

Website Phone Number & Logo Extractor

Goal

For given website url as an input argument, script needs to find, extract and output all phone numbers (eg. phone, mobile, land line, fax) and a logo URL that are present on the website.

Requirements

- Baseline technology should be Python (version 3.5 or above)
- Feel free to combine and use any other technology, library and external resource
- Structure the project and directories however you like, but there should be a script **extract.py** in the root directory that accepts one argument which is a website url
- Installation of the required libraries should be painless, so having a single command to install and setup a project so that extract.py works after it
- Stream standard output (stdout) should be used only to print out the result of extraction which is a single line of comma delimited phone numbers and a single line of a logo URL as an absolute path. To wrap up, there should be two output lines, first one for phone numbers, second one for the logo URL.
- Stream standard error (stderr) can be used for any other custom logs
- If there are no phones on the website, output should be None on first line(check examples)
- If there is no logo on the website, output should be None on second line (check examples)
- The only cleaning that should be done on a phone number is removal of any character that is not a number, plus sign or parenthesis (eg. +385/98 718-2222 to +385 98 718 2222)
- If there is a logo URL, it should be presented in working absolute path (eg. /path/to/logo.png to http://www.website.com/path/to/logo.png)

- There is no need to handle Flash websites, phone numbers that are generated dynamically through Javascript on the website or phone numbers that are actually images on the website (yes, we had seen those)

Process

Once you receive a task, you can ask us anything about the task (or life in general).

Structure and organize your questions wisely because you can send us list of questions only twice during this process.

Once you are done with the project, zip it and send it back to us.

Review

Your project will be review from a logic and structure point of view. Logic includes solution implementation and running your code through 1000 international websites for phone and logo accuracy test. Structure includes code quality and overall project/code organization.

Examples

```
$ python3 extract.py https://www.infobip.com/en/contact
+16 04 566 9031, +551139001300, +54 11 4890 8619, +44 20 7837 4180, +7 495
642 7243, +971 43 687 293, +60 3 8601 0105, +86 1500 1300 206
https://www.infobip.com/assets/img/infobip-logo-white-new.svg
```

```
$ python3 extract.py http://dnb.com.au/
13 23 33
http://dnb.com.au/_img/logo-dnb.png
```

```
$ python3 extract.py http://dnb.com.au/contact-us.html
None
http://dnb.com.au/_img/logo-dnb.png
```

```
$ python3 extract.py https://www.cialdnb.com/pt-br/
+1 (954) 507 1116
https://www.cialdnb.com/imagens/logo-cialdnb-color.svg
```

```
$ python3 extract.py https://www.powerlinx.com/contact/  
+1 (212) 465 9555  
https://www.powerlinx.com/wp-content/uploads/2016/12/plx-logo-nav-dark.png
```

```
$ python3 extract.py https://www.avea.com.au/  
1800 999 977  
https://www.avea.com.au/assets/images/AVEA_logo_rgb.png
```

```
$ python3 extract.py https://www.cmsenergy.com/contact-us/default.aspx  
(517) 788 0550, (517) 768 3376, (800) 390 1220, (517) 788 6538, (517) 788 2395,  
(517) 788 2394  
https://s2.q4cdn.com/027997281/files/design/ClientLogo.png
```

```
$ python3 extract.py https://www.phosagro.com/contacts/  
+7 (495) 232 96 89, +7 (495) 956 19 02, +7 (495) 231 31 15, + 7 (929) 600 46 42, +7  
(495) 956 09 64, (375 17) 392 56 48, (38 044) 537 04 40, +41417470370, + 48 22  
630 60 90, +49 40 999 99 3013, + 357 25 247 797, + 33 78 38 78 078, + 381 11 430  
0060, + 381 61 185 6217, + 65 6697 4560, +65 6884 9451, +55 11 4280 6900  
https://www.phosagro.com/bitrix/templates/inner/images/logo_eng.png
```

```
$ python3 extract.py https://www.zaba.hr/home/en  
08000024  
https://www.zaba.hr/home/static/img/zaba_logo.svg
```