

# 파이썬을 이용한 데이터수집 및 스마트공장 사례 조사

## - 프로젝트 명세 -

1. 개요 : 팀을 이뤄 수업에서 배운 기법들을 활용하여 웹사이트(들)로부터 자료를 수집 및 분석
2. 팀 구성 : 2~3인 1팀, 1월 11일(화) 오전 9시까지 [jshuh@kpu.ac.kr](mailto:jshuh@kpu.ac.kr) 로 받은 메일들에 한해 팀 편성, 나머지 인원은 11일 수업에서 랜덤 편성
3. 대상 웹사이트(들) : 팀 내에서 자율적으로 선정, 가급적 팀마다 웹사이트가 겹치지 않도록 권장 (같은 웹사이트에서 다른 자료를 수집하는건 가능), 중간 발표 때 너무 어렵지 않은 지 난이도 체크를 해야 하니 필수로 발표할 것
4. 평가 요건 : 각 요건들은 난이도에 따라 배점이 다름, 5점 이상 만족 시 Pass, 미만은 Fail

번호	배점	요건
1	1점	방문한 웹페이지 수가 10개가 넘는가 (주소가 다른 페이지거나, '더보기' 단추를 통해 데이터가 동적으로 추가되면 다른 페이지로 인정)
2	1점	조건문을 활용하여 수집 데이터를 필터링했는가
3	1점	함수를 직접 정의 및 사용하여 코드의 재사용성을 높였는가
4	1점	목적을 달성하기 위한 데이터 수집에 적합한 crawling 방법을 사용하였는가
5	1점	Raw data를 사용할 수 있는 데이터로 전처리를 하였는가
6	1점	수집한 데이터들을 시각화 기법을 통해 표현하였는가
7	1점	적절한 예외처리를 통해 코드의 안정성을 높였는가
8	1점	데이터를 사용해서 어떤 insight를 얻으려고 하는가

5. 중간 발표 내용 (12일): 대상 웹사이트, 수집 대상 내용 간단히 공개

### 6. 최종 발표 (20일)

- 실제 자료를 수집하는 코드 시연
- 코드는 1월 19일(수) 23시까지 e-class를 통해 제출, 늦을 경우 1점 감점, 수업에서 다른 패키지 외의 외부 패키지를 사용했을 경우, 제출할 때 명시할 것
- 발표자료 내용 : 평가 요건 중 만족시킨 요건, 수집 대상과 수집 방법에 대한 간단한 설명, 코드 실행 결과, 부록으로 코드 전체 첨부, 로그인할 경우 비번 등은 반드시 지울 것 (감점 대상은 아님, 개인정보 보호 측면)