

Programming project
report

Icarus

My Fancy Web Crawler

이원석, 박우빈, 유상욱, 한치우



01

What is web crawl 1



02

Code 2



03

Result 3



04

impression 4

What is web crawler



웹 크롤러란
웹 크롤러는 조직적, 자동화된 방법으로 월드
와이드 웹을 탐색하는 컴퓨터 프로그램

Please enter a brief summary of or introduction to the
presentation.
Please enter the intro here.

1-123-456-7890
olivia@miridih.com

Code

```
scrapper.py × middlewares.py × pipelines.py × dj settings.py × crawling_spider.py × items.py × __init__.py ×
1 from scrapy.spiders import CrawlSpider, Rule
2 from scrapy.linkextractors import LinkExtractor
3
4
5 class CrawlSpider(CrawlSpider):
6     name = "myfancycrawler"
7     allowed_domains = ["toscrape.com"]
8     start_urls = ["http://books.toscrape.com/"]
9
10    rules = (
11        Rule(LinkExtractor(allow="catalogue/category")),
12        Rule(LinkExtractor(allow="catalogue", deny="category"), callback="parse_item")
13    )
14    PROXY_SERVER = "127.0.0.1"
15    def parse_item(self, response):
16        yield {
17            "title": response.css(".product_main h1::text").get(),
18            "price": response.css(".price_color::text").get(),
19            "availability": response.css(".availability::text")[1].get().replace("\n", "").replace(" ", "")
20        }
21
22
```

Output

```
scrrpper.py × middlewares.py × pipelines.py × dj settings.py × crawling_spider.py × {..} output.json × items.py × __init__.py ×
1
2 {"title": null, "price": "£12.84", "availability": "Instock"},
3 {"title": "It's Only the Himalayas", "price": "£45.17", "availability": "Instock(19available)"},
4 {"title": "Libertarianism for Beginners", "price": "£51.33", "availability": "Instock(19available)"},
5 {"title": "Mesaerion: The Best Science Fiction Stories 1800-1849", "price": "£37.59", "availability": "Instock(19available)"},
6 {"title": "Olio", "price": "£23.88", "availability": "Instock(19available)"},
7 {"title": "Our Band Could Be Your Life: Scenes from the American Indie Underground, 1981-1991", "price": "£57.25", "availability": "Instock(19available)"},
8 {"title": "Rip it Up and Start Again", "price": "£35.02", "availability": "Instock(19available)"},
9 {"title": "Scott Pilgrim's Precious Little Life (Scott Pilgrim #1)", "price": "£52.29", "availability": "Instock(19available)"},
10 {"title": "Changing the Game (Play by Play #2)", "price": "£13.38", "availability": "Instock(4available)"},
11 {"title": "Dark Lover (Black Dagger Brotherhood #1)", "price": "£12.87", "availability": "Instock(4available)"},
12 {"title": "Bridget Jones's Diary (Bridget Jones #1)", "price": "£29.82", "availability": "Instock(1available)"},
13 {"title": "The Perfect Play (Play by Play #1)", "price": "£59.99", "availability": "Instock(4available)"},
14 {"title": "The Edge of Reason (Bridget Jones #2)", "price": "£19.18", "availability": "Instock(1available)"},
15 {"title": "A Girl's Guide to Moving On (New Beginnings #2)", "price": "£31.30", "availability": "Instock(5available)"},
16 {"title": "I've Got Your Number", "price": "£19.69", "availability": "Instock(3available)"},
17 {"title": "Sit, Stay, Love", "price": "£20.90", "availability": "Instock(7available)"},
18 {"title": "The Bridge to Consciousness: I'm Writing the Bridge Between Science and Our Old and New Beliefs.", "price": "£32.00", "availability": "Instock(15available)"},
19 {"title": "The Emerald Mystery", "price": "£23.15", "availability": "Instock(15available)"},
20 {"title": "The Girl on the Train", "price": "£55.02", "availability": "Instock(15available)"},
21 {"title": "The Kite Runner", "price": "£41.82", "availability": "Instock(15available)"},
22 {"title": "Having the Barbarian's Baby (Ice Planet Barbarians #7.5)", "price": "£34.96", "availability": "Instock(1available)"},
23 {"title": "The Psychopath Test: A Journey Through the Madness Industry", "price": "£36.00", "availability": "Instock(15available)"},
24 {"title": "The Last Girl (The Dominion Trilogy #1)", "price": "£36.26", "availability": "Instock(1available)"},
25 {"title": "The Secret (The Secret #1)", "price": "£27.37", "availability": "Instock(15available)"},
26 {"title": "Three Wishes (River of Time: California #1)", "price": "£44.18", "availability": "Instock(1available)"},
27 {"title": "Thirteen Reasons Why", "price": "£52.72", "availability": "Instock(15available)"},
28 {"title": "Do Androids Dream of Electric Sheep? (Blade Runner #1)", "price": "£51.48", "availability": "Instock(3available)"},
29 {"title": "Tracing Numbers on a Train", "price": "£41.60", "availability": "Instock(15available)"},
30 {"title": "Dune (Dune #1)", "price": "£54.86", "availability": "Instock(3available)"},
31 {"title": "A Piece of Sky, a Grain of Rice: A Memoir in Four Meditations", "price": "£56.76", "availability": "Instock(16available)"},
```

Impression

박우빈

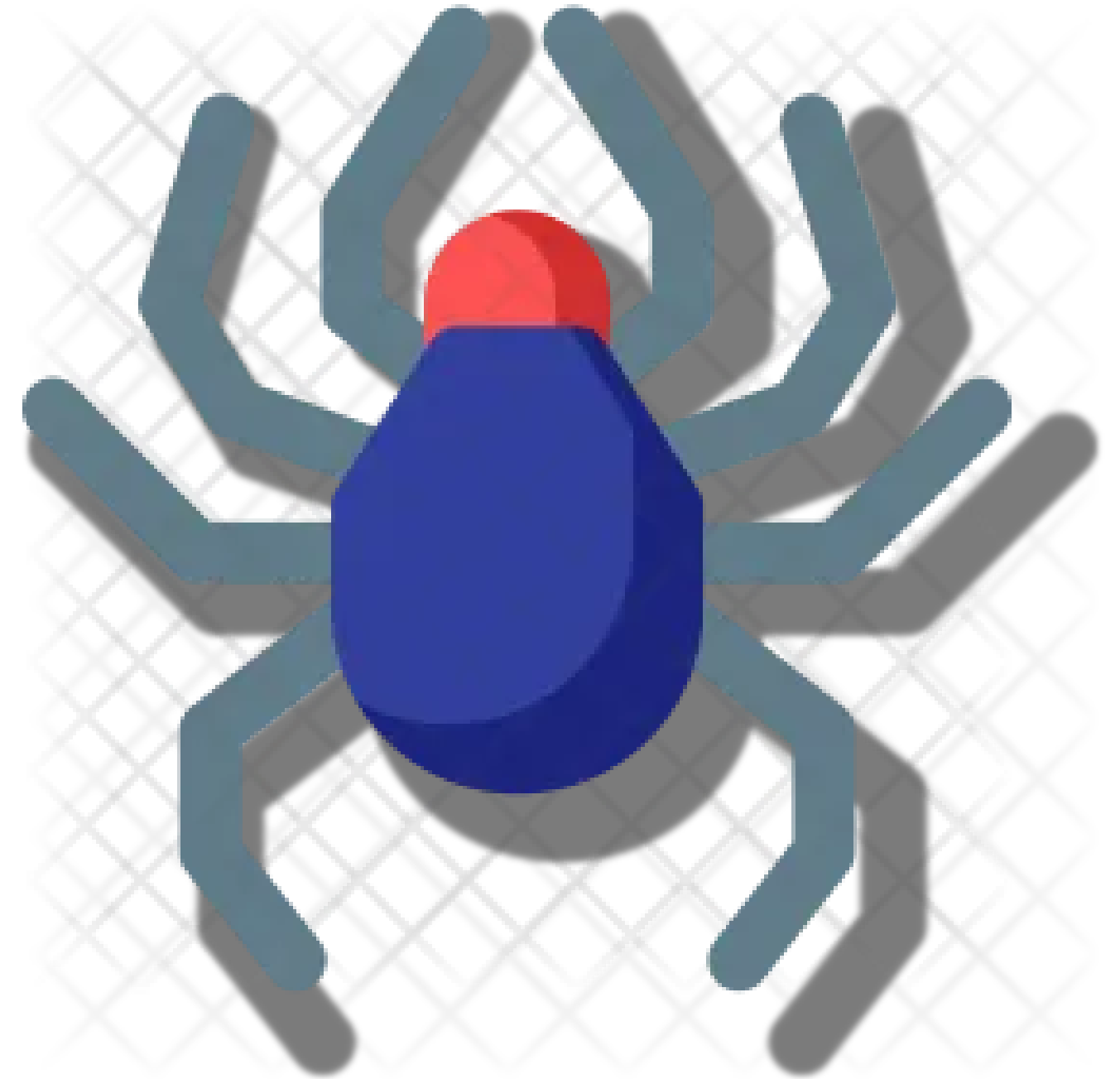
웹 크롤러를 개발하며 데이터를 수집하고 가공하는 방법을 배우고 익혔습니다. 크롤링한 데이터를 분석하고 활용하는 능력이 향상되어, 다양한 분야에서 유용한 정보를 추출할 수 있게 되었습니다.

이원석

웹 크롤러 개발 프로젝트를 통해 실제로 소프트웨어를 구축하고 문제를 해결하는 과정을 경험하게 되었습니다. 이를 통해 개발 역량과 실무에서의 적용 능력이 향상되었습니다.

유상욱

팀으로 하는 프로젝트라는 것이 다같이 하는 것보다보니 합이 잘 맞지 않으면 기한에 맞추는 것이 어렵다는 것을 느꼈고 개개인의 능력이 전체적으로 좋아야지 원활하게 진행된다고 느꼈다.



Luxury must be comfortable, YOUR NAME MIRI UNIV
otherwise it is not luxury.



THANK YOU!

Please enter the subtitle here

Enter any additional details here