

The background features a dark gray field with several light gray geometric elements. On the left, there are two concentric circles. On the right, a large circle is partially enclosed by a dotted diamond shape. A vertical dotted line runs through the center, and a horizontal dotted line crosses the middle of the frame. Small dots are placed at the intersections of these lines and at the vertices of the diamond shape.

유저의 니즈를 파악할 때 유의해야 하는 점 & 데이터 수집을 위한 간단한 웹스크래핑

4기 리서치 시니어_이다솔

프롤로그; PROLOGUE

—
해당 교육세션을 진행하게 된 이유와 배경

01

리서치팀에서 주니어분들이 활동하시면서 조금이나마 도움을 드릴 수 있는 내용이 무엇일까?
다른 시니어분들보다 제가 조금 더 잘 전달해드릴 수 있는 정보가 무엇일까?

02

웹스크래핑과 같은 도구적인 방법론 뿐만 아니라 리서치에서 특별히 유의해야 하는 점은 무엇이 있을까?

03

단순 정보 전달의 교육보다는, 직접 해보고, 찾아보고, 고민해볼 수 있는 가이드라인을 제공하자

목차

01 유저의 니즈를 파악할 때 유의해야 하는 점

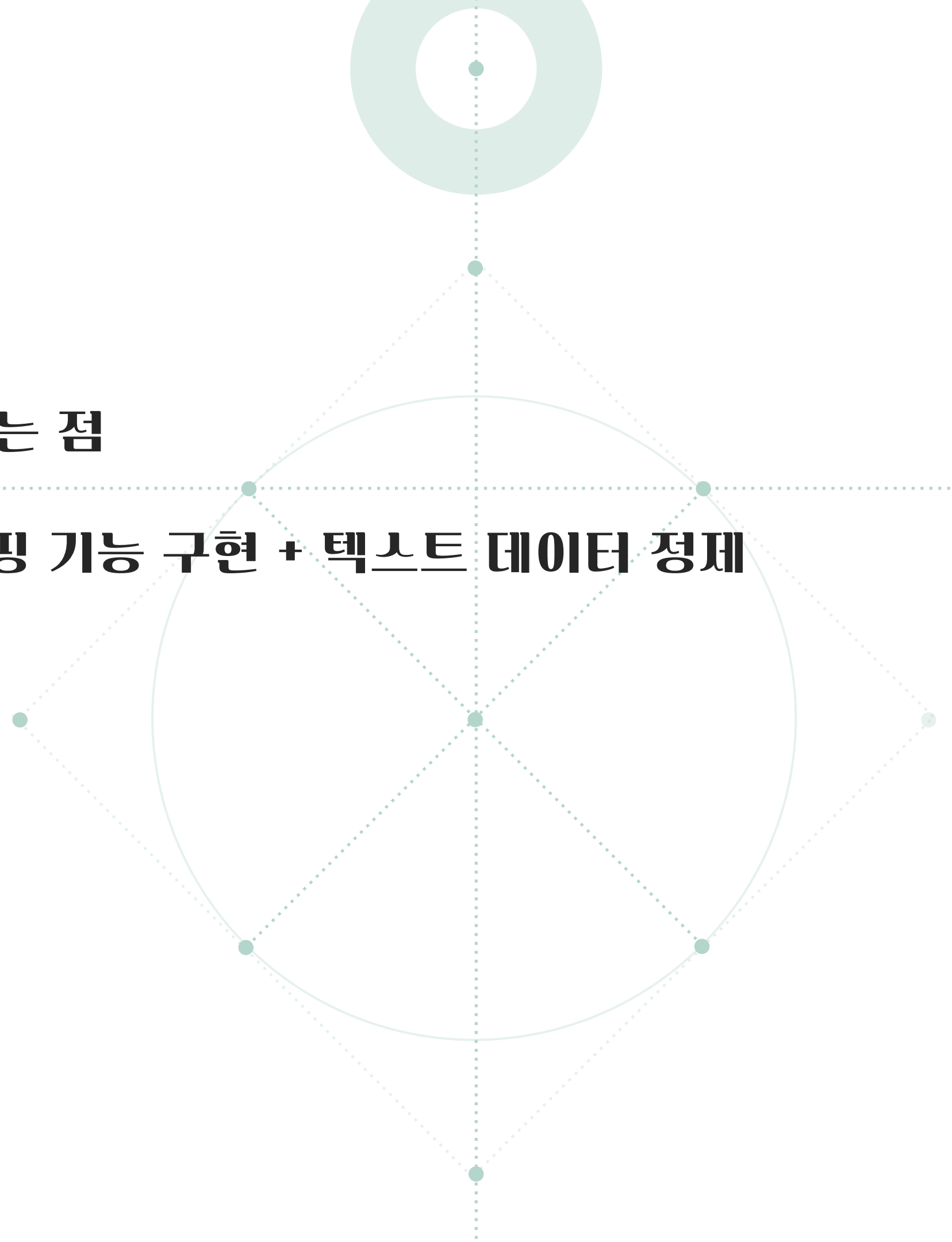
02 데이터 수집을 위한 간단한 웹스크래핑 기능 구현 + 텍스트 데이터 정제

02-1 웹 크롤링 VS 웹 스크래핑

02-2 준비물

02-3 간단한 라이브러리 소개

02-4 실습



01 유저의 니즈를 파악할 때 유의해야 하는 점

가상의 사례

[새로운 프로젝트를 기획할 때 / 창업을 시작할 때]



배경 조사

1인 가구 / 자취생 증가
20대 건강에 대한 관심 증가



문제 파악

금방 상하는 신선식품의
섭취와 보관이 어려움



해결책

앱을 활용한 신선식품
이웃 공동구매 시스템



01 유저의 니즈를 파악할 때 유의해야 하는 점

가상의 사례

[새로운 프로젝트를 기획할 때 / 창업을 시작할 때]



배경 조사

1인 가구 / 자취생 증가
20대 건강에 대한 관심 증가



문제 파악

금방 상하는 신선식품의
섭취와 보관이 어려움



해결책

앱을 활용한 신선식품
이웃 공동구매 시스템



앱 / 웹사이트 기획 및 개발 시작



막상 앱 혹은 웹 사이트를 만들고 나면, 아무도 해당 서비스를 찾지 않는다는 현실과 직면함

기술의 문제가 아닌, 유저의 니즈를 제대로 확인하지 못함

01 유저의 니즈를 파악할 때 유의해야 하는 점

도서 - 유난한 도전

세상 누구도 원하지 않는

울라블라 - 오프라인 만남을 기록하고, 친구들과 공유하는 소셜네트워크서비스

그럼에도 '알려지기만 하면 대박 날 거야'라는 희망은 버리기 어려웠다. 각종 스타트업 경진대회에 나가고 사람들을 만나 앱을 알렸다. 돈을 써서 온라인 광고도 했다. 만남을 울라블라 앱에만 기록하지 않고, 막 흥행하기 시작한 인스타그램과 공유할 수 있도록 하면 사람들이 좋아할 거라 생각해 연동 기능을 개발해 넣었다. 사진 보정 필터도 만들어 넣었다. 아무도 쓰지 않는다는 것만 빼면 완벽한 앱이었다.

초조해진 이승건과 이태양은 울라블라가 세상에 퍼뜨리고자 했던 가치와는 동떨어진 일들을 벌였다. 실패를 똑바로 마주볼 용기가 없었기 때문이다. '사람들이 오프라인 만남을 앱에 기록하고 싶을 것'이라는 첫 번째 가설이 틀렸다는 사실을 인정하지 못했다. 두 사람은 중간에 포기하고, 생각을 바꾸고, 방향 전환하는 걸 최악시켰다. 울라블라가 잘 될 수 밖에 없는 이유를 100가지는 읊을 수 있었다. 이태양은 "우리가 풀고 싶은 문제에 몰두한 나머지, 사람들의 공감을 전혀 얻지 못하고 있다는 것을 너무 늦게 알았다."고 말했다.

"세상이 받아들이는 문제의 크기보다, 우리가 느끼는 문제의 크기가 너무 컸던 거예요. 그래서 사람들에게 이 서비스가 '옳다'고 주장하게 되는거죠. 제품은 계속 발전하고 있는데 쓰는 사람은 전혀 늘지 않았어요."

울라블라의 실패를 인정하고 서비스를 접기까지 1년 4개월이 걸렸다. 자본금 5000만원짜리 비바퍼블리카는 인건비를 포함해 이 앱에 2억 2000만원을 썼다.

01 유저의 니즈를 파악할 때 유의해야 하는 점

— 도서 - 유난한 도전

앱 개발은 시작도 하지 않았다. 하도 실패를 많이 하다보니 앱 제작에 시간과 비용을 들이고 나서 망하는 게 아까웠다. 사람들의 반응을 살피고 난 다음에 서비스를 만들어도 늦지 않았다는 게 3년의 실패에서 얻은 교훈이었다. 이승건이 배경사진을 고르고 문구를 써서 홈페이지를 완성했다.

...토스팀은 홈페이지를 열고, 트위터에 링크를 올렸다. 4시간 만에 1000번 넘게 리트윗됐다. 이후 사흘간 서비스를 써보겠다며 전화번호를 입력한 사람은 2000명에 가까웠다. 3년 동안 한 번도 보지 못한 숫자였다. 짜릿했다. '이게 성장이구나. 소비자들이 원하는 게 맞았구나. 드디어 찾았구나.'

01 유저의 니즈를 파악할 때 유의해야 하는 점

내가 혹은 우리 팀이 세운 가설이 틀릴 수 있다는 것을 받아드려야 한다.
또한, 고객의 니즈를 파악하고 가설을 검증하기 위해 완벽한 서비스를 먼저 만들 필요는 없다.

앞선 경우와 같이 고객의 니즈와 반응을 살피기 위해 아주 간단한 웹사이트를 활용

설정된 가설이 유의미한 지 확인하기 위해 유저에게 인터뷰/설문을 할 때

"마트에서 파는 신선식품은 양이 많아서 보관과 관리가 어려우니까 자취생들을 위한 신선식품 이웃 공동구매 서비스가 있으면 어떨 것 같아? 너 사용할 것 같아?"와 같은 닫힌 질문이 아니라
"요즘 과일이나 야채같은 신선식품을 챙겨먹어?" "자취하면서 음식을 챙겨먹을 때 불편한 점이 있어?"와 같은 열린 질문에서 가설과 부합하는 답이 나올 때 고객의 니즈 분석 정확도가 높다고 할 수 있다.

02 데이터 수집을 위한 간단한 웹크롤링 기능 구현

웹 스크래핑을 통한 데이터 수집의 자동화

교육세션의 목표는 스크래핑을 위한 코딩을 잘하는 것이 아니라,
해당 코드를 전체적으로 이해하고, 필요한 부분을 수정 및 추가하면서 사용해보는 것

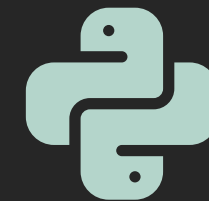
[과정]



웹스크래핑을 통해
데이터 수집



엑셀 형태로
데이터 저장



파이썬으로
데이터 정제



간단한 시각화

02 데이터 수집을 위한 간단한 웹스크래핑 기능 구현

웹 크롤링 VS 웹 스크래핑

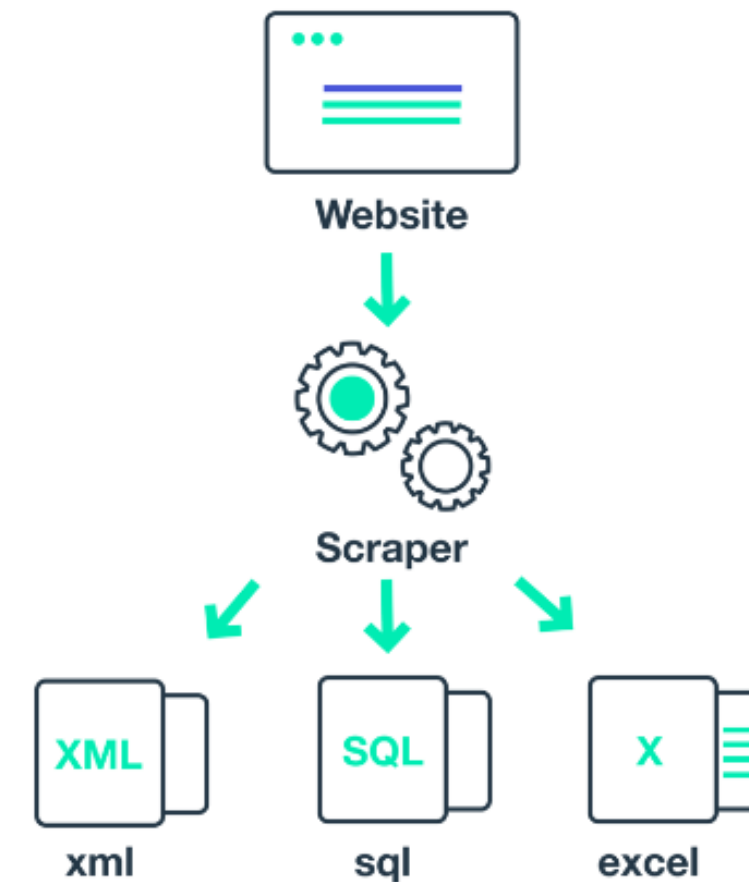
Web Crawler



웹 크롤러

웹 페이지의 내부 링크 및 페이지를 따라
인터넷을 체계적으로 탐색하여 콘텐츠를 검색

Web Scraping



웹 스크래퍼

웹을 포함한 다양한 소스에서
특정 데이터를 추출하는 프로세스

02 데이터 수집을 위한 간단한 웹스크래핑 기능 구현

준비물



CHROME (크롬)
크롬 다운로드 링크

CHROME 드라이버
다운로드 링크

크롬드라이버 설치 블로그 링크

PYTHON(파이썬)

파이썬 다운로드 링크
(처음 설치할 경우 가장 최근 버전 설치)

파이썬 설치 관련 블로그 링크

VSCode

VSCode 다운로드 링크

VSCode 설치 관련 블로그 링크

Python / Visual Studio Code
개발 환경 구축 블로그 링크



EXCEL(엑셀)

아나콘다 설치 +
VSCODE에서 주피터 노트북 사용
블로그 링크

02 데이터 수집을 위한 간단한 웹스크래핑 기능 구현

라이브러리

BeautifulSoup

데이터 추출하는 데 필요한 기능이 들어있는 라이브러리; Parsing Library
→ html 소스코드에서 필요한 내용만을 뽑아내는 것

Selenium

프로그램을 이용해 자동화된 웹 테스트를 수행할 수 있도록 해주는 라이브러리

Selenium 라이브러리를 사용하는 이유

1. 자바스크립트가 동적으로 만든 데이터를 스크래핑 하기 위해
2. 사이트의 다양한 HTML 요소에 클릭, 키보드 입력, 스크롤, 자동 로그인 등 이벤트를 주기 위해

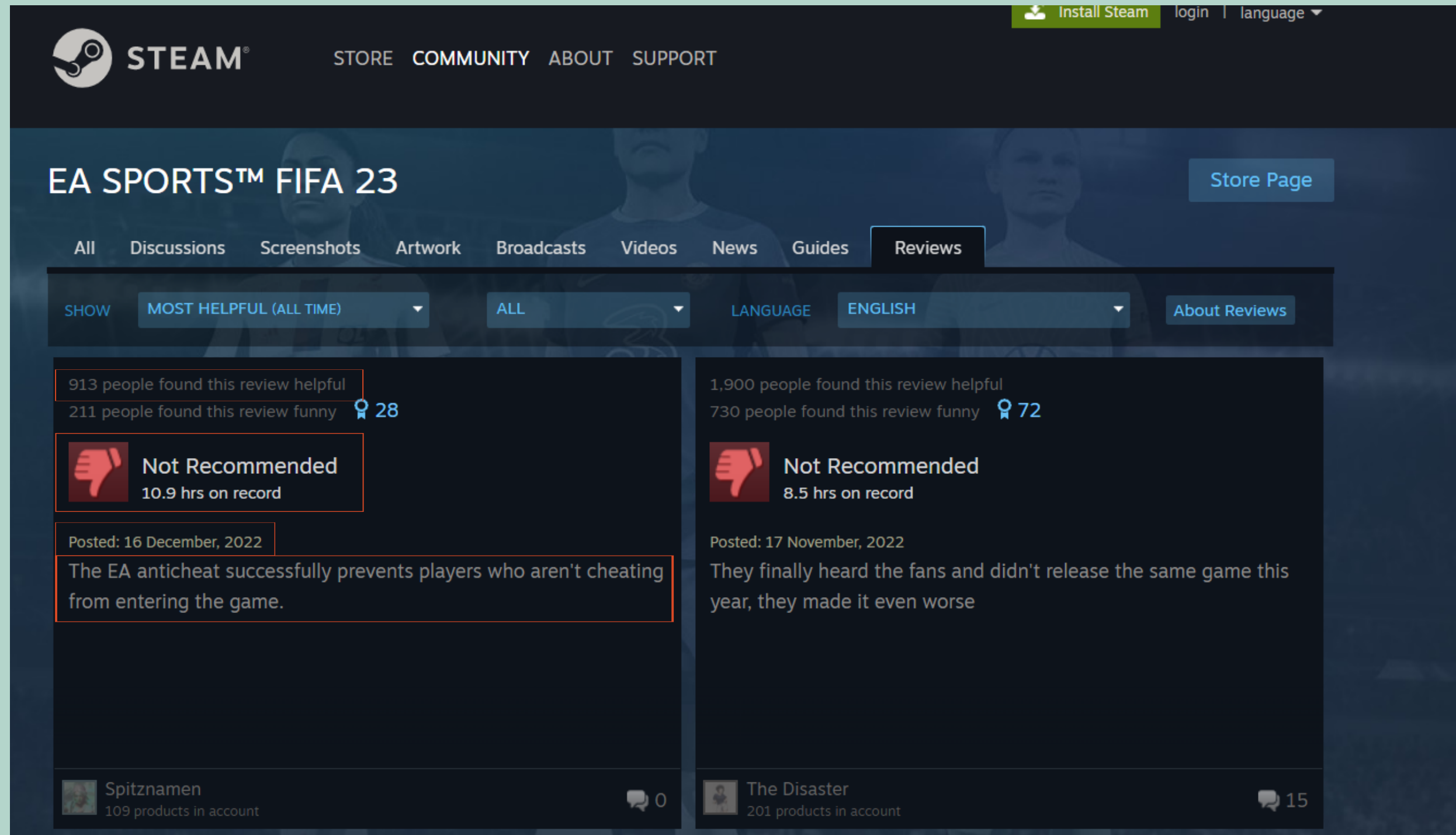
02 데이터 수집을 위한 간단한 웹크롤링 기능 구현

웹스크래핑을 통해
데이터 수집

엑셀 형태로
데이터 저장

파이썬으로
데이터 정제

간단한 시각화



02 데이터 수집을 위한 간단한 웹크롤링 기능 구현

웹스크래핑을 통해
데이터 수집

엑셀 형태로
데이터 저장

파이썬으로
데이터 정제

간단한 시각화

The screenshot displays the Steam website's review page for EA Sports™ FIFA 23. The page shows a list of reviews, including one from 'Spitznamen' (Not Recommended, 10.9 hrs on record) and another from 'The Disaster' (Not Recommended, 8.5 hrs on record). The Chrome DevTools developer console is open on the right, showing the 'Elements' tab with the DOM tree. A red circle highlights the 'Always match Chrome's language' button in the DevTools toolbar. The console also shows the 'Styles' pane for the selected element, displaying CSS rules for the review card content.

STEAM® STORE COMMUNITY ABOUT SUPPORT

EA SPORTS™ FIFA 23 Store Page

All Discussions Screenshots Artwork Broadcasts Videos News Guides Reviews

SHOW MOST HELPFUL (ALL TIME) ALL LANGUAGE ENGLISH About Reviews

913 people found this review helpful
div.title 300 x 28.2 review funny 28

Not Recommended
10.9 hrs on record

Posted: 16 December, 2022
The EA anticheat successfully prevents players who aren't cheating from entering the game.

Spitznamen
109 products in account

1,116 people found this review helpful
109 people found this review funny 172

1,900 people found this review helpful
730 people found this review funny 72

Not Recommended
8.5 hrs on record

Posted: 17 November, 2022
They finally heard the fans and didn't release the same game this year, they made it even worse

The Disaster
201 products in account

1,132 people found this review helpful
494 people found this review funny 78

1,502 people found this review helpful
475 people found this review funny 174

DevTools is now available in Korean!

Always match Chrome's language Switch DevTools to Korean Don't show again

Elements Console Sources Network

class="apphub_Card modalContentLink interactable" style="float: left; width: 460.825px; height: 345px; opacity: 1;" data-modal-content-url="https://steamcommunity.com/profiles/76561198954472404/recommended/1811260/" data-modal-content-size-to-fit="false">

<div class="apphub_CardContentMain" style="height: 287px;">

<div class="apphub_UserReviewCardContent">

<div class="found_helpful"></div>

<div class="vote_header">

<div class="reviewInfo">

<div class="thumb"></div> == \$0

<div class="title">Not Recommended</div>

<div class="hours">10.9 hrs on record</div>

</div>

... centMain div.apphub_UserReviewCardContent div.vote_header div.reviewInfo div.title ...

Styles Computed Layout Event Listeners DOM Breakpoints Properties Accessibility

Filter :hov .cls +

element.style {

}

.apphub_UserReviewCardContent .vote_header .title { apphub.css?...english:1418

font-size: 16px;

color: #d6d7d8;

font-family: 'Motiva Sans', arial, tahoma;

padding: 5px 0 0 0;

}

div { user agent stylesheet

display: block;

}

Inherited from div.apphub_CardContentMain

.apphub_CardContentMain { apphub.css?...english:853

position: relative;

overflow: hidden;

02 데이터 수집을 위한 간단한 웹크롤링 기능 구현

웹스크래핑을 통해
데이터 수집

엑셀 형태로
데이터 저장

파이썬으로
데이터 정제

간단한 시각화

HTML; HyperText Markup Language

웹 사이트를 만들기 위해 필요한 언어 → 뼈대

```
<div class="title">Not Recommended</div>
```

Tag(태그)

속성

속성값

에필로그; EPILOGUE

—
교육세션을 마치며

01

웹 스크래핑은 웹의 구성(HTML 구조)가 바뀌면 코드도 그에 맞게 변경해야 함

02

구글링은 영원한 친구..^^

03

웹 스크래핑을 통해 데이터 수집을 하고자하는 웹페이지가 있다면,
혹은 오류가 난다면 언제든지 연락주세요!