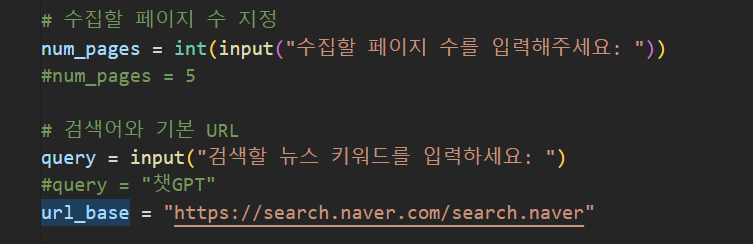


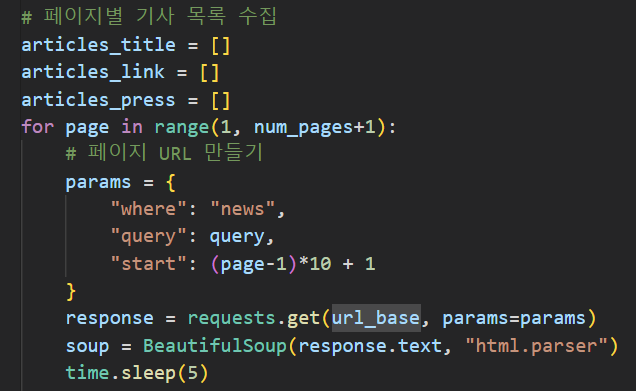
Ip가 막히거나 서버 부하를 방지하기 위한 time.sleep()함수를 사용하기 위해 time을 import 해주었습니다.

그리고 후에 나올 결과를 dataframe을 만들어 엑셀파일로 변환하기 위하여 panadas를 import해주었습니다.

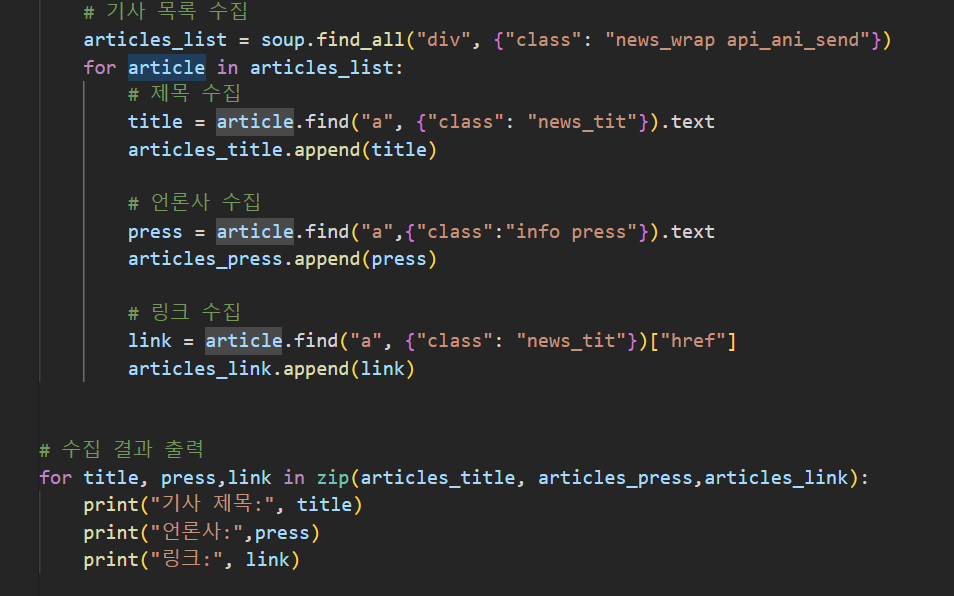


수집할 페이지 수를 num\_pages로 받고 query로 검색하고 싶은 뉴스 키워드를 입력받습니다.

그리고 url\_base를 설정했습니다. 수집한 데이터 셋 파일에는 챗 GPT 키워드를 입력받고 페이지 수를 5로 입력받았습니다.



Articles\_title, articles\_link, articles\_press를 각각 기사제목, 링크, 언론사 제목으로 받고 나중에 활용하기위해 리스트로 받아주기 위해 빈 리스트로 설정하였습니다. 그리고 url\_base에 news페이지의 입력받은 키워드(query),페이지수(start)를 합쳐 requests.get()함수를 사용하여 url을 html로 받아줍니다. 그리고 news\_wrap api\_ani\_send, 클래스를 갖는 모든 div를 뽑아서 articles\_list에 리스트로 저장해주고, 그 안에 있는 요소들을 추출하여 title,press,link 변수에 각각 저장해주었습니다.



그리고 articles\_title,articles\_press,articles\_link 리스트들의 인덱스에 해당하는 제목,언론사,링크 데이터를 묶어서 하나의 튜플로 만들고 하나씩 꺼내 출력하였습니다. .



출력한 요소들을 dataframe으로 만들어주고 보기 편하도록 엑셀 파일로 변환하였습니다.

수집한 데이터 셋은 네이버 뉴스의 제목, 언론사,링크 이고 원하는 페이지 수와 검색하고 싶은 뉴스 키워드를 입력받아 제목,언론사,링크를 한눈에 알아볼 수 있도록 하였습니다.