

Polytech Marseille 2022

# Geolocalized Web Usage MiningRapport

Projet S8

BARAR Bilal / MONIMEAU Paul / MEYA Georges  
20/05/2022

Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

### Historique des révisions

Date	Description et justification de la modification	Auteur	Pages / Chapitre	Edition / Révision
14/05/2022	Création du document	BB	Toutes	0A
16/05/2022	Structuration du rapport	Tous	Toutes	1
17/05/2022	Rédaction des parties	Tous	Toutes	2
20/05/2022	Relecture	Paul	Toutes	2.1

Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

# TABLE DE MATIERE

TABLE DE MATIERE .....	3
1. Introduction.....	4
1.1. Objet du document .....	4
1.2. Responsabilités.....	4
1.3. Outils Utilisés.....	4
2. Etude de l'existant.....	6
2.1. Présentation Général .....	6
3. Analyse .....	7
3.1. Les Acteurs .....	7
4. Conception .....	10
4.1. Arborescence et organisation du code .....	10
4.2. Mapping entre les différents micro services.....	12
4.3. Gestion de la donnée .....	14
5. Réalisation .....	18
6. Tests.....	18
7. Organisation .....	24
8. Conclusion .....	25
9. Terminologie.....	26
10. Repository GITHUB.....	26

Réf : RP-2.1	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges		
Client : Polytech Projet : Géocalized Web Usage Mining		

# 1. Introduction

## 1.1. Objet du document

Ce document sert à présenter le projet de fin d'étude : "Geocalized Web Usage Mining"  
Il présente les différentes fonctionnalités de l'application proposée. Il contient les explications des choix techniques des différents outils utilisés, le déroulement et la réalisation de ce projet.

## 1.2. Responsabilités

La constitution de ce rapport est de la responsabilité des membres du groupe de projet. Chaque membre du groupe s'engage à relire et approuver chaque nouvelle version de ce dernier.

## 1.3. Outils Utilisés

Nous avons utilisé différents types d'outils en fonctions de nos besoins

Pour mettre en place l'application nous avons utilisé différentes technologies et Framework :

- Angular : Framework GOOGLE qui permet de monter des applications web. C'est cet outil qu'on a utilisé.
- Python Flask : C'est avec cet outil qu'on a géré la partie backend du projet sous forme d'API REST.
- Docker : Pour conteneuriser notre application, nous avons fait le choix de prendre la solution Docker.

Nous avons utilisé différents logiciels, que ce soit pour le code ou pour réaliser de la documentation:

Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

- Google Chrome : C'est le choix de navigateur que l'on a pris sachant que Google est aussi le propriétaire d'Angular. Ainsi on maximise la compatibilité.
- Word : Pour éditer notre rapport, nous avons fait le choix de la suite de Microsoft
- Webstorm : C'est l'environnement de développement qui a été retenu. Très complet et le meilleur dans sa catégorie pour les projets Angular.
- Github : Pour le versionnage de l'application, nous avons pris Github sachant que c'est une solution que nous avons déjà utilisée et donc nous en connaissant déjà les principes.

Concernant les librairies nous nous sommes appuyés sur :

- Bootstrap : Il nous a facilité dans la création du design avec du code CSS et HTML pré-embarqué. Par exemple nous l'avons utilisé pour la barre de navigation ou encore la page d'authentification.
- Charts JS : Chart.js est une bibliothèque JavaScript open source gratuite pour la visualisation de données. Il nous aura permis de mettre en place tous les graphiques
- API Google MAP JavaScript : L'API de Google avec lequel on a mis en place la carte
- IP API : Grâce à cette API, on a pu extraire des informations d'une adresse IP comme sa localisation ou encore le type d'appareil.
- GZIP : Librairie python qui permet de compresser le JSON. La quantité de donnée important ainsi compresser permet d'augmenter la fluidité de l'application.
- Panda : Librairie qui permet de gérer et créer des DataFrames en Python, permettant une meilleure structure de la donnée.
- RSA : Permet de chiffrer des données de manières asymétriques avec clé privée et clé publique. Utilisé dans notre projet pour crypter les logins et mots de passe lors de leur stockage.

Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges  Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

## 2. Etude de l'existant

### 2.1. Présentation Général

Le contexte de ces deux dernières années aura permis de connaître une forte digitalisation de notre société et surtout au niveau du monde professionnel : Création de site web pour les commerçants, mise en place de services sur le web, travail à distance et décuplement des visio-conférence,. Cette croissance a généré de très grandes masses de données relatives aux traces d'usage du Web par les internautes. Ces traces d'usages sont enregistrées par les serveurs dans des fichiers journaux, appelés Log Files.

Ces données sont des ressources précieuses pour les propriétaires des sites web qui auront la capacité de mieux comprendre leurs visiteurs afin de mieux répondre à leurs attentes. Le Web Usage Mining est un domaine informatique qui s'intéresse au processus d'extraction des connaissances à partir des données d'usage sur internet.

Ce projet s'inscrit dans le cadre du module PROJET du semestre 8 de la formation Ingénieur filière informatique de Polytech Marseille. Le but étant d'analyser un jeu de données d'un serveur pour en extraire les informations sur le trafic des sites, mais également en établissant une analyse de la localisation géographique des sources de trafic.

Pour mettre en place notre projet, nous avons fait le choix de proposer une solution d'application web Angular qui propose différents services d'analyse des fichiers logs d'un certain site. Cette une application avec une architecture à 3 couches avec une partie Frontend, Backend et Base de données : Voir 4.1.

On retrouvera donc principalement les fonctionnalités suivantes :

- Renseigne sur l'affluence d'un site avec une échelle temporelle
- Affiche le nombre de consultations d'un site web
- Affiche la localisation des adresses IP qui ont visité un site web sur une MAP
- Prédiction et Machine Learning : Prédire le nombre de consultation d'un site sur une durée

Réf : RP-2.1	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges		
Client : Polytech Projet : Géocalized Web Usage Mining		

## 3. Analyse

### 3.1. Les Acteurs

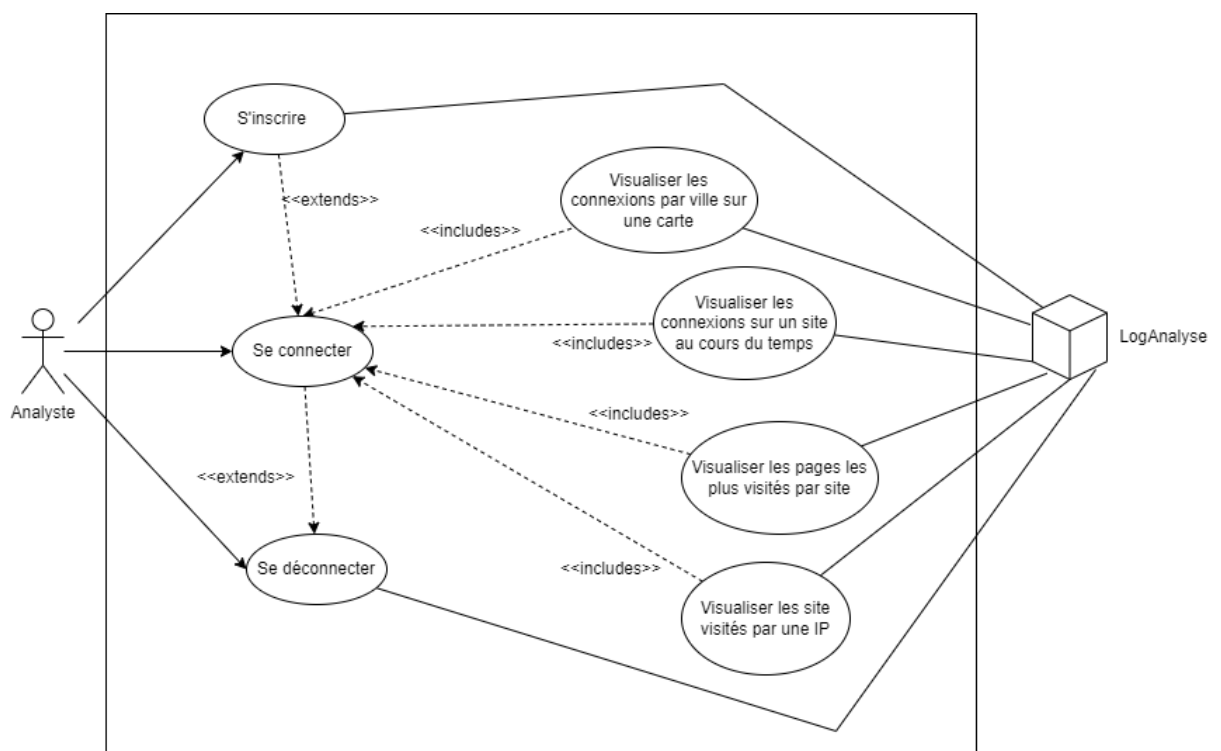
Les fonctionnalités du site web sont destinées à une personne voulant analyser les données d'un site web grâce à l'application LogAnalyse.

#### 3.1.1. Liste des Cas D'utilisation

Les principales fonctionnalités du site web sont listés dans le tableau ci-dessous :

Référence du cas d'utilisation	Description
SPC-01	S'inscrire sur LogAnalyse
SPC-02	Se connecter à LogAnalyse
SPC-03	Se déconnecter de LogAnalyse
SPC-04	Visualiser les connexions sur un site par ville sur une carte
SPC-05	Visualiser les connexions sur un site au cours du temps
SPC-06	Visualiser les pages les plus visités par site
SPC-07	Visualiser les site visités par une IP

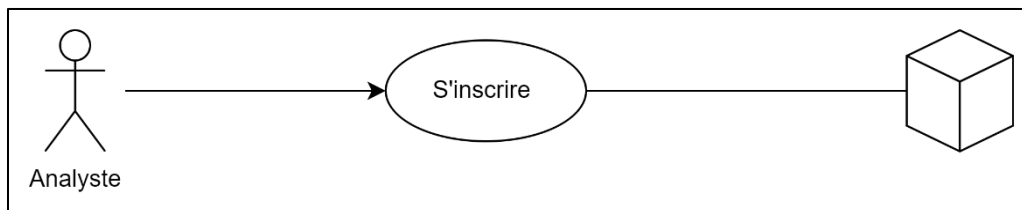
Voici un diagramme des cas d'utilisation du site web :



Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

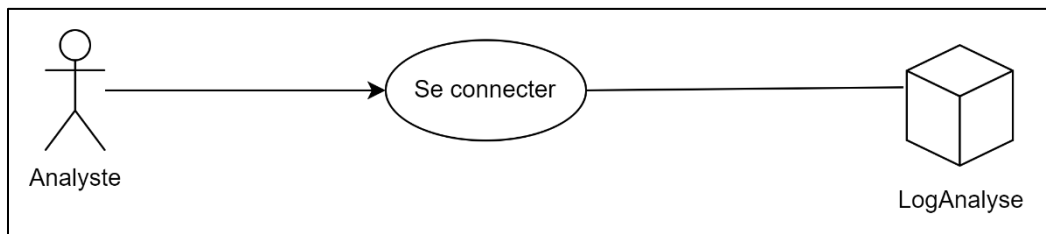
### 3.1.2. Description des cas d'utilisations

- SPC-01 : S'inscrire sur LogAnalyse



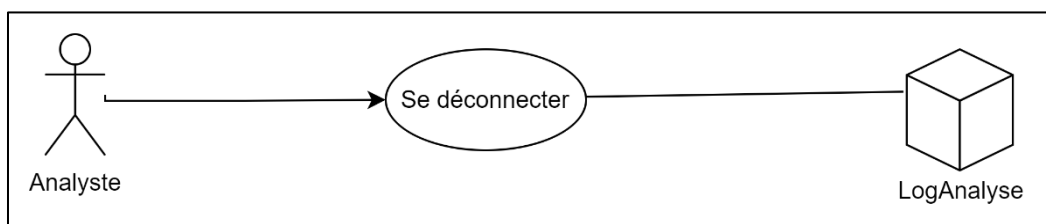
Pour accéder aux fonctionnalités de l'application une authentification est nécessaire. L'inscription au site est donc obligatoire de s'inscrire. L'utilisateur devra saisir un login et un mot de passe. Ces informations seront ensuite cryptées puis sauvegardé dans une base de donnée.

- SPC-02 : Se connecter à LogAnalyse



Pour accéder aux fonctionnalités du site web il est impératif de se connecter. L'utilisateur doit entrer son login et son mot de passe sur la page de connexion du site web.

- SPC-03 : Se déconnecter de LogAnalyse



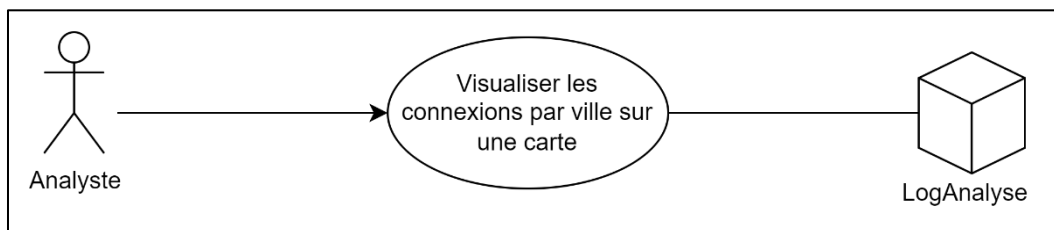


Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

Afin d'assurer la confidentialité du compte et pouvoir permettre à un autre utilisateur de se connecter sur la même machine l'utilisateur en cours doit pouvoir se déconnecter.

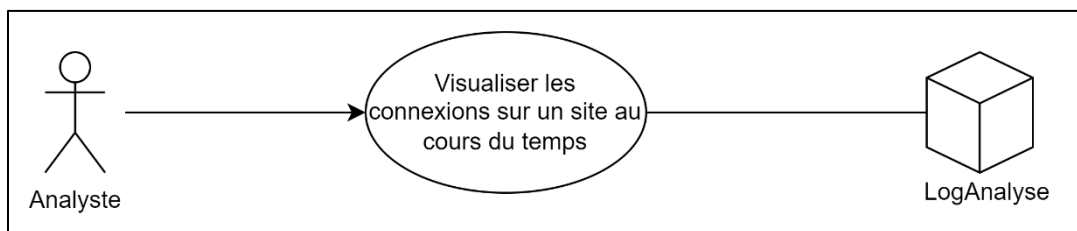
Pour pouvoir sécuriser les navigations sur le site l'utilisateur doit pouvoir se déconnecter du site.

- SPC-04 : Visualiser les connexions par ville sur une carte



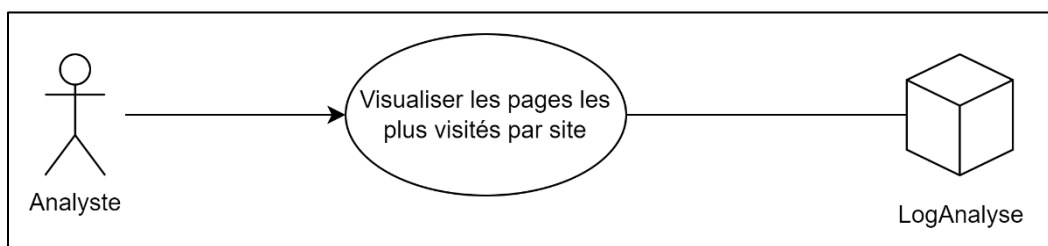
L'utilisateur doit pouvoir visualiser sur une carte les différentes connexions à un site web donné. Avant cela il faut que celui-ci soit connecté au site web.

- SPC-05 : Visualiser les connexions sur un site au cours du temps



L'utilisateur doit pouvoir visualiser à l'aide d'un graphique l'évolution des connexions au cours du temps. Avant cela il faut que celui-ci soit connecté au site web.

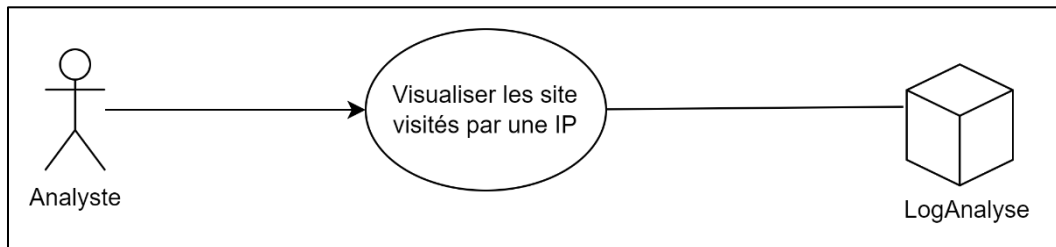
- SPC-06 : Visualiser les pages les plus visitées par site



Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

L'utilisateur doit pouvoir visualiser à l'aide d'un graphique les pages les plus consultés sur un site web donné. Avant cela il faut que celui-ci soit connecté au site web.

- SPC-07 : Visualiser les sites visités par une IP



L'utilisateur doit pouvoir visualiser sur un tableau les différentes connexions à un site web donné par une adresse IP. Avant cela il faut que celui-ci soit connecté au site web.

## 4. Conception

### 4.1. Arborescence et organisation du code

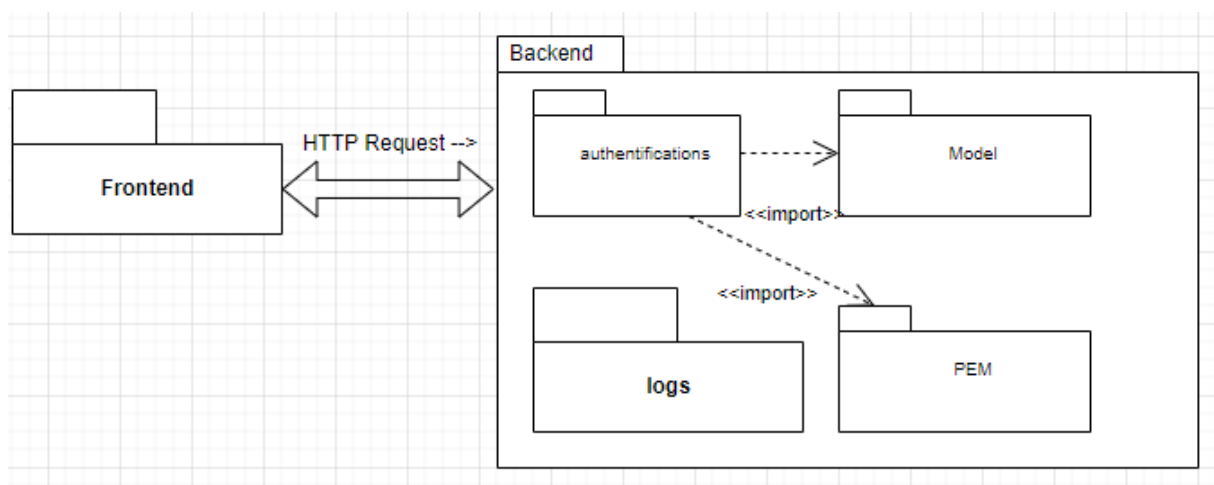
L'organisation du code a été pensée pour permettre, d'une part une meilleure organisation du travail d'équipe sur un même projet et d'une autre part une intégration cloud.

L'organisation et l'arborescence de notre projet se présente sous la forme suivante :

- Un Package Frontend :  
Contenant les ressources nécessaires au bon fonctionnement d'Angular. (HTML/CSS/TS). Le code à l'intérieur de ce package est séparé en Components / Services et chaque component est séparé initialement par Angular en (.html, .scss et .ts).
- Un Package Backend :  
Contenant dans l'ensemble deux applications Python Flask permettant de faire tourner deux micro services (API REST).  
Le premier, « backend.py » (à la racine du package) sert de liaison avec le frontend/Angular, il contient des fonctions de type 'Controller' qui font la liaison avec le frontend en récupérant les requêtes http et en lançant les fonctions de type 'Model' qui permettent d'aller récupérer sous la forme voulue la donnée utilisée pour le projet. Cette application fait aussi la liaison avec la deuxième application pour la gestion des utilisateurs. Le deuxième script se trouve dans le package backend.authentications.

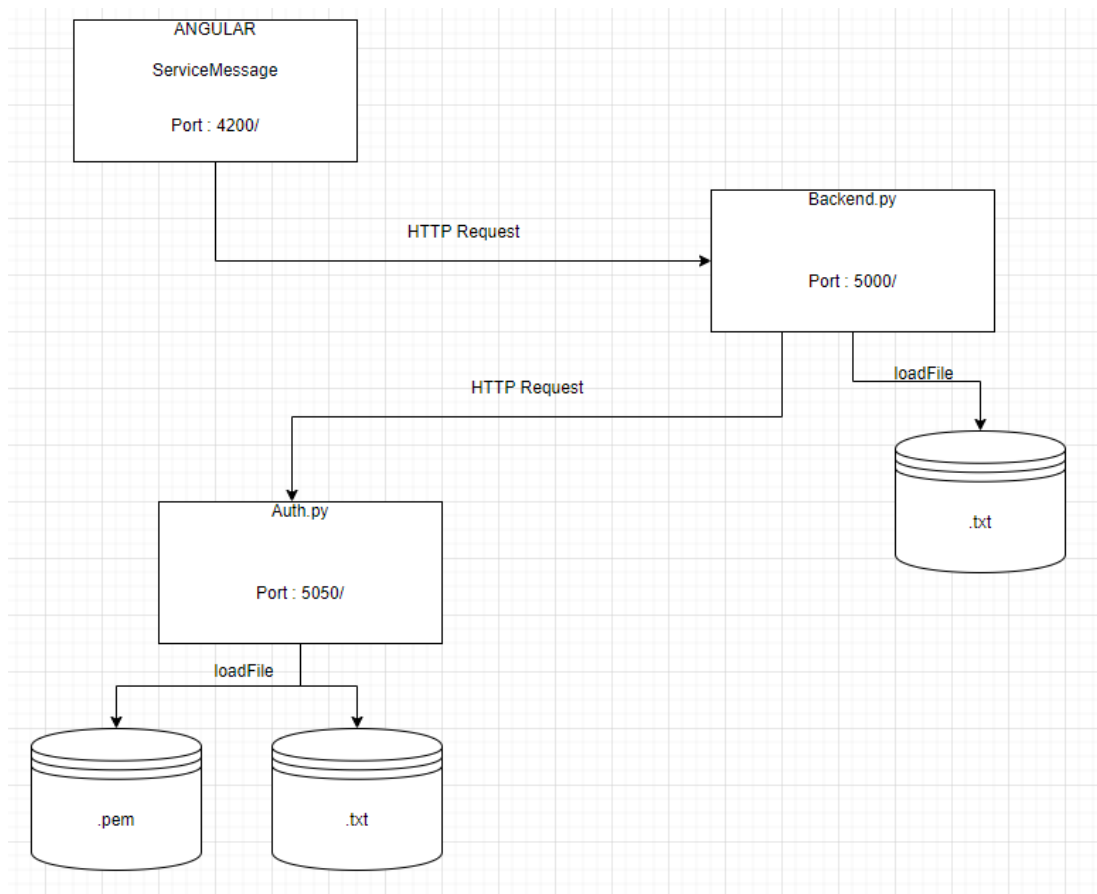
Réf : RP-2.1	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges Client : Polytech Projet : Géocalized Web Usage Mining		

- Un Package Backend.logs  
Ce package contient seulement un fichier texte contenant toute notre donnée utilisée. Ce fichier de logs à été travaillé et nettoyé au préalable par un autre script python indépendant.  
On y retrouve les informations suivantes :
  - Date
  - Heure
  - Page visité
  - IP
  - Site visité
  - Statut
  - Transfère de Bytes
- Un Package Backend.authentications  
Ce package contient la deuxième application Python Flask « auth.py » qui sert à gérer l'authentification des utilisateurs. Il comprend comme le premier, deux types de fonctions. Celles qui récupèrent les requêtes http venant du « backend.py » et celles qui font l'accès à la donnée utilisateur contenue aussi dans un fichier texte.
- Un Package Backend.authentications.Model  
Qui contient la donnée utilisateur (Login, Mdp) dans un dossier texte crypté par un algorithme RSA. Les clefs sont générées par l'app auth.py (ne sont pas régénérées tous les x temps pour l'instant).
- Un Package Backend.authentications.PEM  
Qui contient les clefs publics et privées, stockées au format 'pem', permettant l'utilisation de l'algorithme de cryptage RSA



Réf : RP-2.1	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges		
Client : Polytech		
Projet : Géocalized Web Usage Mining		

## 4.2. Mapping entre les différents micro services



### - Fonction du Backend

- Routing FLASK / Récupération des requêtes http

Les applications flask permettent de récupérer les requêtes http, émises sur l'url de l'application en cours, grâce au code suivant :

```
@app.route("/hello",methods=['GET','POST'])
def hello():
```

Il suffit de préciser le lien « / ... » dans les paramètres de route pour pouvoir récupérer une requête 'GET' ou 'POST'.

Par exemple si on fait un post à l'url : 'http://localhost :5000/hello' alors l'application flask le récupérera et lancera la fonction hello.

Toutes nos applications Flask retournent du JSon et une réponse Jsonify avec comme headers :

Réf : RP-2.1	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges		
Client : Polytech Projet : Géocalized Web Usage Mining		

```
response.headers.add('Access-Control-Allow-Origin',myUrl)
response.headers.add('Access-Control-Allow-Credentials', 'true')
response.headers.add('Access-Control-Allow-Headers', 'Content-Type')
response.headers.add('Content-type','application/json')
response.headers.add('charset','utf8')
```

- Requête Model et Logs
  - Le backend.py va s'occuper d'aller récupérer les logs situé dans un fichier texte et nettoyer au préalable.  
Le tout est récupéré sous forme de DataFrame, simple d'utilisation pour le traitement qui doit suivre.

```
def getTAB():
    url = "./logs/dataset.txt"
    full_df2=pd.read_csv(url, sep=" ", encoding="utf-8")
    full_df2.columns =['Date', 'Heure', 'ConsultedPage', 'IP','VisitedSite',
'StatusCode','DataBytes']
    df2 = full_df2.head(1000)
    return df2
```

- Quant à lui le auth.py va s'occuper de récupérer les logins et mots de passe situé dans un fichier texte aussi.  
Ces données d'utilisateur son crypté par RSA, à chaque récupération on décrypte et recrypte ensuite la donnée.  
Les clefs utilisées pour l'instant sont générées une première fois puis stockées dans des fichiers au format PEM. Il serait bien d'avoir une fonction permettant de régénérer ces clefs tous les laps de temps prédéfinis.

- Authentification

Dans toutes les fonctions de mapping, du backend.py, nécessitant une connexion d'un utilisateur on retrouve un appel de la fonction « isAuthenticate » qui récupère le loginCookie passé en paramètre de chaque post du front vers le back.

```
def getRequestLoginCookie(request):
    try:
        if(request.method == "GET"):
            login = request.args.get("loginCookie")
```

Réf : RP-2.1	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges		
Client : Polytech Projet : Géocalized Web Usage Mining		

```

else:
    login = request.form.get('loginCookie')
    return login
except:
    return ""

```

Une fois récupéré il est comparé avec le login de l'utilisateur actuellement connecté, si ce dernier est le même alors la fonction initialement appelée peut s'effectuer. Sinon aucune donnée n'est renvoyée.

```

@app.route("/json",methods=['POST','GET'])
def visualisation_rooting():
    login = getRequestLoginCookie(request)
    @isAuthenticate #SI VRAI ON LANCE LA FONCTION, SINON NON
    def visualisation(login):#FONCTION RECUPERANT DE LA DONNEE

```

- Fonction du Frontend
  - o Message Service

Ce service va permettre d'avoir qu'une seule fonction générique d'envoi de post vers le backend.py.

Il suffit de passer le service dans le constructeur des classes devant effectuer une requête au back et ainsi elles pourront utiliser la fonction 'sendMessage(url,body :{})' Dans laquelle on passe l'url (/...) à appeler dans le back ainsi que les informations à faire passer sous forme de json. Chaque fonction s'occupe aussi de faire passer le cookie de login afin de vérifier si l'utilisateur est bien connecté.

### 4.3. Gestion de la donnée

- Gestion de la donnée coté backend

La donnée du backend est gérée par une panoplie de fonctions qui prennent en paramètre un dataframe et qui appliquent des modifications particulières dessus. Une fois la fonction « getTab() » appelé, on se retrouve avec toutes les logs sous forme de dataframe, qu'on pourra ensuite passer aux fonctions de gestion pour avoir des informations plus précises.

Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

Voici une liste d'exemple de quelques fonctions principales qui s'occupent d'affiner cette donnée :

Nom	Paramètres	Retour	Description
searchSiteAffluence	Url :string	DataFrame	Retravaille les dataframes pour retourner seulement : la date / l'Heure / l'IP / Le site visite / Le mois
ajoutMois	DataFrame	DataFrame	Rajoute une colonne avec le numéro du mois
groupByMois	DataFrame	DataFrame	Groupe les dataframes par mois et ajoute un compteur
ajoutHeurs	DataFrame	DataFrame	Rajoute une colonne avec l'heure
groupByHeurs	DataFrame	DataFrame	Groupe les dataframes par heure et ajoute un compteur

- Gestion de la donnée coté frontend
  - Du côté du frontend la donnée est gérée par différents services et composants. Chaque composant s'occupe de récupérer et d'utiliser la donnée qu'il utilise sur sa page html, par contre pour faire le lien avec le backend, ils utilisent tous un service générique « messageService » et sa fonction sendMessage qui s'occupe de transmettre la requête avec un POST. Pour ce qui est de la donnée concernant les logins, le service cookieService s'occupe de sauvegarder le login actuel et le service « AthGuardService » et sa fonction canActivate appelée lors de chaque mapping.

Réf : RP-2.1	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges		
Client : Polytech		
Projet : Géocalized Web Usage Mining		

Nom	Paramètres	Retour	Description
sendMessage	Url : string data : any	Observable<DataFormat>	S'occupe d'envoyer un post à l'url stocké dans les variables d'environnement plus le lien passé en paramètre. En passant dans le body le data.
canActivate	next: ActivatedRouteSnapshot, state: RouterStateSnapshot	boolean	Vérifie si le cookie loginCookie existe bien.
searchSite	url:string recherche:string cpt:number	void	Récupère la data du back en fonction du type de recherche voulue. Et remplit la donnée du graph avec celle obtenue.

#### ○ Gestion de l'authentification

Comme précisé ci-dessus, lors de chaque appel de la fonction 'sendMessage' on lui passe le cookie("loginCookie") ce qui nous permet de lancer la fonction une fois authentifié. Mais cela concerne le back. Au niveau d'Angular, on utilise un service de redirection permettant la redirection vers la page login si l'utilisateur essaye d'accéder à toute autre page nécessitant l'authentification.

Ce service, authService possède une fonction canActivate qui vérifie si le cookie de loginCookie existe bien. Sinon on est redirigé vers la page login.

Cette fonction est appelée à chaque changement de route de la façon suivante :

```
{path: 'affluence', component: AffluenceComponent, canActivate: [AuthGuardService]}
```

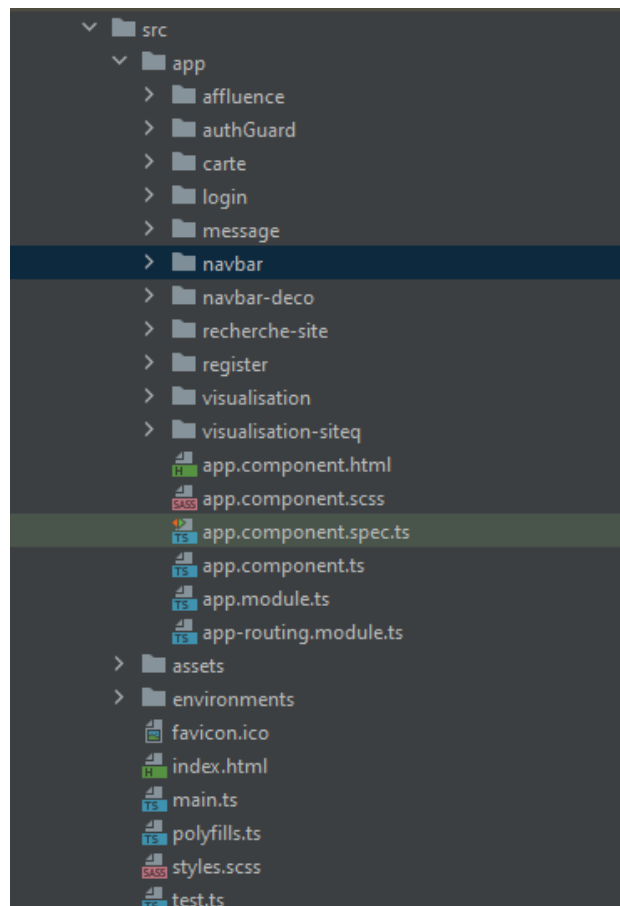
Par exemple ici, lorsqu'on essayera d'accéder à l'url « /affluence » on lancera avant la fonction canActivate d'AuthGuardService.



Réf : RP-2.1	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges		
Client : Polytech Projet : Géocalized Web Usage Mining		

- Fonctions de gestion d’affichage

Angular permet de gérer la liaison entre l’affichage (html/css) et le Type Script à merveille. Comme vu précédemment, tout est séparé en « component » On a donc séparé nos différentes pages et élément html en composant comme ci-dessous.



On utilisera aussi par la suite, les fonctions d’Angular permettant d’utiliser des boucles ou de créer des conditions à l’intérieur du HTML.

On utilisera notamment le FormControl, pour récupérer et gérer les données de nos barres de recherche, de même qu’un matAutocomplete pour l’autocomplétassions de nos url.

Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

## 5. Réalisation

Login :

Visualisation d'une partie de la donnée :

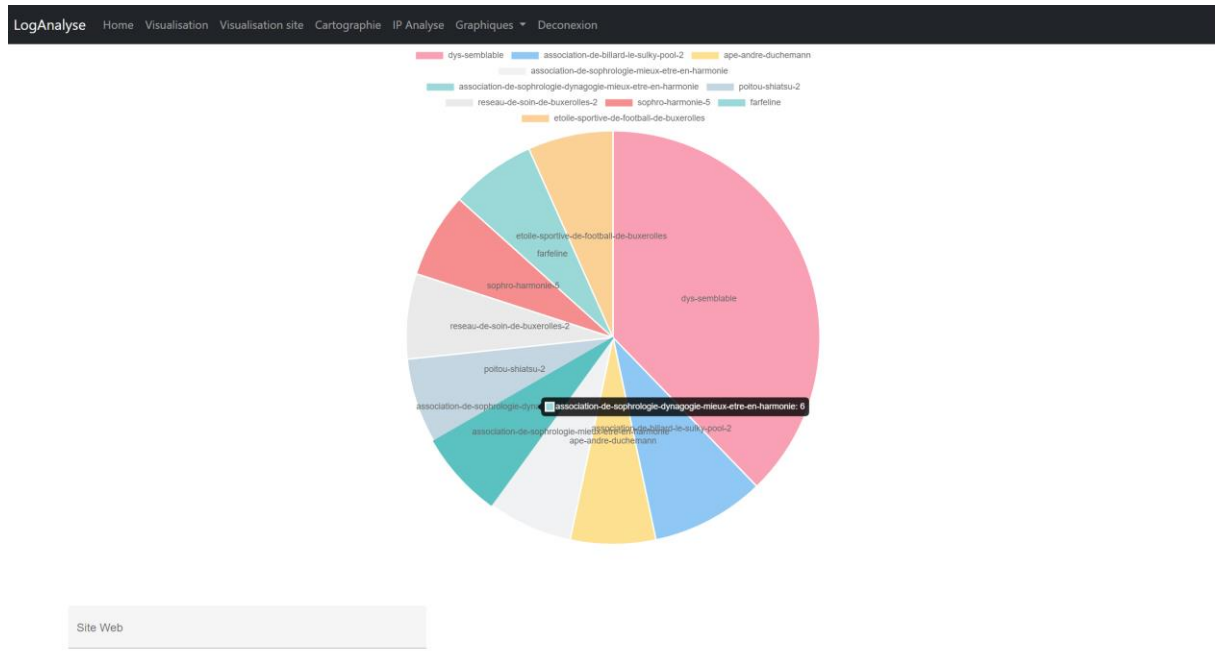
Date	Heure	Consulted Page	IP	Visited Site	Status Code	Data(Bytes)
2015-11-16	22:34:49	usep-pinchinades	68.180.229.29	associations.paysdelaixassociations.org	500	12562
2015-11-16	22:35:03	association-regionale-pour-le-developpement-de-l-enseigne-3	68.180.228.251	agneaux-50.ville.mygaloo.fr	301	171
2015-11-16	22:35:05	association-regionale-pour-le-developpement-de-l-enseigne-3	68.180.228.251	agneaux-50.ville.mygaloo.fr	500	1968
2015-11-16	22:35:14	asclair	68.180.230.237	paris-75.ville.mygaloo.fr	301	156
2015-11-16	22:35:16	asclair	68.180.230.237	paris-75.ville.mygaloo.fr	500	1968

Items per page: 5 1 - 5 of 1000 |< < > >|

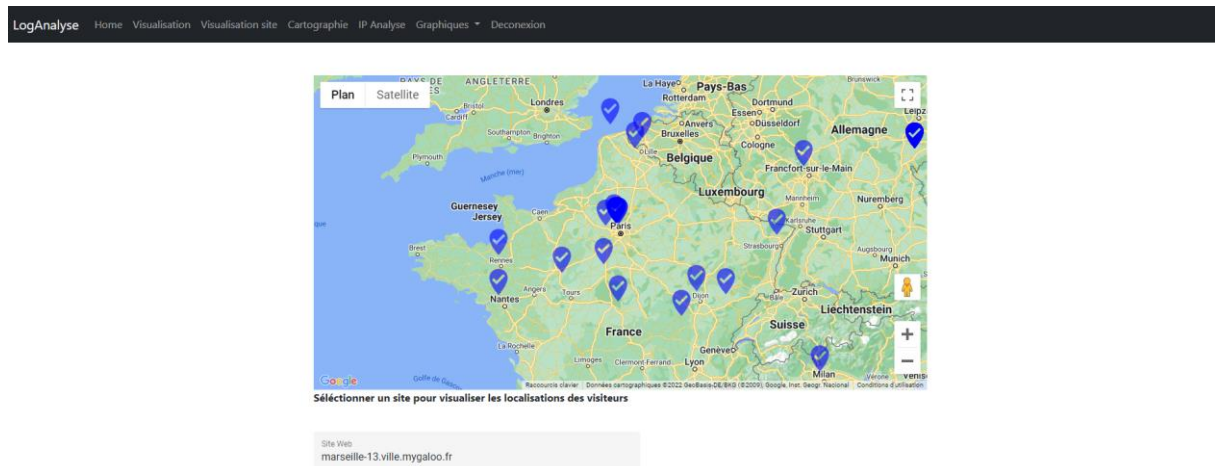
localhost:4200/visualisation/site

Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges  Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

## Visualisation des connexions sur un site

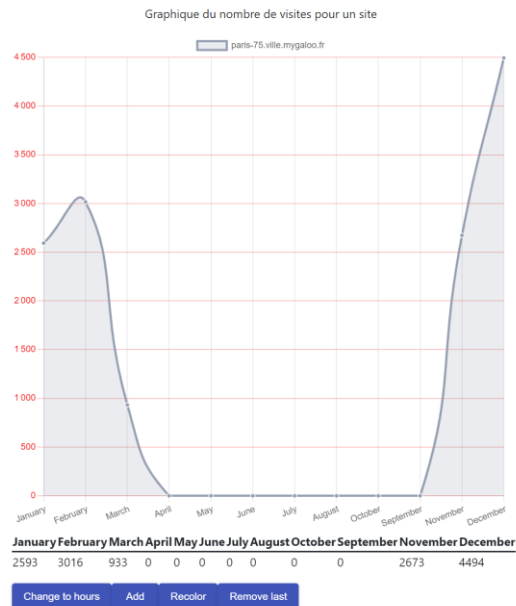


## Visualisation de la localisation des adresses IP ayant accédé à un site donné :



Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges  Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

## Visualisation de l'affluence pour un site donné



Site Web

## Visualisation des sites visité par une IP donnée :



Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

## 6. Tests

Afin de s'assurer du bon fonctionnement du site web il est nécessaire de préparer une batterie de test. Voici une liste non-exhaustive des tests nominatif à exécuter :

Test FCT-01	S'inscrire sur LogAnalyse	
Description	Vérifier que les données de l'utilisateur sont bien inscrit dans la base de données	
Etape	Action	Résultat attendu
1	Cliquer sur « sign up »	La page d'inscription s'ouvre
2	Entrer le login, le mot de passe et la confirmation du mot de passe	Les trois champs sont remplis
3	Cliquer sur « Register »	Un message est affiché dans la console avec le login de l'utilisateur

Test FCT-02	Se connecter à LogAnalyse	
Description	S'assurer que l'utilisateur inscrit puisse se connecter à LogAnalyse avec un compte déjà enregistré dans la base de donnée.	
Etape	Action	Résultat attendu
1	Cliquer sur « sign in »	La page de connexion s'ouvre
2	Entrer le login et le mot de passe	Les deux champs sont remplis
3	Cliquer sur « Connect »	La page du menu est ouverte avec un message « welcome + login »

Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

Test FCT-03	Se déconnecter de LogAnalyse	
Description	Vérifier que cliquer sur le bouton déconnexion déconnecte bel et bien le compte anciennement connecté et que les fonctionnalités ne sont plus accessibles.	
Etape	Action	Résultat attendu
1	Cliquer sur « déconnexion » dans la barre de navigation.	Redirection vers la page de connexion
2	Cliquer sur « visualisation ».	Redirection vers la page de connexion (l'accès est refusé car l'utilisateur n'est pas connecté)

Test FCT-04	Visualiser les connexions sur un site par ville sur une carte	
Description	S'assurer que l'on puisse bien voir les connexions à un site web donné sur la carte	
Etape	Action	Résultat attendu
1	Cliquer sur « Carte ».	La page de visualisation des données sur la carte s'ouvre
2	Entrer l'URL « associations.paysdaixassociations.org » sur la zone de texte dédié.	Le champs est remplis.
3	Cliquer sur « entrer »	La carte affiche les connexions grâce à des cluster et des marqueurs.

Test FCT-05	Visualiser les connexions sur un site au cours du temps	
Description	S'assurer que l'on puisse bien voir les connexions à un site web par périodes sur le graphique	
Etape	Action	Résultat attendu
1	Cliquer sur « Affluences ».	La page de visualisation des données par périodes s'ouvre
2	Entrer l'URL « associations.paysdaixassociations.org » sur la zone de texte dédié.	Le champs est remplis.
3	Cliquer sur « entrer »	Le graphique des visites par site s'affiche
Test	Visualiser les pages les plus visités par site	

Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

FCT-06		
	S'assurer que l'on puisse bien voir les pages les plus consultées par site.	
Etape	Action	Résultat attendu
1	Cliquer sur « Visualisation site ».	La page de visualisation consultations de page par site s'ouvre
2	Entrer l'URL « associations.paysdaixassociations.org » sur la zone de texte dédié.	Le champs est remplis.
3	Cliquer sur « entrer »	Le graphique des visites par page s'affiche

Test FCT-07	Visualiser les site visités par une IP	
	S'assurer que l'on puisse bien voir les connexions à un site web donné sur la carte	
Etape	Action	Résultat attendu
1	Cliquer sur « Consultation IP »	La page de visualisation consultations de page par site s'ouvre
2	Entrer l'IP « 208.115.111.66 » sur la zone de texte dédié.	Le champs est remplis.
3	Cliquer sur « entrer »	Le graphique des visites par IP s'affiche

Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges  Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

## 7. Organisation

Pour la méthodologie de travail, nous nous sommes basés sur la méthodes AGILE avec des réunions régulières et des périodes de codage intensif ou le but était de remplir les objectifs à court terme et à moyen terme. Nous travaillons via Discord soit pour des réunions afin de fixer les objectifs. Nous programmions aussi des séances de codes, ou chacun avançait sur sa partie et en cas de besoin consulter les autres membres du groupe qui l'aidait à résoudre la problématique.

Nous avons aussi ouvert un Drive afin de stocker l'ancien projet et d'autre document comme le rapport. Ce Drive était partagé par tous les membres du projet.

Afin d'organiser les grandes étapes du projet et pour n'oublier aucune action importante, nous avons fait le choix au début de prendre la suite Trello. Nous n'avons pas réussi à utiliser son plein potentiel et l'utilisation de fichier texte aura été plus efficace.

Concernant le découpage du travail, nous n'avons pas fait le choix de faire front/back mais plutôt de répartir en fonction des fonctionnalités à implémenter. Cela nous a permis de ne pas trop nous emmêler et de développer de manier indépendante.

Nous avons fait le choix de répartir le code sur la base du modèle MVC expliquer sur le paragraphe conception . Ainsi les Views peuvent être modifiées au bon vouloir et nous avons bien pris soins de permettre l'adaptation du code en fonction du contexte.

Pour garder un visuel et rester motivé, nous avons défini des deadlines pour les étapes importantes. Ainsi nous avons programmé :

- D'avoir un visuel sur les objectifs de notre projet : 12/04
- D'avoir une première base de code avec les fonctionnalités opérationnelles pour 24/04
- D'avoir une première version du rapport : 17/05
- De finaliser le code pour 20/05
- De finaliser le rapport pour 22/05
- De concevoir la Slide pour : 23/05
- De faire un essaie de notre présentation pour 23/05



Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

## 8. Conclusion

Notre projet a pour but de démontrer ce qu'il est possible de faire avec de la donnée récupérée des logs d'un serveur. Ce n'est qu'un POC (Proof of concept) qui laisse entrevoir les possibilités. Nous avons montré qu'il est possible d'afficher certaines informations sur une carte, comme la géolocalisation des adresses ip s'étant connecté à un certain site web. D'analyser et d'afficher sous forme de graphiques des moyennes ou sommes de connexion vers un site web ou une page en fonction d'une date / heure donnée. Mais les possibilités restent encore très large et avec les technologies et l'organisation que nous avons utilisée, il est très facile de compléter le projet.

Que ce soit en termes de design ou de rajout d'information. Nous avons par exemple pensé à rajouter un modèle de machine learning qu'on entrainerait en amont et qu'on pourrait importer dans notre backend pour prédire des pourcentages de connexions en fonction d'une heure de la journée ou bien d'autre chose. Nous avons aussi pensé à rajouter d'autre logs sous un certain format prédéfinis. Pour que chaque utilisateur puisse accéder à sa propre donnée. Nous avons, grâce au projet parallèle de Cloud Computing, implémenter la gestion de l'authentification, ce qui permettrait de faire ces choses-là. Malheureusement nous avons fortement manqué de temps pour implémenter toutes ces fonctionnalités.

Ce projet nous a fortement appris à travailler en équipe, à utiliser des logiciels pour être plus performant et mieux organisé en équipe. De plus il nous a fait découvrir une large panoplie de nouvelles technologies.

Réf : RP-2.1  Emetteurs : BARAR Bilal MONIMEAU Paul MEYA Georges Client : Polytech Projet : Géocalized Web Usage Mining	Projet Geocalized Web Usage Mining Rapport de Projet	Date: 20 mai 2022 Version : 2.1 Service : Ecole Polytech
---	---	--

## 9. Terminologie

- API : Acronyme pour Application programming interface, c'est une solution informatique qui permet à des applications de communiquer entre elles et de s'échanger mutuellement des services ou des données.
- Framework : ensemble d'outils et de composants logiciels à la base d'un logiciel ou d'une application.
- Backend : programmation au niveau du serveur, pour exécuter les requêtes qui sont réalisées sur le site web par les utilisateurs.
- Frontend : fait référence à l'ensemble des éléments visibles et accessibles directement sur un site web.

## 10. [Repository GITHUB](#)

[Lien de la repo](#)