

Fil Rouge : modalités de notation

Mastère Spécialisé
« Assurance, Actuariat et Big Data »

23 mars 2018

Romain Jouin



ÉCOLE
D'INGÉNIEURS
PARIS-LA DÉFENSE

Contexte

Projet Vélib

Projet Open Data

Projet « Project Centric Mailbox »

Une réorientation des projets fil rouge s'impose

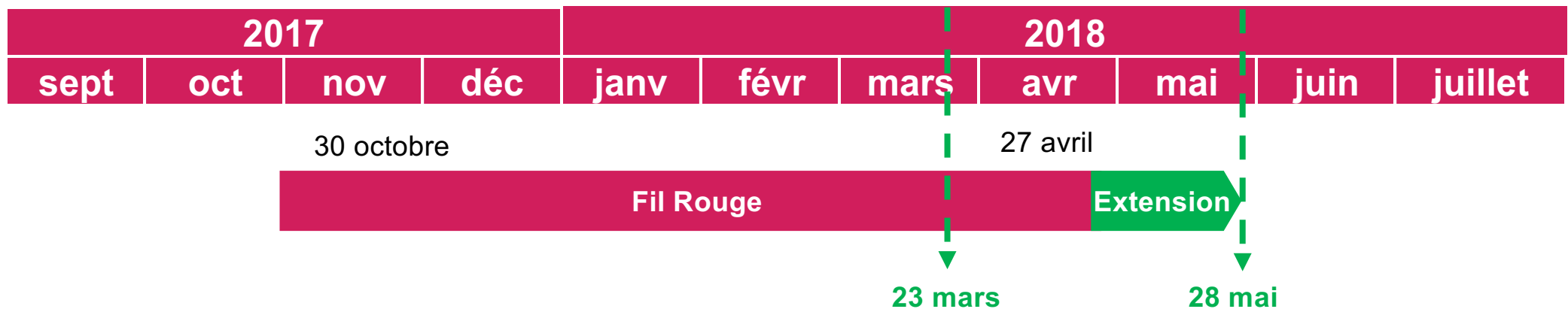
Différentes difficultés imprévues ont rendues difficile l'attente des objectifs annoncés en début d'année sur les projets fil rouges.

Ce document à pour objectif de proposer un cadre adapté pour l'évaluation de cette matière.

Il se compose notamment :

1. D'une extension d'un mois de la durée des projets, qui devront être rendu pour **le lundi 28 mai à 12h00**.
2. D'une **redéfinition des objectifs** tels que décrits dans les pages qui suivent.

Compte tenu de ces adaptations, l'école n'admettra **aucun retard** sur la délivrance des projets et sera d'autant plus attentive à la qualité des rendus.



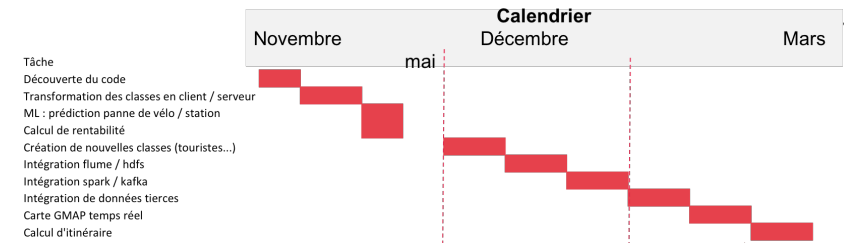
Contexte

Projet Vélib

Projet Open Data

Projet « Project Centric Mailbox »

Projet Vélib :



Tâches ajustées

- 1 - **Documenter** le code avec **Sphinx** (rendre un site web de documentation Sphinx)
- 2 - Modifier le code pour le rendre **"client" "serveur"** :
 - Une classe de serveur centrale "réseau vélib" qui attend la connexion de stations
 - Des serveurs détachés "station" qui se connecte au réseau pour s'annoncer avec leur coordonnées
 - Des clients "cyclistes" qui demandent au "réseau" qu'elle est la station la plus proche de leur position x,y => les vélos récupèrent l'adresse IP et le numéro de port de la station, puis lui demandent de prendre ou laisser un vélo
- 3) « **Dockerisation** » du code (créer un DockerFile qui permet de lancer le code dans un environnement virtuel indépendant)

Notation

Documentation	: 4 points
Code client-serveur	: 10 points
Dockerisation	: 6 points

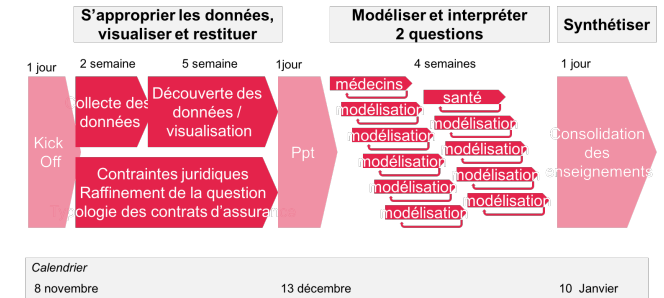
Contexte

Projet Vélib

Projet Open Data

Projet « Project Centric Mailbox »

Projet Open Data:



Tâches ajustées

- 1) **Inventaire détaillé des ressources** (Lister les bases existantes, leur contenu, les historiques, etc..)
- 2) **Cartographie et statistiques** sur des tables choisies par le groupe **sur un historique** temporel choisi/**pertinent** d'au moins trois époques. Il est attendu au moins **8 analyses** d'intérêts. Par exemple : dépenses de santé par département en 2017/2016/205 = 1 point d'analyse. L'intérêt du choix des analyses sera évalué.
- 3) **Réflexion sur l'open data** :
 - **acteurs, enjeux, dynamique d'alimentation**, contraintes, avantages, inconvénients.
 - **Possibilités d'exploitation** pour une société 'assurance : quels usecase ? Quelles assurances ouvrent ?
 - **Comment ouvrir ses données** en open data pour une entreprise d'assurance ? (Création d'un **draft de stratégie open data** : quelles données ouvrir. Exemple démarche axa sur données de sinistre : des données brutes ou des données agrégées, ..)

Notation

- Inventaire : 6 points
Cartographies et statistiques : 6 points
Réflexion sur l'open data : 8 points

Contexte

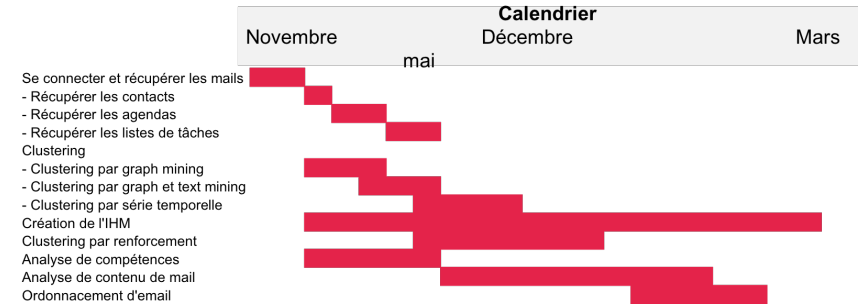
Projet Vélib

Projet Open Data

Projet « Project Centric Mailbox »

PCM :

Tâches ajustées



1) **Comprendre le code** des 5ème année : le **documenter** de code sous **sphinx** (rendre un site web de documentation Sphinx) et **Récupérer un jeu de donnée** (petit, il peut être de l'ordre de 1000 mails) sur vos mails personnels

2) **Proposer deux algorithmes de clustering** de mails améliorant l'algorithme actuel (qui crée les groupes en supprimant le propriétaire du compte de la matrice expéditeur / récepteur) **prenant en compte** :

2.1) **le nombre de mails envoyé entre les personnes** (création d'une matrice qui compte le nombre d'échange entre chaque personne des clusters actuels) => **décider de l'exclusion du cluster les personnes** n'ayant pas suffisamment d'échange avec le reste du groupe)

2.2) **la distance des mails les uns par rapport** aux entres en terme de text mining (cf stop-words / tf-idf / librairie gensim...) => décider de **regrouper dans des sous clusters les textes qui se ressemblent le plus**

3) « **Dockerisation** » du code (créer un DockerFile qui permet de lancer le code dans un environnement virtuel indépendant)

Notation

Documentation : 4 points

Algo sur le nombre de mails : 4 points

Algo sur la distance de mails : 6 points

Dockerisation : 6 points