

정보검색기술 기말프로젝트

KMA Search 프로젝트

- AI 기반 마켓데이터분석 플랫폼



컴퓨터정보통신대학원
디지털융합금융학과 1조
정낙현, 엄호천, 박진수, 김주원

Contents

1. 프로젝트 배경
2. Problem
3. KMA Search
4. 기대효과
5. 진행사항





프로젝트 배경

01

1-1. 주식투자



1-2. 주식 분석



2020 상반기 베스트 애널리스트

부문	성명	회사
반도체	김경민	하나금융투자
통신	김홍식	하나금융투자
통신 및 네트워크 장비 단말기	김복호	하나금융투자
전기전자·가전	김동원	KB증권
디스플레이	김동원	KB증권
인터넷·소프트웨어	황승택	하나금융투자
엔터테인먼트·레저	이기훈	하나금융투자
미디어·광고	이기훈	하나금융투자
유통	박종대	하나금융투자
운송	강성진	KB증권
증권·보험·기타금융	장효선	삼성증권
은행·신용카드	최정욱	하나금융투자
유틸리티	유재선	하나금융투자
자동차·타이어	김준성	메리츠증권
조선·중공업·기계	김현	메리츠증권
제약·바이오	선민정	하나금융투자
석유화학	윤재성	하나금융투자
음식료·담배	김정욱	메리츠증권
생활소비재(화장품 등)·교육	박종대	하나금융투자
섬유·패션	하누리	메리츠증권
철강·금속	박성봉	하나금융투자
건설·시멘트	채상욱	하나금융투자
지주회사	정대로	미래에셋대우
거시경제·금리	나중혁	하나금융투자
투자전략	이재만	하나금융투자
계량분석	이경수	하나금융투자
데일리 시황	김용구	하나금융투자
파생상품	김용구	하나금융투자
글로벌 투자전략(미국·선진국)	이재만	하나금융투자
글로벌 투자전략(중국·신흥국)	김경환	하나금융투자
원자재	황병진	NH투자증권
채권	윤여삼	메리츠증권
신용분석	김상훈	신한금융투자
글로벌 자산배분	김종원	현대차증권
글로벌 ETF	이창환	현대차증권
스몰캡	코스닥 벤처팀	하나금융투자





Problem

02

1-1. 문제점

1) 기존 문제점





KMA Search

03

3-1. KMA Search ?

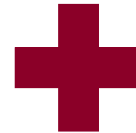
KMA Search
Korea university
MiaeassetDaewoo
AI analyst
Research



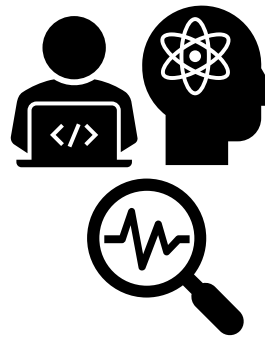
3-2. KMA Search ?



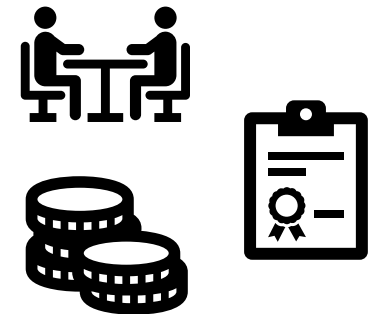
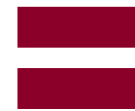
SNS



주가, 재무제표, 뉴스



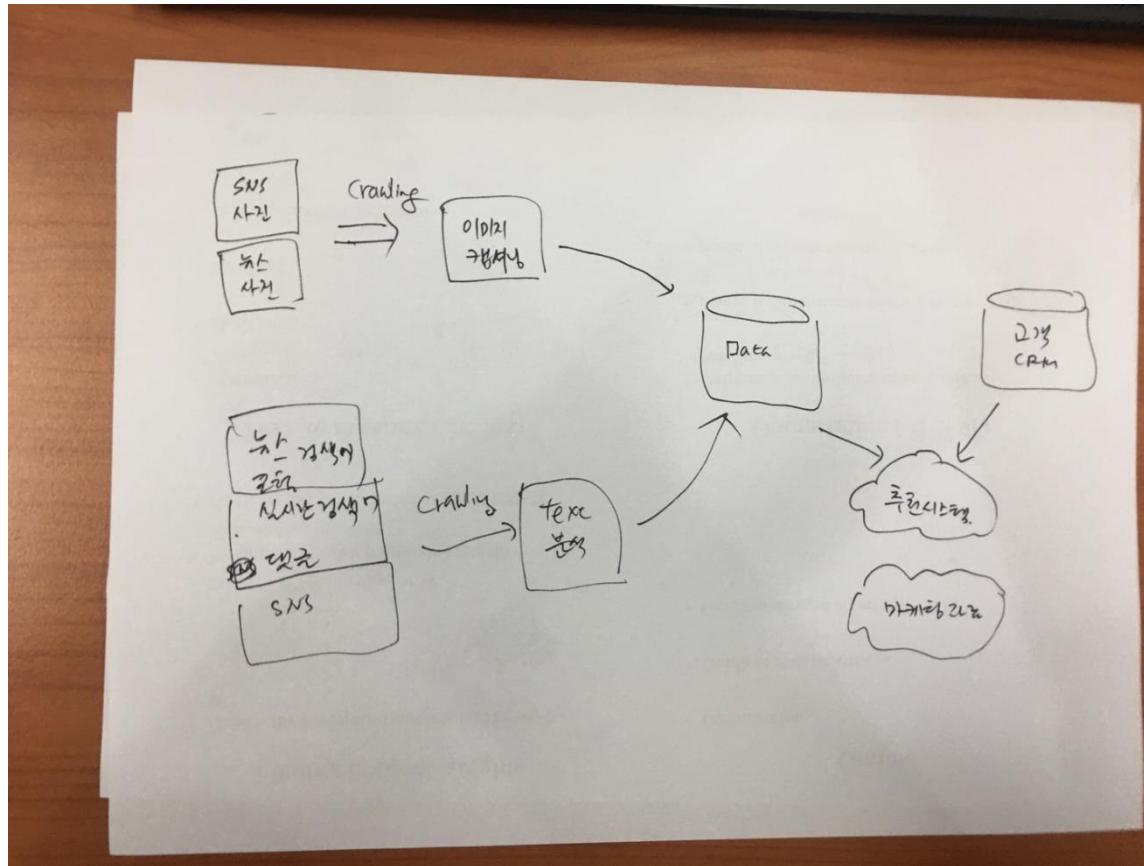
KMA Search



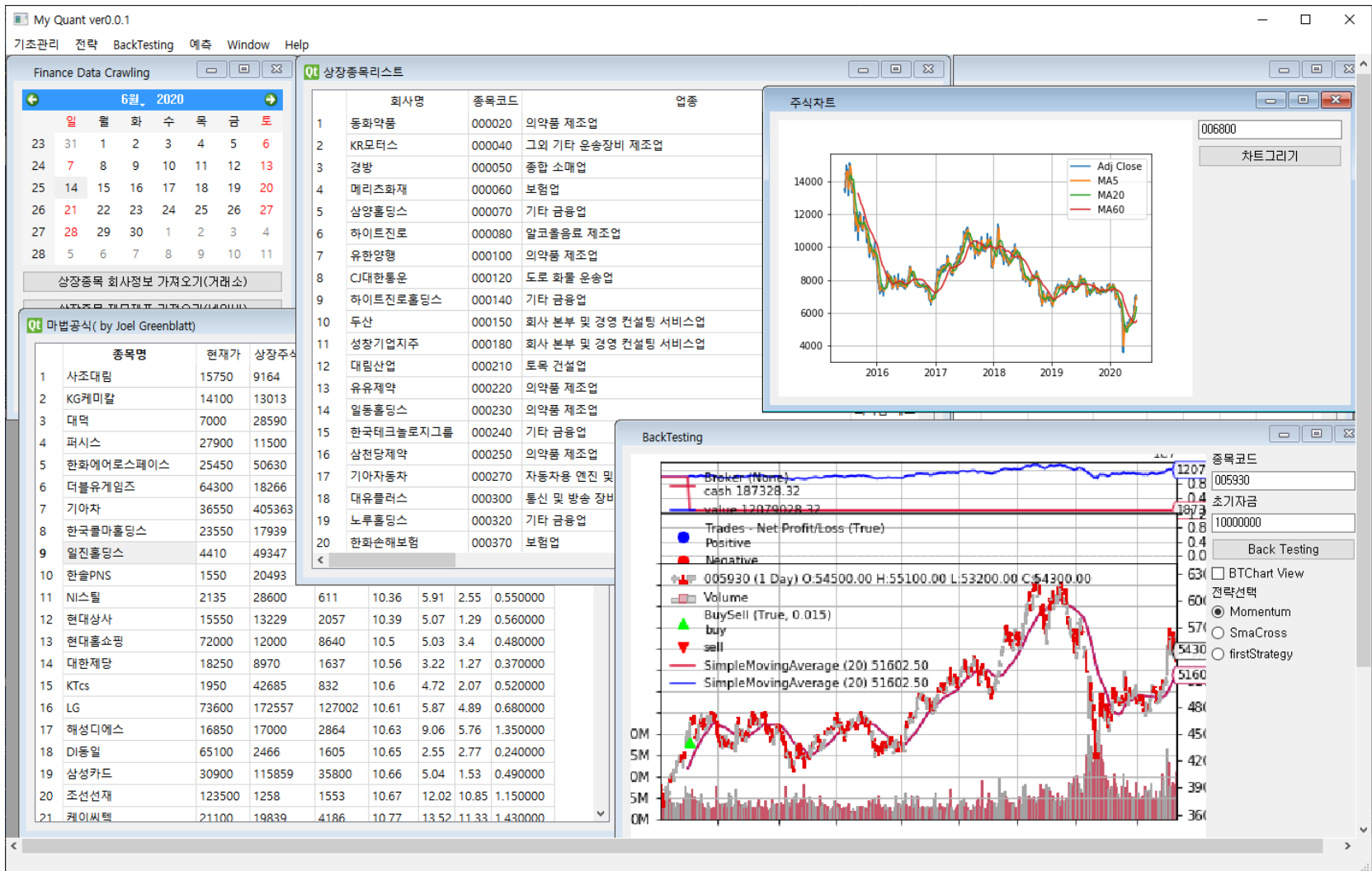
마케팅
투자 추천
리서치



3-3. KMA Search 구성도



3-4. KMA Search



3-5. SNS, 뉴스 이미지 → TEXT 가져오기

김세영, LPGA 펠리컨 챔피언십 우승...상금왕 선두

동료 2020-11-23 09:32 수정 2020-11-23 09:50

f t + -

스포츠파보는 기사

지난달 메이저 대회 이후 시즌 2승
시즌 상금, 올해의 선수 등 선두
별간 색 바지 선호는 우크 따라서
"올해 목표는 세계랭킹 1위다"



김세영이 23일(한국시간) 미국여자프로골프(LPGA) 투어 펠리컨 챔피언십에서 우승을 차지한 뒤 우승 트로피를 들어 올리고 있다. 발매/AP 연합뉴스

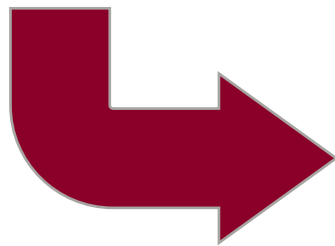
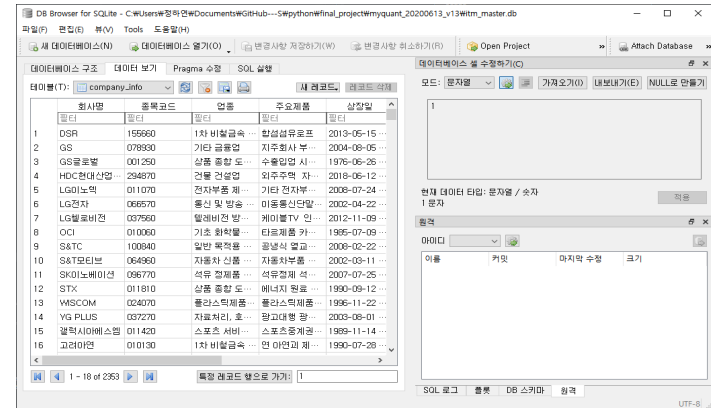
김세영(27)이 미국여자프로골프(LPGA) 투어 펠리컨 챔피언십(총상금 150만 달러) 우승을 차지

1. "핵 주력" 마이크 타이슨, 기회가 없었다
2. SK서 워터닉 집합, 현화 아규스로 간다
3. 양현종-이대호-최형우 위 경인 등-KBO FA전후 팀 단 확정
4. 박세리의 매질내 '명예의 전당' 입성
5. 타이슨 대 존슨 '세계 헤비급 레전드' 맞대결

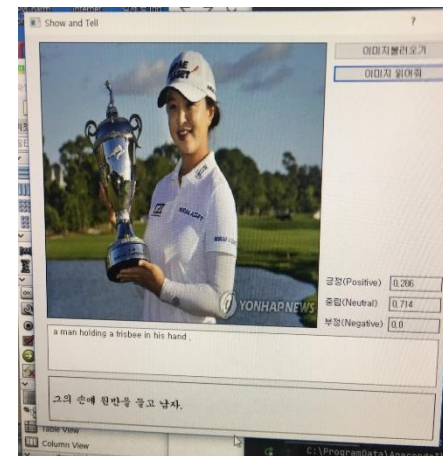
한겨레와 친구하기

f t

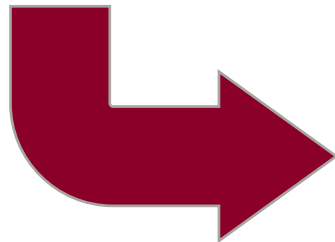
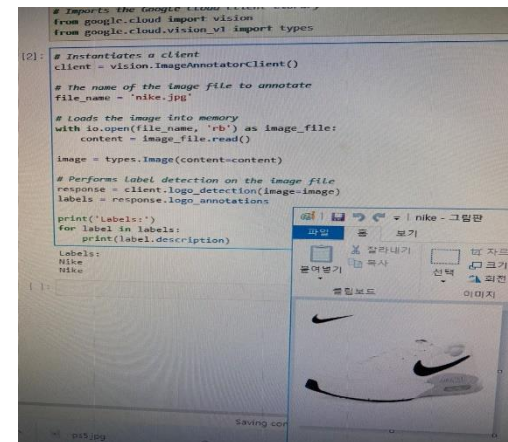
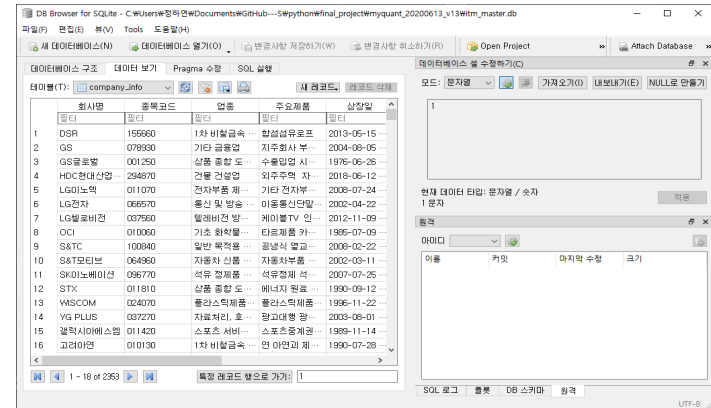
like 2,320



이미지 캡셔닝
긍정, 부정



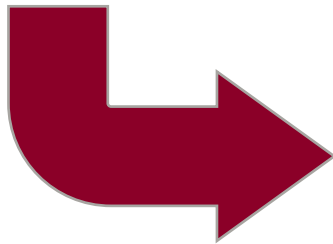
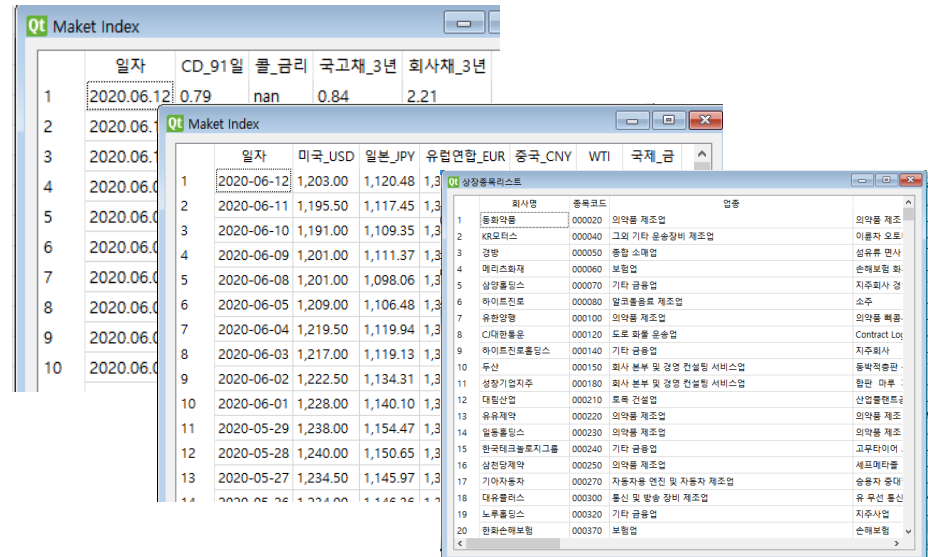
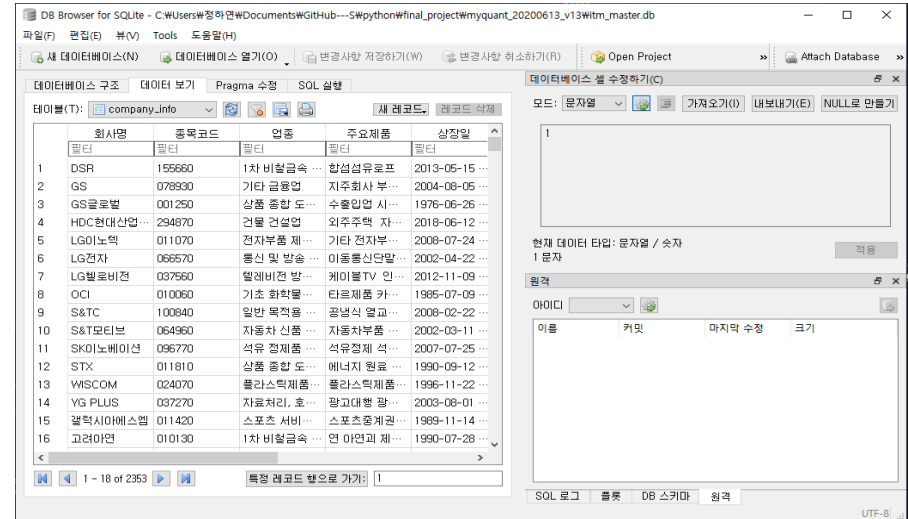
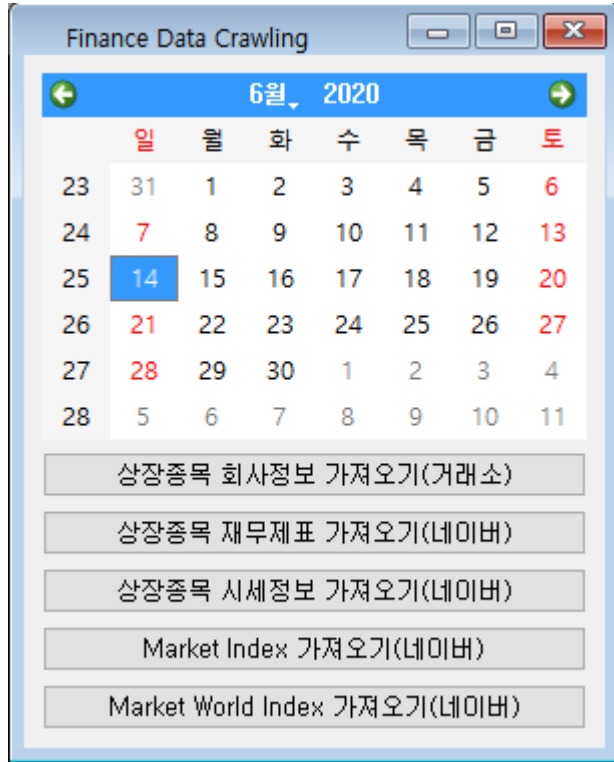
3-6. 이미지 → 로고 데이터 가져오기



기업로고 분석
긍정, 부정



3-6. 재무데이터 가져오기



3-5. 재무데이터 가져오기

Finance Data Crawling

6월 2020

	일	월	화	수	목	금	토
23	31	1	2	3	4	5	6
24	7	8	9	10	11	12	13
25	14	15	16	17	18	19	20
26	21	22	23	24	25	26	27
27	28	29	30	1	2	3	4
28	5	6	7	8	9	10	11

상장종목 회사정보 가져오기(거래소)
상장종목 재무제표 가져오기(네이버)
상장종목 시세정보 가져오기(네이버)
Market Index 가져오기(네이버)
Market World Index 가져오기(네이버)

Market Index

CD 91일 콜 금리 국고채 3년 회사채 3년

일자	미국_USD	일본_JPY	유럽연합_EUR	중국_CNY	WTI	국제_금
2020-06-12	1,203.00					
2020-06-11	1,195.50					
2020-06-10	1,191.00					
2020-06-09	1,201.00					
2020-06-08	1,201.00					
2020-06-05	1,209.00					
2020-06-04	1,219.50					
2020-06-03	1,217.00					
2020-06-02	1,222.50					
2020-06-01	1,228.00					
2020-05-29	1,238.00					
2020-05-28	1,240.00					
2020-05-27	1,234.50					

상장종목리스트

회사명	종목코드	업종
동화약품	000020	의약품 제조업
KR오티스	000040	그외 기타 운송장비 제조업
경방	000050	종량 소재업
메리츠화재	000060	보험업
상원통일스	000070	기타 금융업
하이브진로	000080	알코올음료 제조업
유한양행	000100	의약품 제조업
CJ대한통운	000120	도로 화물 운송업
하이브진로통일스	000140	기타 금융업
두산	000150	회사 본부 및 경영 컨설팅 서비스업
상원기업지주	000180	회사 본부 및 경영 컨설팅 서비스업
대원산업	000210	특수 건설업
유유제약	000220	의약품 제조업
일통통일스	000230	의약품 제조업
한국테크놀로지그룹	000240	기타 금융업
삼천양행	000250	의약품 제조업
가이아지주	000270	자동차용 엔진 및 자동차 부품 제조업
대우물류	000300	통신 및 방송 장비 제조업
노루통일스	000320	기타 금융업
한화손해보험	000370	보험업



DB Browser for SQLite

C:\Users\장하연\Documents\GitHub\---5\python\final_project\myquant_20200613_v13\ltn_master.db

데이터베이스 구조 데이터 보기 Pragma 수정 SQL 실행

데이터베이스 쿼리 수정하기(C)

테이블(T): company_info

회사명	종목코드	업종	주요제품	상장일
필터	필터	필터	필터	필터
DSR	155660	1차 비철금속	합성섬유로프	2013-05-15
GS	078930	기타 금융업	지주회사 부	2004-08-05
GS글로벌	001250	상품 종합 도	수출입업 시	1976-06-26
HDC현대산업	294870	건물 건설업	외주주력 자	2018-06-12
LG이노텍	011070	전자부품 제	기타 전자부	2008-07-24
LG전자	066570	통신 및 방송	이동통신단말	2002-04-22
LG유플러스	037560	멀티미디어	케이블TV 인	2012-11-09
OCI	010060	기초 화학물	타르제품 카	1985-07-09
S&TC	100840	일반 목적용	공냉식 열교환	2008-02-22
S&T모티브	064960	자동차 부품	자동차 부품	2002-03-11
SK이노베이션	096770	석유 정제품	석유정제 석	2007-07-25
STX	011810	상품 종합 도	에너지 원료	1990-09-12
WHISCOM	024070	물리화학제품	플라스틱제품	1996-11-22
YG PLUS	037270	자료처리, 호	광고대행 할	2003-08-01
엘지시애프트	011420	스포츠 장비	스포츠중계권	1989-11-14
고려아연	010130	1차 비철금속	연 아연과 제	1990-07-28

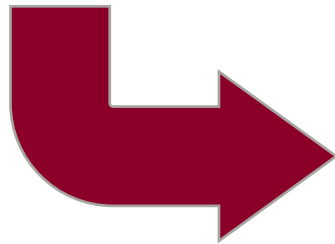
현재 데이터 타입: 문자열 / 숫자
1 문자

출력

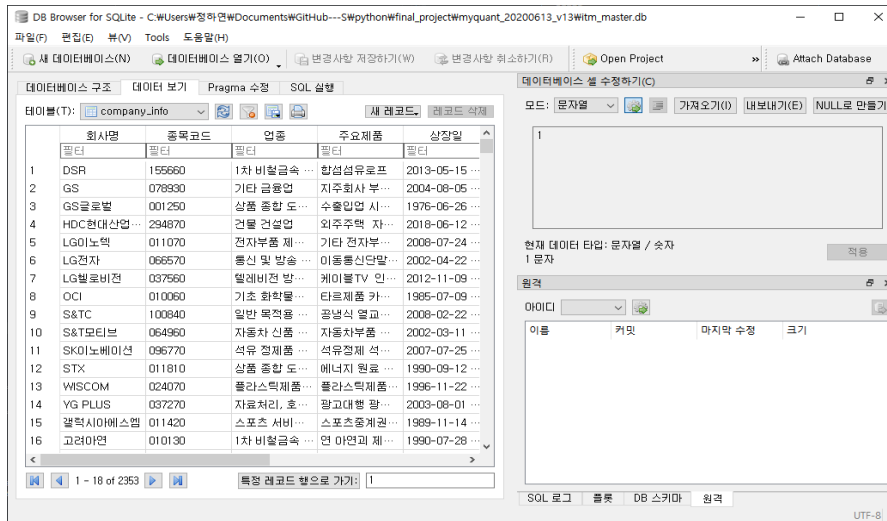
아이디

이름 커밋 마지막 수정 크기

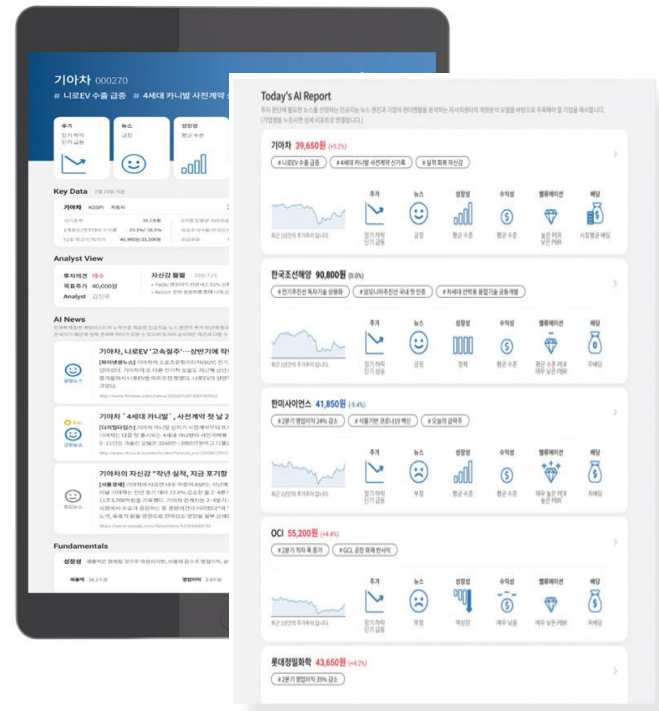
SQL 로그 실행 DB 스키마 환경



3-5. AI 리서치 보고서 생성



BERT, GPT3





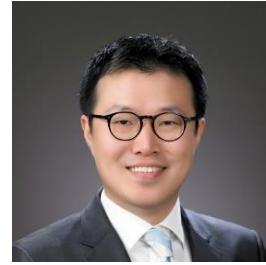
진행사항

04

4-1. 프로젝트 인원



정낙현 (조장)
미래에셋대우
금융상품시스템팀



엄호천
미래에셋대우
전략팀
프로젝트 기획



박진수
미래에셋대우 자금팀
프로그램 개발

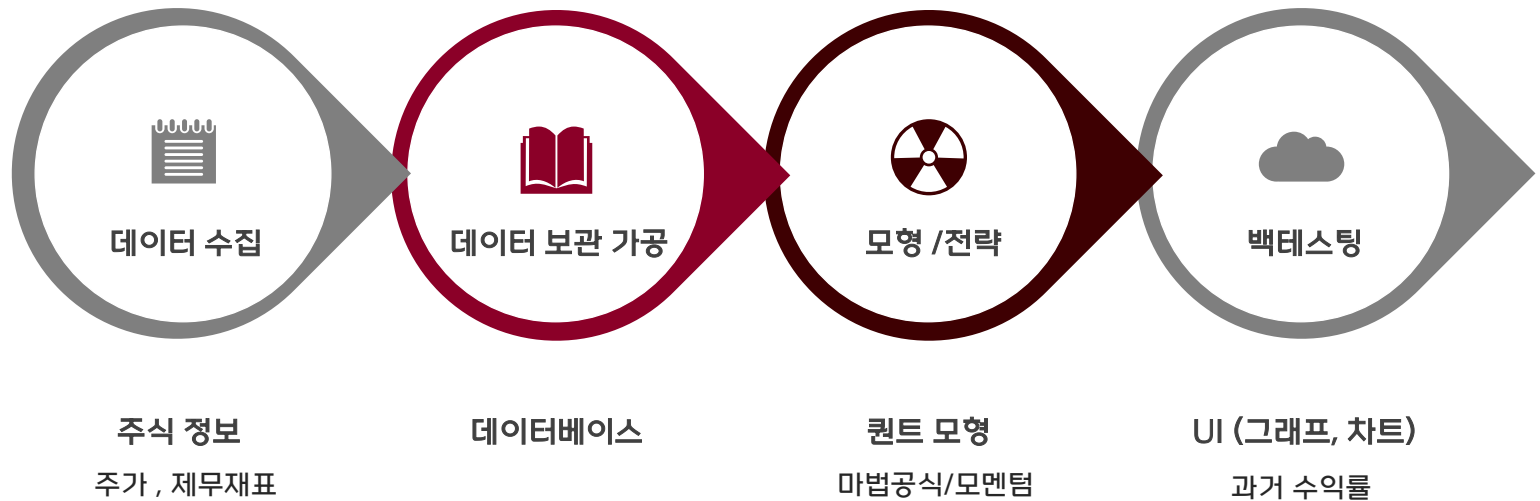


김주원
미래에셋대우 디지털
사례조사 ,업무분석

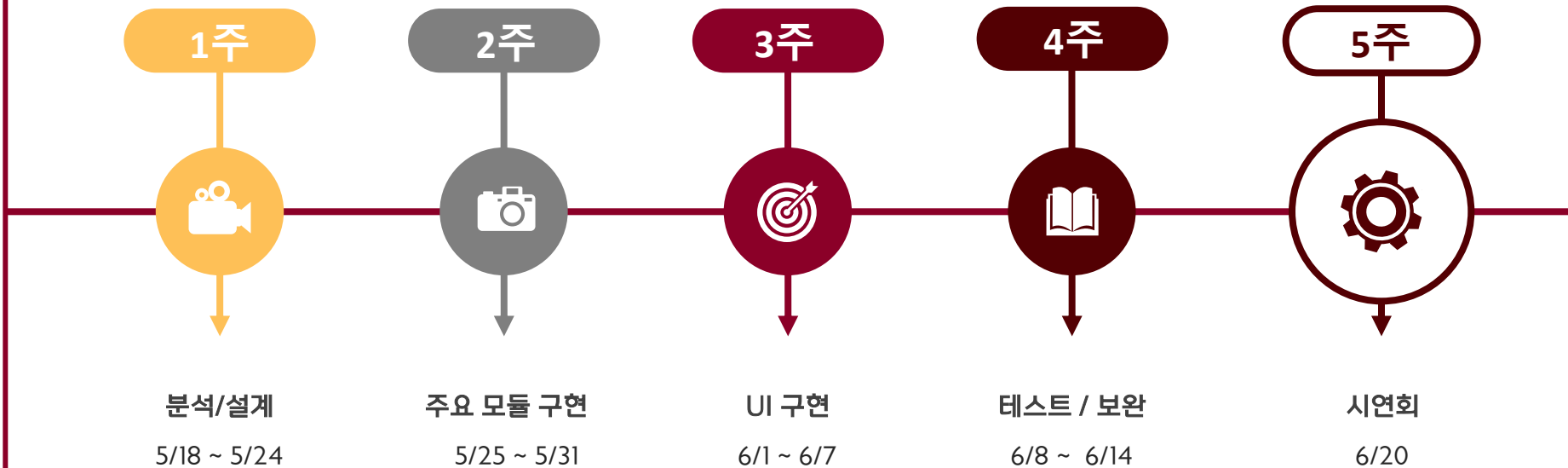


4-2. 프로젝트 개발 내용

1. 데이터 수집 모듈
2. 데이터 보관 및 가공 모듈
3. 모형 및 투자 전략 모듈
4. 백테스팅 모듈



4-3.프로젝트 진행 일정



4-4. 프로그램 목록

프로그램	설명
myquant.py	메인프로그램
company_info.py company_balance.py company_sise_all.py	종목관련 정보 크롤링 클래스 정의 <ul style="list-style-type: none">- 상장회사정보- 상장회사 재무제표- 상장회사 주요시세정보
market_index.py market_world_index.py	마켓데이터 크롤링 클래스 정의 <ul style="list-style-type: none">- 일자별 주요금리: CD,콜,국고채,회사채- 주요환율:USD,JPY,EUR,CNY- WTI, 국제 금시세
myquant_bt_strategy.py	백테스팅 클래스 BackTrader 모듈
itm_master.db	데이터베이스 SQLITE 3



4-5. 진행사항

1) 뉴스 분석

- 뉴스 크롤링 , 뉴스요약

2) 재무 분석

- 재무 데이터 크롤링 , 마켓데이터 수신

3) SNS 분석

- 실시간검색어 , 이미지 분석

4) 마케팅 자료

- 추천종목

5) 리서치 자료

- 시황,매수,매도 의견





기대효과

05

5-1. 기대효과

- 1) 뉴스 분석
- 2) 재무 분석
- 3) SNS 분석
- 4) 마케팅





감사합니다