

# INF344 2019-2020

[Tableau de bord](#) / [Mes cours](#) / [INF344 2019-2020](#) / [Représentation des connaissances, Extraction d'instances](#) / [Mail Extraction](#) / [Description](#)

- Description
- [Devoir rendu](#)
- [Edit](#)
- [Submission view](#)

## Mail Extraction

**Due date:** Monday 1 June 2020, 13:59  
**Requested files:** extractor.py ([Download](#))  
**Nombre maximal de fichiers:** 20  
**Type of work:** Individual work  
**Reduction by automatic evaluation:** 10 **Free evaluations:** 5

### Purpose

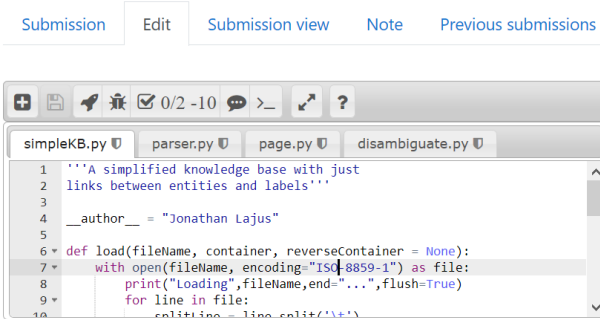
The goal of this lab is to extract mail adresses from a corpus of texts. You do not need to fetch the RFC on what constitutes a correct mail address. You only need to have a score of 90% to pass this test and move to the real one.

### Task

Complete the extractMail(content) function so that it extracts the list of the mail addresses from the content. For example, on "Hi my mail is louis.jachiet@telecom-paris.fr (but this one should also work louis.jachiet@telecom-paristech.fr). This format <louis.jachiet@telecom-paris.fr> should also works as well as louis.jachiet+test@telecom-paris.fr. ".

### Code

This lab can be programmed, run and evaluated, and graded directly in the Moodle.



In the “Edit” tab, you will find the file extractor.py you have to fill. You can directly modify the code in the Web interface. Click the disk symbol to save. Click the rocket to run it. Click the bug to debug it.

If you want to work offline: You can download the code file, modify it, and then upload it (in the tab “Submission”).

### Submitting the lab

When you are done, you can submit the lab (click the checkbox). This will directly compute an estimated grade on our gold standard. You can submit 5 times. If you submit more often, you will start to lose points.

Plagiarism is sanctioned with a grade of 0/20 (plus potentially other sanctions).

## Requested files

extractor.py

```
1  '''Extracts mails addresses
2  usage: extractor.py input.txt output.txt
3
4  Every line of output.txt contains a single mail address.
5  Note: the formatting of the output is already taken care of
6  by our template, you just have to complete the function
7  extractMail below.
8
9  (Public skeleton code)'''
10
11 import sys
12 import re
13
14 if len(sys.argv) != 3:
15     print(__doc__)
16     sys.exit(-1)
17
18 def extractMail(content):
19     # Code goes here
20     # this function should return a list of string
21     return []
22
23 inputContent = ""
24 with open(sys.argv[1], 'r', encoding="utf-8") as input:
25     inputContent = input.read()
26
27 with open(sys.argv[2], 'w', encoding="utf-8") as output:
28     for mail in extractMail(inputContent):
29         output.write(mail + "\n")
30
31
```

[VPL](#)

◀ Captation du cours

Aller à...

Résumé du cours (lexique) ▶

Connecté sous le nom « Romain Legrand » (Déconnexion)  
INF344 2019-2020  
Résumé de conservation de données  
Obtenir l'app mobile