

UNIVERSITE ABDELMALEK ESSAADI

FACULTE DES SCIENCES ET TECHNIQUES DE TANGER



PROJET DE FIN D'ETUDES

Licence Génie Informatique

Sujet :

Vers le développement d'une plateforme web,
référentiel de vente de véhicules basé sur le scraping

ML View Consulting

Réalisé PAR :

- HANAE EL M'RABET
- TASNIME MORSE

Sous l'encadrement de :

- Mr. Pr. Abdelhadi FENNAN
- EL ATTABI Mohamed Amine

Soutenu le : 22/06/2022

Année Universitaire : 2021/2022

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

فَدَلَّاهُ عَلَى الْإِسْلَامِ وَكَانَ أَتَمَّ الْمُسْلِمِينَ

صَدَقَ اللَّهُ الْعَظِيمَ

Dédicaces

A nos parents :

Aucune dédicace ne peut exprimer l'amour, l'estime, et le respect qu'on a toujours eu par vous.

Rien au monde ne vaut les efforts fournis jour et nuit pour notre éducation et notre bien-être.

Vos prières et vos bénédictions ont été pour nous un grand secours pour mener à bien nos études.

On vous dédie ce travail en témoignage de nos profonds amours.

Puisse dieu, le tout puissant, vous préserver et vous accorder santé, longue vie et bonheur.

A nos frère et sœurs :

Qui ont partagé avec nous tous les moments d'émotion lors de la réalisation de ce travail, Ils nous ont chaleureusement supporté et encouragé tout au long de notre parcours, Que dieu les protège et leurs offres la chance et le bonheur.

A nos amis :

A notre ami Mohammed ELBAKKALI ELAMMARI qui nous a aidé dans notre projet.

Et à tous nos amis qui n'ont pas cessée de nous encourager et soutenir tout au long de notre vie.

Et finalement sans oublier mon binôme, pour son soutien moral, sa patience et sa compréhension tout au long de ce projet.

Remerciements

Avant tout développement sur notre expérience professionnel, on offre ces lignes de gratitude à tous ce qui ont contribué de près et de loin à l'élaboration et la réussite de ce travail.

Nous remercions tous d'abord Allah pour la volonté et la patience qui nous a donné pour réaliser ce projet et durant les longues années d'études fréquentes.

Avec un immense plaisir nous exprimons notre profonde gratitude à notre encadrant M. Abdelhadi FENNAN, pour son encouragement et ses précieux conseils tout au long de notre projet de fin d'études. Nous adressons ainsi nos vifs remerciements à notre encadrante professionnelle Mlle. Soumaya LOUKILI, qui a gardé un œil attentif sur le déroulement de ce projet en donnant des remarques constructives, et pour l'intérêt particulière qu'elle a accordé à notre stage malgré leurs préoccupations.

Notre remerciement s'adresse aussi à monsieur Ihab MOUDHICH et Mlle Ihsane OUACHEIKH, pour leurs conseils, et leurs aides précieuses.

Nous tenons à remercier vivement tout le personnel de MI View Consulting ayant contribué à la réalisation de ce présent travail.

ET on adresse nos fidèles remerciements au département génie informatique, et à tous les enseignants de la FST de Tanger, et tout le corps d'administratif et professoral pour leur efforts considérables. Que les membre de jury d'avoir acceptés de consacrer leurs précieux temps à l'évaluation et le jugement de ce travail.

Résumé

Le travail proposé à travers ce rapport est de répondre à une problématique bien précise, c'est l'extraction des données existant en utilisant une stratégie automatique et efficace, afin de savoir visualiser ces données via une application développée nommé AutoMaroc.

En effet, notre challenge c'est d'implémenter une application web qui puisse agréger plusieurs annonces de différentes sites web du véhicule tel que Avito, moteur, Annonce... et aussi permet aux utilisateurs de partager leurs annonces au sein de notre site. AutoMaroc est une application qui collecte les informations du plusieurs sites et les affiche, Ce qui en fait un référentiel de produits grace à un programme minutieux pour garantir le bon traitement de données.

Abstract

The work proposed in this report of license is to answer a very precise problem, is the extraction of existing data using an automatic and effective strategy, in order to be able to visualize these data through an application named “Auto Maroc”.

As a matter of fact, our challenge is to implement a search engine that draws several ads from different websites of the vehicles such as Avito, Moteur, Annonce ... and allows users to share their ads within our site. “Auto Maroc” is an application that collects information’s from several sites and displays it, and this answer that the customer can find what he is looking for in a quick way. By dint of a meticulous program to ensure the correct processing data. Alternatively, in other technical terms is to use web scraping

Avant-propos

**Nom et Prénom de
l'Etudiant Stagiaire :**

Hanae El M'rabet

Tasnime Morsse

Intitulé du sujet :

**Application de la gestion des
ventes des véhicules**

Etablissement d'accueil :

ML View Consulting

Encadrant Pédagogique :

- Mr. Pr. Abdelhadi Fennan

Encadrant Technique de

Stage :

-Mr. EL ATTABI Mohamed Amine

Période de projet :

Du 18 avril 2022 au 18 juin

2022

Cadre du Stage :

*Projet de Fin d'Etudes présenté en vue de
l'obtention d'une Licence en Sciences et
Techniques en génie informatique.*

LISTE DES FIGURES

| | |
|--|----|
| FIGURE 1 : DIAGRAMME DE GANTT | 16 |
| FIGURE 2 : DIAGRAMME DES CAS D'UTILISATION | 20 |
| FIGURE 3 : DIAGRAMME DE CLASSE..... | 21 |
| FIGURE 4 : DIAGRAMME DE SEQUENCES D'AUTHENTIFICATION | 22 |
| FIGURE 5 : DIAGRAMME DE SEQUENCES SUPPRESSION D'ANNONCES. | 22 |
| FIGURE 6 : DIAGRAMME DE SEQUENCES PARTAGE D'ANNONCE | 23 |
| FIGURE 7 : DIAGRAMME D'ACTIVITE 1..... | 23 |
| FIGURE 9 : DIAGRAMME D'ACTIVITE 2..... | 24 |
| FIGURE 9 : DIAGRAMME D'ACTIVITE 3..... | 25 |
| FIGURE 10 : SCHEMA DU SCRAPING | 27 |
| FIGURE 11 : SCHEMA DU SCRAPING ET CRAWLING..... | 28 |
| FIGURE 12 : SCHEMA DE L'ACHITETCURE SCRAPY | 29 |
| FIGURE 13 : GUIDE D'ANALYSE HTML AVEC BEAUTIFULSOUP EN PYTHON | 30 |
| FIGURE 14 : ILLUSTR FRAMWORK SELENIM DE CES DIFFERENTS ETAPES | 31 |
| FIGURE15 : ARCHITECUTRE MVC DE LARAVEL | 34 |
| FIGURE 16 : PRINCIPALE AXES DU MODELE DU DONNEE XPATH | 35 |
| FIGURE 17 : REQUETES GET AND POSTS à L'AIDE DE PYTHON | 35 |
| FIGURE 18 : SCHEMA DE DEVELOPPEMENT | 39 |
| FIGURE 19 : SCHEMA DE SCRAPING | 39 |
| FIGURE 20 : INTERFACE PRINCIPALE DU SITE WEB (interface client) | 40 |
| FIGURE 21 : FILTRAGE PAR MOTO | 40 |
| FIGURE 22 : ACCEDER AU INFORMATIONS DES ANNONCES | 41 |
| FIGURE 23 : INTERFACE D'ENREISTREMENT | 41 |
| FIGURE 24 : INTERFACE DE CONNEXION | 42 |

| | |
|---|----|
| FIGURE 25 : INTERFACE PROFIL..... | 42 |
| FIGURE 26 : AJOUTER ANNONCE | 43 |
| FIGURE 27 : FORMULAIRE | 43 |
| FIGURE 28 : LISTE DES ANNONCES | 44 |
| FIGURE 29 : ENVOYER MESSAGE | 44 |
| FIGURE 30 : LISTE DES FAVORIS | 45 |
| FIGURE 31 : AUTHENTIFICATION D'UN ADMIN | 45 |
| FIGURE 32 : PROFIL ADMIN | 46 |
| FIGURE 33 : GERE LES ANNONCES PARTAGER | 46 |
| FIGURE 34 : GERE LES MEMBRES | 47 |

Table des matières

| | |
|--|----|
| Dédicace | 1 |
| Remerciement..... | 2 |
| Résumé..... | 3 |
| Abstract..... | 4 |
| Introduction générale | 11 |
| Chapitre1 : Contexte générale du projet | 12 |
| 1. Présentation de l'organisme d'accueil | 13 |
| 2. Cadre général du projet | 14 |
| a. Cahier des charges | 14 |
| i. Cadre du projet | 14 |
| ii. Acteurs du projet | 14 |
| iii. Contraintes liées au projet | 14 |
| b. Problématique | 14 |
| a. Objectif du projet | 15 |
| 3. Planification..... | 15 |
| a. Phase de déroulement du projet | 15 |
| b. Diagramme de Gantt | 16 |
| 4. Conclusion..... | 16 |
| Chapitre 2 : Analyse et conception | 17 |
| 1. Spécification des besoins | 18 |
| a. Besoins fonctionnels | 18 |
| b. Besoins Non-fonctionnels | 18 |
| 2. Conception..... | 19 |
| a. Identification des acteurs et des rôles..... | 19 |
| b. Diagramme des cas d'utilisation | 20 |
| c. Diagramme de classe..... | 21 |
| d. Diagramme de séquence | 21 |
| i. Authentification..... | 22 |
| ii. Supprimer annonce | 22 |
| iii. Partager annonce | 23 |
| e. Diagramme d'activité | 23 |
| i. Gestion d'application par un membre | 23 |

| | | |
|--|--|-----------|
| ii. | Gestion d'application par administrateur | 24 |
| iii. | Partager annonce | 25 |
| 3. | Conclusion | 25 |
| Chapitre 3 : Scraping..... | | 26 |
| 1. | Introduction | 27 |
| 2. | Fonctionnement | 27 |
| 3. | Les types de scraping | 27 |
| 4. | Le web scrapping vs le web crawling | 28 |
| 5. | Etude comparative des outils du web scraping | 28 |
| a. | Scrapy | 28 |
| b. | Beautifulsoup | 29 |
| c. | Sélénium..... | 30 |
| b. | Dédution | 31 |
| 6. | Conclusion | 31 |
| Chapitre 4 : Architecture et technologie de développement | | 32 |
| 1. | Technologie de développement..... | 33 |
| a. | Côté Frontend | 33 |
| b. | Cote Backend | 33 |
| c. | Coté Scraping | 34 |
| 2. | Outil de développement | 36 |
| a. | Outil de conception | 36 |
| b. | Base de Données | 36 |
| c. | Visual Studio | 37 |
| 3. | Conclusion | 37 |
| Chapitre 5 : Mise en œuvre de l'application | | 38 |
| 1. | Architecture de développement | 39 |
| 2. | Scraping | 39 |
| 3. | Interface d'accueil | 40 |
| 4. | Partie client | 41 |
| 5. | Partie Administrateur | 45 |
| 6. | Conclusion | 45 |
| Conclusion | | 48 |

Introduction générale

L'informatique a connu une évolution remarquable depuis quelques années, surtout dans le domaine du développement web. Les sites web sont parmi ces évolutions qui permet au plusieurs personnes de chercher et accéder à des contenus selon leur besoin et leur centre d'intérêt. De ce fait les sites devenus riche en données, et pour cela nous avons été amenés à réaliser un projet de fin d'études au sein de l'entreprise Ml View Consulting, pour savoir améliorer les traitements de la recherche, à l'objectif d'effectuer une solution informatique qui facilite le partage de ces données.

Dans cette perspective, on a utilisé le scraping qui consiste à extraire des contenus d'un ou plusieurs site web d'une manière automatique. Cette technique peut être effectuée à l'aide des programmes, outils, ou script spécifique.

En effet, dans le cadre de notre projet nous allons offrir une application qui facilite la gestion des ventes des véhicules.

Au long de ce rapport, on va résumer le cycle de développement du projet en 5 chapitres principaux. Le premier chapitre présente le contexte général du projet, par une présentation générale de la société Ml view Consulting, qui comprend son domaine d'activité, il traitera, en deuxième lieu le cadrage de projet. Le deuxième chapitre est consacré à l'étude conceptuelle, qui consiste à identifier la problématique, et présenter les différents diagrammes de l'UML. Suivi d'un troisième chapitre qui présente la définition et les principes de fonctionnement du scraping.

Le quatrième chapitre sera consacré aux différentes technologies et outils de développement utilisés dans le projet, tel que LARAVEL et VUE.JS.... Quant au cinquième chapitre sera dédié aux plusieurs parties, qui présente l'interface de site web, partie admin et partie user.

Finalement on terminera par une conclusion qui résumera les objectifs atteints, et une perspective.

Chapitre 1

Contexte générale du projet

Dans cette partie, on va présenter la société dans laquelle nous avons effectué notre stage de fin d'études, Ensuite on va présenter le cadrage du projet en déclarant la problématique, l'étude de l'existant et l'objectif.

1. Présentation d'organisme d'accueil

MI View Consulting est une société spécialisée dans le marketing sur Internet, que ça soit dans l'email marketing ou au niveau des réseaux sociaux. Parmi ses activités, on peut citer la gestion des listes des abonnés, l'analyse des campagnes en temps réels, ou encore la création et maintenance des sites web et applications d'e-commerce. Elle se compose des deux départements suivants :

➤ **Département IT :**

Ce département est chargé de veiller au bon fonctionnement du réseau et matériel informatique au sein de la société, ainsi que de développer des applications qui facilitent le travail pour les deux départements. On cite parmi leurs missions :

Développement des applications internes de la société.

Maintenance du matériel informatique et du réseau.

Mise à jour et maintenance des applications pour qu'elles soient compatibles avec les données affluant de l'environnement de l'entreprise, en parallèle des changements que cet environnement subit.

Développement des sites web et applications d'e-commerce, ainsi que des outils de gestion de ces derniers.

➤ **Département marketing :**

L'équipe appartenant à ce département est chargée de la mission principale de l'entreprise : l'envoi des campagnes publicitaires. Pour ce faire, ils ont plusieurs missions :

- Analyse du marché et de la compétition, en termes de niches, produits, créatives, etc.
- Création des campagnes publicitaires sur multiples canaux, ainsi que leur optimisation en temps réel.
- Comparaison des résultats obtenus selon différents paramètres.

2. Cadre général du projet

a. Cahier des charges

i. Cadre du projet :

Ce projet est réalisé dans le cadre du projet de fin d'études, afin d'obtenir le diplôme de licence en génie informatique de la faculté de sciences et techniques de Tanger.

ii. Acteurs du projet :

Maitre d'ouvrage : Le maitre d'ouvrage est la société MI View Consulting.

Maitre d'œuvre : c'est la faculté des sciences et techniques de Tanger, représenté par Hanae ELM'RABET et Tasnime MORSE élèves en licence génie informatique

Acteurs relais : le projet a été réalisé sous l'encadrement de :

- Tuteur pédagogique : M. FENNAN Abdelhadi, professeur à la faculté de sciences et techniques de Tanger.
- Tuteur technique : Mme. Soumaya LOUKILI, ingénieur.

iii. Contraintes liées au projet

La gestion de ce projet doit tenir en compte des contraintes suivant :

Les contraintes pédagogiques :

- Appliquer les technique et méthodes de gestions des projets.
- Apprendre à être autonome dans la réalisation d'un projet.
- Enrichie des nouvelles connaissances informatiques.

Les contraintes temporelles :

- Le projet doit être terminé dans l'intervalle de temps défini entre le 18 Avril 2022 et le 20 Juin 2022.

b. Problématique

Nous avons été confrontés à des situation ou ; on cherche des informations au sein de l'internet, et on trouve beaucoup de source, ce qui fait l'accès à ces données représente un challenge, qui nécessite d'appliquer des méthodes

traditionnelles comme le copier-coller, capture ou accéder aux bases données par des requêtes. Dans ce cadre et autour de cette problématique, pour essayer de pallier ce problème, nous avons essayé de développer une application qui récapitule, et fait une synthèse des informations utiles disponibles sur les sites web.

c. Objectif du projet

Notre projet se focalise sur le développement d'une interface pour les ventes des véhicules, en récapitulant les données des trois sites choisis : Avito, Moteur, Maroc annonces. Une application qui doit gérer le système d'affichage des annonces de ces sites, et aussi permet à l'utilisateur de partager au sein de notre application des annonces. Principalement, on doit évoluer une solution informatique qui permet de faciliter la recherche des véhicules, en économisant le temps d'utilisateur.

3. PLANIFICATION

a. Phase de déroulement du projet

Tout en longue du duré de stage, nous avons pris soin que les taches du projet se fait en parallèles, pour ne pas perdre le temps sans aucune négligence. Notre projet se compose de cinq principales étapes :

La collection des données : cette partie est très essentiel avant de commencer n'importe quel projet. Cela aide de ne pas perdre beaucoup de temps dans le projet, grâce à la bonne documentation.

Analyse et conception : cette partie consiste à l'organisation de travail, et donne une vue globale sur le projet.

Le scraping : cette étape consiste à récapituler les informations qu'on veut afficher dans notre site.

L'apprentissage de laravel et vuejs : dans cette partie, on a enrichi nos connaissances sur des nouveaux Framework, qu'on doit les utiliser au sein de la réalisation de notre application

Développement : on a réalisé la partie backend et la partie frontend de notre application, dans cette partie.

Rédaction du rapport : cette tâche nous a accompagné tout au long de notre travail, il consiste à présenter les différentes étapes qu'on a réalisé.

Le projet a pour délai 72 jours, il commence le 18/04/2022, et il finit le 13/06/2022.

b. Diagramme de Gantt

Ce diagramme permet de répartir les activités et visualiser l'affectation des ressources aux tâches. Donc la conduite de notre projet est comme ci-dessous :

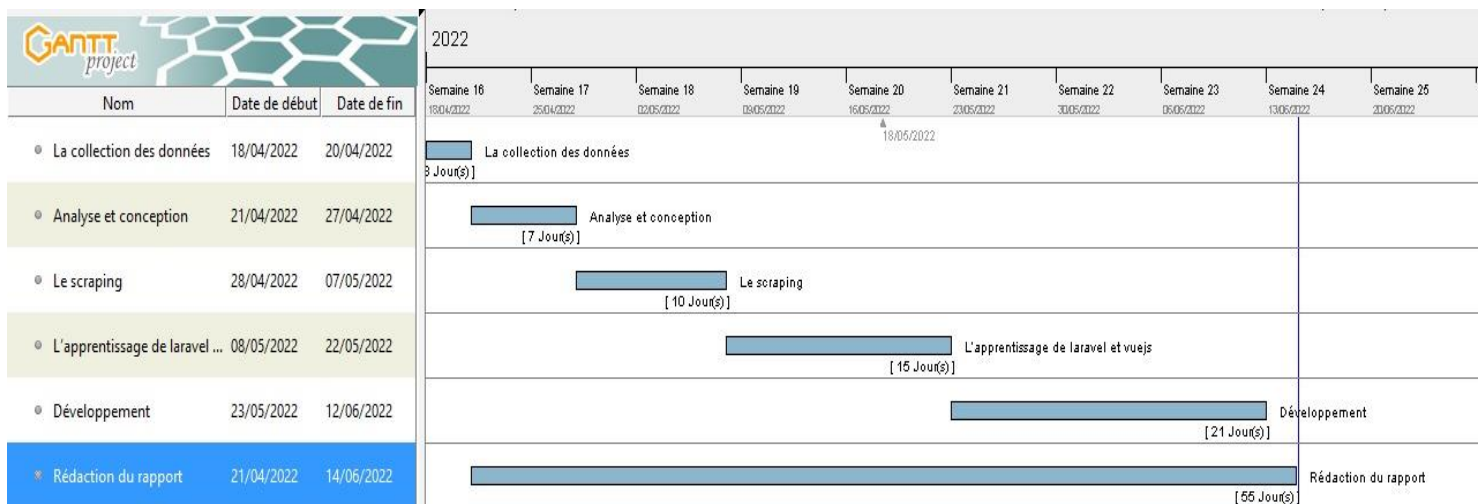


FIGURE 1 : DIAGRAMME DE GANTT

4. Conclusion

Ce chapitre en ayant décrit l'organisme d'accueil d'une manière générale, nous avons abordé le cahier de charge relatif à notre projet de fin d'études, le nous avons parlé du contexte général du projet et les risque qu'on peut trouver.

Dans le chapitre suivant nous décrivons d'une manière non exhaustive les besoins fonctionnels et la partie conception adoptée pour la réalisation du travail et les outils utilisés pour traiter ce sujet

Chapitre 2

Analyse et conception

Ce chapitre est consacré à l'étude des besoins fonctionnels, il présente en premier partie la problématique, ensuite il élabore la phase de conception UML à travers des différents diagrammes.

1. Spécifications des besoins

a. Besoin fonctionnel

Les besoins fonctionnels expriment des conditions réaliser par le système pour répondre à une demande, dans le contexte de la réalisation de notre projet notre application web doivent offrir les services suivants :

Authentification : cette étape permet au visiteur de s'authentifier, pour pouvoir exploiter des autres services de notre site. Il s'agit de saisir un email et un mot de passe.

Contacter le propriétaire : l'utilisateur peut tout simplement contacter un autre utilisateur dans notre site.

Visualiser une annonce : l'application offre la possibilité d'accéder aux annonces par le visiteur soit que dans notre site ou bien dans le site officiel.

Gérer l'application : l'administrateur peuvent gérer le site web à travers plusieurs fonctions, tel que supprimer une annonce, et désactiver un compte d'utilisateur.

Raffiner les recherches : l'application permet aux visiteurs d'utiliser les filtres, pour limiter le contenu selon leur besoin.

Ajouter aux favoris : le système donne aux membres la possibilité de sauvegarder les annonces, pour les revisiter ultérieurement.

Partager des Annonces : le membre peut déposer et publier une annonce au sein de l'application.

b. Besoin non fonctionnel

Ces fonctionnalités sont considérées comme des exigences identifient des contraintes de l'application.

Les besoins non fonctionnels de notre projet se résument comme suivant :

Sécurité : l'application doit être sécurisé au niveau de la protection des données. D'un autre côté, les comptes des utilisateurs sont sécurisés par un mot de passe.

Fiabilité : Bon fonctionnement d'application sans avoir perdre aucun détail.

Performance : l'application doit avoir un temps de réponse rapide en répondant au besoin de manipulateur.

Configuration : la configuration de site ne doit présenter aucune difficulté pour un utilisateur non expert.

Gestion des erreurs : Les erreurs doivent être signalées par des messages d'erreur explicites.

a. Conception

a. Identification des acteurs et des rôles

Un acteur est une entité qui définit le rôle joué par un personne externe, un processus ou une chose qui interagit avec le système.

Les acteurs qui peuvent interagir avec notre application :

- Visiteur
- Membre
- Administrateur

Chaque acteur a un rôle spécial dans le système :

➤ **Administrateur :**

- S'authentifier
- Supprimer les annonces
- Gestion des utilisateurs : désactiver le compte d'un utilisateur.

➤ **Visiteur :**

- Visualiser le contenu du site web
- Utiliser les filtres pour raffiner la recherche

➤ **Membre :**

- S'authentifier

- Ajouter des annonces à ses favoris
- Créer et partager des annonces au sein de l'application.
- Contacter le propriétaire du véhicule par message.

b. Diagramme des cas d'utilisation

Le diagramme de cas d'utilisation est utilisé pour donner une vision globale sur le comportement de système. Dans notre projet, on a besoin d'un diagramme pour l'utilisateur et le visiteur, et un autre pour l'administrateur.

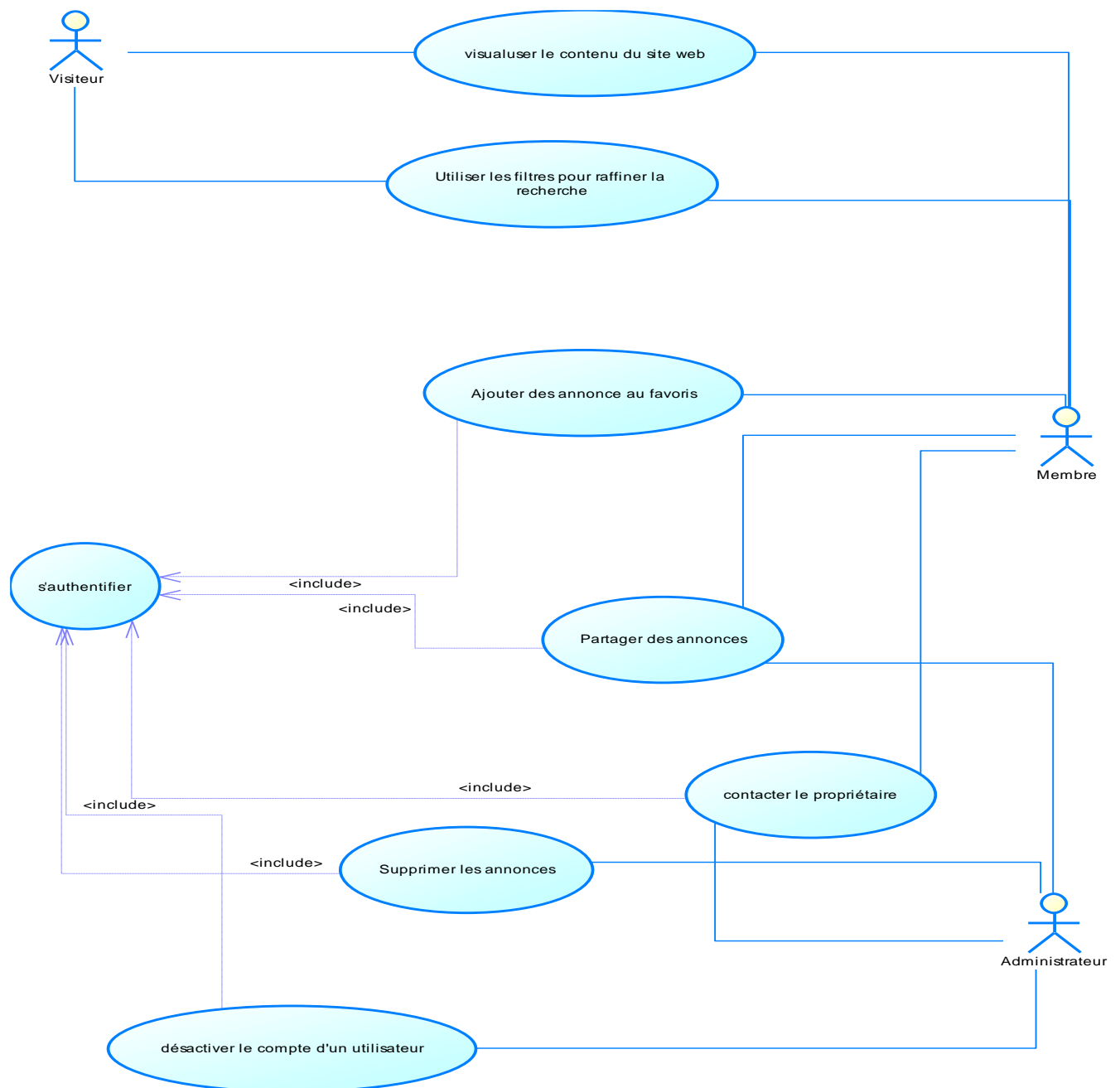


FIGURE 2 : DIAGRAMME DES CAS D'UTILISATION

c. DIAGRAMME DE CLASSE

Représente les différents entités composants le système ainsi que les relations qui relient entre eux.

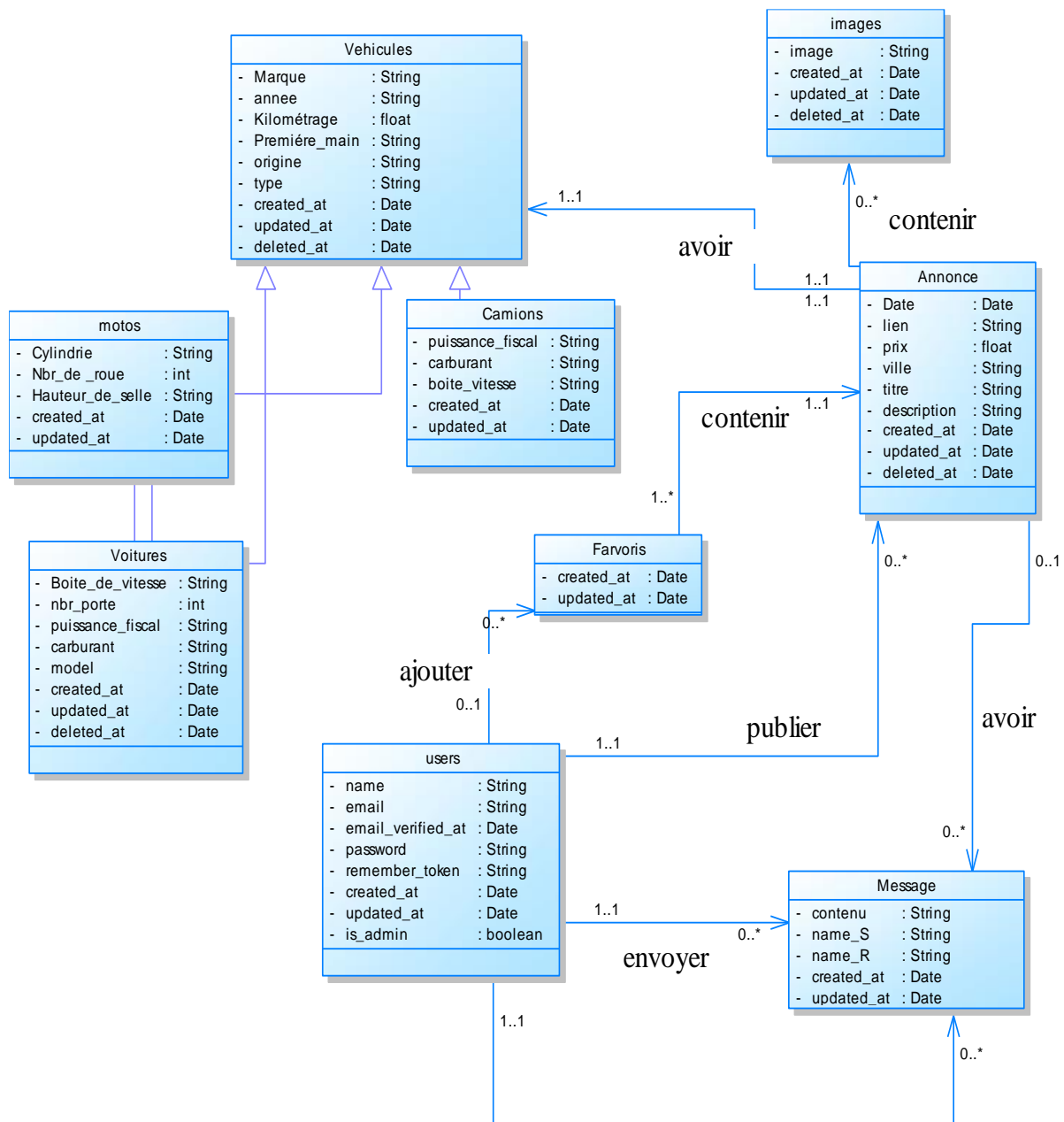


FIGURE 3 : DIAGRAMME DE CLASSE

d. DIAGRAMME DE SEQUENCES

Il montre la participation d'objets (instance de classe ou acteur) dans les interactions et les messages qu'ils échangent lors de l'interaction.

- *Authentication :*



FIGURE 4 : DIAGRAMME DE SEQUENCES D'AUTHENTIFICATION

○ *Supprimer annonce :*

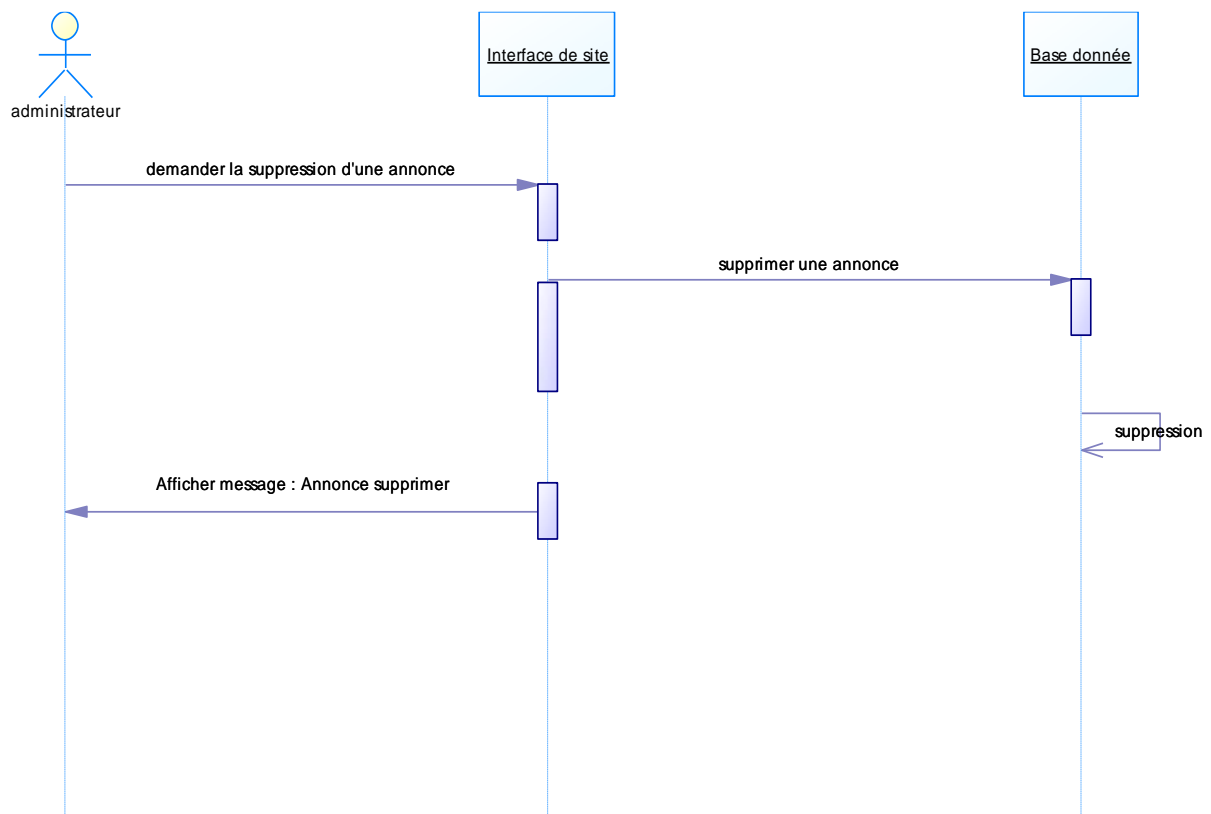


FIGURE 5 : DIAGRAMME DE SEQUENCES SUPPRESSION D'ANNONCES

○ **PARTAGER ANNONCE :**

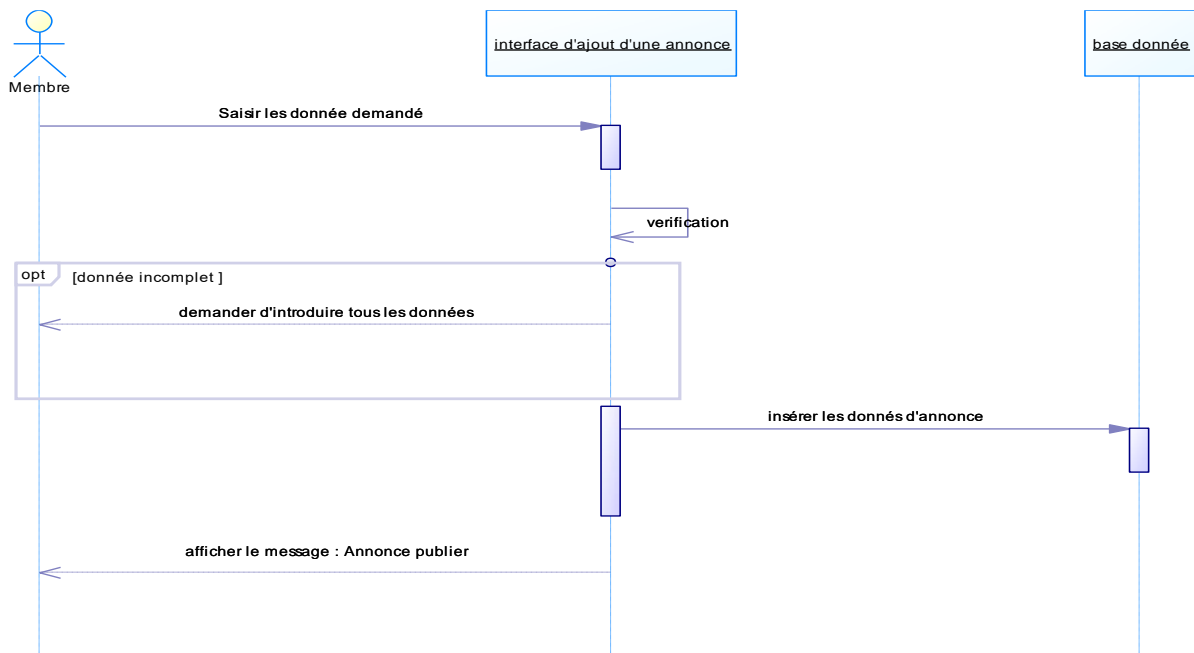


FIGURE 6 : DIAGRAMME DE SEQUENCES PARTAGE D'ANNONCE
e. **DIAGRAMME D'ACTIVITE**

Tous simplement, ce diagramme est utilisé pour décrire un flux de travail, et présenter le déclenchement d'événement en fonction de système.

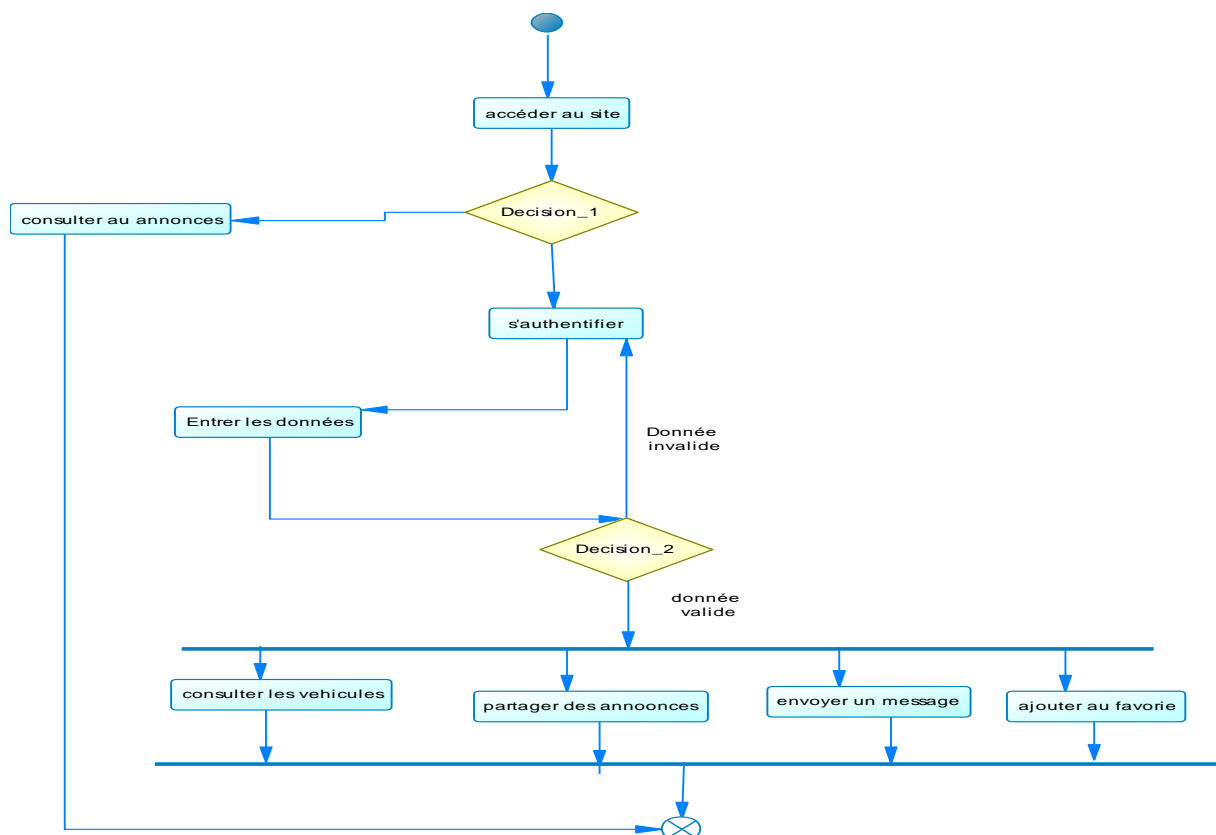


FIGURE 7 : DIAGRAMME D'ACTIVITE 1
Le diagramme précédent décrit la gestion d'application par un membre, le

membre doivent d'abord s'authentifier pour avoir y accéder aux fonctionnalités du notre site.

Tous d'abord, si le test des données est invalide l'utilisateur ne pourra pas accéder au système.

Sinon une liste des utilisateurs serra ajouter, et l'administrateur (après une authentification réussite) peut se contenter de consulter cette liste, et il peut supprimer ou désactiver un compte d'utilisateurs.

Le diagramme si dessous explique la gestion d'application par l'administrateur :

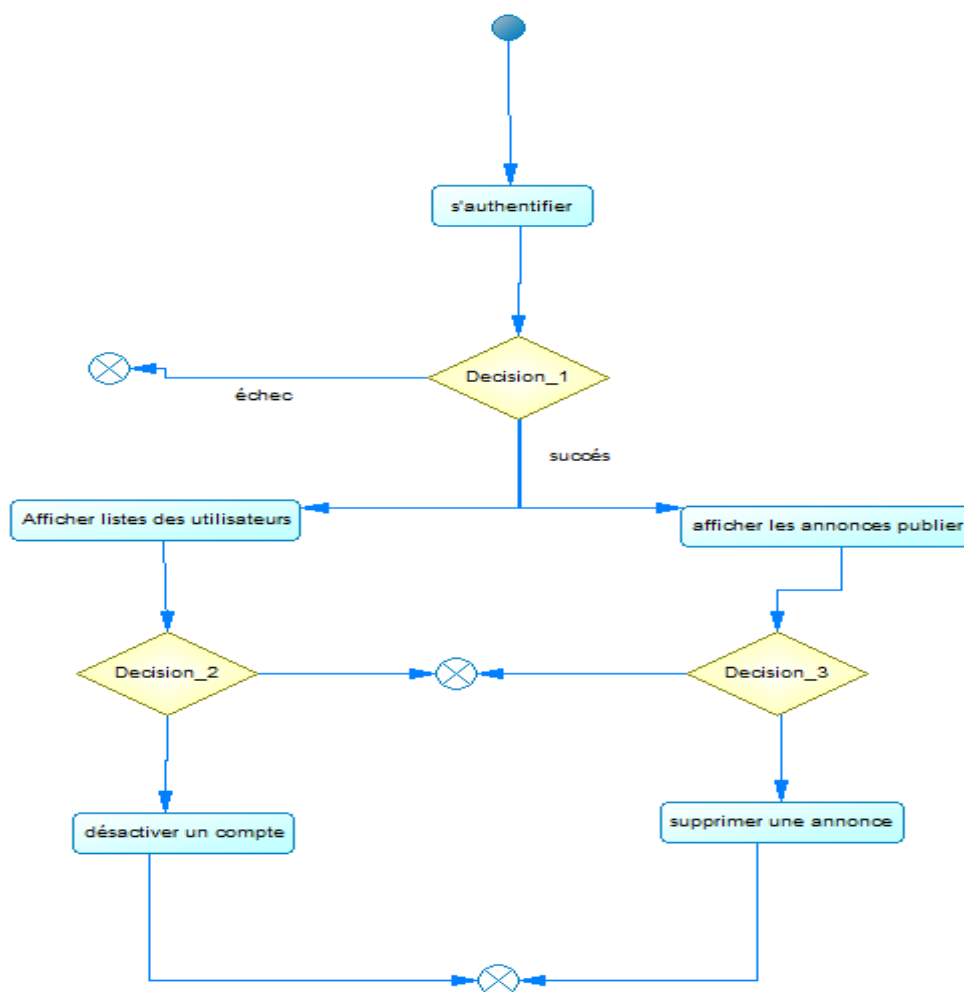


FIGURE 8 : DIAGRAMME D'ACTIVITE 2

L'utilisateur de notre site s'authentifier, et accède à l'interface du partage d'annonce :

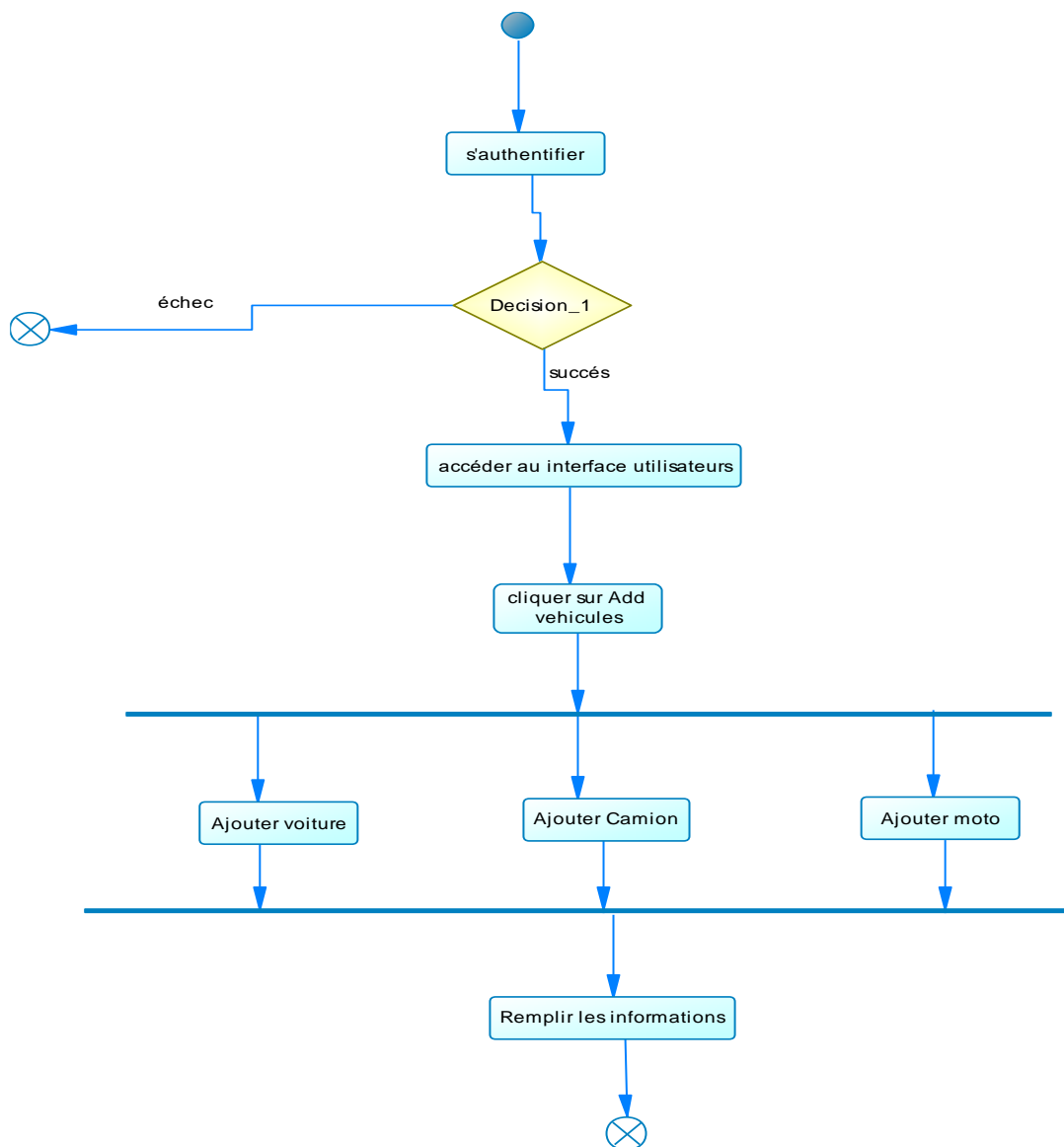


FIGURE 9 : DIAGRAMME D'ACTIVITE 3

b. Conclusion

Ce chapitre a donné une vision sur notre travail, par l'élaboration des spécifications, qui déterminent les principales fonctionnalités de notre système à travers la présentation des besoins fonctionnels et non fonctionnels, et l'aspect conceptuel par la présentation des différents diagrammes décrits en UML.

Chapitre 3

Scraping

Ce chapitre sera dédié à la représentation du web scraping, en décrivant fonctionnement et son domaine d'activité.

1. Introduction

Le web scraping est une technique d'extraction du contenu de sites Web, et de les convertir dans un format structuré utilisable, par exemple csv. Json ou les stocker dans une base de données.

En effet, le scraping pourra s'effectuer avec plusieurs outils tels que BeautifulSoup, Selenium, Scrapy, et chacun de ces derniers est livré avec ses propres ensembles d'avantages et d'inconvénients.

2. Fonctionnement

Principalement, le scraping c'est de transformer des données non structurées dans des pages en données structurées, qui seront utilisées d'une façon simple. Il se fait en 3 étapes : demander au site web le contenu d'un URL spécifique, en téléchargeant le code HTML de la page à scraper, et après on va utiliser un processus qui consiste à prendre le code d'une manière que l'ordinateur peut comprendre et utiliser.

En effet, pour obtenir le code HTML il suffit d'émettre une requête http avec la fonction get de requests, et après l'obtention du code, il faut récupérer les informations en utilisant un parser de code html qui sert à identifier le code. Et finalement télécharger les données dans CSV, JSON ou dans une base de données afin qu'elles puissent être utilisées manuellement ou dans un autre programme.



FIGURE 10 : SCHEMA DU SCRAPING

3. Les types de scraping

Le scraping web manuelle : c'est le fait l'insertion manuelle d'informations et de données, effectué lorsqu'on souhaite trouver et enregistrer des informations d'une façon irrégulière. Ce processus est rarement appliqué par des grandes quantités de données.⁵

Le scraping web automatique : Dans ce cas, on utilise un logiciel ou un

algorithme qui explore plusieurs sites Internet afin d'extraire un maximum d'informations pertinentes. Selon le contenu que vous souhaitez extraire, il faut utiliser un logiciel automatisé adapté.⁵

4. Le web scrapping vs le web crawling

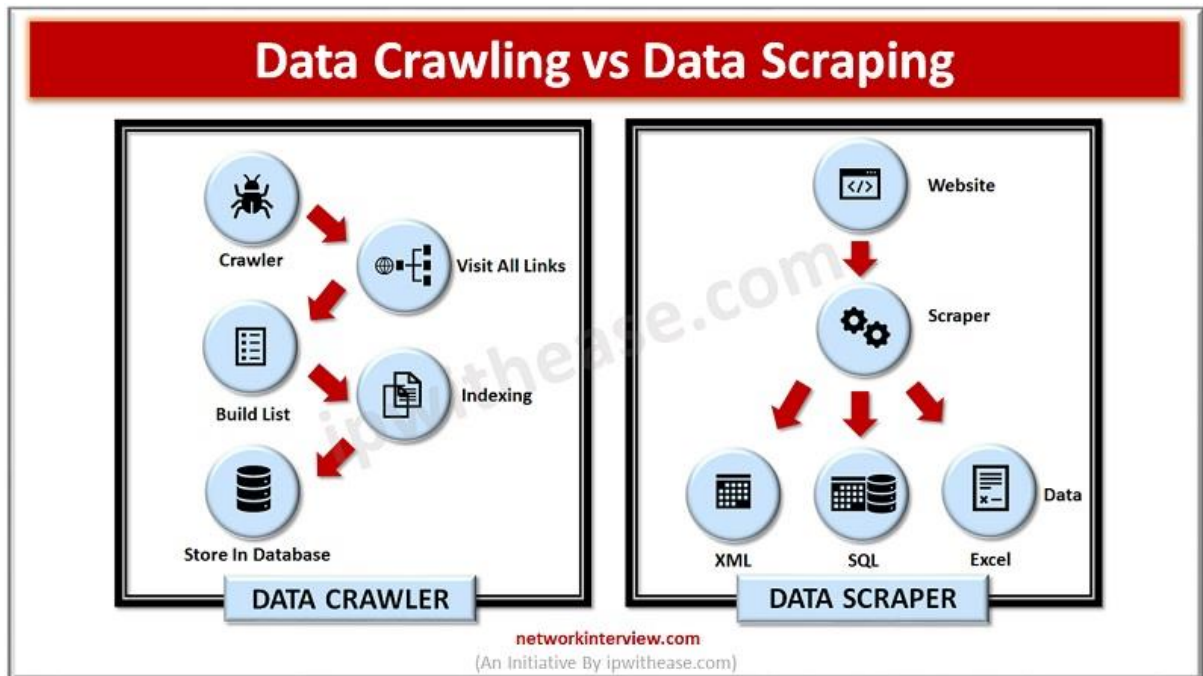


FIGURE 11 : SCHEMA DU SCRAPING ET CRAWLING

Comme il est illustré dans la figure précédente, le processus du « web crawling » consiste à indexer le web en suivant des hyperliens. Au contraire le « web scraping » c'est le processus d'extraction d'informations structurés d'une page. De toute façon le web scraping et le web crawling sont deux techniques d'extractions de données d'une manière différente et bien évidemment avec des résultat différents. Mais dans la majorité des cas le web crawling est inclus dans le processus du scraping⁶

5. Etude comparative des outils du web scraping

a. Scrapy



Scrapy est un Framework d'exploitation web gratuit et open-source écrit en Python, utilisé pour explorer les sites web et extraire des données structurées de leurs pages, c'est l'un des bibliothèques les plus puissantes disponibles.

Ce Framework fournit un ensemble de scripts Python qui contiennent la plupart de code requis pour utiliser python pour le web scraping⁷. Sinon il est compatible Python 2.7 et Python 3.6 ou au-dessous, sur la majorité des plateformes

Après l'apparition de la première version en 2009, l'équipe de développement publie des nouvelles versions à l'objectif d'enrichir le Framework en fonctionnalité.

Parmi ces caractéristiques, c'est que le Framework est rapide avec une gestion d'actions en parallèles, et en plus qu'il est compatible Linux, Windows, et Mac.

D'une autre part il est extensible, ce qui signifie que chaque robot peut être personnalisés via des extensions, modifiant son comportement.

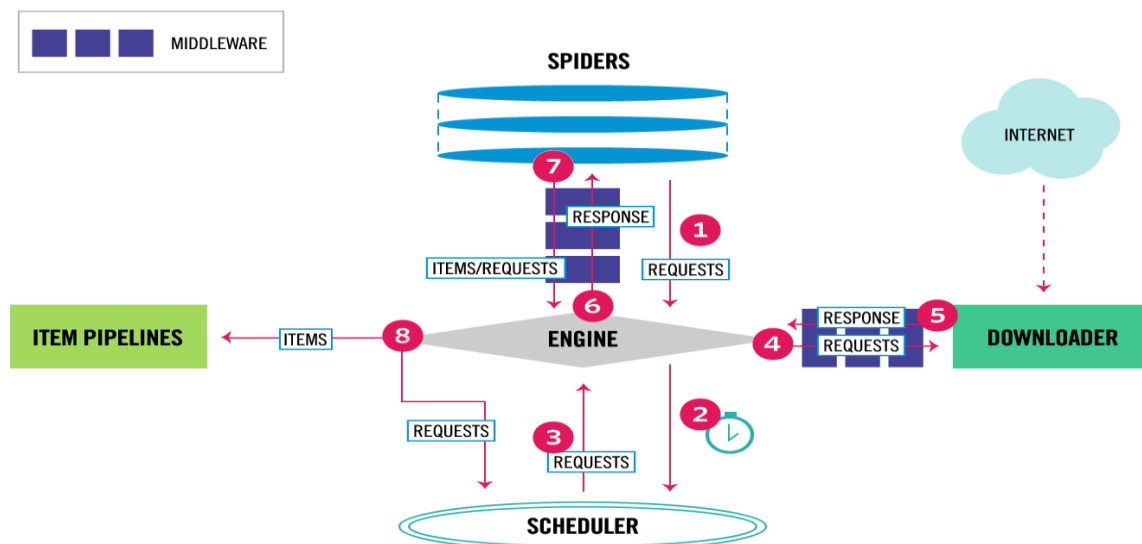


FIGURE 12 : SCHEMA DE L'ACHITETCURE SCRAPY

b. BeautifulSoup



Beautiful Soup est une bibliothèque Python pour extraire des données de fichiers HTML et XML⁹. Développée par Leonard Richardson, et soutenu par Tidelift.

C'est grâce à leur bonne documentation il est convivial pour les débutants. BeautifulSoup prend en charge l'analyseur HTML inclus dans la bibliothèque standard en Python, mais aussi plusieurs analyseurs python tiers comme lxml ou html5lib. ce Framework est facile à l'utiliser, ainsi qu'il est très

populaire auprès des développeurs, également chez les Data Engineer.

Mais on doit souligner que vous devez être vraiment bon avec la programmation multithread Python pour pouvoir effectuer plusieurs tâches à la fois efficacement en utilisant BeautifulSoup.

BeautifulSoup est un excellent choix pour les travaux de grattage Web uniques ou plus petits, où vous n'avez pas besoin d'extraire systématiquement des données à grande échelle¹¹.

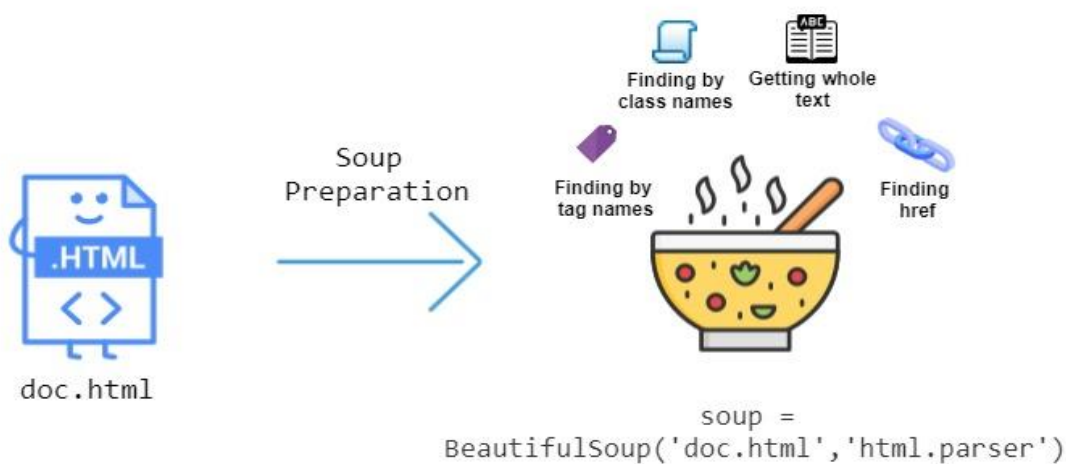
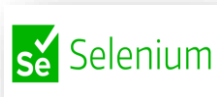


FIGURE 13 : GUIDE D'ANALYSE HTML AVEC BEAUTIFULSOUP EN PYTHON

c. Selenium



Selenium est un Framework pour les pages dépendantes de javascript, crée par Software Freedom Conservancy. Son principal objectif c'est pour les tests automatisé des application Web.

C'est un Framework polyvalent qui fonctionne avec plusieurs langages de programmation tel que, Python, Java, Ruby et node.js. Ainsi qu'il a une grande communauté.

En effet, les autres scrapeurs qui utilisent beautifulsoup ou scrapy utilisent selenium quand ils ont besoin de données qui ne peuvent être disponible que lorsque les fichiers Javascript sont chargés

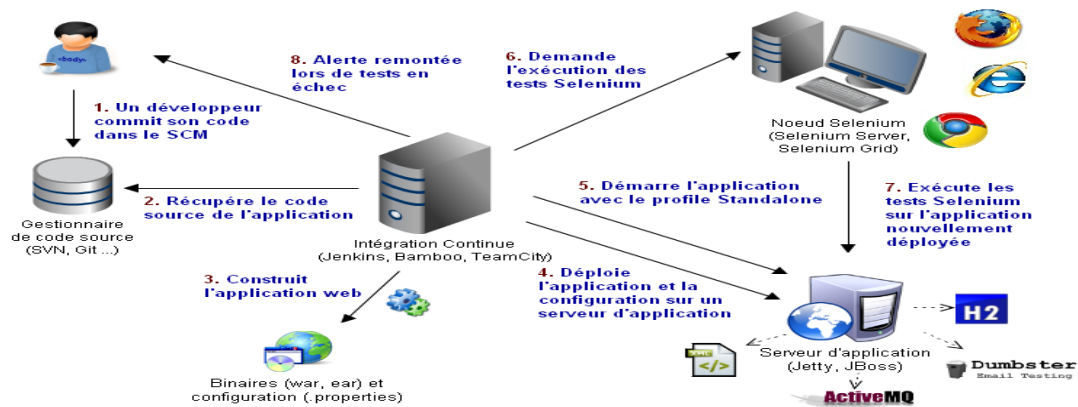


FIGURE 14 : ILLUSTRATION DU FRAMEWORK SELENIM DE CES DIFFÉRENTES ÉTAPES

d. Déduction

Dans l'ensemble, Scrapy est utile pour les grands projets de web scraping, car il s'agit d'un cadre conçu pour gérer des tâches complexes de grattage Web. Par contre BeautifulSoup, c'est un excellent choix pour les petits travaux de grattage, en tous cas il a besoin seulement d'installer l'analyseur HTML et le module de demande.

En ce qui concerne Selenium, il est pratique pour tous les sites web en vedette Javascript.

6. Conclusion

Dans ce chapitre, nous avons défini le concept du web scraping, en parlant sur leur fonctionnement, et avec une étude comparative des outils du web scraping.

Chapitre 4

Architecture et technologie de développement

Dans ce chapitre, on va présenter les outils qu'on a utilisé durant la réalisation du notre projet

1. Technologie de développement

a. Côté Frontend



Vue.js est un Framework pour construire des interfaces utilisateurs, développé par Evan You à l'été de 2013. Ce Framework est également concentré sur la partie vue ainsi qu'il est simple à l'intégrer avec autres bibliothèque ou projets existant.

À la différence des autres Framework monolithiques, Vue a été conçu et pensé pour être adaptable et évolutif. D'un autre côté, Vue est tout à fait capable de faire tourner des applications web monopages quand il est couplé avec des outils modernes et des bibliothèques complémentaires¹³

En effet, vue.js simplifie la communication entre les hiérarchies de composant, par exemple on peut utiliser `v-on` pour recevoir l'évènement émis par le composant parent, ou bien on utilise l'évènement `$emit` qui servira déclencher d'après la méthode `onclick`, qui sert à émettre l'évènement `add` vers un autre composant Ce Framework fournit aussi des composants globaux où on peut enregistrer un composant une fois et l'utiliser partout, aussi des composants de fichier unique qui permettent d'étendre les éléments HTML, CSS et Javascript essentiels pour encapsuler du code réutilisable.

b. Côté Backend



Laravel est un Framework MVC de langage de programmation PHP, créé par Taylor Otwell, ce Framework regroupe les meilleures bibliothèques utiles pour créer un site web, et fournit des fonctionnalités en termes de routage de requête, de mapping objet-relationnel, d'authentification, de migration de base de données.

En ce qui concerne l'option relative aux bases de données, Laravel crée automatiquement une table dans la base de données. Cette dernière sert en effet à enregistrer les informations de migrations des tables. Avec la commande « `php artisan` », il est possible d'enregistrer des migrations, de pouvoir migrer ou encore

de revenir à une version précédente d'une table. En outre, avec cet outil, on peut définir des classes ou « **seeder** ». Lorsqu'elles sont appelées avec la commande « `php artisan db:seed` », ces classes vont remplir la base de données avec des données prédéfinies.¹⁴

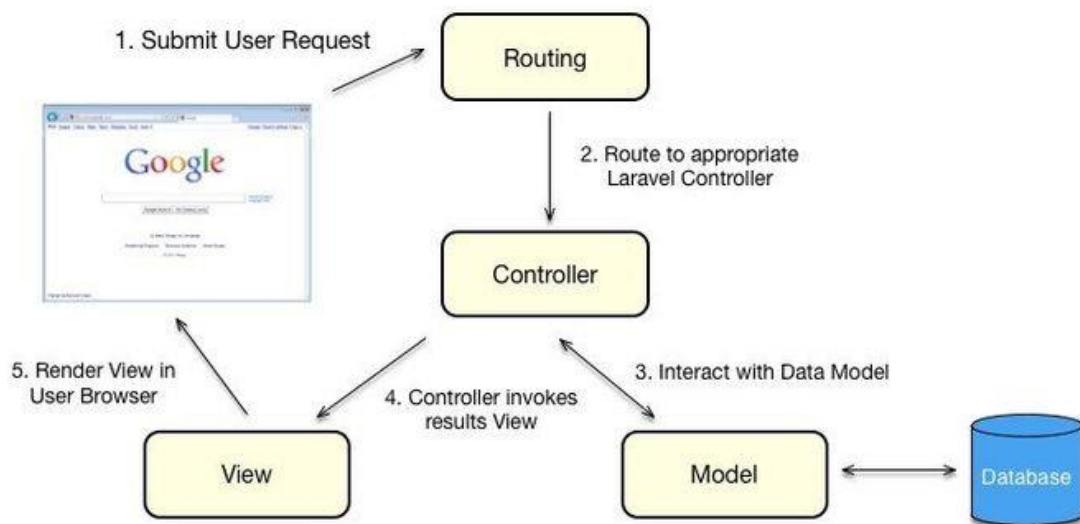


FIGURE15 : ARCHITECTURE MVC DE LARAVEL

c. Cote scraping

- Python :



Python est le langage de programmation open source le plus employé par les informaticiens. Ce langage s'est propulsé en tête de la gestion d'infrastructure, d'analyse de données ou

dans le domaine du développement de logiciels. En effet, parmi ses qualités, Python permet notamment aux développeurs de se concentrer sur ce qu'ils font plutôt que sur la manière dont ils le font. Il a libéré les développeurs des contraintes de formes qui occupaient leur temps avec les langages plus anciens.¹⁵

- Xpath :



L'xml path est un langage de requêtes pour les documents XML développé par le consortium w3, XPath met à disposition

des utilisateurs une syntaxe non basée sur XML permettant d'adresser les éléments

d'un document XML de façon ciblée.

Il est basé sur un modèle de données qui interprète les documents XML comme une séquence d'éléments disposés dans une arborescence. L'arborescence du modèle de données XPath est comparable à celle du Document Object Model (DOM) qui agit comme une interface entre le HTML et le JavaScript dynamique dans les navigateurs web.

La localisation des éléments XML est effectuée sous la forme de chemins, à la manière du système de répertoires Unix. Les éléments de base de ces chemins de localisation sont les nœuds, les axes, les tests de nœud et les prédicats.¹⁶

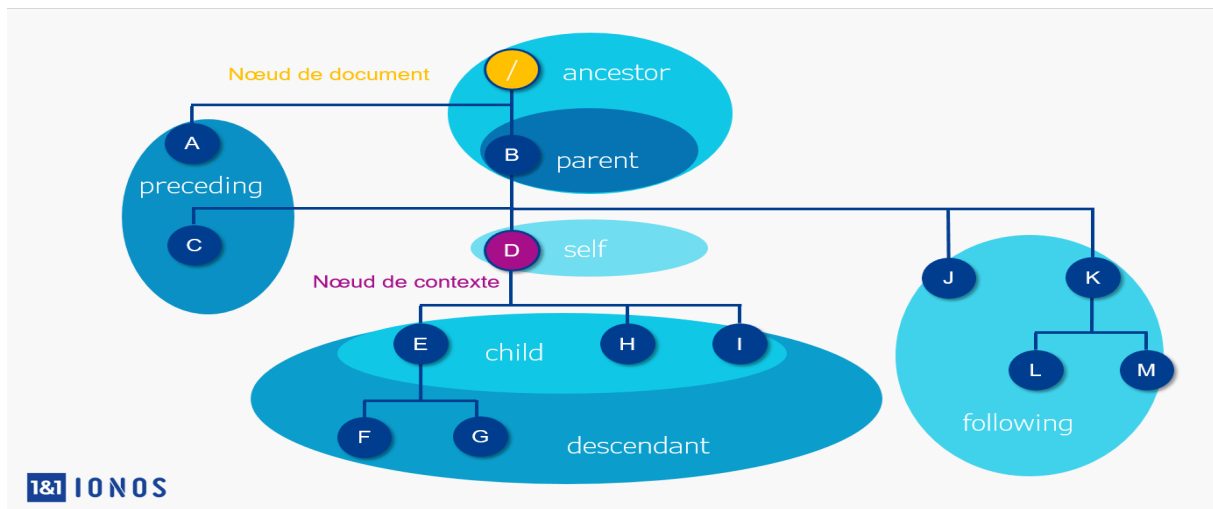
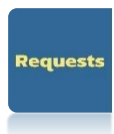


FIGURE 16 : PRINCIPALE AXES DU MODELE DU DONNEE XPATH

- Requests :



La bibliothèque requests sert à envoyer des requêtes HTTP en utilisant Python. Elle est simple et facile à utiliser avec de nombreuses fonctionnalités allant de la transmission de

paramètres dans les URL à l'envoi d'en-têtes personnalisés et à la vérification SSL. Pour l'installer : il suffit d'exécutez la commande suivante dans le terminal : `pip install requests`.¹⁷

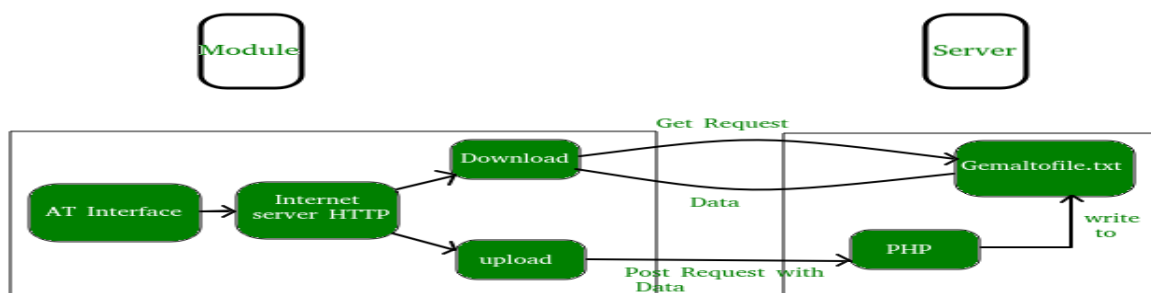


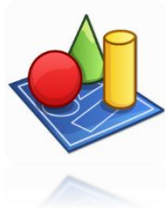
FIGURE 17 : REQUETES GET AND POSTS à L'AIDE DE PYTHON

Remarque : Et on a utilisé aussi le Framework BeautifulSoup {voir le chapitre 3 partie 5 – b}

2. Outil de développement

a. Outil de conception

- PowerAMC



PowerAMC ou bien « AMC designer », c'est un programme qui propose des différentes techniques de modélisation, il permet de réaliser tous les types de modèles informatiques et leurs bases de données associées. Il a été créé par SPD, racheté par Power Soft et produit par Sybase depuis 1995.

C'est un logiciel basé sur le langage de modélisation UML, et il gère la plupart des diagrammes spécifiés dans la norme de ce langage qu'on a choisi pour la conception de notre projet

b. Base de Données

- PhpMyAdmin



PhpMyAdmin est un outil logiciel gratuit écrit en PHP, destiné à gérer l'administration de MySQL sur le Web. C'est une interface web intuitive, qui pris en charge la plupart des fonctionnalités de MySQL tel que, parcourir, supprimer, créer, copier, ainsi que modifier des tables, des vues, des champs et des index.

Il a plusieurs autres fonctionnalités comme la transformation des données stockées dans n'importe quel format à l'aide des ensembles de fonctions prédéfinis, importer ou exporter des données vers différents formats (CSV, SQL, PDF, XML ...)

- MySQL



Le terme MySQL, pour My Structured Query Language, désigne un serveur de base de données distribué sous licence libre GNU (General Public License). Il est, la plupart du temps, intégré dans la suite de logiciels LAMP qui comprend un système d'exploitation (Linux), un serveur web (Apache) et un langage de script (PHP). Le serveur MySQL peut être utilisé sur de nombreux systèmes d'exploitation (Windows, Mac OS, etc.). Il supporte les langages informatiques SQL et SQL/PSM.²⁰

c. Visual Studio



Visual Studio Code est un éditeur de code open-source développé par Microsoft supportant un très grand nombre de langages grâce à des extensions. Il supporte l'autocomplétions, la coloration syntaxique, le débogage, et les commandes git.²¹ Quel que soit le langage : Javascript, PHP, JAVA, C#, C++ ou autres, **VS code** permet de développer soit via les fonctionnalités par défaut (pour le HTML, CSS, Javascript, Typescript...) ou en ajoutant des extensions disponibles selon les besoins de chacun.²²

3. Conclusion

Ce chapitre a défini les différents outils et technologies qu'on a utilisées pour réaliser notre application.

Chapitre 5

Mise en oeuvre de l'application

Ce chapitre, va illustrer les différentes fonctionnalités de la réalisation de notre application web.

1. Architecture de développement

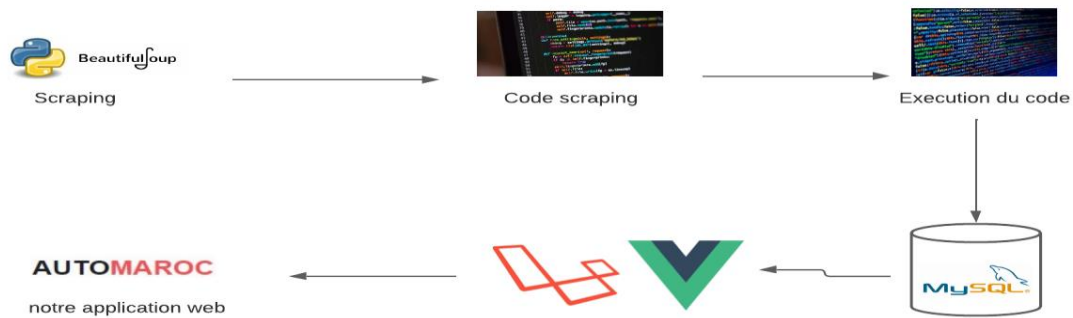


FIGURE 18 : SCHEMA DE DEVELOPPEMENT

2. Scraping

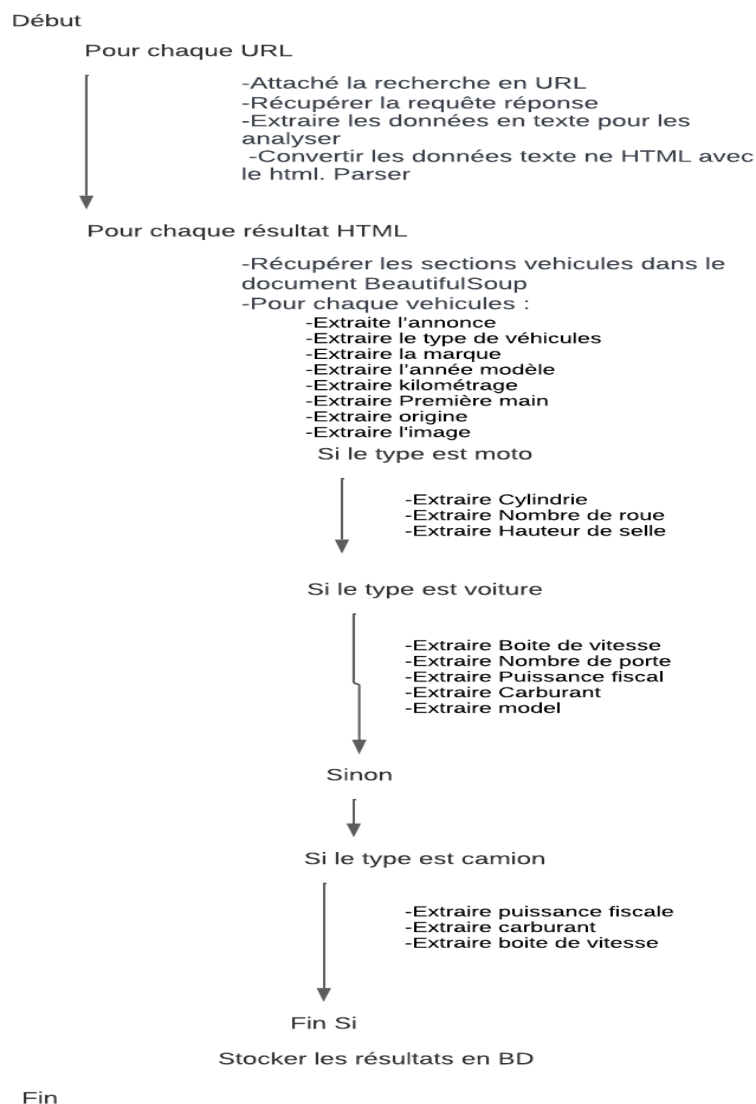


FIGURE 19 : SCHEMA DE SCRAPING

3. Interface d'accueil

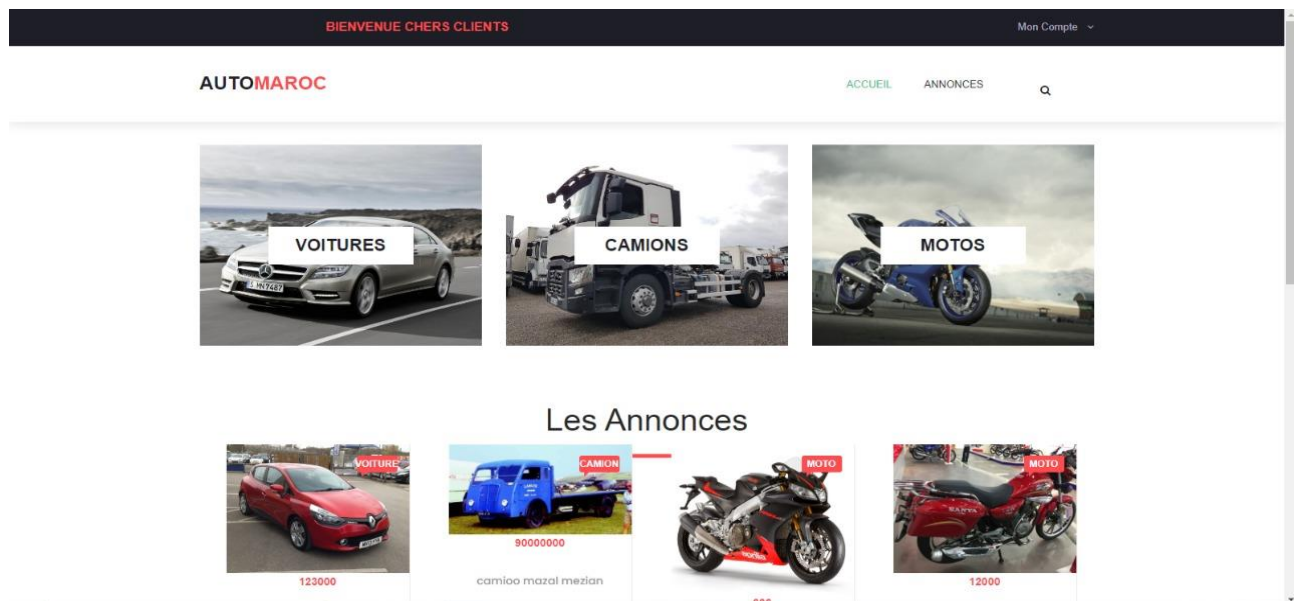


FIGURE 20 : INTERFACE PRINCIPALE DU SITE WEB (interface client)

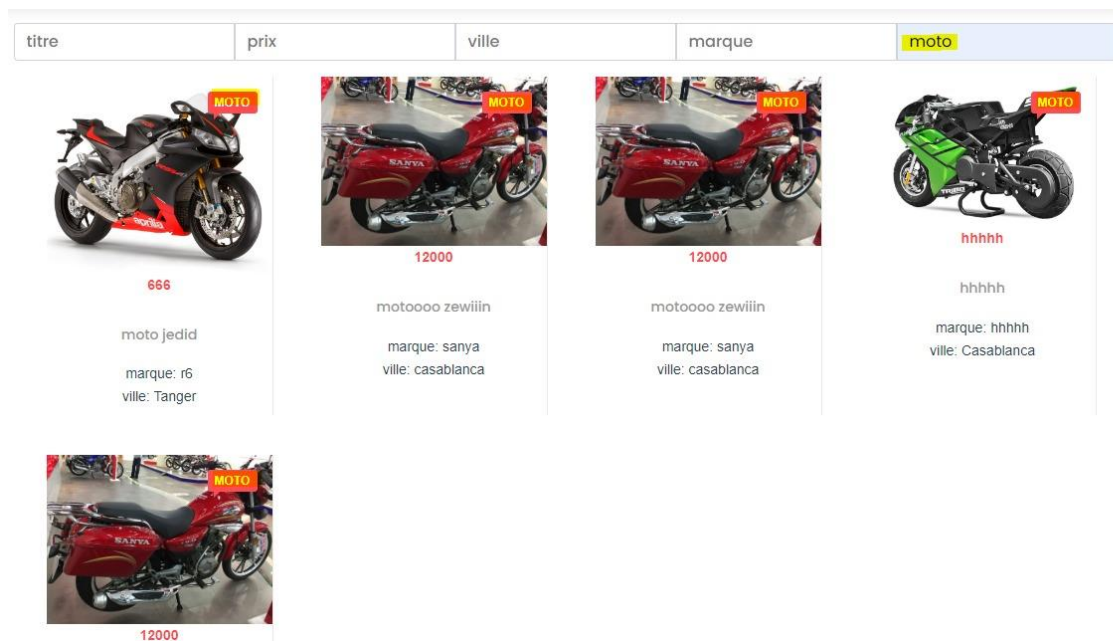



FIGURE 21 : FILTRAGE PAR MOTO

Dans la page principale du site on a plusieurs annonces partager par des clients, ou les annonces des autres sites qu'on a scrapper.

camioo mazal mezan



| | | | |
|-------------------|----------|----------------|--------|
| Marque: | camioo | Annee: | 2014 |
| kilometrage: | 10000 | premiere main: | non |
| Origine: | maroc | type: | camion |
| puissance_fiscal: | 8cv | carburant: | diesel |
| boite_vitesse: | manuelle | | |

Description: je partage mon camion pour le vendre 3 days ago

Donec id elit non mi porta.

Envoyer un message

si vous avez une question contacter M/Mme: hanae
hanae@example.com

ENVOYER

FIGURE 22 : ACCEDER AU INFORMATIONS DES ANNONCES

4. Partie client

I. Authentification

Chaque visiteur peut s'authentifier, pour qu'il pourra publier, recevoir des messages, ajouter aux favoris des annonce.

I. Register

Cr  er votre compte

Votre Nom:

Votre Email:

Mot de passe:

S'INSCRIRE

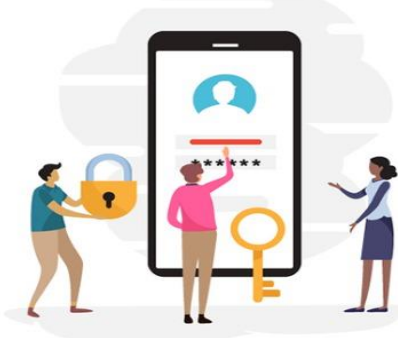


FIGURE 23 : INTERFACE D'ENREISTREMENT

II. Login



FIGURE 24 : INTERFACE DE CONNEXION

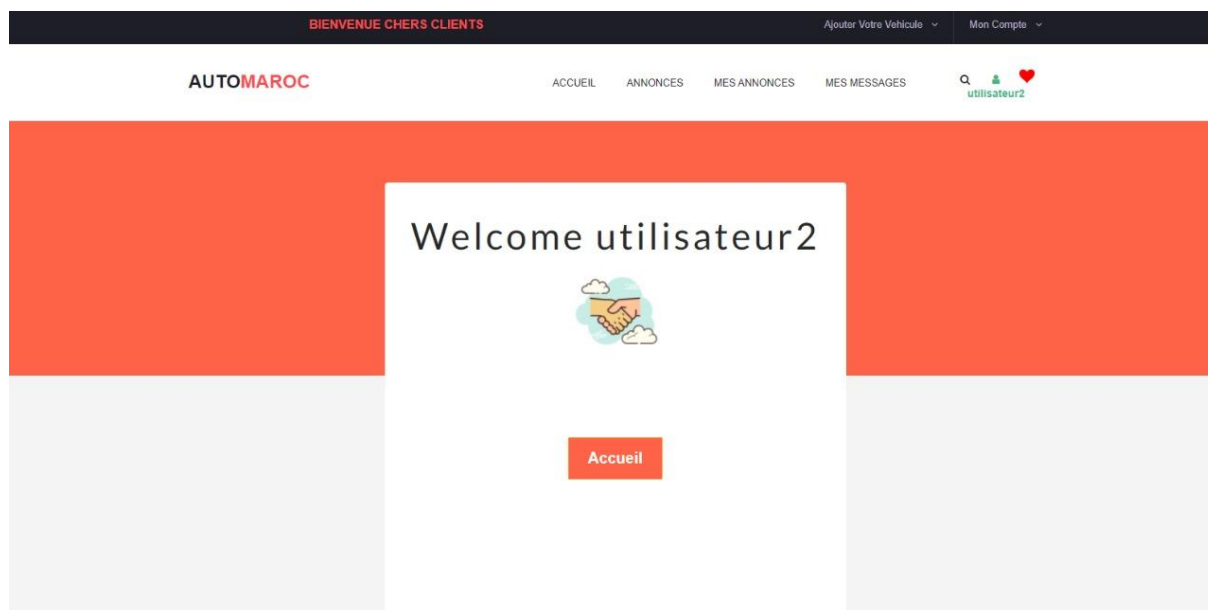


FIGURE 25 : INTERFACE PROFIL

II. Déposer annonce

Le membre inscrit pourra aussi ajouter des annonces pour vendre leur véhicule quel que soit leur type.



FIGURE 26 : AJOUTER ANNONCE

Publiez gratuitement votre annonce

| | |
|---|--------------------------------|
| Prix | Première main |
| <input type="text" value="0 Dhs"/> | <input type="text" value=""/> |
| Ville | Origine |
| <input type="text" value=""/> | <input type="text" value=""/> |
| Titre | Type |
| <input type="text" value="Titre de l'annonce"/> | <input type="text" value=""/> |
| Marque | Cylindrée (cm3) |
| <input type="text" value=""/> | <input type="text" value="0"/> |
| Année-Modèle | Nombre de roue |
| <input type="text" value=""/> | <input type="text" value=""/> |
| Kilométrage | Hauteur de selle (cm) |
| <input type="text" value=""/> | <input type="text" value="0"/> |
| Description | |
| <input type="text" value=""/> | |
| Ajouter une photo | |
| <input type="text" value="Choisir des fichiers"/> Aucun fichier n'a été sélectionné | |
| <input type="button" value="ENVOYER"/> | |

FIGURE 27 : FORMULAIRE







| Mes annonces | | | | | | |
|--------------|--------------------|---------|--|------------------------|------------------------|----------------------|
| id | titre | type | image | modifier | supprimer | voire |
| 71 | moto jedid | moto |  | update | delete | show |
| 72 | motoooo zewiiln | moto |  | update | delete | show |
| 74 | motoooo zewiiln | moto |  | update | delete | show |
| 76 | hhhhh | moto |  | update | delete | show |
| 78 | camion nv | camion |  | update | delete | show |
| 80 | modifier | voiture |  | update | delete | show |

FIGURE 28 : LISTE DES ANNONCES

III. Envoyer et recevoir les messages

| Les Messages reçu | Les Messages envoyer |
|--|---|
| <p>utilisateur2 Contenu :</p> <p>salam utilisateur ana bghit ncheri mn 3and had voiture</p> | <p>Contenu :</p> <p>bonsoir anass bechehal had i moto</p> |
| <p>admin Contenu :</p> <p>bonsoir utilisateur il faut mettre une annonce correcte</p> | |
| <p>anass Contenu :</p> <p>salam utilisateur wach la voiture clio 4 3amra wla khawya</p> | |
| <p>mohammed Contenu :</p> <p>svp wach had clio mazal dispo</p> | |

FIGURE 29 : ENVOYER MESSAGE

IV. Ajouter aux favoris

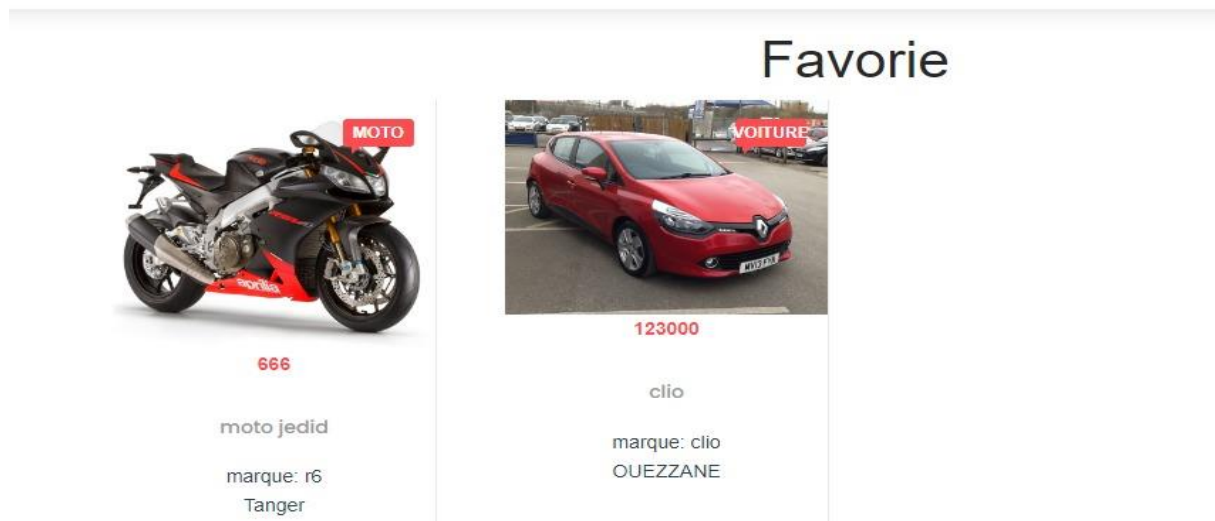


FIGURE 30 : LISTE DES FAVORIS

5. Partie administrateur



FIGURE 31 : AUTHENTIFICATION D'UN ADMIN

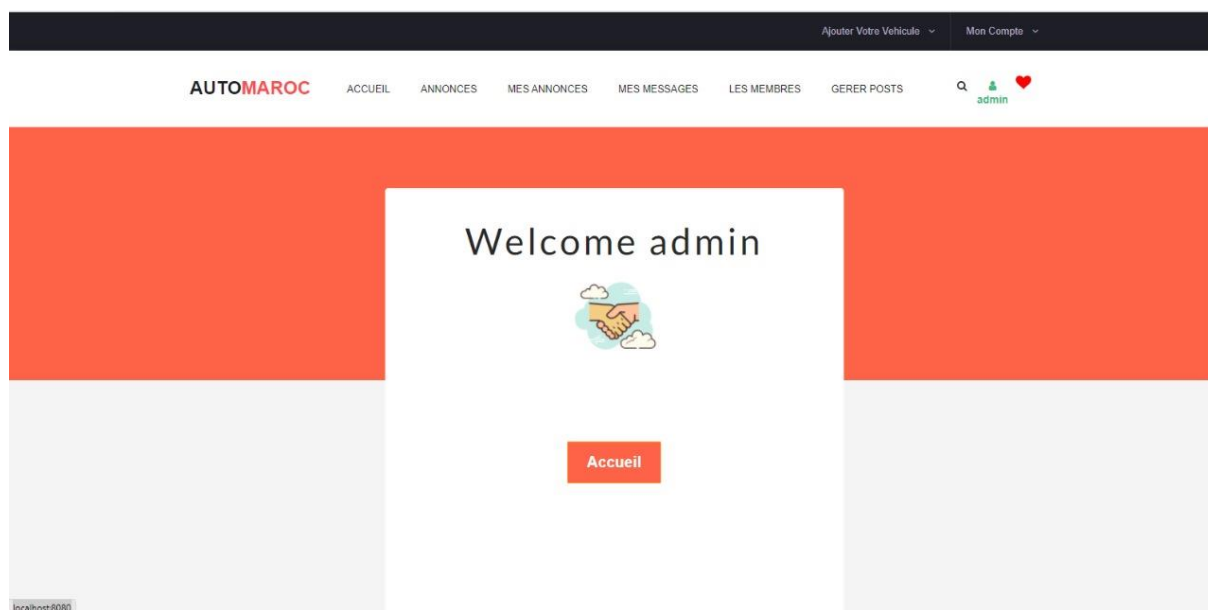


FIGURE 32 : PROFIL ADMIN

Administrateur pourra gérer les annonces et les utilisateurs, ainsi qu'il peut ajouter des annonces et envoyer des messages comme un membre.

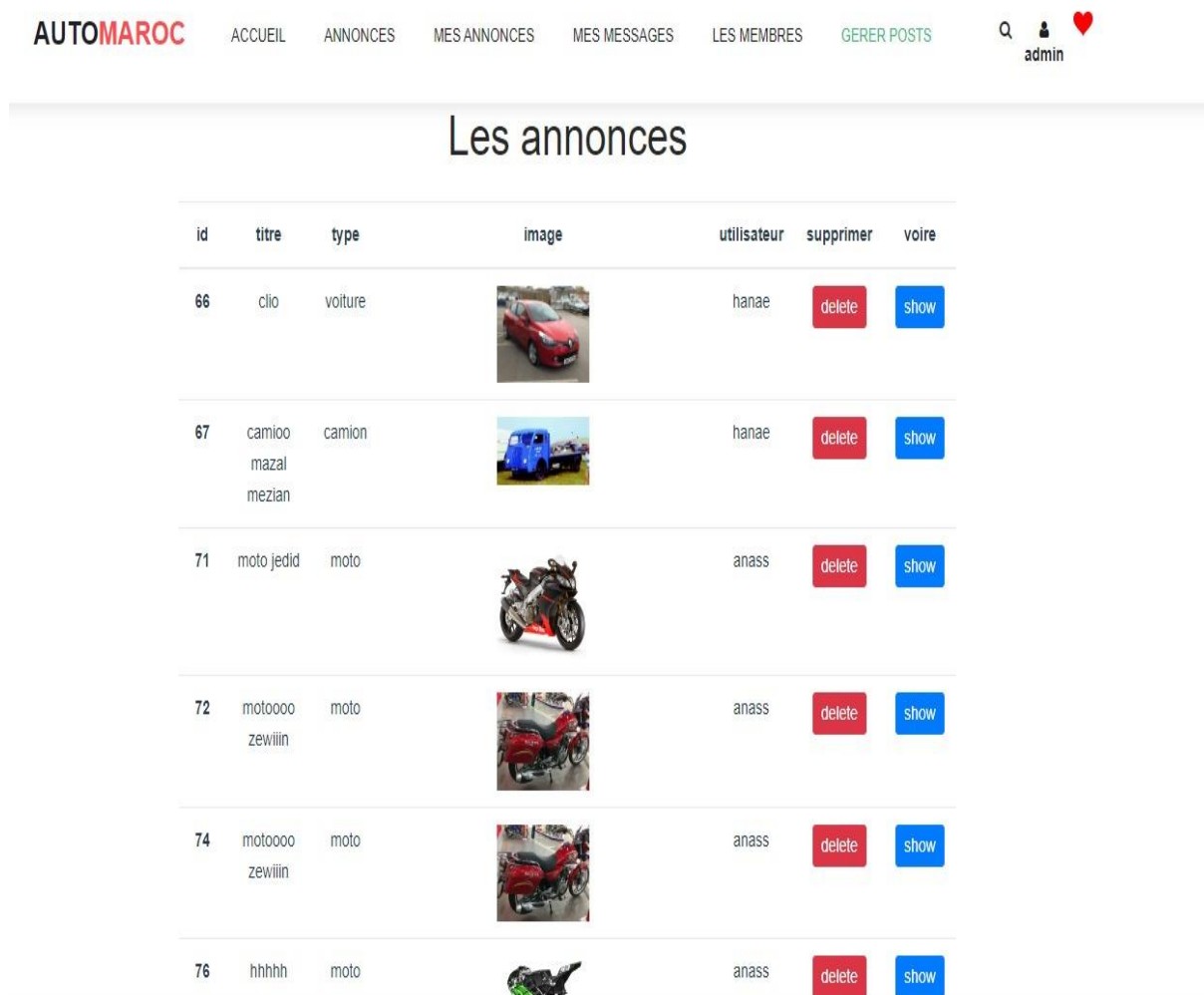


FIGURE 33 : GERE LES ANNONCES PARTAGER

| | | | | | | | | | | | | |
|-------------|----------|---------------------|--|---------|----------|--------------|--------------|-------------|-------------|---|-------|--|
| AUTOMAROC | | | | ACCUEIL | ANNONCES | MES ANNONCES | MES MESSAGES | LES MEMBRES | GERER POSTS | Q | admin | |
| Les Membres | | | | | | | | | | | | |
| id | name | email | | actions | | | | | | | | |
| 1 | hanae | hanae@exemple.com | | delete | | | | | | | | |
| 2 | tasnime | tasnime@exemple.com | | delete | | | | | | | | |
| 3 | mohammed | mohammed@gmail.com | | delete | | | | | | | | |
| 4 | anass | anass@gmail.com | | delete | | | | | | | | |
| 5 | asmae | asmae@gmail.com | | delete | | | | | | | | |
| 10 | hamza | hamza@gmail.com | | delete | | | | | | | | |
| 11 | darine | darine@gmail.com | | delete | | | | | | | | |
| 12 | youssef | youssef@gmail.com | | | | | | | | | | |

FIGURE 34 : GERER LES MEMBRES

6. Conclusion

Ce chapitre a représenté notre site web et ses fonctionnalités.

Conclusion générale

Le présent rapport présente une synthèse de notre expérience vécue au sein de Ml View Consulting, cette expérience nous a permis d'avoir une vision bien détaillée et de mieux découvrir le secteur des entreprises en passant de la théorie au pratique en appliquant les connaissances requises durant notre formation en licence.

En effet, notre principale occupation de notre projet est de penser comment extraire et traiter des données spécifiques d'internet en utilisant plusieurs techniques, afin de savoir visualiser tous ces données dans une application web. Ceci a été une occasion pour bien maîtriser des nouveaux outils, notamment Laravel, Vue.js et Python.

Cependant le scraping et le e-commerce sont dans le début d'une croissance certaine, il faut faire face que le web n'est pas exploité à son maximum car la quantité d'information est telle qu'il serait très facile de s'y perdre, ça a été par ailleurs notre mission, qui visait à faire le grattage d'une quantité d'informations et savoir les visualiser dans un autre site qui stocke tous ces données, ainsi d'avoir une capacité de les structurer et les récupérer.

Références

- [1] <https://www.antevenio.com/fr/quest-ce-que-le-web-scraping-et-a-quoi-sert-il/> 20-05-2022
- [2] <https://blog.oceanproxy.com/scraping-web> 02-06-2022
- [3] <https://www.leadgenerator.fr/blog/le-scraping-des-donnees-pour-generer-des-leads> 16-05-2022
- [4] <https://www.ionos.fr/digitalguide/sites-internet/developpement-web/quest-ce-que-le-web-scraping/> 28-04-2022
- [5] [Qu'est-ce que le Web scraping et comment ça marche ? - IONOS](#) 03-06-2022
- [6] [Exploration de données vs récupération de données : comparaison détaillée » Entretien réseau \(networkinterview.com\)](#) 03-06-2022
- [7] [Introduction au web scraping : Web scraping à l'aide de Python et Scrapy \(librarycarpentry.org\)](#) 04/06/2022
- [8] [Scrapy Vs Selenium Vs Belle Soupe pour le Web Scraping. | par Sri Manikanta Palakollu | Analyse Vidhya | Douleur moyenne \(medium.com\)](#) 04/06/2022
- [9] [Documentation Beautiful Soup — Documentation Beautiful Soup 4.4.0 \(beautiful-soup-4.readthedocs.io\)](#) 04/06/2022
- [10] [Guide d'analyse HTML avec BeautifulSoup en Python \(stackabuse.com\)](#) 05-06-2022
- [11] [Scrapy vs Beautiful Soup vs Selenium - Lequel utiliser? | Proxyway](#) 05- 06- 2022
- [12] [Scrapy Vs. BeautifulSoup Vs. Selenium pour web scraping | Meilleurs avis sur proxy \(bestproxyreviews.com\)](#) 05-06-2022
- [12] [Introduction — Vue.js \(vuejs.org\)](#) 07-06-2022
- [13] [Création interface VueJS, Agence développement web Alsace | 2le](#) 07-06-2022
- [14] [Définition de Laravel : un framework web open-source en PHP \(pappleweb.com\)](#) 07-06-2022
- [15] [Python : définition et utilisation de ce langage informatique \(journaldunet.fr\)](#) 07-06-2022
- [16] [XPath | Introduction au langage de requête - IONOS](#) 07-06-2022
- [17] [La bibliothèque requests pour interagir avec le Web – Pythonforge](#) 07-06-2022
- [18] [PowerAMC — Wikipédia \(wikipedia.org\)](#) 08-06-2022
- [19] [phpMyAdmin](#) 08-06-2022
- [20] [MySQL \(My Structured Query Language\) : définition \(journaldunet.fr\)](#) 08-06-2022
- [21] [Visual Studio Code | Framalibre](#) 08-06-2022
- [22] [Visual Studio Code - Le blog technique Webnet](#) 08-06-2022



Université Abdelmalek Essaadi
Faculté des Sciences et Techniques de Tanger
Département Génie Informatique



Année Universitaire : 2021/2022

PV de Projet de Fin d'Etude (PFE)
Filière : LST Génie Informatique

| Nom et prénom de l'étudiant (ou des étudiants) | Note Finale (Individuelle ou du Groupe) |
|--|--|
| 1. Hanae El M'rabet | |
| 2. Tasnime Morsse | |

Sujet du projet :

Vers le développement d'une plateforme web,
référentiel de vente de véhicules basé sur le scraping

Date de la soutenance : 22/06/2022

| Membres du Jury | Signature |
|---|-----------|
| Pr Abderrahim GHADI : Examineur | |
| Pr Abdelhadi FENNAN : Encadrant Interne | |
| Mr Soumaya LOUKILI : Encadrant entreprise | |

Recommandations & Appréciations du jury

.....

