

# Report

김준호

## 코드 설명

#hankyung\_scraper.py

init: 클래스의 초기화 메서드. 시작 날짜, 종료 날짜, 출력 파일 경로 등을 초기화한다. 또한, Chrome 드라이버를 설정하고 WebDriverWait를 통해 드라이버가 요소를 기다리도록 설정한다.

scrape: 한경닷컴의 모든 뉴스를 스크랩하는 메서드다. URL을 열고 경제 섹션으로 이동한다.

load\_more\_button.click(), last\_date:

"더보기" 버튼 클릭: 더보기 버튼을 반복적으로 클릭하여 추가 기사를 로드한다.

날짜 확인: 로드된 기사들의 마지막 날짜를 확인하여 시작 날짜 이전인지 확인한다. 시작 날짜 이전이면 루프를 종료한다.

all\_article\_links = [] ~ all\_article\_links.append(link)

기사 링크 수집: 로드된 기사들의 링크를 수집하여 리스트에 저장한다.

for link in all\_article\_links:

기사 내용 스크랩: 각 기사의 링크를 열어 HTML을 파싱하고, 제목, 날짜, 수정 날짜, 기사 내용을 추출하여 데이터 리스트에 추가한다.

except Exception as e: , finally:

예외 처리: 오류 발생 시 메시지를 출력한다.

결과 저장: 수집된 기사 데이터를 JSON 파일로 저장하고 드라이버를 종료한다.

```
#main.py
```

```
def create_parser() -> ArgumentParser:
```

명령줄 인자를 정의하고 처리하는 ArgumentParser 객체를 생성한다. 시작 날짜, 종료 날짜, 출력 파일 경로를 인자로 받는다

```
if __name__ == "__main__":
```

ArgumentParser를 통해 인자를 파싱하고, HanKyungScraper 객체를 생성하여 스크래핑 작업을 수행한다.