

| 자유 게시판 | >

## 웹사이트 파싱할때 주의하세요!



**changmun** 챗봇 하수 ★ 1:1 채팅  
 2020.08.02. 10:15 조회 202

댓글 10 URL 복사

웹 크롤링을 누구나 쉽게 할 수 있는 시대지만, 그만큼 악용되기 쉽고 본의 아니게 범죄가가 될 수도 있습니다.

국내 최초 유튜브 빅데이터 관련 사업을 하던 작은 제 스타트업이 대량의 크롤링 피해를 입었습니다. 자료확보과정, 검찰 고소와 수사 진행, 범인과의 합의까지를 글로 담아 또다른 피해자를 예방하고 여러분들의 의견을 청취하고자 합니다.

작년 모 포털이 당근마켓 U.I. 카피했다가 시끄러웠던 적이 있죠. 그 때도 많은 분들의 지지가 큰 도움이 됐던 걸로 기억합니다. 제 글을 통해 크롤링/저작권/고소에 대한 지식도 접하시고 널리 공유와 관심도 부탁드립니다.

감사합니다.

**[YouTube 서비스 이용 약관]**

유튜브에 가입하여 동영상을 업로드한 youtube.com 회원은 유튜브의 서비스 약관(<https://www.youtube.com/static?gl=KR&template=terms>)을 적용받고 있습니다. 해당 서비스 약관 6. C 에 의해 사용자는 콘텐츠의 소유권 전부를 보유하나, YouTube에게 콘텐츠를 제출함으로써, 선전하고 재배포하는 것에 동의하였습니다.

**[YouTube API Services Terms of Service]**

소셜러스의 1차 raw Data는 YouTube API Services Terms of Service (<https://developers.google.com/youtube/terms/api-services-terms-of-service-apac>)를 준수하여 수집됩니다.

✓ **[소셜러스 데이터의 생산, 저작권 및 소유]**

YouTube에서 수집된 1차 데이터는 (주)필로비즈의 수집/분석 시스템을 통해 통계학 및 독자적인 알고리즘으로 재가공되며, 재가공된 콘텐츠/데이터에 대한 모든 저작권 및 소유권은 PhiloBiz Inc.에 귀속됩니다.

이는 Media Index, Social Index, Ranking Index 및 독자적인 방법론으로 수집/분석/제공되는 좋아요/댓글/싫어요 등 소셜러스 시스템을 통해 계산되고 웹을 통해 제공되는 모든 콘텐츠에 해당됩니다.

구독자, 조회수는 유튜브의 원시(raw) 데이터를 수집하지만, 수집하는 시점과 방법에 따라 다른 데이터가 추출되며, 좋아요/댓글/싫어요 및 동급 데이터 평균, 성장률 데이터는 '소셜러스만의 알고리즘'이 '특정 범위'의 데이터를 가공하여 생산됩니다.

소셜러스는 '한국저작권보호원'의 전문가 검수를 거쳐 '저작권 OK' 인증을 받았으며 저작권 및 콘텐츠산업진흥법 등 관련 법령의 보호를 받고 있습니다.

✓ **[크롤링 및 소셜러스 콘텐츠/데이터의 수집 거부]**

socialerus.com 을 통해 제공되는 모든 정보와 데이터의 크롤링은 금지됩니다.

philosophiren.com

크롤링과 저작권 침해 그 시 지혜 이대기

## 그들강과 저작권 침해 고소 진행 중...

흔해빠진 크롤러로 누구나 피해자, 피의자가 될 수 있다 !! 지금 시대는 누구...

위 사진은 2일전에 올라온 글입니다..

100개가 넘는 아이피로 지속적인 크롤링(파싱)을 감지하였고, 그로인해 서버 응답량과 DB사용량이 폭주해 서버 응답속도가 많이 느려졌다고합니다

이로인해 정상적으로 사이트 운영을 못하자, 수많은 아이피를 가지고 경찰에 넘겼고 실제로 그 아이피로 크롤링을 한 사람이 누군지 잡혔답니다....

다들 크롤링 할때 조심하세요!!



changmun님의 게시글 더보기 >

❤ 좋아요 2 💬 댓글 10

🔗 공유 신고

댓글 등록순 최신순

댓글알림



마일로

ㄷㄷ 이제 직접 만들면 됩니다(?)

2020.08.02. 10:17 답글쓰기



changmun

작성자

한국공공데이터에서 JSON받아와서 파싱해야 안전할거같네요 ㅜㅜ  
다른 사이트는 위험해욧....

2020.08.02. 10:18 답글쓰기



마일로

changmun 흠 그래도 한강사이트 잘 쓰고있습니다

2020.08.02. 10:19 답글쓰기



changmun

작성자

마일로 <https://hangang.life/api>

여기가 더 파싱하게 좋을거예요!!

메인페이지 구조가 좀 복잡하니..

2020.08.02. 10:19 답글쓰기



마일로

changmun 오호

<http://kmilo.pe.kr/한강.php>

이거 저사이트 크롤링으로 바꿔야지

2020.08.02. 10:20 답글쓰기



Hibot

우리가 평소에 쓰는 크롤링은 조직적이지도않고 수십만개의 쿼리를보내는방식이 아닌지라...  
저 경우는 정말 최악의 경우입니다  
(혹여나 크롤링이 금지일것이라는 사람이있을까...)

2020.08.02. 11:48 답글쓰기



changmun

작성자

그래도 조심하라는 뜻으로 올린글입니다...

2020.08.02. 11:48 답글쓰기



인디벨

robots.txt에 막 써놔도 안 지키는 경우가 과반수

2020.08.02. 12:09 답글쓰기



사로로

이게 정답

2020.08.02. 12:24 답글쓰기



Kiri

헉 ㄷㄷ

2020.08.02. 16:20 답글쓰기



Hibot

댓글을 남겨보세요



등록

글쓰기

답글

목록

▲ TOP

## '자유 게시판' 게시판 글

이 게시판 새글 구독하기

사자성어 출력 api 🤖 [3]

EliF

2020.08.02.

비어있어요 [6]

imroy09

2020.08.02.

웹사이트 파싱할때 주의하세요! 🤖 [10]

changmun

2020.08.02.

옛 카페등급이 내려가진 않겠네요! 🤖 [2]

마일로

2020.08.02.

형들 주소찍지마요 ㅌㅌ 🤖 [5]

맨들

2020.08.02.

1 2 3

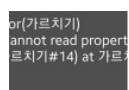
전체보기

## 이 카페 인기글

간단한 검색결과 가져오기

마일로

♡1 💬5



도미님의 학습소스 질문

바탕화면에 많은 기능이 추가되었습니다.

♡0 💬8

nolbo

♡0 💬7



[스압] loco는 도대체 왜, 언제 시작되었으며, 아닥로충은 언제 나타났는가? 카카오 계정은 하는데 전화번호를 몰라서;;; 지민 OtakidTory

♡0 💬7

카링 사용하는데

CiBot

♡0 💬10



사자성어 퀴즈 소스



카톡을 한폰에서 두개쓰는데 답장도 두계정 전부 다 합니다 ㅌㅌ 카페가 터지고 있다 범이야

AlphDo 16  
♡2 💬12

