

## Contents

<b>1</b>	<b>Prototype 1</b>	<b>1</b>
1.1	Tool to look for jobs . . . . .	1
1.1.1	static backend . . . . .	1
1.1.2	front end . . . . .	2
1.2	potfolio . . . . .	2
1.2.1	blog : Gandi / git . . . . .	2
1.2.2	Portability . . . . .	3

## 1 Prototype 1

For this first version of the program that should be ready for use in a portfolio, objectives are to select jobs depending on word apparitions, browse offers, and have some basic stats of word usage.

### 1.1 Tool to look for jobs

#### 1.1.1 static backend

1. scrapping
  - (a) **DONE** check result of today 2018-09-14 ven. 20:20
  - (b) **TODO** launch everyday from query list
    - i. **DONE** make query list out of already made queries
    - ii. cron
  - (c) add source query to features
  - (d) spped up
    - i. remove self set limits ?
    - ii. server bans ?
    - iii. vpn ?
    - iv. tor ?
  - (e) test : make scrap of the whole website
2. data selection
  - (a) org : form interface
  - (b) **DONE** add tags to actual dataset / save working dataset

3. filtering
  - (a) **DONE** resolve ipython output
4. internal correlations
  - (a) nlp
    - i. nltk
    - ii. similarity / clustering
    - iii. speed !
      - A. numpy
      - B. sql
      - C. cython
  - (b) historic
    - i. redundancy / new
    - ii. database sorted per date
5. external correlations
  - (a) identify big firms / small ventures
  - (b) geolocalize headquarters on a map
6. querying system

### 1.1.2 front end

1. **TODO** queries management
  - (a) **DONE** query bag
    - i. simple plain text file
    - ii. prompt : "add to query bag ?" when running a crawl
    - iii. autofill : get all previously used queries
2. fiche job
3. analysis viz

## 1.2 potfolio

### 1.2.1 blog : Gandi / git

1. article présentation
2. extensions futur

### **1.2.2 Portability**

1. docker / makefile package
2. Test with a friend