

텍 마 - 크롤링

원래 피피티 잘만드는데 시간 없어서 ㅎ..

✓ 크롤링의 종류

1. OPEN API(페북, 트위터 등등)
 2. 정적 웹페이지 – BS4
 3. 동적 웹페이지 – Selenium + BS4
 4. 나만의 크롤러 만들기 – scrapy 등등
- 조금 복잡함

❖ 파이썬 기준입니다

1. OPEN API(페북, 트위터 등등)

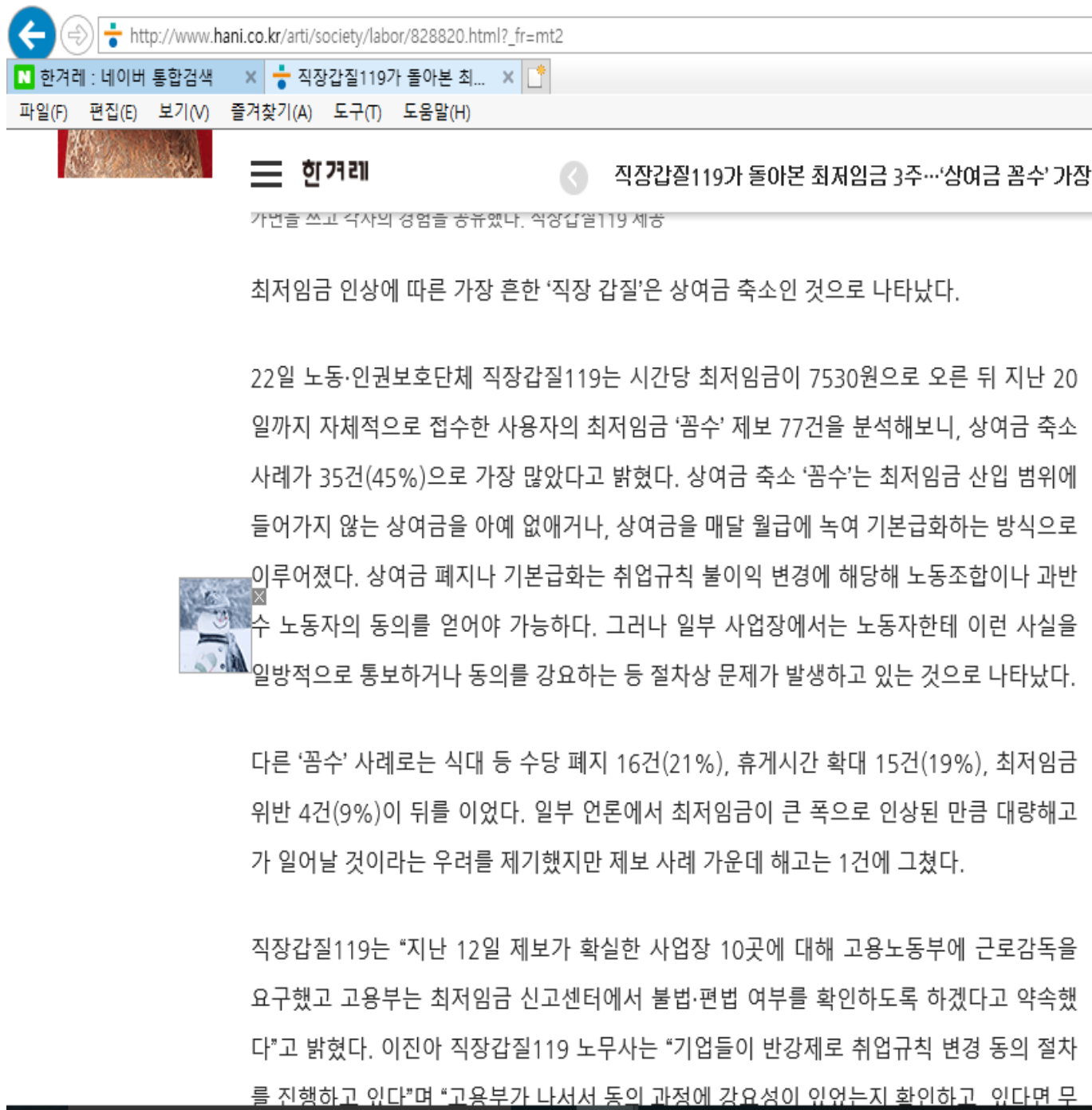
- 페북이나 트위터에서 자체적으로 제공하는 코드

장점 : HTML, CSS 보는 법 같은 거 몰라도 된다. 코드 다루기가 매우 쉽다.

단점 : 개수 제한, 시간 제한 등 제약이 많다. OPEN API에서 제공하는 기능만 써야 되므로 코드를 변형할 수 없다. 일반 웹페이지는 못 긁는다.

2. 정적 웹페이지 – BS4

따로 마우스 클릭이 필요 없는 평범한
웹페이지를 긁을 때 사용
단, HTML 구조를 봐야함~



The screenshot shows a web browser window with the URL http://www.hani.co.kr/arti/society/labor/828820.html?_fr=mt2. The browser has two tabs: '한겨레 : 네이버 통합검색' and '직장갑질119가 돌아본 최저임금 3주...'. The page title is '직장갑질119가 돌아본 최저임금 3주... '상여금 꼼수' 가장'. The article content discusses the impact of minimum wage increases on overtime pay (상여금). It states that the minimum wage has risen to 7,530 won per hour, and over 77 cases of overtime pay reduction have been reported. The article notes that while some companies have reduced overtime pay, others have increased it, and some have used loopholes to avoid paying overtime. It also mentions that some companies have used the minimum wage increase as a pretext to change employment terms without the employee's consent.

한겨레

직장갑질119가 돌아본 최저임금 3주... '상여금 꼼수' 가장

가변을 쓰고 각자의 경험을 공유했다. 직장갑질119 제공

최저임금 인상에 따른 가장 흔한 '직장 갑질'은 상여금 축소인 것으로 나타났다.

22일 노동·인권보호단체 직장갑질119는 시간당 최저임금이 7530원으로 오른 뒤 지난 20일까지 자체적으로 접수한 사용자의 최저임금 '꼼수' 제보 77건을 분석해보니, 상여금 축소 사례가 35건(45%)으로 가장 많았다고 밝혔다. 상여금 축소 '꼼수'는 최저임금 산입 범위에 들어가지 않는 상여금을 아예 없애거나, 상여금을 매달 월급에 녹여 기본급화하는 방식으로 이루어졌다. 상여금 폐지나 기본급화는 취업규칙 불이익 변경에 해당해 노동조합이나 과반수 노동자의 동의를 얻어야 가능하다. 그러나 일부 사업장에서는 노동자한테 이런 사실을 일방적으로 통보하거나 동의를 강요하는 등 절차상 문제가 발생하고 있는 것으로 나타났다.


다른 '꼼수' 사례로는 식대 등 수당 폐지 16건(21%), 휴게시간 확대 15건(19%), 최저임금 위반 4건(9%)이 뒤를 이었다. 일부 언론에서 최저임금이 큰 폭으로 인상된 만큼 대량해고가 일어날 것이라는 우려를 제기했지만 제보 사례 가운데 해고는 1건에 그쳤다.

직장갑질119는 "지난 12일 제보가 확실한 사업장 10곳에 대해 고용노동부에 근로감독을 요구했고 고용부는 최저임금 신고센터에서 불법·편법 여부를 확인하도록 하겠다고 약속했다"고 밝혔다. 이진아 직장갑질119 노무사는 "기업들이 반강제로 취업규칙 변경 동의 절차를 진행하고 있다"며 "고용부가 나서서 동의 과정에 강요성이 있었는지 확인하고, 있다면 무

직장갑질119가 돌아본 3 x

www.hani.co.kr/arti/society/labor/828820.html?_fr=mt2

한겨레 직장갑질119가 돌아본 최저임금 3주... '상여금 꼼수' 가장 많아



div.desc | 643 x 44

지난해 12월7일 저녁 서울 마포구의 한 커피숍에 직장 '갑질' 피해자 20여명이 모였다. 이들은 종이봉투로 만든 가면을 쓰고 각자의 경험을 공유했다. 직장갑질119 제공

최저임금 인상에 따른 가장 흔한 '직장 갑질'은 상여금 축소인 것으로 나타났다.

22일 노동·인권보호단체 직장갑질119는 시간당 최저임금이 7530원으로 오른 뒤 지난 20일까지 자체적으로 접수한 사용자의 최저임금 '꼼수' 제보 77건을 분석해보니, 상여금 축소 사례가 35건(45%)으로 가장 많았다고 밝혔다. 상여금 축소 '꼼수'는 최저임금 산입 범위에 들어가지 않는 상여금을 아예 없애거나, 상여금을 매달 월급에 녹여 기본급화하는 방식으로 이루어졌다. 상여금 폐지나 기본급화는 취업규칙 불이익 변경에 해당해 노동조합이나 과반수 노동자의 동의를 얻어야 가능하다. 그러나 일부 사업장에서는 노동자한테 이런 사실을 일방적으로 통보하거나 동의를 강요하는 등 절차상 문제가 발생하고 있는 것으로 나타났다.

다른 '꼼수' 사례로는 식대 등 수당 폐지 16건(21%), 휴게시간 확대 15건(19%), 최저임금 위반 4건(9%)이 뒤를 이었다. 일부 언론에서 최저임금이 큰 폭으로 인상된 만큼 대량해고가 일어날 것이라는 우려를 제기했지만 제보 사례 가운데 해고는 1건에 그쳤다.

직장갑질119는 "지난 12월 7일 저녁 서울 마포구의 한 커피숍에 직장 '갑질' 피해자 20여명이 모였다. 이들은 종이봉투로 만든 가면을 쓰고 각자의 경험을 공유했다. 직장갑질119 제공"

Elements Console Sources Network Performance 53

```

<!-- 기사툴바메뉴 -->
<div class="article-tools">...</div>
<!-- //기사툴바메뉴 -->
<div class="article-text">
  <div class="article-text-font-size" style="line-height:2em;"
  itemprop="articleBody">
    <div class="subtitle">...</div>
    <div class="text"> == $0
      <div class="image-area">
        <div class="imageL" style="width:643px;">
          <div class="image">...</div>
          <div class="desc" style="width:643px;">...</div>
        </div>
      </div>
      "
      최저임금 인상에 따른 가장 흔한 '직장 갑질'은 상여금 축소인 것으로 나타났다."
      <p align="justify"></p>
    "
    22일 노동·인권보호단체 직장갑질119는 시간당 최저임금이 7530원으로 오른 뒤 지난 20일까지 자체적으로 접수한 사용자의 최저임금 '꼼수' 제보 77건을 분석해보니, 상여금 축소 사례가 35건(45%)으로 가장 많았다고 밝혔다. 상여금 축소 '꼼수'는 최저임금 산입 범위에 들어가지 않는 상여금을 아예 없애거나, 상여금을 매달 월급에 녹여 기본급화하는
  
```

... #a-left-scroll-in div div div.text div.image-area div.imageL div.desc

Styles Event Listeners DOM Breakpoints Properties

Filter :hov .cls +

element.style.f margin -

Console What's New x

Highlights from the Chrome 63 update

Multi-client remote debugging
Use DevTools while debugging your app from an IDE or testing framework.

Workspaces 2.0
Save changes made in DevTools to disk, now with more helpful UI and better auto-mapping.

3. 동적 웹페이지 – Selenium, BS4

마우스 클릭, 로그인 등등 어떤 행동을 취해야만 크롤링이 가능한 웹페이지일 경우에 사용합니다.

(ex – 페이스북 로그인, 더 보기 버튼 클릭, 게시물 검색 등등)

정적웹페이지처럼 html을 볼 줄 알아야하고
필요에 따라 css도 볼 줄 알아야해요

[단독] 센터 영상평가 | 활민현  Hands on Me ▶ 769,881 ♥ 26,451 | 7달 전



facebook

이메일 또는 휴대폰

비밀번호

로그인

계정을 잊으셨나요?

Facebook에서 전세계에 있는 친구, 가족, 지인들과 함께 이야기를 나눠보세요.



가입하기

항상 지금처럼 무료로 즐기실 수 있습니다.

성(姓)

이름(성은 제외)

휴대폰 번호 또는 이메일

새 비밀번호

생일

1993

1월

21

왜 생년월일을 입력해야 하나요?

☐ 여성 ☐ 남성

계정 만들기 버튼을 클릭하면, Facebook의 [약관](#)에 동의하며 [쿠키 사용](#)을 포함한 Facebook [데이터 정책](#)을 읽었음을 인정하게 됩니다. Facebook으로부터 SMS 알림을 받을 수 있으며 알림은 언제든지 옵트 아웃할 수 있습니다.

계정 만들기

4. 나만의 크롤러 만들기 – scrapy 등등

자신이 원하는 방식으로, 원하는 요소만 크롤링을 할 수 있도록 커스터마이징한다고 생각하시면 됩니다. 그러나 이 방법은 코딩을 어느정도 잘하고 html이나 css를 조금 능숙하게 하셔야 해요.

반복되는 크롤링이 필요한 기업들은 이렇게 크롤러를 따로 만들어서 사용해요