

## **PFE – Web Usage Mining**



## **Documentation du projet**

**Réalisé par :**  
Ismael MEBROUKI  
Ali AZZOUG

**Polytech INFO 5A**  
**Professeur : M. Quafafou, N. Durand**

# 1. Description et analyse de l'environnement

## 1.1 Environnement

L'application web est compatible avec tous les navigateurs : Chrome, Mozilla, Opera, Safari. Ce site est en version bureau, et il ne dispose pas de version mobile. L'utilisation du site sur une tablette ou un smartphone peut engendrer des erreurs d'affichage. Le temps de réponse d'une requête vers la base de données (recherche, ajout, suppression et modification) ne doit pas dépasser 3 secondes.

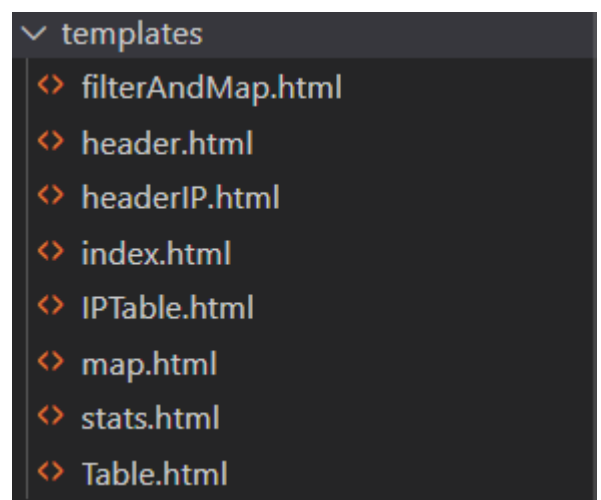
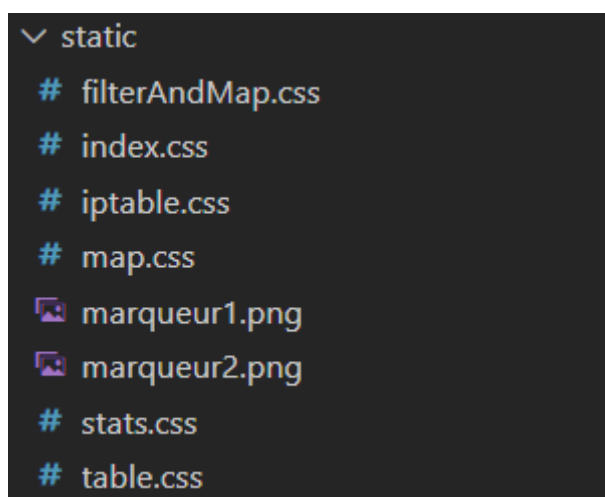
## 1.2 Description des ressources logicielles

- Le code et les fonctions sont réalisés en Python
- Les bibliothèques utilisées pour Python sont disponibles su github
- Le Framework utilisé côté serveur est Flask
- Utilisation d'IP-API pour la géolocalisation
- Le front end est codé en utilisant uniquement Html et CSS
- La base de données est un fichier log situé dans le dossier log
- Pour utiliser le programme, une connexion Internet est obligatoire.

# 2. Architecture globale

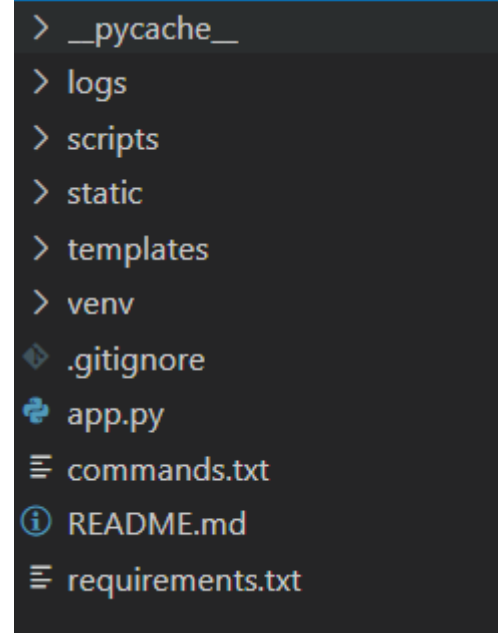
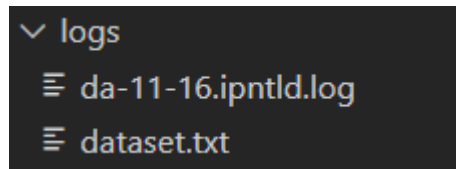
## 2.1 Décomposition côté affichage

Les fichiers Html sont situés dans la partie templates de l'application et leur fichier style correspondant ainsi que les images utilisées sont dans static.



## 2.2 Décomposition côté serveur

Le code de l'application est situé dans `app.py` à la racine du dossier et récupère les données dans le dossier logs contenant les logs que l'on souhaite utiliser pour l'application.



## 2.3 Structure d'un fichier log

Un fichier log est disposé comme suit :

Date | Heure | Pages consultées | IP | Site Visité | Status Code | Données échangés

```
2015-11-16 22:34:49 usep-pinchinades 68.180.229.29 associations.paysdaixassociations.org 500 12562
2015-11-16 22:35:03 association-regionale-pour-le-developpement-de-l-enseigne-3 68.180.228.251 agneaux-50.ville.mygaloofr 301 171
2015-11-16 22:35:05 association-regionale-pour-le-developpement-de-l-enseigne-3 68.180.228.251 agneaux-50.ville.mygaloofr 500 1968
2015-11-16 22:35:14 ascair 68.180.230.237 paris-75.ville.mygaloofr 301 156
2015-11-16 22:35:16 ascair 68.180.230.237 paris-75.ville.mygaloofr 500 1968
2015-11-16 22:35:29 arc-animation-entrainement-a2e-2 208.115.111.66 associations.paysdaixassociations.org 301 171
2015-11-16 22:35:31 aclap-action-de-coordination-de-lieux-et-d-accueil-aux-pers 208.115.111.66 associations.paysdaixassociations.org 3
2015-11-16 22:35:33 renesens-2 208.115.111.66 associations.paysdaixassociations.org 301 171
2015-11-16 22:37:40 les-archers-de-l-abbaye-2 82.227.35.120 massy-91.ville.mygaloofr 301 46
2015-11-16 22:37:42 les-archers-de-l-abbaye-2 82.227.35.120 massy-91.ville.mygaloofr 500 1218
2015-11-16 22:37:46 tja-kon-fesi-avenir-2 68.180.229.161 maripasoula-973.ville.mygaloofr 301 171
2015-11-16 22:37:46 association-sport-et-musique-de-bras-fusil-2 146.185.234.48 saint-benoit-974.ville.mygaloofr 301 109
2015-11-16 22:37:48 tja-kon-fesi-avenir-2 68.180.229.161 maripasoula-973.ville.mygaloofr 500 2000
2015-11-16 22:37:52 association-sport-et-musique-de-bras-fusil-2 146.185.234.48 saint-benoit-974.ville.mygaloofr 500 2750
2015-11-16 22:38:09 union-economique-commerciale-de-l-agglomeration-rochelaise- 146.185.234.48 la-rochelle-17.ville.mygaloofr 301 46
2015-11-16 22:38:10 union-economique-commerciale-de-l-agglomeration-rochelaise- 146.185.234.48 la-rochelle-17.ville.mygaloofr 500 131
2015-11-16 22:38:39 les-films-du-minotaure-2 146.185.234.48 le-creusot-71.ville.mygaloofr 301 93
2015-11-16 22:38:43 les-films-du-minotaure-2 146.185.234.48 le-creusot-71.ville.mygaloofr 500 3671
2015-11-16 22:38:49 alizee-club 68.180.230.188 marseille-13.ville.mygaloofr 301 171
2015-11-16 22:39:00 alizee-club 68.180.230.188 marseille-13.ville.mygaloofr 500 2062
2015-11-16 22:40:50 association-des-commerçants-de-ville-port-2 146.185.234.48 saint-nazaire-44.ville.mygaloofr 301 78
2015-11-16 22:40:59 yogaroma-2 68.180.228.96 bonzac-33.ville.mygaloofr 301 171
```

### 3. Analyse statique

La page carte : Carte et filtres

#### **sendCity() :**

Cette fonction récupère les formulaires présents dans la page filterAndMap.html sous forme de POST et vérifie leurs existences.

*Les variables :*

Data : contient les valeurs des POST

Table : Table contenant les IPs et leur localisation

ipAdress : contient l'IP entré en formulaire dans la partie analyse

ipdata et ipdata\_html : contient le tableau d'information pour une IP

topPages et topPages\_html : contient les pages les plus consultés par une IP

site : contient le site entré en formulaire dans la partie analyse

vues : contient le nombre de vue pour un site précis

Retourne la page qui va être afficher à l'URL : /map et les données à envoyer à cette page

#### **searchBySite(Vsite) :**

Cette fonction prend en paramètre un site et permet d'obtenir toutes les IPs qui ont visités ce site.

*Les variables :*

Tab : contient les données du fichier log

Filter : contient le filtre de recherche dans le fichier log

Research : contient le résultat du filtre appliqué au fichier log

Retourne un tableau qui contient le résultat de la recherche avec filtre

#### **searchByIP(ipAdress) :**

Cette fonction prend en paramètre une adresse IP et permet d'obtenir toutes les informations d'une IP en particulier.

*Les variables :*

Tab : contient les données du fichier log

Filter : contient le filtre de recherche dans le fichier log

Research : contient le résultat du filtre appliqué au fichier log

Retourne un tableau qui contient le résultat de la recherche avec filtre

**searchByDate(date) :**

Cette fonction prend en paramètre une date et permet d'obtenir toutes les IP qui ont été actives à cette date.

*Les variables :*

Tab : contient les données du fichier log

Filter : contient le filtre de recherche dans le fichier log

Research : contient le résultat du filtre appliqué au fichier log

Retourne un tableau qui contient le résultat de la recherche avec filtre

**PagesLesPlusConsultés(ipAdress) :**

Cette fonction prend en paramètre une adresse IP et permet d'obtenir les pages les plus consultés par une IP.

*Les variables :*

df : contient les données du fichier log avec que des adresses IP uniques

res : contient le résultat du filtre appliqué au fichier log

Retourne un tableau qui contient toutes les pages les plus consultés par une IP

**numClickSite(site):**

Cette fonction prend en paramètre un site et permet d'obtenir le nombre clique sur le site passé en paramètre.

*Les variables :*

df : contient les données du fichier log

Filter : contient le filtre de recherche dans le fichier log

Research : contient le résultat du filtre appliqué au fichier log

Retourne un tableau qui contient le résultat de la recherche avec le filtre

**convertTabtoIPT(tab) :**

Cette fonction prend en paramètre un tableau et convertis les adresses IP du tableau en latitude, longitude.

*Les variables :*

Df2 : contient les données du fichier log

data : définition du tableau qui contient les conversions

api\_url : contient l'url de l'api pour convertir IP en géolocalisation

convertedDF : contient le tableau convertis en dataframe

Retourne un dataframe qui associe une latitude longitude à une IP

**deleteDoublon(tableau,param):**

Cette fonction prend en paramètre un tableau et un paramètre et supprime les lignes contenant le même paramètre.

*Les variables :*

Tab : contient les données du fichier log et supprime les lignes en double

Retourne un dataframe qui ne contient aucune ligne en double

**getallTAB():**

Cette fonction permet de charger les données du fichier log dans un dataframe pandas.

*Les variables :*

url : contient l'url du fichier log

full\_df2 : contient le dataframe qui récupère les données du fichier log dans un tableau

Retourne un dataframe qui contient toutes les données d'un fichier log

## La page Statistiques

### **stats():**

Cette fonction récupère les formulaires présents dans la page stats.html sous forme de POST et vérifie leurs existences.

*Les variables :*

clientsActifs: contient le tableau des clients les plus actifs

mostVisitedSite: contient le tableau des sites les plus visités

pageConsultes: contient le tableau des pages les plus consultés

joursA: contient le tableau des jours avec le plus d'activités

Retourne la page qui va être afficher à l'URL : /stats et les données à envoyer à cette page

### **clientPlusActifs():**

Cette fonction permet d'obtenir les IP qui sont le plus actifs dans le fichier log.

*Les variables :*

df : contient les données du fichier log

res : contient le résultat du filtre appliqué au fichier log

Retourne un tableau qui contient tous les clients les plus actifs et leur nombre de requêtes

### **siteLesPlusVisités():**

Cette fonction permet d'obtenir les sites avec le plus de visites dans le fichier log.

*Les variables :*

df : contient les données du fichier log

res : contient le résultat du filtre appliqué au fichier log

Retourne un tableau qui contient les sites avec le plus d'IP ayant visités ce site et le nombre de visites

### **pageLesPlusConsultes():**

Cette fonction permet d'obtenir les pages qui ont été le plus consultés par les IP.

*Les variables :*

df : contient les données du fichier log

res : contient le résultat du filtre appliqué au fichier log

Retourne un tableau qui contient les pages les plus consultés et leurs nombres

### **joursActif():**

Cette fonction permet d'obtenir les jours avec le plus d'activité (plus de requêtes).

*Les variables :*

df : contient les données du fichier log

res : contient le résultat du filtre appliqué au fichier log

Retourne un tableau qui contient les jours et le nombre de requêtes dans ce jour.

## La page IP

### **headerIPdef() :**

Cette fonction récupère les formulaires présents dans la page headerIP.html sous forme de POST et vérifient leurs existences.

*Les variables :*

table : contient le tableau de conversion d'une IP en latitude longitude

data : contient les valeurs des POST

Retourne la page qui va être afficher à l'URL : /iptable et les données à envoyer à cette page.

### **convertIPtoCoord(ipadress) :**

Cette fonction prend en paramètre un tableau et convertis les adresses IP du tableau en latitude, longitude.

*Les variables :*

Df2 : contient les données du fichier log

data : définition du tableau qui contient les conversions

api\_url : contient l'url de l'api pour convertir IP en géolocalisation

convertedDF : contient le tableau convertis en dataframe

Retourne un dataframe qui associe une latitude longitude à une IP