Technical
assignment

Hi,

As part of your assessment, we have prepared a small assignment for you. If you have any questions, feel free to contact us via dominik.susmel@cialdnb.com
Task is to be completed within seven days from when you start working on it.

Once you are done with the assignment, push your code to github and email us the link to
dominik.susmel@cialdnb.com
fabiola.rozendo@cialdnb.com
silvia.tomaz@cialdnb.com

Git(Hub) repo should contain a README.md with clear instructions how to run the code with a (brief) technical overview.

Good luck!

Dominik Susmel
dominik.susmel@cialdnb.com
Engineering coordinator | CIAL dun&bradstreet

Goal

Create a command-line application that, given a list of website URLs as input, visits them and finds, extracts and outputs the websites' logo image URLs and all phone numbers (e.g. mobile phones, land lines, fax numbers) present on the websites.

Requirements

- Write the application in Python (3.5 or above).
- Structure the Python project according to best practices.
- Commit all code into a git repository.
- Feel free to combine and use any other technology, library and external resource.
- The application visits a given input website and outputs the website's logo, and any found phone numbers. The application should receive a list of website URLs via standard input, one per line. It should write its results to standard output, one per line, unordered, in JSON format. Any logs should be output to standard error. Find examples of the input and output formats in the Examples section below.
- The application should process the inputs concurrently. Running it given 50 websites should output most results significantly faster than running it 50 times given 1 website.
- Create a Dockerfile that builds an image containing the ready-to-run command-line application. One should be able to invoke "docker run" and give it input via stdin instead of having to manually install and run a Python package.
- Light cleaning of the found phone numbers should be performed: replace any characters that are not digits, a plus sign (+) or parentheses with whitespace, e.g. clean "+(385)/98 718-2222" into "+(385) 98 718 222".
- If a logo image URL is found, it should be output as an absolute URL, e.g. "http://www.website.com/path/to/logo.png", not "/path/to/logo.png".
- Don't worry about handling Flash websites, websites generated dynamically through Javascript, phone numbers represented as images on the website or any other similar scenario that significantly ramps up the project's complexity.

Process

Once you receive the task, you can ask us anything about it. Structure and organize your questions wisely because you can send us a list of questions only twice during this process. You have 7 days to finish the task, starting from the moment you received it. Once you are done with the project, zip up your local clone of the git repository and send it to us.

Review

Your project will be reviewed from a logic and structure point of view. Logic includes solution implementation and running your code through 1000 international websites as a

phone number and logo accuracy test. Structure includes code quality and overall project/code organization.

Examples

An example of an input stream, in this case a file called websites.txt.

```
https://www.cmsenergy.com/contact-us/default.aspx
https://www.illion.com.au
https://www.phosagro.com/contacts
https://www.powerlinx.com/contact
https://www.cialdnb.com/en
https://www.illion.com.au/contact-us

$ cat websites.txt | python3 -m yourmodule
$ cat websites.txt | docker run -i yourimage
```

Result

{"logo":"https://www.illion.com.au/wp-content/uploads/2019/03/ION-RGB-Gradient-64.png","phones":[],"website":"https://www.illion.com.au"}
{"logo":"https://www.illion.com.au/wp-content/uploads/2019/03/ION-RGB-Gradient-64.png","phones":["+61 3 9828 3200","+61 8 7122 9452","1800 233 533","13 23 33","+64800836268"],"website":"https://www.illion.com.au/contact-us"}
{"logo":"https://www.cialdnb.com/imagens/logo_color.jpg","phones":["+1 (954) 507 1116","+54 11 5984 2700","+55 11 4933 7500","+52 55 4166 3000"],"website":"https://www.cialdnb.com/en"}
{"logo":"https://www.powerlinx.com/wp-content/uploads/2016/12/plx-logo-nav-dark.png","phones":["+1 (212) 465 9555"],"website":"https://www.powerlinx.com/contact"}
DeadPage
{"logo":"https://s2.q4cdn.com/027997281/files/design/ClientLogo.png","phones":["575 01 07 20","(517) 788 0550","(517) 768 3376","(800) 390 1220","(517) 788 6538","(517) 788 2395","(517) 788 2394","1 855 598 2714","651 450 4064","517 768 3114"],"website":"https://www.cmsenergy.com/contact-us/default.aspx"}
{"logo":"https://www.phosagro.com/local/templates/.default/img/fosagro-logo.svg","phones":["8 (495) 956 09 64","+ 7 (495) 232 96 89","+7 (495) 231 31 15","+ 7 (929) 600 46 42","+7 (925) 586 24 05","+375 173 92 56 48","+55 11 4280 6900","+7 (495) 956 09 64","+7 (495) 956 09 65","65 6697 4560","+381 61 185 6217","+381 11 430 0060","+ 357 25 247 797","+49 40 999 99 3013","+48 22 203 4500","8 (8202) 59 37 96","8 (81531) 55 590","8 (8202) 59 24 09","8 (8202) 59 37 57","8 (8202) 55 50 34","8 (8453) 62 48 72","8 (4752) 49 25 91","8 (4752) 49 25 92","8 (4752) 49 25 93","8 (863) 203 65 00","8 (863) 203 65 02","8 (863) 203 65 07","8 (4712) 999 014","8 (4712) 999 015","8 (4712) 999 016","+7 (4722) 32 16 09","+7 (4722) 32 42 83","8 (473) 200 84 48","8 (861) 220 36 70","8 (861) 220 32 25","8 (861) 220 44 51","8 (4862) 495 450","8 (4862) 495 340","8 (8652) 66 55 51","8 (8652) 66 55 52","8 (8652) 24 73 34","8 (8652) 24 73 38","8 (4742) 48 88 15","8 (4742) 77 17 61","8 (4742) 47 88 55","8 (8442) 37 22 66","8 (8442) 33 65 95","8 (8202) 59 41 25","8 (8202) 59 31 27","8 (8202) 65 01 84","8 (8202) 59 43 65","8 (8202) 59 38 92","8 (8202) 59 33 40","8 (8202) 59 35 94","8 (831) 216 22 90","8 (831) 216 22 91","8 (8342) 24 22 63","8 (8342) 24 27 23","8 (843) 202 07 31","8 (843) 202 07 48","+7 (820) 259 32 32","+74959561902","+33783878078","88136364902","+74952312747"],"website":"https://www.phosagro.com/contacts"}