

단감소프트 구르미랑 App 이벤트 정보 크롤러

2023. 07. 06

신 주 용

1. 프로젝트 개요
2. 요구사항 및 R&R
3. 진행 내역
4. 총평 및 향후 계획



1. 프로젝트 개요

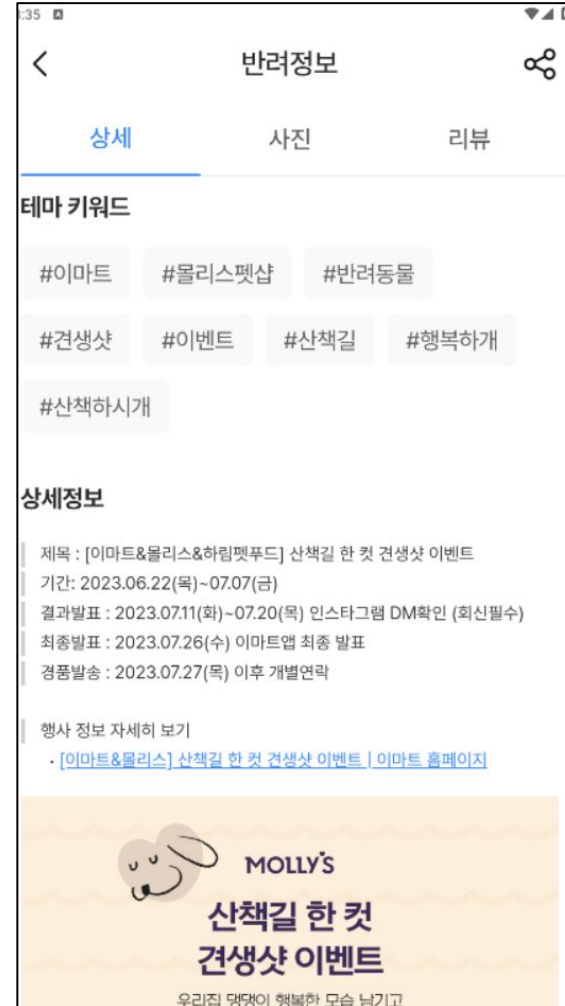
✓ (주)단감소프트 솔루션인 “구르미랑 App “ 에 [\[반려정보\]](#) 서비스 중

- 반려 동식물 관련 [이벤트/교육](#) 정보 제공
- 현재는 주요 정보를 선별하여, 시스템에 [직접 등록](#) 방식으로 운영 → [크롤링을 통한 방대한 정보 자동화 필요](#)

반려정보



상세정보



2. 요구사항 및 R&R

✓ 프로젝트 완료후, 운영 실서버에 적용하기위해 요구사항 도출

- 요구사항 : 향후 효율적 관리 측면의 요구사항 도출
- R&R : (단감소프트) 교육생 필요역량 강화, 단계별 검수를 통한 이슈 파악 및 해결책 제공 → (개발자) 일정 및 품질 확보

요구사항

- 크롤링 대상 사이트의 테이블 관리
 - 테이블로 사이트주소 및 크롤링 대상 태그들 관리를 통해, 변화관리 용이
 - 크롤링 대상 사이트 접속 불가 시 Error Log 정보 관리를 통해, 판별 용이
- 크롤링 결과 테이블 관리
 - 1차 테이블에 결과 기록 → 관리자기능으로 운영반영 선별 및 자동 반영
 - 이미 가져온 정보는 제외 (제목, 시작날짜로 구분)
 - 이벤트/교육 진행여부 확인 → 시작, 종료날짜로 나눠 저장, 판별 용이
- 크롤러 실행시 파라미터값 관리
 - 크롤링 대상 사이트에서 대상 선정을 위한 파라미터 처리

R&R

- (주)단감소프트
 - 크롤링 대상 사이트 정보 제공
 - 크롤러 웹 실행 및 결과 관리를 위한 웹사이트 구축 (향후)
 - 필요 기술 역량 강화 (Database 설계 역량, 프로그램 설계 역량 등)
 - 단계별 검토 및 가이드 제공(설계검토, 개발검토, 이슈검토)
 - 프로젝트 완료후 운영 반영
- (주)글로벌 소프트웨어 캠퍼스
 - 크롤링 대상 사이트 DataBase 구축
 - 크롤러 설계 / 개발 / 테스트
 - 크롤링 결과 DataBase 구축
 - 산출물 작성

3. 진행 내역

√ 개발 방법론에 따른 프로젝트 진행 (착수 > 분석 > 설계 > 개발 > 테스트 > 종료)

- 현업에서 진행하는 방법론 절차 경험, 특히 **요구사항분석, 분석, 설계 방법론 경험**
- 각 단계별 **리뷰를 통해, 부족한 부분 도출** 및 해결책에 대한 가이드 받음
- 기술적으로는 **DataBase 경험이 전무**했으나, **프로젝트를 통해 경험 확보**

일정	진행 내역	얻은점
5월 2주 (착수)	<ul style="list-style-type: none"> - 개발환경 구성 - 구르미랑 앱 소개 - 요구사항 및 개발 방향 리뷰 진행 	- 개발전 요구사항 분석단계의 노하우 경험
5월 2주 (분석)	<ul style="list-style-type: none"> - 크롤러 대상 사이트 분석 . 분석후 사이트별 특이사항 리뷰 진행 	- 사이트 분석 방법 및 효과적 이슈사항 도출 방법 경험
5월 3주 ~ 5월 4주 (설계)	<ul style="list-style-type: none"> - DataBase 설계 및 생성 . 기본 설계후 리뷰를 통해 보강 	- 요구사항을 Database에 반영하는 방법 경험
6월 1주 ~ 6월 3주 (개발/테스트)	<ul style="list-style-type: none"> - 크롤러 개발 . 주차별 이슈 리뷰 진행 	- 개발시 이슈 도출/공유 방법 및 효과적인 해결책 도출 방법 경험
6월 4주 (종료 보고)	<ul style="list-style-type: none"> - 산출물 작성 - 단감소프트 내부 발표 - 글로벌 소프트웨어 캠퍼스 발표 자료 작성 	<ul style="list-style-type: none"> - 산출물 작성 역량 경험 - 완료 보고서 작성 역량 경험

3.1. 크롤링 대상 사이트

- √ 크롤링 대상 정보 확정 완료 (5월)
- √ 크롤링 대상 사이트 URL 확정 완료 (6월, 총 20개 사이트)

크롤링 대상 사이트 목록

ID	기관명	URL
1	코엑스(COEX)	http://www.coex.co.kr/event-performance/total-schedule
2	킨텍스(Kintex)	https://www.kintex.com/web/ko/event/clist.do
3	세텍(SETEC)	https://www.setec.or.kr/front/schedule/list.do

⋮

18	서울특별시 공공서비스예약	https://vevak.seoul.go.kr/web/search/selectPageListDetailSearchImg.do?code=T200&dCode=T205
19	서울특별시 공공서비스예약	https://vevak.seoul.go.kr/web/search/selectPageListDetailSearchImg.do?code=T200&dCode=T206
20	경기문화재단	https://www.ggcf.kr/

크롤링 대상 화면



3.2. 테이블 설계

✓테이블 설계 완료 (5월)

✓ Sources 테이블 작성 완료 (6월)



3.3. 크롤러 설계/개발

√ Sources 테이블과 Events 테이블 생성 코드 작성 완료 (5월)

√ 크롤러 코드 작성 완료 (6월)

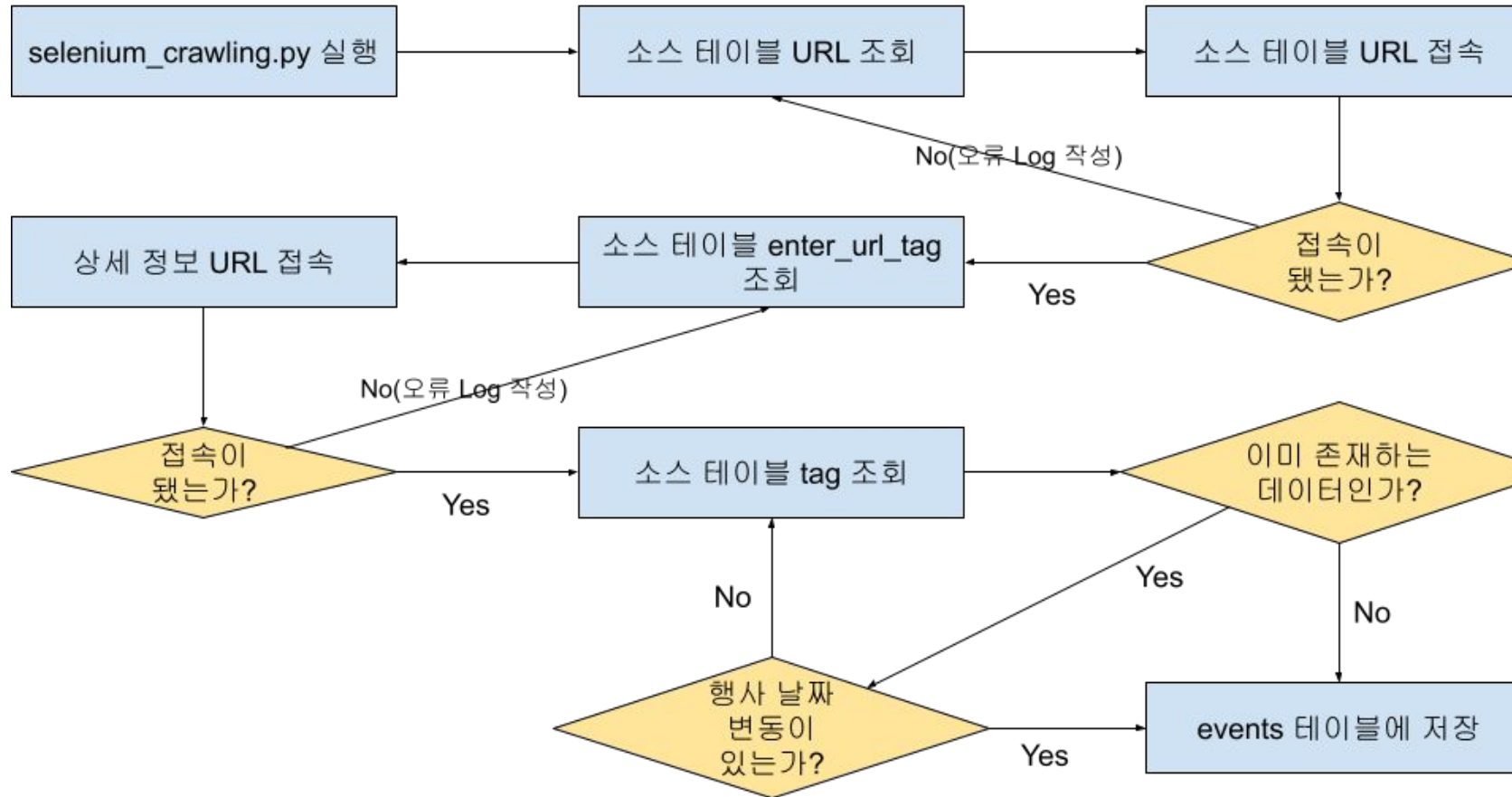
create_tables.py

- Mysql 생성
 - mysql 서버 생성, – 기본 DB생성
- create_tables.py 코드 작성
 - sources 테이블 생성
 - events 테이블 생성
 - sources.xlsx 엑셀 파일을 불러와서 sources 테이블에 값 추가

selenium_crawling.py

- URL 접속
 - sources 테이블의 URL 주소 접속
- 각 Event 별로 상세정보 URL을 List로 저장
 - sources 테이블의 Tag를 참조하여 상세 정보 URL을 List로 저장
 - sources 테이블의 Tag를 참조하여 행사 이미지 URL을 List로 저장
- 상세 정보 페이지 크롤링
 - 저장된 상세 정보 URL에 접근하여 양식에 맞는 정보들을 크롤링
- 요구 데이터 분류
 - 크롤링 된 행사들을 동식물 관련된 행사들만 추출
 - 중복된 행사는 제거
- events 테이블에 저장

3.4. 크롤러 진행도



3.5. 크롤링 결과물 (MySQL.Events Table)

event_id	source_id	event_type	class_type	img_url	title	overview	start_date	end_date	playtime	place
1	1	0	0	https://www.coex.co.kr/wp-content/uploads/2...	2023 케이팻페어	국내 최대 반려동물 산업 전시회 최신 반려동물 ...	2023-07-21	2023-07-23	10:00 ~ 18:00(17:30 입장마감)	Hall D
2	3	0	0	https://www.setec.or.kr/file/viewImg.do?fidx=...	코리아 렵타일 쇼	※ 예매정보 ※ - 기간:2023.07.29~2023.0...	2023-07-29	2023-07-30		제3전시
3	3	0	0	https://www.setec.or.kr/file/viewImg.do?fidx=...	2023 서울팻쇼	반려동물 산업의 마케팅 활성화와 발전방향을 ...	2023-09-15	2023-09-17		제1전시
4	3	0	0	https://www.setec.or.kr/file/viewImg.do?fidx=...	희귀반려동물 박람회		2023-10-07	2023-10-08		제3전시
5	4	0	0		서울 FCI 국제 도그쇼	상세 정보 홈페이지 참조	2023-07-14	2023-07-16		제1전시
6	10	0	0	https://d5bvmkxgb6q.doudfront.net/wp-cont...	2023 케이팻페어 코엑스	※ 사전등록 기간 : ~2023년 7월 20일(목) 23시 ...	2023-07-21	2023-07-23	10:00~18:00 (입장마감 17:30)	코엑스
7	10	0	0	https://d5bvmkxgb6q.doudfront.net/wp-cont...	2023 케이팻페어 송도	※ 사전등록 기간 : ~2023년 8월 24일(목) 23시 ...	2023-08-25	2023-08-27	10:00~18:00 (입장마감 17:30)	송도컨
8	10	0	0	https://d5bvmkxgb6q.doudfront.net/wp-cont...	2023 케이팻페어 대구	※ 사전등록 기간 : ~2023년 10월 05일(목) 23...	2023-10-06	2023-10-08	10:00~18:00 (입장마감 17:30)	대구 엑
9	11	1	1	https://yeyak.seoul.go.kr/web/common/file/File...	[서울식물원/식물문화학교/화요일/성인/7월] #행...		2023-07-04	2023-07-25		서울식
10	11	1	1	https://yeyak.seoul.go.kr/web/common/file/File...	[서울식물원/식물문화학교/목요일/성인/7월] 채...		2023-07-04	2023-07-25		서울식
11	11	1	1	https://yeyak.seoul.go.kr/web/common/file/File...	[서울식물원/식물문화센터/수요일/성인/7월] ...		2023-07-05	2023-07-26		서울식
12	11	1	1	https://yeyak.seoul.go.kr/web/common/file/File...	[서울식물원] 초록놀이터 (주말) (만5~7세 유아...		2023-06-27	2023-07-31		서울식
13	11	1	1	https://yeyak.seoul.go.kr/web/common/file/File...	[서울식물원] 해설사와 함께하는 온실 로봇타닉...		2023-06-23	2023-07-31		서울식
14	11	1	1	https://yeyak.seoul.go.kr/web/common/file/File...	[서울식물원] 꼬물꼬물 텃밭체험 (목) (유아) [어...		2023-06-27	2023-07-31		서울식
15	11	1	1	https://yeyak.seoul.go.kr/web/common/file/File...	[서울식물원/식물문화센터/금요일/가족/7월] ...		2023-07-07	2023-07-28		서울식
16	11	1	1	https://yeyak.seoul.go.kr/web/common/file/File...	[서울식물원] 어린이탐험대 "하늘의 비행사, 잠...		2023-06-27	2023-07-31		서울식
17	12	1	1	https://www.bdna.or.kr/board/get_img.do?a_n...	특별전시 "나자식을 그리고, 그러다."	3월 7일부터 국립백두대간수목원 방문자센터 1...	2023-03-07	2023-07-09		국립백
18	13	0	1		2023년 반려식물전시회 및 체험운영	상세 정보 홈페이지 참조	2023-05-19			제주한
19	14	1	1	https://reserve.koagi.or.kr/data/edusat2_cate/...	(전 연령대상)야간개장 체험 프로그램 '여름밤, ...	상세 정보 홈페이지 참조	2023-06-03	2023-09-23	30분	
20	14	1	1	https://reserve.koagi.or.kr/data/edusat2_cate/...	(셀프교육) 특별전시 연계 '식물미술사 강희안'	상세 정보 홈페이지 참조	2023-05-24	2023-08-27	30분	
21	14	1	1	https://reserve.koagi.or.kr/data/edusat2_cate/...	(유아 단체) 꼬마탐험가 - 정원편	상세 정보 홈페이지 참조	2023-05-12	2023-11-24	60분	
22	14	1	1	https://reserve.koagi.or.kr/data/edusat2_cate/...	(셀프교육) 온실로의 초대 [지중해편, 열대편 2종]	상세 정보 홈페이지 참조	2023-03-01	2023-12-31	30분	
23	14	1	1	https://reserve.koagi.or.kr/data/edusat2_cate/...	나의 작은 화분	상세 정보 홈페이지 참조	2023-03-17	2023-11-30	120분 내외	
24	14	1	1	https://reserve.koagi.or.kr/data/edusat2_cate/...	(초등 4-6학년 단체) 사계절온실 대탐험	상세 정보 홈페이지 참조	2023-04-04	2023-11-28	120분	
25	18	1	1	https://yeyak.seoul.go.kr/cmsdata/web_upload...	(공일)2023년 8월 불암산나비정원 단체 관람		2023-08-01	2023-08-31		불암산
26	18	1	1	https://yeyak.seoul.go.kr/cmsdata/web_upload...	2023년 태국수(유아수) 가족 산림학습서비스		2023-07-01	2023-08-31		부천시

4. 총평 및 향후 계획

4.1 프로젝트 진행 중 문제점 및 해결

① Mysql 연동

- Python과 Mysql의 연동이 가능한 SQLAlchemy 라이브러리를 사용. 쿼리문과 ORM 모두 가능

② 상세 페이지 URL 접속

- 상세 페이지 URL이 Javascript 코드로 되어있을 때, requests로 접근이 어려움
- 해결방안으로 Selenium의 execute_script()로 상세 페이지 URL로 접근

③ 게시글의 HTML 위치 변화 or class="name" 변화

- CSS_SELECTOR로는 변화하는 페이지 크롤링이 어려움
- CSS_SELECTOR를 XPATH로 전부 변경

4.2 프로젝트를 통해 확보한 경험

- 요구사항 “분석”과 아키텍처 “설계”의 중요성
- DB를 활용하여 프로그램 구성
- 향후 유지보수 용이
- 프로그램 완성 및 사용자 우선

4.3 향후 목표

- 단순히 결과물을 만들어 내는거에서 끝나게 아니고, 사용자가 편하게 사용할 수 있어야하고, 유지보수가 잘 되도록 만들어 내는게 완성된 결과물이라는 것을 배웠습니다. 앞으로 보다 완성된 결과물을 만들어내는 개발자가 되고자 합니다.