

PJT명	관심 종목 토론 데이터 수집	
단계	04 PJT	
진행일자	2025.09.26	
예상 구현 시간	필수기능	6H
	심화기능	2H

1. 목표

- 웹 크롤링의 기본 개념과 프로세스를 이해한다.
- Selenium 과 BeautifulSoup 을 활용해 동적/정적 페이지에서 데이터를 추출할 수 있다.
- 관심 종목의 토론 데이터를 크롤링하여 데이터베이스에 저장하고, Django 웹 애플리케이션을 통해 출력·삭제 기능을 구현한다.
- OpenAI API 를 활용한 텍스트 분석을 통해 댓글 여론을 분류하는 기능을 추가적으로 구현할 수 있다.

2. 준비사항

1) 개발언어 및 툴

- Python, Django
- Selenium, BeautifulSoup4, Requests
- SQLite
- Visual Studio Code

3. 작업 순서

- 1) 페어와 같이 요구사항(기본/심화)을 분석한다.
- 2) 페어와 함께 개발 시 역할을 분담한다.
 - 팀장/네비게이터 등 역할 분담
- 3) [선택] 추가적인 아이디어를 적용하여 구현하고 프로젝트를 완성한다.
- 4) 코드 및 산출물을 정리하여 Git에 반영한다.

4. 요구사항

본 프로젝트는 웹 크롤링과 Django 웹 애플리케이션 구현을 통해 관심 종목의 토론 데이터를 자동으로 수집하고 관리하는 것을 목표로 한다. Selenium과 BeautifulSoup 같은 라이브러리를 활용하여 토스증권 웹사이트에서 특정 종목의 토론 댓글을 크롤링하고, 이를 데이터베이스에 저장하여 웹 화면에서 확인할 수 있도록 한다.

이를 통해 정적/동적 웹 크롤링 기법, 데이터베이스 연동, 웹 서비스 구현 과정을 체계적으로 경험할 수 있으며, 단순한 데이터 수집을 넘어 출력·삭제 기능 구현과 같은 실무적인 데이터 관리 능력을 익힐 수 있다. 추가적으로 OpenAI API를 활용한 댓글 여론 분석 기능을 구현함으로써 AI 기반 텍스트 분석 확장성도 체험할 수 있다.

● 요구사항 명세

- 아래에 제시된 기능 명세에 맞는 기능을 구현한다.
- 필수 기능은 반드시 구현해야 하며, 심화 기능은 팀 내 기획에 따라 선택적으로 구현한다.

번호	분류	요구사항명	요구사항 상세	우선순위
기능적 요구사항				
F01	종목 입력 화면	종목 입력 UI	"/crawlings/index/" 경로에서 사용자로부터 회사명을 입력 받을 수 있는 화면 구성	필수
F02	데이터 크롤링	종목 데이터 수집	사용자가 입력한 종목을 토스증권에서 검색 후 커뮤니티 탭의 댓글 데이터를 Selenium으로 크롤링하여 DB에 저장	필수
F03	데이터 출력	댓글 목록 출력	저장된 댓글 데이터를 웹 화면에 출력	필수
F04	데이터 삭제	댓글 삭제	각 댓글 옆에 삭제 버튼을 배치하고 DB에서 삭제할 수 있도록 구성	필수

F05	심화 기능	OpenAI API 분석	수집한 댓글을 OpenAI API에 전달하여 긍정/부정/중립 등 여론 분석 결과를 출력	심화
...
비 기능적 요구사항				
NF01	문서화	README 작성	구현 과정, 학습 내용, 어려웠던 점, 새로 배운 것들을 상세히 기록	필수
NF02	Git 관리	GitLab 제출	Git에 소스코드 및 산출물을 정리하여 업로드. 프로젝트명은 04_pjt로 지정	필수
NF03	유지보수	코드 정리	.gitignore를 추가하여 불필요한 파일 제외	필수
...

1) 기본(필수) 기능

본 프로젝트는 관심 종목의 토론 데이터를 크롤링하여 수집·관리할 수 있는 웹 애플리케이션 구현을 목표로 한다.

사용자가 입력한 종목명을 바탕으로 토스증권 사이트에서 관련 토론 댓글을 수집하고 데이터베이스에 저장한다.

저장된 댓글은 웹 화면에 출력되며, 필요 시 삭제할 수 있도록 기능을 제공한다.

A. 종목 입력 화면 구성하기 (요구사항 F01)

- 사용자로부터 회사 이름(종목명)을 입력받는 화면을 구성한다.
- /crawlings/index/ URL에 접속했을 때 입력 폼이 표시되도록 한다.
- 세부 기능:
 - **입력 폼**: 회사 이름을 입력할 수 있는 input 필드를 배치한다.
 - **버튼**: "검색" 버튼 클릭 시 종목명을 서버로 전달한다.
 - ✓ 요구사항 F02에서 구현
 - **화면 레이아웃**: 직관적인 UI로 종목 입력 → 검색이 가능하도록 구성한다.
- 결과물 예시

주식에 대한, 투자자들의 목소리

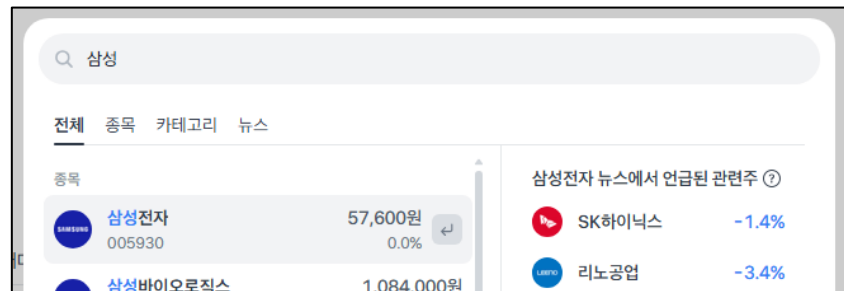
주식 댓글 분석기

회사 이름을 입력하세요:

B. 종목 데이터 크롤링하기 (요구사항 F02)

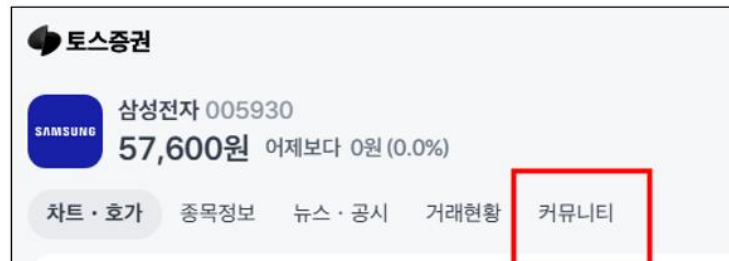
- 사용자가 입력한 종목명을 기준으로 토스증권 사이트에서 해당 종목의 토론 데이터를 크롤링한다.
- 세부 기능:
 - **검색 실행:** Selenium을 통해 토스증권 메인 페이지에 접근 후, 종목명 검색을 수행한다.

✓ 토스 증권 삼성 검색 예시



- **종목 선택:** 동일한 이름의 회사가 여러 개일 경우, 가장 상단의 회사를 선택한다.
- **댓글 수집:** 커뮤니티 탭에 접근하여 댓글 내용을 모두 수집한다.

✓ 실제 토스 증권에서 댓글 위치 파악



- **DB 저장:** 수집된 데이터를 데이터베이스에 저장한다. (테이블 구성: 회사 이름, 종목 코드, 댓글 내용, 저장 일자)

- **결과물 예시**

- 데이터베이스에 저장된 댓글 정보

C. 크롤링 데이터 출력하기 (요구사항 F03)

- 데이터베이스에 저장된 댓글 데이터를 웹 화면에 출력한다.
- 세부 기능:
 - **목록 출력:** 수집된 댓글들을 리스트 형식으로 나열한다.
 - **종목 정보 표시:** 댓글과 함께 회사 이름, 종목 코드도 함께 표시한다.
 - **결과 화면:** 사용자에게 직관적으로 보이도록 표 또는 카드 형태로 디자인한다.

- **결과물 예시**

주식에 대한, 투자자들의 목소리

주식 댓글 분석기

회사 이름을 입력하세요:

삼성전자 (005930)

수집된 댓글:

- 삼성아 . 나 . 또 들어가게하지말고 올라가 !!!
- 내일 주주총회니까 버터볼게?
- 젠슨황 믿는다?
- 여러분 다같이 영차영차 해보죠
- 영차영차
- 개미 99퍼가 털릴때까지 안올리는게
- 외국인들 기를쓰고 매집하고 있는데....조금 남기겠다고 낮은 금액에
- 버거들 입속에 넣어주는 개미들.....-.-;
- 외계인들 매수 멈췄는데
- 팔지도 않고 내일 어떻게 되려나

D. 댓글 삭제하기 (요구사항 F04)

- 출력된 댓글 옆에 삭제 버튼을 배치하여, 필요 시 데이터를 제거할 수 있도록 한다.
- 세부 기능:
 - **삭제 버튼 구성:** 각 댓글마다 삭제 버튼을 둔다.
 - **DB 연동:** 버튼 클릭 시 해당 댓글이 데이터베이스에서 삭제되도록 한다.
 - **화면 반영:** 삭제 완료 후 웹 화면에서 즉시 반영되도록 한다.
- 결과물 예시

주식에 대한, 투자자들의 목소리

주식 댓글 분석기

회사 이름을 입력하세요:

삼성전자 (005930)

수집된 댓글:

- 삼성아 . 나 . 또 들어가게하지말고 올라가 !!!
- 내일 주주총회니까 버텨볼게?
- 젠슨황 믿는다?
- 여러분 다같이 영차영차 해보죠
- 영차영차
- 개미 99퍼가 털릴때까지 안올리는게
- 외국인들 기를쓰고 매집하고 있는데....조금 남기겠다고 낮은 금액에
- 버거들 입속에 넣어주는 개미들.....-,-;
- 외계인들 매수 멈췄는데
- 팔지도 않고 내일 어떻게 되려나

E. OpenAI API를 활용한 종목 분석 (요구사항 F05, 심화)

- 수집된 댓글 데이터를 바탕으로 전체 여론을 분석한다.
- OpenAI API를 활용하여 긍정, 부정, 중립 등으로 분류 결과를 도출한다.
- 세부 기능:
 - **데이터 전달**: 크롤링한 댓글들을 OpenAI API로 전달한다.
 - **프롬프트 구성**: 댓글 내용이 여론 분석이 가능하도록 프롬프트를 설계한다.
 - **모델 활용**: 생성형 AI를 통해 분석 결과를 생성한다.
 - **결과 출력**: 페이지 하단에 “여론 분석 결과” 영역을 구성하여, 전체적인 종목 토론 분위기를 사용자에게 제공한다.
- 결과물 예시

주식에 대한, 투자자들의 목소리

주식 댓글 분석기

회사 이름을 입력하세요: 삼성전자 검색

삼성전자 (005930)

수집된 댓글:

- 삼성아. 나. 또 들어가게하지말고 올라가 !!! 삭제
- 내일 주주총회니까 버티볼게? 삭제
- 전상황 만는다? 삭제
- 여러분 다같이 영차영차 해보죠 삭제
- 영차영차 삭제
- 개미 99퍼가 탈릴때까지 안올리는게 삭제
- 외국인들 기를쓰고 매집하고 있는데...조금 남기겠다고 낮은 금액에 삭제
- 배가들 입속에 넣어주는 개미들.....? 삭제
- 외국인들 매수 멈췄는데 삭제
- 팔지도 않고 내일 어떻게 되려나 삭제

'gpt-4o-mini' 분석:

이 댓글들은 삼성전자 주식에 대한 투자자들의 감정과 기대를 반영하고 있습니다. 다음과 같은 몇 가지 주요 포인트로 분석할 수 있습니다: 1. "투자의 불안정성": "삼성아. 나. 또 들어가게하지말고 올라가 !!!"와 같은 요청은 투자자들이 주가 상승을 간절히 바라는 심리를 보여줍니다. 이는 주식 시장의 불안정함에 대한 우려를 나타내기도 합니다. 2. "존재하는 기대": "내일 주주총회니까 버티볼게?"라는 댓글은 주주총회가 주가에 긍정적인 영향을 미칠 것이라는 기대를 내포하고 있습니다. 이는 특정 이벤트에 대한 기대감이 여전히 존재하고 있음을 시사합니다. 3. "시장 경쟁에 대한 의견": "개미 99퍼가 탈릴때까지 안올리는게 외국인들 기를쓰고 매집하고 있는데..."와 같은 댓글은 소규모 투자자(개미)와 외국인 투자자 간의 힘의 불균형을 우려하는 목소리입니다. 이는 시장의 불균형성을 지적하며, 소규모 투자자들이 손해를 볼 가능성에 대한 우려를 나타냅니다. 4. "감정의 혼재": "영차영차" 같은 응원 의 표현이 보이지만, 뒤따르는 부정적인 표현은 투자자들의 혼란과 고민을 반영합니다. 특히 외국인 투자자들의 매수 중단에 대한 언급은 시장에 대한 신뢰와 불안정성이 복합적으로 나타나는 지점입니다. 종합적으로, 삼성전자에 대한 댓글들은 투자자들의 간절한 희망과 함께 불안감이 혼재된 모습을 보여줍니다. 주주총회에 대한 기대감과 더불어, 대규모 투자자와 소규모 투자자 간의 불균형에 대한 우려가 지배적입니다. 이러한 요소들을 종합적으로 고려할 때, 현재 여론을 부정적으로 판단할 수 있습니다.

5. 참고자료

1) Selenium 공식 문서

- <https://www.selenium.dev/documentation/>

2) BeautifulSoup4 공식 문서

- <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>

3) Requests 공식 문서

- <https://requests.readthedocs.io/en/latest/>

4) Django 공식 문서

- <https://docs.djangoproject.com/>

5) OpenAI API 문서 (심화 기능용)

- <https://platform.openai.com/docs/>

6. 결과

최종적으로 제출해야 할 항목은 아래와 같다

1. 구현 소스 코드

- Django 프로젝트 및 crawlings 앱 소스 코드

2. 실행 결과 캡처본

- DB 및 웹 페이지 결과 캡처본

3. README.md

- 구현 기능 설명, 학습 내용, 느낀 점 포함

위 내용들을 모두 포함하여 GitLab에 업로드 한다.

GitLab 프로젝트 이름은 프로젝트 번호 + _pjt 로 지정한다 (04_pjt)