/view/20200423n11245?mid=n0100', 'http://news.nate.com/view/20200423n11244?mid=n0100',  
'http://news.nate.com/view/20200423n11243?mid=n0100', 'http://news.nate.com/view/20200423n11242?mid=n0100',  
'http://news.nate.com/view/20200423n06012?mid=n0100', 'http://news.nate.com/view/20200423n11240?mid=n0100',  
'http://news.nate.com/view/20200423n11239?mid=n0100', 'http://news.nate.com/view/20200423n11238?mid=n0100' ]  
코드 실행이 완료되었습니다.
```

크롤링한 결과는 다음과 같습니다.

이전 실습과 비슷한 내용이니 연습해보세요.

sbs 뉴스 최신 기사 목록의 내용 수집하기

SBS NEWS

비디오매그

△브스뉴스

·VOICE

SBS MOBILE 24

f

twitter

youtube

SBS | 로그인 | 회원가입

분야별

다시보기

취재파일

팟캐스트

이슈

끝까지판다

SBS 8 뉴스

라이브

제보

전체메뉴

검색어를 입력하세요

최신

최신

정치

경제

사회

국제

생활·문화

연예

스포츠



[속보] 오거돈 부산시장 전격 사퇴...부산시 공식 확인

오거돈 부산시장 전격 사퇴...부산시 공식 확인 (SBS 뉴미디어부/사진=부산시 제공, 연합뉴스)

2020.04.23 10:42

마스크 쓴 시진핑, 초등학교 방문 '강한 체력' 강조

중국의 코로나19 확산이 진정된 가운데 시진핑 국가주석이 마스크를 쓴 채 초등학교를 방문해 강한 체력을 강조하고 나섰다. 23일 중국중앙TV 등에 따르면 시진핑 주석은 지난 21일 산시 시찰에서 마스크를 착용하고 초등학교 5학년 교실에 들어가 코로나19 속 학생들의 수업 상황 등을 살펴보았다.

2020.04.23 10:42

sbs 뉴스의 최신 기사 목록의 href를 추출하고

href로 접근할 수 있는 기사들의 내용을 추출해봅니다.

sbs 뉴스 최신 기사 목록의 내용 수집하기

각 기사의 href 주소를 얻는 get_href 함수와
기사의 내용을 얻는 crawling 함수를 각각
올바르게 구현해주세요.

sbs 뉴스 최신 기사 목록의 내용 수집하기

오거돈 부산시장이 사퇴를 결정했습니다.부산시는 "오 시장이 23일 오전 11시 기자회견을 열어 공식적으로 사퇴 의사를 밝힐 예정"이라고 밝혔습니다.오 시장은 일신상의 사유를 들어 사퇴 의사를 밝힌 것으로 알려졌습니다.오 시장은 최근 건강 이상설이 나왔습니다.21대 총선 하루 전인 14일 연가를 냈고, 선거 당일인 15일도 비공개 투표를 했습니다.이후에도 부산 시청으로 출근은 했지만, 외부활동을 하지 않았습니다.오 시장 한 측근은 "일신상 이유로 사퇴하는 것은 맞지만 자세한 내용은 알 수 없다"고 말했습니다.부산지역 정가와 시청 안팎에서는 오 시장 사퇴 사유를 두고 다른 배경이 있는 것 아닌가 하는 얘기들이 나오고 있습니다.(사진=부산시 제공, 연합뉴스)

중국의 코로나19 확산이 진정된 가운데 시진핑 국가주석이 마스크를 쓴 채 초등학교를 방문해 강한 체력을 강조하고 나섰다.23일 중국중앙TV 등에 따르면 시진핑 주석은 지난 21일 산시(陝西)성 안강시(安康市) 평리현(平利縣) 시찰에서 마스크를 착용하고 초등학교 5학년 교실에 들어가 코로나19 속 학생들의 수업 상황 등을 살펴봤다.시진핑 주석은 이들 학생에게 학습 및 생활 상황을 물어보면서 "요새 아이들은 대부분 안경을 쓰고 있다"면서 "이게 은근히 걱정되는 부분"이라고 말했다.시 주석은 "체력 단련 부족으로 아이들의 신체 및 건강 수준도 낮아졌다"면서 "아이들은 밝은 정신과 강건한 체력이 있어야 하며 강건한 체력이란 바로 신체를 튼튼히 하는 것"이라고 언급했다.시진핑 주석은 이날 한 이주민의 집도 방문해 마스크를 쓴 채 주민들과 소파에 같이 앉아 담소를 나눴다.이 자리에서 "거주가 안정되려면 안정된 취업이 있어야 한다"면서

크롤링한 결과는 다음과 같습니다.

다양한 섹션의 속보 기사 href 추출하기

속보

정치

경제

사회

생활/문화

세계

IT/과학

오피니언

연합뉴스 속보

모바일 메인에서
보고싶은 뉴스
구독하세요!
바르가기 >

정치 속보

· 전체

· 청와대

· 국회/정당

· 행정

· 국방/외교

· 북한

· 정치 일반

전체

신문게재기사만 | 제목형 | 요약형 | 포토만



여야 '경제통'이 말하는 코로나19發 경제위기

최운열 "보수정권 10년 구조적 위기가 한몫" 신세돈 "국난에 돈 제대로 쓸 생각 안 해" ●...

신동아 | 1분전



당정 "전 국민에 재난지원금 합의...고소득자는 자발적 기부 유도"

【앵커멘트】 더불어민주당과 정부가 코로나19 재난지원금을 전국민에게 지급하되, 고...

MBN | 1분전



[뜬별★] 黨命 받들어 험지에서 생환한 김두관

'재선 고지' 오르며 大權 교두보 구축 김두관의 생환(生還). 김두관(61) 더불어민주당 의원...

신동아 | 1분전

네이버 뉴스 속보 페이지에는 여러 섹션이 있습니다.

다양한 섹션의 속보 기사 href 추출하기

<https://news.naver.com/main/list.nhn?sid1=100>

URL의 쿼리 부분에서 sid1의 값에 따라 섹션이 결정됩니다.

어떤 섹션을 크롤링할지는 input 함수로 입력하세요.

다양한 섹션의 속보 기사 href 추출하기

쿼리와 함께 get 요청을 담고 있는 **requests** 객체를 반환하는

get_request 함수와,

섹션별로 나뉘어진 목록에 있는 기사들의 **href**를 **추출**하는

get_href 함수를 올바르게 구현하세요.

다양한 섹션의 속보 기사 href 추출하기

"정치", "경제", "사회", "생활", "세계", "과학" 중 하나를 입력하세요.

> 정치

```
['https://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=100&oid=277&aid=0004666877',  
'https://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=100&oid=277&aid=0004666877',  
'https://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=100&oid=047&aid=0002267478',  
'https://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=100&oid=047&aid=0002267478',  
'https://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=100&oid=018&aid=0004626432',  
'https://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=100&oid=018&aid=0004626432',  
'https://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=100&oid=003&aid=0009831311',  
'https://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=100&oid=003&aid=0009831311',  
'https://news.naver.com/main/read.nhn?mode=LSD&mid=sec&sid1=100&oid=001&aid=0011567544',
```

크롤링한 결과는 다음과 같습니다.

다양한 섹션의 속보 기사 내용 추출하기

속보

정치

경제

사회

생활/문화

세계

IT/과학

오피니언

연합뉴스 속보

모바일 메인에서
보고싶은 뉴스
구독하세요!
바르가기 >

정치 속보

· 전체

· 청와대

· 국회/정당

· 행정

· 국방/외교

· 북한

· 정치 일반

전체

신문게재기사만

제목형

요약형

포토만



여야 '경제통'이 말하는 코로나19 경제위기

최운열 "보수정권 10년 구조적 위기가 한몫" 신세돈 "국난에 돈 제대로 쓸 생각 안 해" ●...

신동아 | 1분전



당정 "전 국민에 재난지원금 합의...고소득자는 자발적 기부 유도"

【앵커멘트】 더불어민주당과 정부가 코로나19 재난지원금을 전국민에게 지급하되, 고...

MBN | 1분전



[뜬별★] 黨命 받들어 험지에서 생환한 김두관

'재선 고지' 오르며 大權 교두보 구축 김두관의 생환(生還). 김두관(61) 더불어민주당 의원...

신동아 | 1분전

이번에는 섹션별 기사의 **내용**을 추출해보도록 하겠습니다.

다양한 섹션의 속보 기사 내용 추출하기

기사의 href에서 내용을 추출할 때, 본문 영역은 태그가 없습니다.

따라서, 기사 본문을 싸고 있는 div 태그의 children을 얻고,

각 children의 name이 None인 요소만 추출해야 합니다.

다양한 섹션의 속보 기사 내용 추출하기

그 뿐만 아니라 HTML 문서에 적혀있는 주석도

걸림돌이 될 수 있으므로 어려울 수 있는 실습입니다.

어려움을 겪으신다면 정답 코드가 담긴 solution.py 파일을

확인해보시거나, 튜터에게 문의해주세요.

다양한 섹션의 속보 기사 내용 추출하기

"정치", "경제", "사회", "생활", "세계", "과학" 중 하나를 입력하세요.

> 세계

['KOSDAQ 643.79 UP 8.63 points (close)(END) ', 'KOSPI 1,914.73 UP 18.58 points (close)(END) ', '[서울=뉴시스] 김예진 기자 = 23일 일본 도쿄증시에서 닛케이225지수(닛케이평균주가)는 전일 대비 291.49 포인트, 1.52% 상승하며 1만 9429.44에 장을 마감했다. 4 거래일 만에 상승 마감했다.JPX 닛케이 인덱스 400지수는 전일 대비 157.30 포인트, 1.25% 오른 1만 2782.45에 거래를 마쳤다. 토픽스(TOPIX)지수는 전일 대비 19.08 포인트, 1.36% 상승한 1425.98에 시장을 마무리했다.니혼게이자이 신문에 따르면 미국에서 경제 활동 재개 기대가 고조되고 국제유가가 상승하자 투자자의 심리가 개선됐다. 해외 투자자들에 의해 일본 증시가 상승했다. aci27@newsis.com<© 공감언론 뉴시스통신사. 무단전재-재배포 금지> ', '[서울=뉴시스] 김예진 기자 = 23일 일본 도쿄증시에서 닛케이225지수(닛케이평균주가)는 전일 대비 291.49 포인트, 1.52% 상승하며 1만 9429.44에 장을 마감했다. 4 거래일 만에 상승 마감했다.JPX 닛케이 인덱스 400지수는 전일 대비 157.30 포인트, 1.25% 오른 1만 2782.45에 거래를 마쳤다. 토픽스(TOPIX)지수는 전일 대비 19.08 포인트, 1.36% 상승한 1425.98에

크롤링한 결과는 다음과 같습니다.

특정 영화 리뷰 추출하기

네이버 영화 페이지의 영화 리뷰의 제목을 크롤링하겠습니다.

이번에는 정보를 얻고자 하는 영화의 제목이 입력으로 주어지고,

해당 영화에 대한 리뷰 결과를 보여주어야 합니다.

특정 영화 리뷰 추출하기

이를 위해 `get_url`, `get_href`, `crawling`
세 개의 함수를 올바르게 구현해주셔야 합니다.

특정 영화 리뷰 추출하기

get_url은 영화 제목을 입력받고,
해당 제목을 **검색하였을 때 나오는 URL**을 반환해야 합니다.

requests.get 메소드의 params 매개변수를 이용해도 되지만,
이번에는 문자열의 결합을 이용하는 편이 더 간편하기 때문에
get_url은 **문자열의 결합**을 이용하여 URL을 만듭니다.

특정 영화 리뷰 추출하기

'스타워즈'에 대한 영화 통합검색결과 입니다.

☒ 전체 ☐ 영화 ☐ 영화인 ☐ 영화제 ☐ 영화사 ☐ 극장 ☐ 이미지 ☐ 동영상 ☐ 배역

영화 (1-5 / 26건) : [정확도순](#) · [제작년도순](#)



[스타워즈: 라스트 제다이 \(Star Wars: The Last Jedi\)](#)

★★★★★ 6.44 (참여 7110명)

액션, 모험, 판타지, SF | 미국 | 152분 | 2017

감독 : 라이언 존슨 | 출연 : 데이지 리들리, 마크 해밀, 오스카 아이삭, 아담 드라이버, 캐리 피셔, 존 보예가

다운로드



[스타워즈 에피소드 3 - 시스의 복수 \(Star Wars: Episode III - Revenge Of The Sith\)](#)

★★★★★ 9.09 (참여 3024명)

SF, 모험, 액션 | 미국 | 139분 | 2005

감독 : 조지 루카스 | 출연 : 이완 맥그리거, 나탈리 포트만, 헤이든 크리스텐슨, 이언 맥디어미드, 프랭크 오즈

다운로드

네이버 영화 페이지에서 키워드를 검색했을 때 화면입니다.

get_url 함수는 이 때의 URL을 반환하면 됩니다.

특정 영화 리뷰 추출하기

get_href는 검색 결과,

가장 위에 있는 영화의 href를 반환합니다.

특정 영화 리뷰 추출하기

마지막으로 crawling 함수를 구현하여
get_href에서 얻은 영화의 href로 접근하고,
해당 영화의 리뷰 목록을 크롤링하세요.

특정 영화 리뷰 추출하기

영화 제목을 입력하세요.

> 스타워즈

['루카스 없는 스타워즈 = 영혼 없는 라이트 세이버 (스포주의)', '시리즈의 절점을 찍다 <스타워즈:라스트 제다이>', '갈수록 재미가 없어져...', '(스포)최고의 소재로 최악의 영화를 만들다', '(스포,분노) 이제부터 이 영화는 포스닥이 입니다.', "[영화 분석(스포일러有)] '모두'를 위한 '마지막' 제다이", '역대 최악의 스타워즈 영화.', '스타워즈를 3류 영화로 전락시킨 쓰레기', "한마디로 '개망작' 이라 생각합니다. (스포 많아요. 영화 보신분만 리뷰 보세요...)"], '스타워즈: 라스트 제다이 분노해서 쓸수밖에 없었던 리뷰']

코드 실행이 완료되었습니다.

크롤링한 결과는 다음과 같습니다.



`/*elice*/`

contact@elice.io