

Contents

| | | |
|----------|---|----------|
| 1 | Information sources | 2 |
| 1.1 | stathead / qualitative | 2 |
| 1.2 | scout / quantitative | 2 |
| 2 | Problem description | 2 |
| 2.1 | objectives | 2 |
| 2.1.1 | needs | 2 |
| 2.1.2 | will | 2 |
| 2.2 | definition questions | 2 |
| 2.2.1 | Target to reach | 2 |
| 2.2.2 | Problems to overcome | 3 |
| 3 | Project management | 3 |
| 3.1 | Time constraints | 3 |
| 3.1.1 | run test / small steps / prototypes | 3 |
| 3.2 | Limits | 3 |
| 3.2.1 | quantitative unknown | 3 |
| 3.2.2 | is it useful ? | 3 |
| 3.2.3 | data acquisition is costly | 3 |
| 3.3 | Crossovers | 4 |
| 3.3.1 | test features / ideas | 4 |
| 3.3.2 | build prototype project to show | 4 |
| 4 | Prototype 1 | 4 |
| 4.1 | outil pour chercher des jobs | 4 |
| 4.1.1 | backend statique | 4 |
| 4.1.2 | front end | 6 |
| 4.2 | moyen de présenter mes compétences | 6 |
| 4.2.1 | site : jekyll / git | 6 |
| 4.2.2 | docker / makefile package | 6 |

1 Information sources

1.1 stathead / qualitative

1.2 scout / quantitative

2 Problem description

2.1 objectives

2.1.1 needs

1. need to find a job quickly
2. need to be able to have a market overview
3. need to be able to propose a data product

2.1.2 will

1. want to have alerts for job opportunities
2. want to reveal career paths / market trends

2.2 definition questions

2.2.1 Target to reach

what makes a job interesting ?

1. easy access
 - (a) new ?
 - (b) formation ?
 - (c) entrance cost ?
 - (d) success stories
2. demand > supply
3. loan
4. re-usability
 - (a) tremplin
 - (b) multiple use case
5. connection with your parcours

2.2.2 Problems to overcome

what is it you want to solve / answer ?

1. can there be a tool to help me find a job / seek for opportunities ?
 - (a) can this be quicker the reading / regular digging ?
 - i. have an overview of field to dig further in
 - ii. needs to be quick to browse : efficient summary
 - iii. needs to be reliable : data quality / second check

3 Project management

3.1 Time constraints

3.1.1 run test / small steps / prototypes

3.2 Limits

what is shaky about this approach ?

3.2.1 quantitative unknown

1. former predictions / insights : find and evaluate
2. trending/ubiquitous ideas/concepts

3.2.2 is it useful ?

1. what would be the user case ?
 - (a) individual : orientation
 - (b) firm : market analysis

3.2.3 data acquisition is costly

1. if the approach is not efficient : pb
2. **TODO** can only be tested if large amount of data
3. need to gain faith / enthusiasm
 - (a) take small step achievements
 - i. small jobs market

- ii. data science / computing
- (b) select exciting data sources
- (c) readings
 - i. signal and the noise
 - ii. use cases
 - iii. data culture / world
- (d) methods / milestones
 - i. doubt is good
 - ii. **TODO** testing is good

3.3 Crossovers

3.3.1 test features / ideas

to gain confidence / go in the right direction

1. UI / UX
 - (a) possible on small amounts of data
2. broader analysis
 - (a) only possible on larger amount
3. application to other fields
4. sponsoring

3.3.2 build prototype project to show

4 Prototype 1

4.1 outil pour chercher des jobs

4.1.1 backend statique

1. scrapping
 - (a) **DONE** check result of today 2018-09-14 ven. 20:20
 - (b) **TODO** launch everyday from query list
 - i. make query list out of already made queries
 - ii. cron

- (c) **TODO** make a whole scrap : test
 - (d) add source query to features
 - (e) speeded up
 - i. own set limits ?
 - ii. server bans ?
 - iii. vpn ?
 - iv. tor ?
2. data selection
- (a) org : form interface
 - (b) add tags to actual dataset / save working dataset
3. filtering
- (a) resolve ipython output
4. internal correlations
- (a) nlp
 - i. nltk
 - ii. similarity / clustering
 - iii. speed !
 - A. numpy
 - B. sql
 - C. cython
 - (b) **TODO** 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

4.1.2 front end

1. **TODO** queries management
 - (a) 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

4.2 moyen de présenter mes compétences

4.2.1 site : jekyll / git

1. portage
2. article présentation / extensions / futur / Gab application

4.2.2 docker / makefile package