

Sharing Alumni Experiences



Sigit Dewanto
(Web Crawling Engineer, ScrapingHub)

About Me



- S1 Ilmu Komputer UGM
- S2 Ilmu Komputer UGM



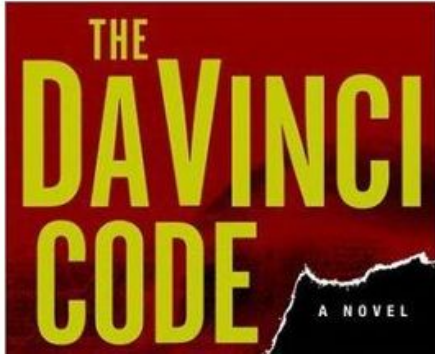
- Web Crawling Engineer (2014 - now)



- Community Organizer (2019 - now)

The Jobs

Web scraping: how to extract data from web page



The Da Vinci Code (Robert Langdon #2)

£22.96

✓ In stock (3 available)



```
{  
  "title": "The Da Vinci Code",  
  "author": "Robert Langdon",  
  "price": "22.96",  
  "stock": 3,  
  "rating": 2.0,  
  ...  
}
```

Web scraping approaches

1. Programming
2. Supervised learning
3. Automatic extraction (heuristics)

AutoExtract: Deep learning powered

- Web R-CNN
- Implemented using PyTorch
- <https://www.scrapinghub.com/automatic-data-extraction-api/>

AI-Enabled Automatic Data Extraction APIs

Try it now and say goodbye to writing scripts or managing code.



News & Article
Data



Product
Data



Job Postings
Data



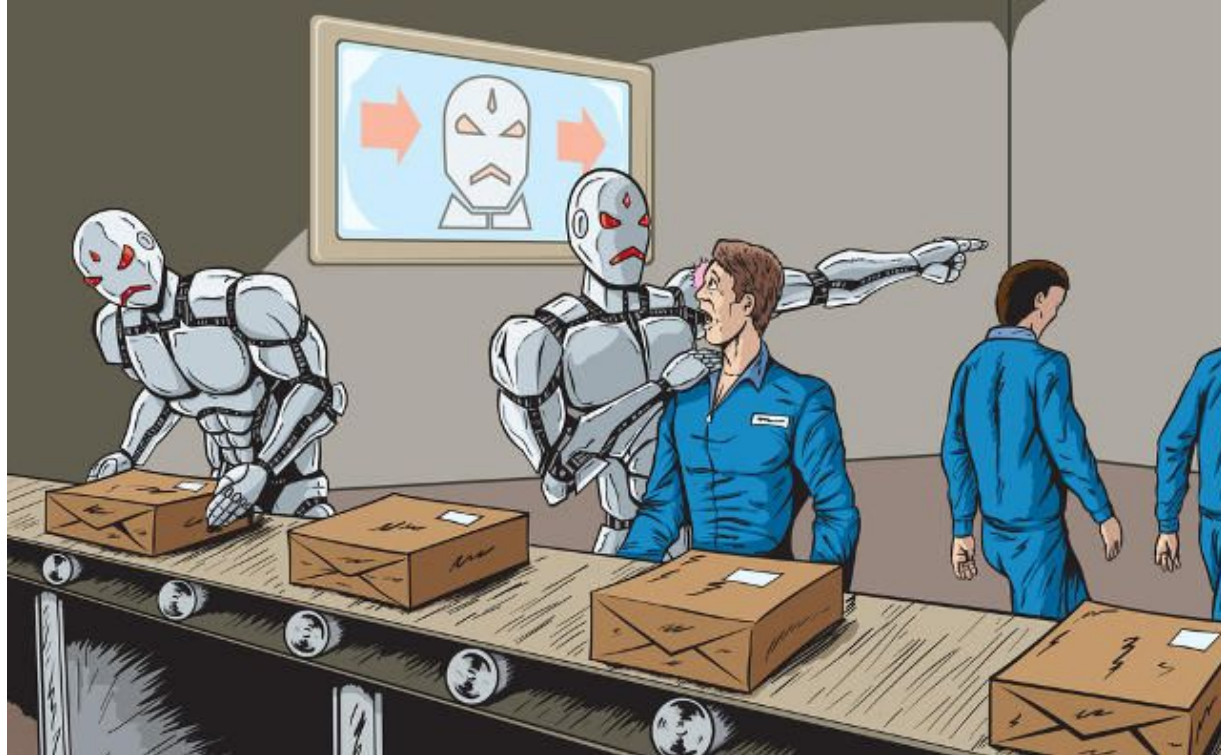
Vehicle
Data



Product
Reviews Data

They took our jobs!!!

<https://www.forbes.com/sites/cognitiveworld/2018/08/07/job-loss-from-ai-theres-more-to-fear/#42c3805823eb>



The Culture

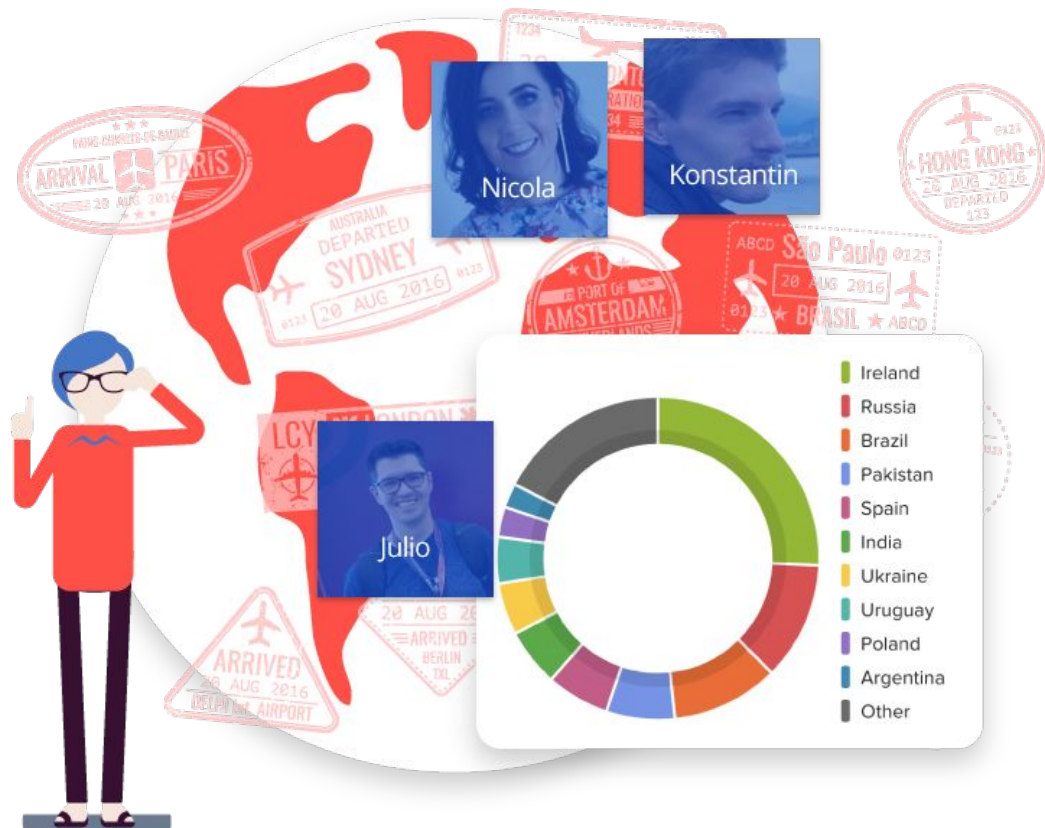
Open source is in our DNA

<https://www.scrapinghub.com/open-source/>



Scrapy

Distributed team



Remote working in ScrapingHub



8 hours/day, flexible working hour



Trust & Responsibility

The perks: annual company gathering



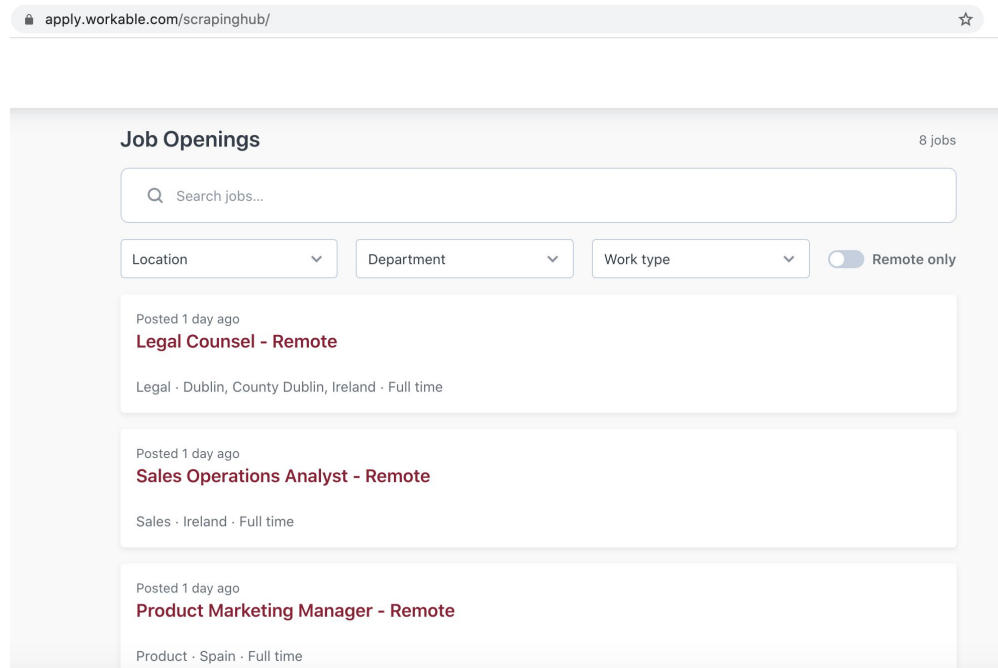
Blameless culture



The Application

How to apply?

<https://seagatesoft.wordpress.com/2019/03/25/kisah-mendapatkan-pekerjaan-remote/>



The screenshot shows a web browser window with the URL `apply.workable.com/scrapinghub/`. The page displays a "Job Openings" section with 8 jobs. It includes a search bar and filters for Location, Department, and Work type, along with a "Remote only" toggle. Three job listings are visible:

- Legal Counsel - Remote**
Posted 1 day ago
Legal · Dublin, County Dublin, Ireland · Full time
- Sales Operations Analyst - Remote**
Posted 1 day ago
Sales · Ireland · Full time
- Product Marketing Manager - Remote**
Posted 1 day ago
Product · Spain · Full time

Open source your Final Assignment software

github.com/seagatesoft/sde

Search or jump to... Pull requests Issues Marketplace Explore

seagatesoft / sde

Unwatch 7 Star 49 Fork 24

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

master 1 branch 0 tags

Go to file Add file Code

About

seagatesoft Added API usage documentation. 24a6708 on Jun 9, 2012 10 commits

SDE	Added API usage documentation.	8 years ago
.gitignore	Initial commit	8 years ago
README.md	Synchronized this file content with README-EN.HTML.	8 years ago

Structured Data Extractor. An application to extract structured data from web pages. It uses Data Extraction Based on Partial Tree Alignment (DEPTA) method. (UPDATE: I implemented a newer algorithm: <https://github.com/seagatesoft/webdext>)

README.md

Important technical skills

- UNIX command line
- Git (version control system)
- Any programming language, popular ones are:
 - Python
 - JavaScript
 - Go
 - Ruby
 - Java
- Company/position-specific skill

Important non-technical skills

- English
- Communication
- Problem solving
- Learn how to learn
- Leadership

Important subjects

- Fundamentals
 - Algorithm and Programming
 - Statistics
- Artificial Intelligence
- Distributed System

Established Remote Companies

Curated list: <https://github.com/yanirs/established-remote>



AUTOMATTIC



The Salary



How much is the salary?



MisterEko Today at 7:31 AM

1k/month = noob

2k/month = beginner

3k/month = intermediate

>4k/month = now you're talking



Salary Calculator

- **Buffer**
 - <https://buffer.com/resources/salary-formula-changes-2019/>
- **Gitlab**
 - <https://about.gitlab.com/handbook/total-rewards/compensation/compensation-calculator/calculator/>
- **For other companies, you can use Glassdoor.com**

The Challenges

Working From Home



What my friends
think I do.



What my mom
thinks I do.



What my neighbors
think I do



What my boss
thinks I do.



What I think I do.



What I actually do.



Tax, Insurance, & Pension



Consult to local tax office



BPJS Kesehatan

Pekerja Bukan Penerima Upah



BPJS Ketenagakerjaan



BAZNAS

Badan Amil Zakat Nasional

Tax deduction

The End



<https://t.me/seagatesoft>



<https://t.me/pyjogja>



<https://instagram.com/pyjogja>



<https://www.youtube.com/c/PythonIDJogja>