

Programmation et projet encadré - L7TI005

Vue d'ensemble

Yoann Dupont yoann.dupont@sorbonne-nouvelle.fr

Pierre Magistry pierre.magistry@inalco.fr

2023-2024

Université Sorbonne-Nouvelle
INALCO
Université Paris-Nanterre

Généralités

Se référer au syllabus du cours

- Les horaires et lieux
 - actuellement → 9h - 12h, campus Nation, salle B326
 - github: <https://github.com/pmagistry/PPE1-2024>
 - iCampus (P3) :
<https://icampus.univ-paris3.fr/course/view.php?id=61395>
 - clé d'autoinscription icampus : **PPE1@2425**

Se référer au syllabus du cours

- Les horaires et lieux
 - actuellement → 9h - 12h, campus Nation, salle B326
 - github: <https://github.com/pmagistry/PPE1-2024>
 - iCampus (P3) :
<https://icampus.univ-paris3.fr/course/view.php?id=61395>
 - clé d'autoinscription icampus : **PPE1@2425**
- Modalités de contrôle de connaissance (MCC)
 - **40%** - travail individuel
 - Exercices
 - Journal de bord individuel

Se référer au syllabus du cours

- Les horaires et lieux
 - actuellement → 9h - 12h, campus Nation, salle B326
 - github: <https://github.com/pmagistry/PPE1-2024>
 - iCampus (P3) :
<https://icampus.univ-paris3.fr/course/view.php?id=61395>
 - clé d'autoinscription icampus : **PPE1@2425**
- Modalités de contrôle de connaissance (MCC)
 - **40%** - travail individuel
 - Exercices
 - Journal de bord individuel
 - **60%** - projet en groupe (idéalement 3 personnes) → pensez-y tout de suite !
 - Une de suivi de projet
 - Une sur le résultat final

Allez ici : <https://lite.framacalc.org/plurital2425-a9yd>

- Systèmes d'exploitation (OS) attendus : (K)Ubuntu 24.04 (version d'avril 2024 : <https://ubuntu.com/download/desktop>, <https://kubuntu.org/getkubuntu/>) ou MacOS

- Systèmes d'exploitation (OS) attendus : (K)Ubuntu 24.04 (version d'avril 2024 : <https://ubuntu.com/download/desktop>, <https://kubuntu.org/getkubuntu/>) ou MacOS
 - idéalement : votre seul OS
 - acceptable : dual-boot (deux systèmes cohabitent sur l'ordinateur), WSL? (Windows)
 - bon courage : machine virtuelle, chromeOS, tout le reste en fait.

- Systèmes d'exploitation (OS) attendus : (K)Ubuntu 24.04 (version d'avril 2024 : <https://ubuntu.com/download/desktop>, <https://kubuntu.org/getkubuntu/>) ou MacOS
 - idéalement : votre seul OS
 - acceptable : dual-boot (deux systèmes cohabitent sur l'ordinateur), WSL? (Windows)
 - bon courage : machine virtuelle, chromeOS, tout le reste en fait.
- Si vous n'entrez pas dans ces cases ou que vous avez des problèmes spécifiques, venez nous voir !

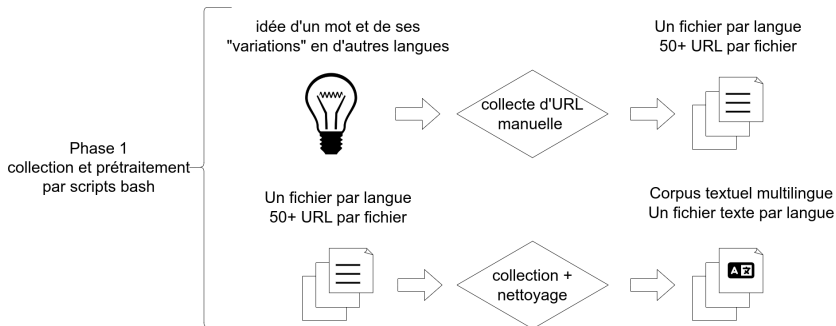
Vision globale du cours

Projet encadré ?

- Projet : *La vie des mots sur le web*
- projet initiatique à la linguistique de corpus et à Unix
 - projet à 3 (idéalement de chaque université)
 - étudier l'usage un mot et ses "variations" en plusieurs langues sur le web
 - constituer un corpus multilingue → ensemble de textes compilés
 - constitution à l'aide d'outils Unix → occasion d'apprendre les outils avec une utilisation concrète de récolte des données textuelles et leur nettoyage

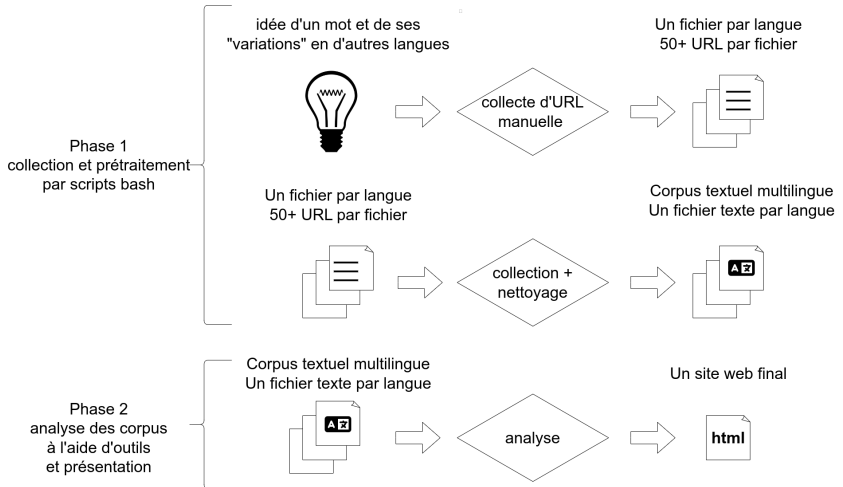
Encadré car nous allons vous guider dans les premières étapes et ensuite suivre votre progression.

Projet en deux phases



Crédit image ampoule : OpenClipart-Vectors (pixabay.com)

Projet en deux phases



Crédit image ampoule : OpenClipart-Vectors (pixabay.com)

Les pipelines Unix - petit exercice !

Que peut-on dire sur ce groupes d'éléments ?

2 2 4 4 5 6

Les pipelines Unix - petit exercice !

Que peut-on dire sur ce groupes d'éléments ?

2 2 4 4 5 6

Comment trouve-t-on le plus grand nombre dans cet ensemble ?

Les pipelines Unix - petit exercice !

Comment trouve-t-on le plus grand nombre dans :

2 4 2 6 5 4

?

Les pipelines Unix - petit exercice !

Comment trouve-t-on la taille du plus long mot dans :

je suis un mignon petit chat

?

Les pipelines Unix - petit exercice !

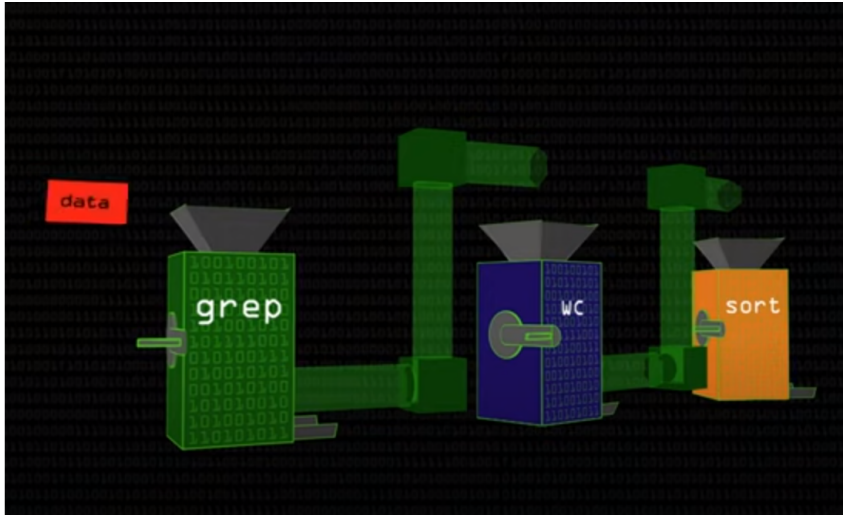
Comment trouve-t-on la taille du plus long mot dans :

je suis un mignon petit chat

?


On a appliqué plusieurs traitements à la suite pour obtenir différents résultats. La plupart des outils sont réutilisables pour peu que les données soient mises dans un état propice → c'est ce qu'on va faire avec les *pipelines* Unix dans ce cours.




Les pipelines Unix






source : "Unix Pipeline (Brian Kernighan) - Computerphile" (Youtube)




Gestion de projet avec GitHub - le dépôt de vos fichiers








 **steeve** / **france.code-civil** Public


 Notifications  Fork 149  Star 2.7k

[Code](#) [Issues](#) 21 [Pull requests](#) 9 [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#)

 master  3 branches  0 tags [Go to file](#) [Code](#)

 **steeve** Merge pull request #18 from Kcchouette/patch-1 ...  2 159d951 on 31 Mar 2015  281 commits

 Livre II	Modifié par LOI n°2015-177 du 16 février 2015 - art. 1 (V)	8 years ago
 Livre III	Modifié par LOI n°2015-177 du 16 février 2015 - art. 1 (V)	8 years ago
 Livre IV	Modifié par Ordonnance n°2014-326 du 12 mars 2014 - art. 100	8 years ago
 Livre Ier	Modifié par LOI n°2013-403 du 17 mai 2013 - art. 1 (V)	8 years ago
 Livre V	Modifié par LOI n°2015-177 du 16 février 2015 - art. 1 (V)	8 years ago
 Titre préliminaire	Créé par LOI n°2013-404 du 17 mai 2013 - art. 13	10 years ago
 README.md	Improve README	8 years ago

 **README.md**


Le Code civil français, sous Git


« Nul n'est censé ignorer la loi. »


Et pourtant nous autres citoyens (moi y compris) ne connaissons pas les lois qui construisent et gouvernent notre République. Pour faire passer nos idées, nos textes sont bien connus et connus...


About

Le code civil français sous git

 Readme

 2.7k stars

 217 watching

 149 forks


Releases


No releases published


Packages

No packages published

Contributors 3


 **steeve** Steeve Morin

 **emerick42** Emerick Kasbarian


 **Kcchouette**


Projet "france.code-civil" : <https://github.com/steeve/france.code-civil>


Gestion de projet avec GitHub - historique d'un projet




 **steeve** / **france.code-civil** Public


[Code](#) [Issues](#) 21 [Pull requests](#) 9 [Actions](#) [Projects](#) [Wiki](#) [Security](#)



 master ▾


 Commits on Mar 31, 2015




Merge pull request #18 from Kcchouette/patch-1 ...
 **steeve** committed on 31 Mar 2015

 2  [159d051](#) 

Improve README ...
 **Kcchouette** committed on 31 Mar 2015

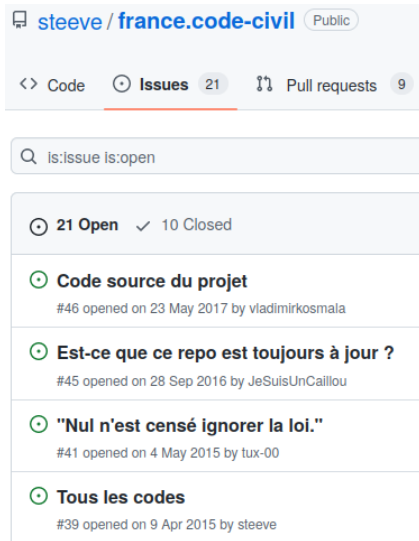
 [d069119](#) 

"Nul n'est censé ignorer la loi." ...
 **steeve** committed on 31 Mar 2015

 1 Unverified  [0fe3371](#) 

Projet "france.code-civil" : <https://github.com/steeve/france.code-civil>

Gestion de projet avec GitHub - communiquer



The screenshot shows the GitHub interface for the repository 'steeve / france.code-civil', which is marked as 'Public'. The navigation bar includes links for 'Code', 'Issues' (which is selected and underlined in red), and 'Pull requests'. Below the navigation bar is a search bar containing the text 'is:issue is:open'. The main content area displays the status of issues: '21 Open' and '10 Closed'. A list of four issues is shown, each with a green circular icon containing a white dot. The issues are: 'Code source du projet' (#46, opened on 23 May 2017 by vladimirkosmala), 'Est-ce que ce repo est toujours à jour ?' (#45, opened on 28 Sep 2016 by JeSuisUnCaillou), '"Nul n'est censé ignorer la loi."' (#41, opened on 4 May 2015 by tux-00), and 'Tous les codes' (#39, opened on 9 Apr 2015 by steeve).

steeve / france.code-civil Public

<> Code Issues 21 Pull requests 9


Q is:issue is:open

21 Open ✓ 10 Closed

- Code source du projet
#46 opened on 23 May 2017 by vladimirkosmala
- Est-ce que ce repo est toujours à jour ?
#45 opened on 28 Sep 2016 by JeSuisUnCaillou
- "Nul n'est censé ignorer la loi."
#41 opened on 4 May 2015 by tux-00
- Tous les codes
#39 opened on 9 Apr 2015 by steeve

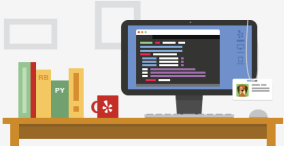
Projet "france.code-civil" : <https://github.com/steeve/france.code-civil>

Page web avec GitHub


 **Engineering** [Home](#) [Blog](#) [Open Source](#) [AWE](#) [What's it like?](#)

Yelp Open Source

- Power your app with the same code Yelp uses every day
- We love open source and want to give back with projects like mrjob
- Get started today with 99 projects from 66 Yelp developers



Popular



paasta

★ 1619 ▼ 222 **Python**

An open, distributed platform as a service

[Blog post](#)


```
16 query_key: username
17 timeframe:
18 days: 1
19 filter:
20 - query:
21   query_string:
22     query: "document_ty
```

elastalert

★ 7830 ▼ 1771 **Python**

Easy & Flexible Alerting With Elasticsearch

[Blog post](#)



undebt

★ 1638 ▼ 62 **Python**

A fast, straightforward, reliable tool for performing massive, automated code refactoring

[Blog post](#)

```
root@and27f6@dc8f:/code$ ps auxf
PID  STAT  TIME  COMMAND
1  ss    0:00  dumb_init
7  ss    0:00  runsvdir
8  S      0:00  \ runsv
9  S      0:00  \ logger
10 S     0:15  \ python3
14 -S    0:28  \ net
```

dumb-init
















★ 5861 ▼ 322 **Python**

A minimal init system for Linux containers

[Blog post](#)

Page GitHub des projets *open source* de Yelp : <https://yelp.github.io>

Organiser son travail

 YoannDupont Better counting of entities in corpus. 0fe35f0 on 23 Jul 2019  8 commits		
	2016	Adding political affairs as events. 4 years ago
	2017	Adding political affairs as events. 4 years ago
	2018	Adding political affairs as events. 4 years ago
	.gitignore	Initial commit 4 years ago
	changelog.md	corrections on products such as Wikipedia or Twitter. 4 years ago
	concatenate_dir.py	corrections on products such as Wikipedia or Twitter. 4 years ago
	concatenate_subdirs.py	corrections on products such as Wikipedia or Twitter. 4 years ago
	count.py	Better counting of entities in corpus. 3 years ago
	get_inconsistencies.py	Wikinews 2018 annotated. 4 years ago
	get_month_articles.py	Wikinews 2018 annotated. 4 years ago
	license.txt	Added the Wikinews 2016, 2017 and the first three months of 2018 doc... 4 years ago
	make_matches.py	Wikinews 2018 annotated. 4 years ago
	readme.md	Added: 4 years ago

Source : WiNER-fr <https://github.com/YoannDupont/WiNER-fr>

Organiser son travail

```
|— year
|   |— month1
|   |   |— *.ann
|   |   |— *.txt
|   |— month2
|   |   |— *.ann
|   |   |— *.txt
|   |— year-month1-urls.txt
|   |— year-month2-urls.txt
```

Source : WiNER-fr <https://github.com/YoannDupont/WiNER-fr>

Instant démonstration

Et maintenant la pratique