(개발일지)cssSelector 사용 간 xpath 로 넘어가게된 계기

일시: 3월 4주차
소프트웨어학부 김동욱

웹 사이트의 정보들을 크롤링 하기 위해 selenium에서 html 태그를 크롤링하는 방법은 여러가지가 있다. classname, id, css tag, xpath 등 각각을 크롤링하는 방식이다. 유튜브 trend 페이지의 소스에는 특이하게 classname, id, css tag는 중복되는 경우가 많다. 이는 내가 생각했을 때 javascript의 map을 이용해 동일한 컴포넌트를 계속해서 생성해내기 때문에 중복이 되는 것 같다.

따라서 정규식을 활용한 xpath는 단 하나씩만 존재하기 때문에 크롤링을 하는것이 빠르다는 판단을 했고, 이를 계기로 각 정보들의 xpath로 크롤링을 하게 됐다.