

취미 클래스 플랫폼 (클래스101, 탈잉, 클래스톡) 별 비교

:취미·자기계발 트렌드 반영 검색/추천시스템

Enjoy your stylish business and campus life with BIZCAM

Crawling Project



고원진

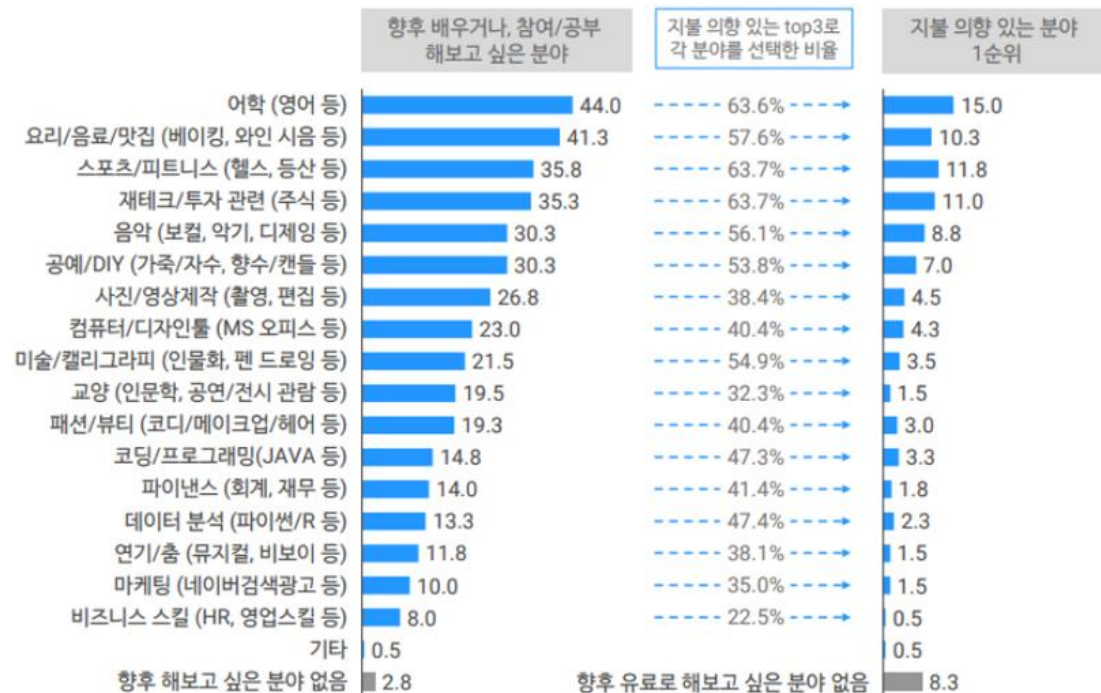


장지혜

취미생활/자기계발 수요

Enjoy your stylish business and campus life with BIZCAM

향후 배우거나, 참여/공부 해보고 싶은 분야 및 지불 의향 있는 분야



[Base: 정기적으로 취미생활/자기계발 하고 있는 사람, N=400, 복수응답/순위형, %]

자료: 오픈서베이 취미생활·자기계발 트렌드 리포트 2019(16p)

향후 유료로 취미생활/자기계발 분야 참여시 선호하는 방식



[Base: 취미생활/자기계발 분야에 지불 의향 있는 사람, N=367, 단수응답, %]

*하늘색 음영: 전체 평균 대비 +3%p 이상인 데이터

자료: 오픈서베이 취미생활·자기계발 트렌드 리포트 2019(17p)

대표 플랫폼 세 곳 비교

Enjoy your stylish business and campus life with BIZCAM



취미·자기계발 클래스 플랫폼

- 주 52시간 근무제 시행으로 늘어난 여가시간을 풍요롭게 하는 취미플랫폼의 급성장
- 코로나로 늘어난 '집콕족'들의 라이프스타일 변화
- 다양한 관심사와 자기계발을 돕는 온라인/오프라인 플랫폼 수요증가
- N잡러, 프리랜서 마켓으로서의 역할

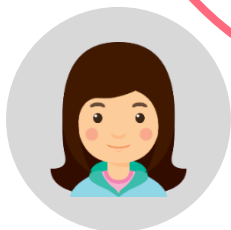
예상 시나리오

Enjoy your stylish business and campus life with BIZCAM

필요한 클래스 정보를 추천/비교 해 볼 수 있는 서비스가 있다면 좋겠다!



"간단하게 파이썬을 이용한
자동투자에 대해 알고 싶어! 강의는
어디서 찾아들어야하지?"



"홈트로 필라테스
배우고 싶은데 튜터추천을
받고 싶어!"



"신사임당 같은 유튜버의
노하우를 알고 싶은데
커리큘럼을 확인할 방법은 없을까?"



"글쓰기를 배우고 싶은데
키워드 검색하면 추천해주는
시스템이 있으면 좋겠어."

CLASS | IOI

taling

ClassTok

사이트 구조 소개

Enjoy your stylish business and campus life with BIZCAM

CLASS | IOI

클 래 스 101

taling

탈 잉

ClassTok

클 래 스 톡

CLASS IOI

1. 반응형 웹페이지

- window 사이즈마다 xpath 경로 다름
- `//*[@id="wrapper"]/div[1]/main/div/div[1]/div/div[2]/div/div/h2`
- `//*[@id="wrapper"]/div[1]/main/div/div[2]/div/div[2]/div/div[1]/h2`

2. 상세 페이지별 다른 구조

- `//*[@id="wrapper"]/div[1]/main/div/div[2]/div/div[2]/div/div[1]/h2`
- `//*[@id="wrapper"]/div[1]/main/div/div[1]/div/div[2]/div/div[1]/h2`

3. Get방식

- POST 방식 데이터도 존재

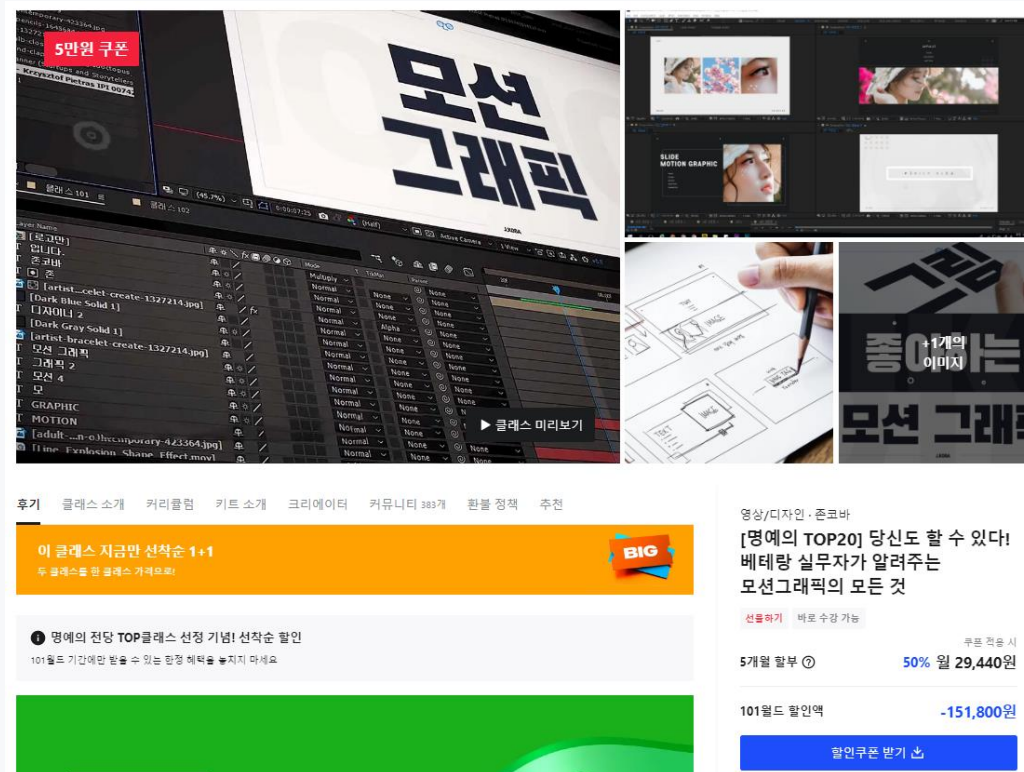
<https://api.channel.io/front/elastic/plugins/5728/events>

4. BeautifulSoup css-selector로는

- 제목 : 크롤링 될 때 안될 때 있음
 - 가격 : 크롤링 안됨
- => Selenium 사용

5. Selenium으로 크롤링 안되는 페이지 있음

- 260 / 1436개



탈잉

카테고리 : 취미/공예

- 총 125 페이지
- 페이지 당 15개 클래스
- likes, link, location, number, price, teacher, title
- 'location'의 경우 타 사이트와의 비교 후 tile과 합칠 예정 '[' + location + ']' + title
- 해당 카테고리 내에 있는 1873개(오후기준: 1875개) 클래스 중 모두 크롤링 완료
- 추가 내용: taling-test7.csv
 - 'soldout'클래스 (soldout box 유무) 미표기
 - 중간에 생겼다 사라지는 클래스의 경우 축적된 데이터로 어떻게 활용할 지 생각

사이트 구조

- 프레임워크: scrapy
 - get 방식
 - html : 정적페이지
 - 반응형
- 카테고리 분류
 - 메인 카테고리
 - 서브 카테고리 (질문: 서브 카테고리로 들어갔을 때 url 규칙x, 메인->서브로의 연결)
 - 클래스 페이지(상세페이지)

- spider.py -> __init__ (url 규칙성 {}) 수정하는 파일
 - start_urls -> page, cateMain 부분 질문하기

- 위 방법이 안되면 크롤링한 데이터 모두 합산한 후 마지막에 카테고리 분류 할 지 고민중

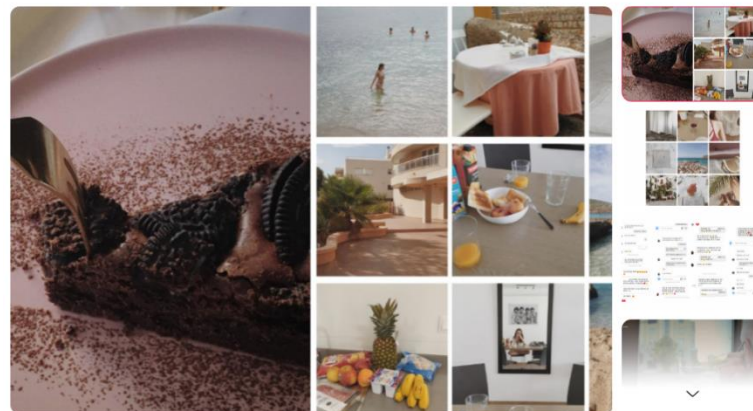
```
In [9]: 1 import scrapy, requests, re
        2 from scrapy.http import TextResponse
        3
        4 import pandas as pd
```

원데이·성수·최대 5명

[원데이] 인스타그램 감성? 황금비율 사진 + 보정 레시피! 내 피드도 분위가 있게 변신!

나화 튜터
★ 4.9(150)

✓ 리뷰 100건+@



클래스 일정

03월 11일 (목) 19:00-22:00

성수

상세장소: 성수동 스튜디오

25,000원 / 시간
75,000원 / 총 1회 3시간



수업 신청하기

```
In [43]: 1 pd.read_csv("taling/taling-test7.csv")
        2
        3 5.0 https://taling.me/Talent/Detail/3503 강남·논리인 총 1회 2 50,000 원 / 남세원 튜터 [강남·1회/90분 원데이] 최고의 감미...
        4 5.0 https://taling.me/Talent/Detail/20164 구로 총 4회 4 150,000 원 / 신성우 튜터 [작곡클래스] 방구석 건반/작곡으로 골장 보자!! [구로디지털 단지][신림]
        ... ..
        1870 (0) https://taling.me/Talent/Detail/13311 대구중앙 총 1회 2 20,000 원 / 김미선 튜터 인스타사진 예쁘게 찍는 꿀팁/촬영/보정/ 필름카메라로 인생샷 남기기까지 단 하루
        1871 (0) https://taling.me/Talent/Detail/8358 구월동 수원 총 1회 3 84,000 원 / 임은영 튜터 [인천/원데이] 특별한 날 특별한 문채 드리는 클라워 앙금케이크
        1872 (0) https://taling.me/Talent/Detail/7426 서현 잠실 총 5회 10시간 120,000 원 / 김해진 튜터 《발송가능》여행,영화의 특별한 순간들을 나만의 일러스트로 남기기!
        1873 (0) https://taling.me/Talent/Detail/25979 신촌충대 총 1회 3 90,000 원 / 강세미 튜터 레터링 케이크 만들기 실전반 클래스
        1874 (0) https://taling.me/Talent/Detail/14610 부산부경대 총 1회 2 40,000 원 / 공계현 튜터 [부산]2시간만에 나만의 음원제작해볼까요??^^

        1875 rows x 7 columns
```


클래스톡

start_now : 현재 수강가능한 클래스

- 카테고리별 분류 페이지x, search_key 로 어디서든 접근 가능
- likes, link, location, number, price, teacher, title, tag
 - 사이트 내 해시태그로 어느 링크로든 접속 가능
 - 상세페이지

사이트 구조

- 프레임워크: scrapy
 - selenium을 사용할 지 고민중
 - get 방식
 - html :
 - 반응형
- 메인 페이지
- 상세 페이지

```
1 import scrapy, requests, re
2 from scrapy.http import TextResponse
```

```
1 # 계획수립: 각 상세페이지 내에서 크롤링 x, start_now 페이지에서 필요한 정보만 크롤링하기로
```

```
1 # 관측정보 xpath, 상세페이지
```

```
1 link = links[3]
2 req = requests.get(link)
3 response = TextResponse(req.url, body=req.text, encoding="utf-8")
4 response
```

```
c200 https://www.classtok.net/class/classDetail/2402
```

```
1 title = response.xpath('//*[@id="videoContainer"]/div/div[1]/h2/text()')[0].extract()
2 #teacher_link = response.xpath('//*[@id="classDetail"]/div/div[2]/div/div/div/a/@href')[0].extract()
3 #teacher = response.xpath('//*[@id="classDetail"]/div/div[2]/div/div/div/a/text()')[0].extract()
4 #link2 = 'https://www.classtok.net'
5 #teacher_link = list(map(response.urljoin, links2))
6 #tag = response.xpath('//*[@id="videoContainer"]/div/div[1]/ul/li/a/@href')[0].extract()
7 discount = response.xpath('//*[@id="sidebar"]/div/div/div[1]/span[1]/text()')[0].extract()
8 price = response.xpath('//*[@id="sidebar"]/div/div/div[1]/span[2]/text()')[0].extract()
9 #curriculum = response.xpath('//*[@id="classDetail"]/div/div[2]/div/div[4]/ul/li[2]/strong/text()')[0].extract()
10 #member = response.xpath('//*[@id="classDetail"]/div/div[3]/div/span[2]/text()')[0].extract()
11 #like = response.xpath('//*[@id="videoContainer"]/div/div[3]/div/span/text()')[0].extract()
12
13 title, discount, price
```

```
('Resume, CV, Cover letter, 해외기업 취업 성공하기!', '83% 할인', '89,000원')
```

하고 싶은 클래스를 찾아보세요
지금 신청 가능한 클래스

전체보기



운동 · Judy
박스필라테스로 만드는 아름다운 라인

원 30,000원 62% 할인

3명 추천



그림 · 석재원
힐링 풍경 펜화 클래스

원 30,000원 62% 할인



수화 · 어구리광
더드림 수화 (베이직싸인)

원 30,000원 66% 할인

4명 추천



비즈니스 · 콘텐츠만드는아빠
콘텐츠로 돈 벌기의 정석! [게.콘.돈.잘.자.직강]

원 20,000원 75% 할인



유튜브 · 한정혜
슬기로운 유튜브 생활

원 39,000원 51% 할인

85명 추천



마케팅 · 이종범
매출을 올려주는 페이스북, 인스타그램 마케팅 완전정복

원 10,000원 89% 할인



마케팅 · 이종범
스타트업을 위한 마케팅 전략

원 10,000원 66% 할인

5명 추천

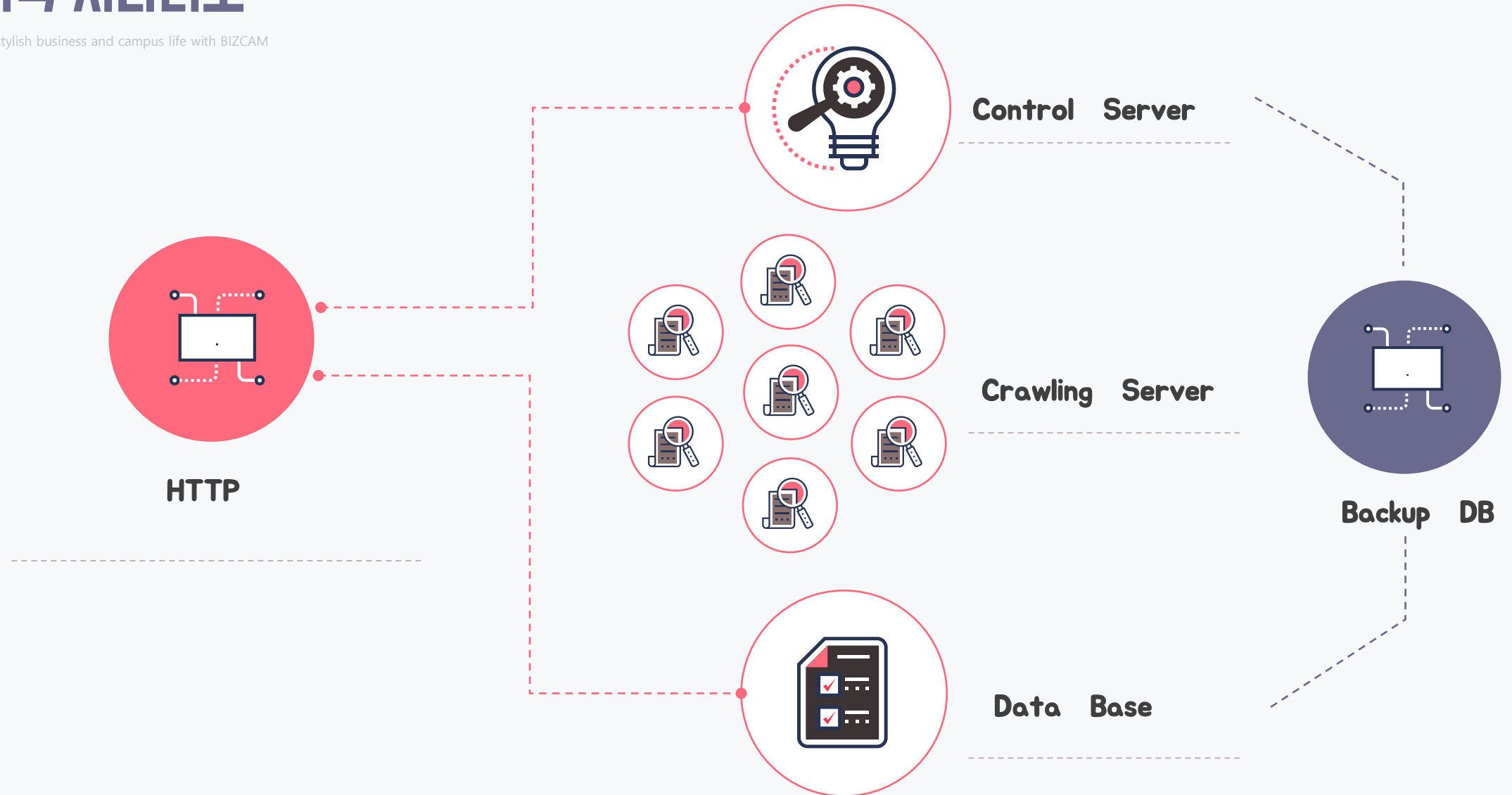


운동 · 유미
기초공사 탄탄! 우리 몸을 리셋하는 필라테스

원 17,000원 78% 할인

물리적 시나리오

Enjoy your stylish business and campus life with BIZCAM



Trouble Shooting 요청사항

Enjoy your stylish business and campus life with BIZCAM

1. Post 방식 From data 가져오는 방법?

<https://api.channel.io/front/elastic/plugins/5728/events>

1. Selenium을 사용하지 않을 수 있는지

- 현재 크롤링 120분 이상 소요

1. Selenium으로 크롤링 안되는 페이지 있음

- 260 / 1436개

1. 서버간 데이터(link 등) / DB 전송 방법?

- link 분산하여 여러 서버로 크롤링 시도 예정
(현재 서버에서 동작시 서버 다운)

1. Flask를 통해서 어느 수준까지 서비스 구현 가능한지? / 검색? 추천?

질의·응답



Thank you!

