

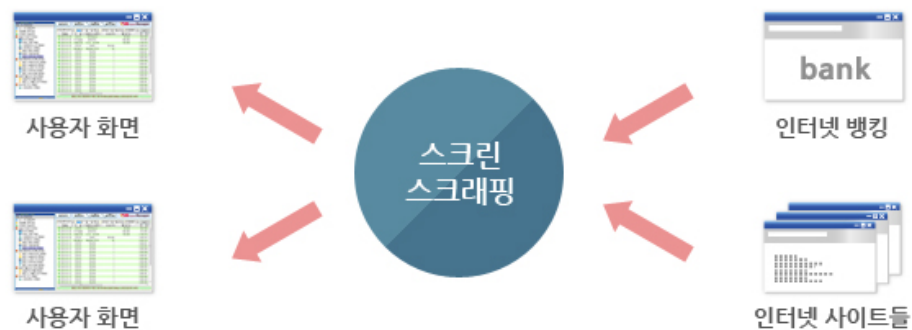
코멘토 직무부트캠프

실무 과제 안내서

프로젝트 과제 선정 이유

2주차 스크래핑을 통해 정보 가져오기

- * 회사내의 없는 데이터를 가져와서 가공하기
 - 파이썬을 활용한 스크래핑으로 코스피 종목 가져오기 수행
 - 심화과정 크롤링을 통해 관리종목 가져오기(과제 수행 후, 추가적으로 필요하며 설명 예정)



과제 제출 안내



제출 기한

1/3차 과제 : 1/3주차 세션일 기준 5일 이내

2/4차 과제 : 1/3차 과제 제출 마감일 기준 7일 이내

*클래스룸 내 '캠프 일정'에서 정확한 일정 확인 가능



제출 방법

캠프 클래스룸 내 '실무 과제'에 업로드

*마이페이지>신청한캠프 에서 클래스룸 입장 가능

| 2 주차 과제 |

2 주차 과제 안내

[주제]

새로운 데이터를 가져와서 가공해서 기존의 데이터와 정보를 보여주기
(파이썬으로 활용하여, 스크래핑해서 상장 회사 가져오기)

[작성 방법]

1. 코스피 종목 수 출력하기
2. 코스피 종목명/종목코드로 출력하기
3. 코스닥 종목 수 출력하기
4. 코스닥 종목명/종목코드 출력하기

[참고 내용]

- <http://kind.krx.co.kr>(한국거래소)
- <https://www.python.org/>(파이썬)

파이썬을 통해 스크래핑으로 정보 가져오기

[절차]

1. 파이썬을 설치
2. Pandas 라이브러리로 엑셀파일 가져와기
3. 코스피/코스닥 종목 가져오기
4. 가져온 데이터에서 특정 데이터만 뽑아내기

[최종결과 출력 내용]

1. 코스피 종목수
2. 코스피 종목명/종목코드(6자리)
3. 코스닥 종목수
4. 코스닥 종목명/종목코드(6자리)

[참고 내용]

- * 코스피 스크래핑 사이트 : <http://kind.krx.co.kr/corpgeneral/corpList.do?method=download&searchType=13&marketType=stockMkt>
- * 코스피 스크래핑 사이트 : <http://kind.krx.co.kr/corpgeneral/corpList.do?method=download&searchType=13&marketType=kosdaqMkt>
- * 파이썬 참고 라이브러리 : pandas, ssl
- * 파이썬 크롤링 라이브러리 : selenium(심화과정)

2 주차 과제 안내

[과제 결과]

```
코스피 종목 수: 813
      code_name  code
0      DRB동일  004840
1        DSR   155660
2    GS글로벌  001250
3  HDC현대산업개발  294870
4        KEC   092220
...
808  현대코퍼레이션홀딩스  227840
809        화신   010690
810      휴켄스   069260
811    흥국화재   000540
812    흥아해운   003280

[813 rows x 2 columns]
<class 'pandas.core.frame.DataFrame'>
코스닥 종목 수: 1506
      code_name  code
0         3S   060310
1      EDGC   245620
2    ES큐브   050120
3  HK이노엔   195940
4      JW신약   067290
...
1501   한라IMS   092460
1502  한스바이오메드  042520
1503   현대사료   016790
1504  홈센타홀딩스  060560
1505      힘스   238490

[1506 rows x 2 columns]
<class 'pandas.core.frame.DataFrame'>
```

2 주차 과제 안내

[과제 코드]

최종결과 코드