예상 프로그램 전달 일자 – 2024.05.06(월)

대상 홈페이지 주소

-<https://www.hoopcity.co.kr/>

-<https://www.kasina.co.kr/>

의뢰 요구 사항

1. 두 쇼핑몰에서 최신 상품 정보를 크롤링
2. 크롤링한 정보를 가공해 디스코드 웹훅 메시지 전송
3. 크롤링한 정보를 DB (DB or 엑셀 파일)에 저장
4. 사용자가 지정한 시간 간격으로 신제품을 크롤링

상세 요구 사항

1. 쇼핑몰 웹사이트의 IP 벤 방지를 위해 프록시 사용을 희망 (프록시 제공)
2. 디스코드 웹훅 내 제품의 사진과 이름이 표시되고 이름을 클릭하면 해당 제품의 링크로 연결
3. 재입고는 품절된 사이즈가 구매가능할시 (버튼활성화 or API나 json으로 활성화되었을시) 웹훅 전송
4. PC 프로그램으로 제작

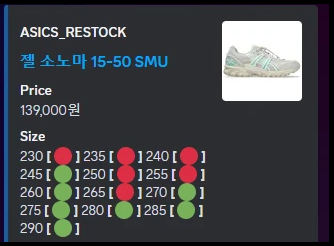
추가 질문 사항

1. 크롤링 해야하는 신상품의 상품 정보를 어디까지 크롤링 하야 하나요? (예를 들면 가격, 사이즈 등등)
   1. 크롤링해야하는 상품정보는 브랜드, 제품이름, 가격 이렇게 3가지 요소면 좋습니다
2. 사진은 제품의 상세사진 모두를 크롤링 하나요? 아니면 대표 사진 하나만 크롤링 하나요?
   1. 사진은 가장 첫 대표이미지로 진행하면 좋습니다
3. 프록시 적용 테스트를 위해서 샘플 프록시 주소 1~2개를 제공 해주실수 있나요?
4. 재입고 정보 크롤링에 대해서 조금 더 구체적으로 말씀해주실수 있을까요?

프록시 주소 목록

124.175.161.152:11089:6071GGAASSeafec:ULbkSz29 124.175.161.153:11089:6071GGAASSeafec:cpR1LcWv 124.175.161.154:11089:6071GGAASSeafec:oujCHEgn 124.175.161.155:11089:6071GGAASSeafec:AaiVs4Xi 124.175.161.156:11089:6071GGAASSeafec:5WE5GuBE

예시 사진



프로그램 플로우

1. PROXY 주소를 이용해 두 웹사이트의 가장 최근 크롤링 한 마지막 상품 (프로그램 최초 실행 후 첫 크롤링에서는 마지막 상품의 url만 받아오고 5번으로 이동)의 url을 기준으로 그 이후에 올라온 상품 정보를 크롤링
2. 상품 정보를 DB (상품 대표 사진, 엑셀 파일, 이름 예시 : 202404221030\_웹사이트이름 – 여기서 년,월,일,시간,분 순서)에 저장.
3. 크롤링한 정보를 디스코드 웹 훅으로 보내기 위한 포맷으로 가공
4. 가공된 정보를 디스코드 웹 훅을 통해 메시지 전송
5. 사용자 설정 시간 (예를 들어 30분 간격)이 지난 후 다시 1부터 반복