EmailScrapper: Scrap and send emails.

EmailScrapper will scrap the Torre company account from time to time on the Un mejor empleo, jooble, jobomas, El empleo and Computrabajo job search platform and will obtain the email address of the applicants to the offers published by those accounts. Once the applicants' emails have been obtained, the app will send a direct message, by email, with the following steps to the Torre job application.

Team

- **Arturo Victoria Rincon**: He is the team leader and will lead the Backend and Database development. He has large experience dealing with web services
- **Jose Luis Saldarriaga Bravo**: He will lead the Frontend development. He is well versed in javascript technologies.
- **Sebastián López Herrera**: He will lead the DevOps development. He is highly skilled providing continuous delivery with high software quality.

These roles were created to divide the overall project load work. However, everyone will work supporting each other in solving the problem with a specific technology. The Backend area was created to develop the solution of scrap and send email. The Frontend area was created to develop the final user interface. The DevOps area was created to provide the CD(continues delivering). The DataBase area was created to handle the data acquired when the app scraps the information.

Technologies

Main technologies:

Front-end technologies:

- React.js

Back-end technologies:

- Node.js

- Puppeteer API

DevOps technologies:

- Docker
- Lambda AWS
- Amazon RDS
- Amazon SES
- S3 BUCKET
- Amazon API Gateway
- GIT

Databases:

- Mysql

Test:

- Travis CI

Alternative technologies:

Back-end technologies:

- Selenium: It is another functional automation, it's low cost of acquisition and high level of integration with .Net, Python, Node.js, and of course, Java.

Databases:

- MongoDB: We can store objects and it is fast to query.

Challenge statement

Problem to solve:

Currently, the Torre company has to hire a freelancer to search the people who apply to a Torre job offer posted in a third party job platform and send them an email from Torre to create a direct channel with the candidates. However, this process takes a lot of time because the freelancer has to enter those third party jobs platforms and search the applicants. In some cases a job offer vacancy needs to be occupied quickly but the freelancer is notified after 2 days, by email, that a person applied to the job. So, is there a way to make this process faster and automatic?

The Portfolio Project will not solve:

EmailScraper will not create a database with the candidates' information to take them into count later.

The Portfolio Project will help and/or who the users will be:

This service will help to improve the Torre recruiter time to contact a candidate after this one has applied to the job offer.

Is this project relevant or dependent on a specific locale?

This project is not dependent on a specific location or current time.

Risks

Technical risks:

We depend on the html content displayed on third-party job platforms like "Un mejor empleo", to populate our data. If the html content is changed we could not scrape the candidate's email. This kind of project does not have a safeguard so it has to be upgraded if the job platform is changed.

Non-technical risks:

Anyone who uses this scraper can be accused of bad handling personal information because the scraper does not ask formally for permission to handle that data.

Infrastructure

Process for branching and merging:

We'll use Git Feature Branch Workflow. As main branches, we will use master and develop during the whole life cycle of the project. The master branch will be the one used in production. The development branch will be considered as the one that always has the last development deliveries for the next release. We will use feature branches to develop our new feature in the development branch.

When the development branch reaches a stable point and is ready to be released, all changes should be merged back into the master.

Strategy for deployment

We are going to use lambda AWS to deploy, every few hours, the scraper.

How will you populate your app with data?

Scraping the third party job platforms.

What tools, automation, or process you will use for testing?

We are going to use Travis CI.

Existing Solutions

Zoominfo: A full-featured email scraping platform with a comprehensive database. You can directly search for titles and companies within their platform. It is more like a directory system that covers professionals in all industries with contact information. Email lists are the assets.

Skrapp.io: An integrated platform that provides from Linkedin integration, a website search, domain search to leads directory. It can discover the business email address of one's Linkedin profile.

Octoparse: A powerful web scraping tool that is worth having for any respect for research. You can extract any information from the websites and save them into a structured format effortlessly.