

RÉPUBLIQUE DU SÉNÉGAL



Un peuple - un but - une foi

XXXXXXXXXXXXXXXXXXXXXXXXXXXX

MINISTÈRE DE L'ENSEIGNEMENT SUPÉRIEUR, DE LA RECHERCHE ET DE
L'INNOVATION

XXXXXXXXXXXXXXXXXXXXXXXXXXXX



L'excellence ma référence

XXXXXXXXXXXXXXXXXXXXXXXXXXXX

Mention : Management des systèmes d'information Automatisé

Département : Économie Gestion

UFR : Science Économique Sociale

XXXXXXXXXXXXXXXXXXXXXXXXXXXX

SUJET :

**Etude et conception d'un comparateur d'hébergements
hôteliers par crawling et scraping web : Cas des destinations
Sénégal**

XXXXXXXXXXXXXXXXXXXXXXXXXXXX

Présenté par :

Salimata SANE

Sous la direction de ;

Dr Edouard Ngor SARR

Sous la supervision de :

Prof Ousmane SALL

Membres du jury :

Prof Abdou Aziz NIANG (**Président**)

Dr Abel DIATTA (**Examineur**)

Dr Bala Moussa BIAYE (**Examineur**)

Dédicace

Je dédie ce mémoire à :

- Mon papa Katime SANE ;
- Ma maman Mbassiré DIATTA ;
- Toute ma famille ;

Merci du fond du cœur.

Remerciements

J'adresse mes sincères remerciements :

- Au bon Dieu de m'avoir donné le courage et la force d'accomplir ce travail ;
- À mon directeur de mémoire, Dr Edouard Ngor SARR, pour sa disponibilité, ses conseils avisés, et son accompagnement constant tout au long de ce travail. Son expertise et sa rigueur ont grandement contribué à l'aboutissement de ce mémoire ;
- Au Professeur Ousmane SALL pour le suivi et l'évaluation de mon travail. Merci à vous cher professeur ;
- Au personnel de S@RRIS GROUP qui m'ont bien accueilli et bien accompagné dans ce travail ;
- Aux membres du jury, pour l'honneur qu'ils me font en acceptant d'évaluer ce travail et pour leurs observations constructives ;
- À l'ensemble des enseignants de la filière MIO (Management Informatisé des Organisations) , dont les enseignements et les orientations m'ont permis de développer les compétences nécessaires à la réalisation de ce mémoire ;
- Au personnel administratif et technique de l'Université Assane Seck de Ziguinchor pour leur accueil, leur disponibilité et leur appui logistique ;
- À ma petite sœur Fatoumata SANE pour son soutien indéfectible ;
- À mes proches, amis et mentors pour leurs encouragements et leur soutien infaillible ;
- À mes camarades de promotion, notamment Saliou Diang DIALLO, Fatimata Asta SARR, Mame Fagueye LO, Abdoul-Hakim Saindou HALIDI, Fatimata Bintou SENE, pour leur esprit d'entraide. Merci pour tout ;
- À ma famille, pilier de mon existence, pour leur amour, leur patience et leur soutien indéfectible. Leur confiance et leur présence ont été essentielles tout au long de ce parcours ;
- À toutes celles et ceux qui, de manière visible ou discrète, ont contribué à la réussite de ce travail.

Je vous adresse à tous ma profonde gratitude.

Résumé

Dans ce mémoire, nous abordons la problématique liée à la promotion de la destination Sénégal par la mise en place d'un comparateur d'hébergements hôteliers par crawling et scraping web de données issues de plusieurs comparateurs en ligne tels que Booking, TripAdvisor ou AirBNB. En effet, grâce aux crawling et scraping web sur ces sites, notre système fouille et récupère automatiquement les informations clés telles que les noms des structures, leurs descriptions, les prix, les disponibilités, les avis des clients ou encore les autres caractéristiques des hébergements. L'objectif principal visé ici est de fournir aux voyageurs une interface unique et pratique pour comparer facilement les options d'hébergement au Sénégal. Plusieurs espaces sont proposés dans la plateforme (Espace administration, Espace Client et Espace Gestionnaire). Des modules de recherche filtrée et d'authentification sécurisée y sont intégrés afin d'améliorer l'expérience voyageur et d'accompagner la prise de décision.

Mots Clés : Hébergements hôteliers, Crawling, Scraping, Comparateur d'hébergement, Plateforme de comparaison

Abstract

In this thesis, we address the issue of promoting Senegal as a destination by setting up a hotel accommodation comparator using web crawling and scraping of data from several online comparators such as Booking, TripAdvisor and AirBNB. Thanks to web crawling and scraping on these sites, our system automatically searches and retrieves key information such as the names of the structures, their descriptions, prices, availability, customer reviews and other characteristics of the accommodation. The main objective here is to provide travellers with a single, practical interface for easily comparing accommodation options in Senegal. Several areas are available on the platform (Administration area, Customer area and Manager area). Filtered search and secure authentication modules have been integrated to improve the traveller's experience and support decision-making.

Keywords: Hotel accommodation, Crawling, Scraping, Accommodation comparator, Comparison platform

Sommaire

Dédicace	i
Remerciements	ii
Résumé	iii
Abstract.....	iv
Sommaire.....	v
Liste des figures.....	vi
Liste des tableaux	ix
Sigle et abréviations.....	x
Introduction générale	1
Chapitre 1 : Généralités	4
Chapitre 2 : Analyse de l'existant et positionnement	24
Chapitre 3 : Cahier des charges	42
Chapitre 4 : Conception et réalisation	50
Chapitre 5 : Présentation de la solution	85
Conclusion et perspectives	96
Références.....	97
Tables des matières.....	100
Annexe 1 : Scraper Booking.....	103
Annexe 2 : Crawler Booking	105

Liste des figures

Figure 1.	Schéma d'un Crawler (Bot d'indexation).....	11
Figure 2.	Schéma du Web Scraping	17
Figure 3.	Schéma de fonctionnement du Web Scraping	18
Figure 4.	Schéma du Web Crawling & du Web Scraping	22
Figure 5.	Tableau de bord de Booking. com.....	25
Figure 6.	Tableau de bord de HotelsCombined	26
Figure 7.	Tableau de bord de Trivago	26
Figure 8.	Tableau de bord de Kayak	27
Figure 9.	Tableau de bord de TripAdvisor :.....	28
Figure 10.	Tableau de bord de Expedia :	28
Figure 11.	Tableau de bord de Agoda.....	29
Figure 12.	Tableau de bord de Amadeus Hotel Plus.....	29
Figure 13.	Tableau de bord de Etraviatx	30
Figure 14.	Tableau de bord de Hotel Reservation Service.....	30
Figure 15.	Tableau de bord de Priceline	31
Figure 16.	Tableau de bord de Amadeus Value Hotels	31
Figure 17.	Tableau de bord de AirBnb.....	32
Figure 18.	Tableau de bord de Cozycozy.....	32
Figure 19.	Tableau de bord de Skyscanner	33
Figure 20.	Tableau de bord de Liligo	33
Figure 21.	Tableau de bord de eDreams	34
Figure 22.	Tableau de bord de Momondo	34
Figure 23.	Tableau de bord de Hotels.com	35
Figure 24.	Répartition des avis en ligne des hôtels par site d'évaluation dans le monde en 2017	38
Figure 25.	Modélisation Architecture générale du système.....	51
Figure 26.	Exemple de page Booking	54
Figure 27.	Figure en étoile pour l'organisation de la source de Booking	55
Figure 28.	Exemple de page Tripadvisor	56
Figure 29.	Page detail tripadvisor	57
Figure 30.	Page detail redirection tripadvisor vers booking.....	57
Figure 31.	Figure en étoile pour l'organisation de la source de TripAdvisor :.....	57
Figure 32.	Page AirBNB	58

Figure 33.	Page detail AirBNB	60
Figure 34.	Figure en étoile pour l'organisation de la source de Airbnb :	61
Figure 35.	Diagramme de cas d'utilisation Gestion des Réservations	63
Figure 36.	Diagramme de Cas d'utilisation Alimentation Base de Données	64
Figure 37.	Diagramme de Séquence d'Authentification	65
Figure 38.	Diagramme de Séquence gestion de réservation	66
Figure 39.	Diagramme de Séquence des Paiements	67
Figure 40.	Diagramme de Classe Gestion Hébergements	68
Figure 41.	Modèle Physique de Données gestion hébergement	69
Figure 42.	Création de la base de Données	70
Figure 43.	Importation des données du fichier SQL	70
Figure 44.	Base de données destination_senegal	71
Figure 45.	Chargement des données dans la base de données	72
Figure 46.	Interface PHPRAD	73
Figure 47.	Création du Projet	73
Figure 48.	Enregistrement du projet	74
Figure 49.	Enregistrement du fichier ppm	75
Figure 50.	Choix du thème	75
Figure 51.	Gestion des pages	76
Figure 52.	Gestion du design	77
Figure 53.	Liste Menu Icon	78
Figure 54.	Gestion du menu	78
Figure 55.	Gestion des utilisateurs	79
Figure 56.	Gestion des rôles et privilèges	80
Figure 57.	Gestion du monitoring ou des audits	81
Figure 58.	Affichage du monitoring	81
Figure 59.	Publication du projet Destination Sénégal	82
Figure 60.	Fichier configuration du projet	83
Figure 61.	Affichage du fichier app	83
Figure 62.	Affichage du fichier Errors	84
Figure 63.	Page d'accueil de la plateforme	86
Figure 64.	Page de connexion	86
Figure 65.	Création de compte	87
Figure 66.	Gestion hébergements	87

Figure 67.	Fiche de logement.....	88
Figure 68.	Liste des réservations.....	88
Figure 69.	Ajout d'une nouvelle réservation.....	88
Figure 70.	Gestion du monitoring	89
Figure 71.	Gestion des comptes	89
Figure 72.	Gestion des rôles.....	89
Figure 73.	Gestion des permissions	90
Figure 74.	Page d'accueil gérant.....	90
Figure 75.	Gestion hébergements.....	91
Figure 76.	Liste des logements.....	91
Figure 77.	Ajout d'un logement	92
Figure 78.	Page d'accueil voyageur	92
Figure 79.	Affichage des hébergements.....	93
Figure 80.	Page de Réservation.....	93
Figure 81.	Fiche de réservation	94
Figure 82.	Gestion des avis	94
Figure 83.	Affichage des réservations émises.....	94

Liste des tableaux

Tableau 1.	Tableau des Zones Côtières du Sénégal	7
Tableau 2.	Tableau des Zones non côtières du Sénégal	10
	Tableau comparatif des types de web crawling	16
Tableau 3.	Tableau des différentes formes de scraping.....	21
Tableau 4.	Tableau de comparaison entre le web scraping et le web crawling.....	22
	Tableau comparatif des différentes plateformes leaders :	36
	Plateforme.....	36
Tableau 5.	Tableau Comparatif des meilleurs tarifs trouvés par les 4 sites sur 6 Séjours : 39	
Tableau 6.	Présentation des types d'hébergements dans Booking :	52
Tableau 7.	Présentation des types d'hébergements dans TripAdvisor :	55
Tableau 8.	Présentation des types d'hébergements dans Airbnb.....	58

Sigle et abréviations

- **HTTP:** HyperText Transfer Protocol
- **URL:** Uniform Resource Locator
- **HTML:** HyperText Markup Language
- **XML:** eXtensible Markup Language
- **JSON:** JavaScript Object Notation
- **CSV:** Comma-Separated Values
- **OCR:** Optical Character Recognition
- **API:** Application Programming Interface
- **DOM:** Document Object Model
- **SQL :** Structured Query Language
- **OTA :** Online Travel Agency
- **GDS :** Global Distribution System
- **OMT :** Organisation Mondiale pour le Tourisme
- **SEO :** Search Engine Optimization
- **UNESCO:** United Nations Educational, Scientific and Cultural Organization

Introduction générale

De nos jours, le tourisme est un pilier important de l'économie mondiale, et le Sénégal s'affirme comme une destination phare en Afrique de l'Ouest. En effet, le **tourisme** peut être défini comme étant « *l'ensemble des activités liées aux déplacements des individus hors de leur environnement habituel, principalement à des fins de loisirs, d'affaires ou de découverte culturelle* ». Le Sénégal, avec ses 700 km de littoral atlantique, offre une grande diversité de destinations touristiques qui séduisent aussi bien les touristes locaux qu'internationaux. Il inclut la recherche, le choix et l'utilisation d'hébergements, comme les hôtels, qui occupent une place essentielle dans l'expérience touristique en fournissant confort, services et un accès facile aux attractions locales. Le pays¹ attire des visiteurs grâce à des sites emblématiques comme l'île de Gorée, le Lac Rose, Saint-Louis, les plages de la Petite Côte et de Cap Skirring en Casamance. Ce tourisme diversifié, mêlant des activités balnéaires, historiques, culturelles et écotouristiques, permet au Sénégal de répondre aux attentes d'un large éventail de visiteurs [1].

Toutefois, le secteur hôtelier des destinations au Sénégal, en particulier en Casamance, reste peu structuré, notamment en termes de visibilité et d'accès à l'information pour les voyageurs. Les informations concernant les établissements hôteliers dans cette région sont souvent dispersées, ce qui complique leur consultation et la comparaison des offres par rapport à d'autres zones du pays mieux organisées et mieux connectées. De plus, le développement hôtelier est inégal le long du littoral sénégalais : certaines destinations bénéficient d'une promotion plus soutenue et d'infrastructures plus développées. Parmi ces zones à fort potentiel, la Casamance se démarque par la beauté exceptionnelle de ses paysages naturels, ses plages idylliques, notamment à Cap Skirring, ainsi que par la richesse de son patrimoine culturel [2].

La principale difficulté réside, la plupart, dans l'accès à certaines destinations du Sénégal. Par ailleurs, certains hébergements locaux manquent de visibilité sur les différentes plateformes en ligne. Souvent, les informations concernant les offres hôtelières sont dispersées, ce qui complique la planification des séjours pour les touristes. Face à ces enjeux, il devient indispensable de développer un outil technologique capable de valoriser ces zones. Pour cela, un comparateur de destinations hôtelières basé sur le crawling et le web scraping serait pertinent afin de centraliser les données issues des différents sites et ainsi faciliter la prise de décision des voyageurs. Cela soulève la question suivante : ***comment un comparateur de***

¹ [Tourisme au Sénégal : les lieux à visiter - Destination Afrique](#), (Consulté le 25 Février 2025)

destinations hôtelières utilisant le crawling et le web scraping peut centraliser et mettre en valeur les offres hôtelières des destinations du Sénégal, tout en simplifiant la décision des touristes ? La problématique de recherche peut donc être formulée ainsi : en quoi l'utilisation du crawling et du web scraping permet-elle de concevoir un comparateur de destinations Sénégal capable de faciliter les choix des voyageurs ? Pour répondre à cette question, plusieurs interrogations seront abordées :

- Quels sont les besoins des voyageurs pour comparer les offres hôtelières dans les destinations Sénégal ?
- Quelles données hôtelières collecter via le crawling et le web scraping et comment les organiser ?
- Quels outils du crawling ou du web scraping utiliser pour extraire les informations des plateformes hôtelières ?
- Comment concevoir une interface utilisateur simple et efficace pour ce comparateur ?

Les comparateurs de destinations hôtelières jouent un rôle crucial pour le tourisme au Sénégal car ils centralisent des informations souvent difficiles d'accès, améliorant ainsi la visibilité des hébergements et facilitant le choix des touristes. C'est pour cela que nous nous sommes intéressés au développement d'un comparateur de destinations hôtelières basé sur le web scraping et le crawling pour améliorer l'accès aux offres touristiques dans les destinations du Sénégal. Autrement dit, cette étude permettra aux destinations touristiques Sénégalaises de :

- Améliorer leur visibilité ;
- Faciliter la comparaison pour les clients ;
- Accroître leur attractivité ;
- Stimuler la demande touristique ;
- Renforcer la compétitivité des hôtels locaux.

Nous nous sommes fixés comme objectif général d'explorer l'utilisation du grattage de contenus web pour développer une meilleure visibilité et compétitivité des destinations hôtelières du Sénégal, dans un environnement de concurrence croissante entre acteurs du secteur touristique. Autrement dit, améliorer, à travers le web scraping et le crawling, la gestion de ces destinations hôtelières en développant des outils efficaces, permettant ainsi aux hôtels d'améliorer leur visibilité et de faciliter la planification des séjours pour les voyageurs. En d'autres termes, nous cherchons à :

- Comprendre les attentes des voyageurs concernant la recherche et la comparaison des offres hôtelières ;

- Collecter et organiser les données hôtelières locales grâce au web scraping et au crawling ;
- Créer une plateforme intuitive pour faciliter la comparaison et le choix des hébergements ;
- Améliorer la visibilité des destinations hôtelières du Sénégal pour attirer davantage de touristes.

Cette étude présente plusieurs intérêts pour les hôtels mais aussi pour les touristes :

- **Pour les Hotels :** Le comparateur aidera les hôtels, les auberges à mieux se positionner face à la concurrence nationale et internationale ;
- **Pour les Touristes :** La centralisation des données sur les hébergements simplifie la recherche et la planification des séjours des touristes ;
- **Pour la prise de décision :** Ce projet offre aux décideurs touristiques un outil pour mieux comprendre les tendances du marché et adapter leurs offres ;
- **Valorisation des zones côtières :** Cette étude contribue à mettre en lumière les richesses de ces zones en rendant leurs offres hôtelières plus accessibles et attractives pour les touristes.

Pour répondre aux interrogations soulevées par le sujet, nous avons élaboré une méthodologie de recherche structurée en plusieurs étapes :

- Réaliser une analyse détaillée du sujet ;
- Identifier les données à collecter ;
- Choisir les méthodes de crawling et de web scraping les plus appropriées ;
- Développer une plateforme web pour centraliser et comparer les données collectées ;
- Explorer les options de solutions concrètes.

Pour répondre à la problématique de recherche, ce travail est subdivisé en cinq grands chapitres :

- D'abord, pour le premier chapitre, nous allons faire la généralité sur les concepts clés du sujet, pour une meilleure compréhension;
- Ensuite, pour le deuxième chapitre, nous allons mettre l'accent sur l'état de l'art et définir notre positionnement ;
- Puis, dans le troisième chapitre, nous allons mettre l'accent sur le cahier des charges ;
- Dans le quatrième chapitre, nous allons nous concentrer sur la conception, la réalisation du projet ;
- Enfin, dans le cinquième chapitre, nous allons présenter la solution de notre système.

Chapitre 1 : Généralités

1.1. Tourisme au Sénégal

Les arrivées de touristes internationaux ont progressé de 5,4 % en 2018 selon l'Organisation Mondiale du Tourisme (OMT). Avec une croissance moyenne annuelle de 3,6 %, le tourisme représentait près de 7 % des exportations mondiales cette même année, se positionnant ainsi au troisième rang derrière les exportations de produits chimiques et de carburants. En 2018, le continent africain a accueilli 7 % des arrivées touristiques mondiales, tout en générant 3 % des recettes liées au tourisme international [3].

Au Sénégal, le marché touristique se caractérise par une offre riche et diversifiée. La composante la plus connue reste le tourisme balnéaire, grâce à ses 718 km de côtes, complétée par le tourisme d'affaires et une offre culturelle variée. Cependant, la demande touristique [8] demeure peu diversifiée, avec environ 1 365 000 visiteurs enregistrés en 2017 selon l'Agence Nationale de la Statistique et de la Démographie (ANSD). Ce chiffre représente 2,1 % des arrivées touristiques en Afrique et seulement 0,105 % des arrivées mondiales selon l'OMT. Par ailleurs, le tourisme sénégalais contribue en moyenne à hauteur de 27,8 % aux exportations totales du pays chaque année. Le Sénégal, avec sa richesse culturelle et naturelle, offre une mosaïque d'expériences touristiques adaptées à tous les goûts. Le pays se distingue par sa diversité de formes de tourisme, permettant aux visiteurs d'explorer et de vivre des univers variés. Parmi les différentes formes de tourismses, nous pouvons citer [4] :

- **Le Tourisme culturel** : Cette forme de tourisme permet aux voyageurs de plonger dans l'histoire du Sénégal, à travers l'exploration de ses sites historiques inscrits au patrimoine mondial de l'UNESCO et de ses musées enrichissants. C'est une invitation à voyager dans le temps, à la rencontre des civilisations qui ont marqué le pays ;
- **Le Tourisme ethnologique** : Pour ceux désireux de comprendre la richesse humaine du Sénégal, le tourisme ethnologique offre une immersion profonde dans le quotidien et les traditions des communautés locales. Participer aux fêtes traditionnelles et échanger avec les habitants ouvrent des fenêtres sur l'âme sénégalaise ;
- **Le Tourisme balnéaire** : Avec ses plages idylliques et son climat ensoleillé, le Sénégal est une destination de choix pour les amateurs de détente au bord de l'eau et de sports nautiques ;
- **Le Tourisme de nature** : Cette forme de tourisme invite à découvrir les paysages époustoufflants du Sénégal, sa faune et sa flore diversifiées, dans des réserves naturelles et des parcs nationaux préservés ;

- **Le Tourisme ornithologique** : Le Sénégal, avec ses zones humides et ses parcs, est un paradis pour les passionnés d'observation des oiseaux, offrant la possibilité de rencontrer des espèces rares et colorées ;
- **Le Tourisme cynégétique** : Pour les amateurs de chasse, le Sénégal propose des expériences dans des cadres éthiques et réglementés, permettant la traque de la faune sauvage dans le respect de l'environnement ;
- **Le Tourisme halieutique** : La pêche en mer offre des moments de détente et de défis, dans les eaux poissonneuses de l'Atlantique, pour une aventure à la fois paisible et excitante ;
- **Le Tourisme sportif** : Que ce soit en mer, avec la voile, la plongée, la planche à voile, le ski nautique ou le jet-ski, ou sur terre, avec le quad, le VTT, le Sénégal est une destination de choix pour les passionnés de sports et de sensations fortes ;
- **Le Tourisme d'affaires** : Le Sénégal accueille également les voyageurs d'affaires, proposant des infrastructures adaptées à l'organisation d'événements professionnels, conférences et séminaires dans un cadre exceptionnel.

Chacune de ces formes de tourisme contribue à l'expérience unique que le Sénégal offre à ses visiteurs, combinant aventure, culture, détente et découvertes, pour un voyage inoubliable au cœur de l'Afrique de l'Ouest. Parmi les différentes formes de tourisme, **le tourisme balnéaire et côtier** occupe une place prédominante, avec des destinations comme Saly, Cap Skirring et l'Île de Gorée qui captent une grande partie des flux touristiques.

1.2. Les destinations du Sénégal

1.2.1. Destinations côtières

Les **destinations côtières du Sénégal** désignent l'ensemble des zones touristiques situées le long du littoral atlantique du pays, qui s'étend sur environ 700 kilomètres, de Saint-Louis au nord jusqu'à la région de Ziguinchor au sud, en passant par Dakar, la Petite Côte et la Casamance. Ces destinations se caractérisent par leur accessibilité, la diversité de leurs paysages marins, et la concentration d'infrastructures touristiques telles que les hôtels, auberges, campements, restaurants et activités de loisirs balnéaires.

Elles sont considérées comme les pôles les plus dynamiques du tourisme sénégalais en raison de leur forte attractivité pour les touristes nationaux et internationaux, notamment européens. En effet, les stations balnéaires de Saly, Cap Skirring, ou encore les plages de Yoff et Ngor à Dakar, attirent chaque année des milliers de visiteurs en quête de soleil, de détente, mais aussi de découvertes culturelles et naturelles [5].

D'ailleurs, nous distinguons deux types de zones dans les destinations côtières du Sénégal qui sont : les zones côtières et les zones non côtières du Sénégal.

1.2.2. Les zones côtières du Sénégal

Les zones côtières du Sénégal correspondent à l'ensemble des régions et localités bordant l'océan Atlantique, sur environ 700 kilomètres de littoral, du nord au sud du pays.

Le tableau suivant propose une synthèse des principaux zones côtières touristiques, précisant leur département d'appartenance ainsi que leur localisation géographique :

Tableau 1. Tableau des Zones Côtières du Sénégal

Destination touristique	Département	Zone côtière
Saint-Louis	Saint-Louis	Grande Côte
Gandioul	Saint-Louis	Grande Côte
Kayar	Thiès	Grande Côte
Lompoul	Kébémér (Louga)	Grande Côte
Dakar(Yoff, Ngor, Ouakam)	Dakar	Presqu'île du Cap-Vert
Île de Gorée	Dakar	Presqu'île du Cap-Vert
Rufisque	Rufisque (Dakar)	Grande Côte / Cap-Vert
Saly-Portudal	Mbour (Thiès)	Petite Côte
Somone	Mbour (Thiès)	Petite Côte
Ngaparou	Mbour (Thiès)	Petite Côte
Popenguine	Mbour (Thiès)	Petite Côte
Joal-Fadiouth	Mbour (Thiès)	Petite Côte
Palmarin	Fatick	Delta du Saloum
Djifffer	Fatick	Delta du Saloum
Toubacouta	Foundiougne (Fatick)	Delta du Saloum
Cap Skirring	Oussouye (Ziguinchor)	Casamance littorale
Kafountine	Bignona (Ziguinchor)	Casamance littorale
Diembering	Oussouye (Ziguinchor)	Casamance littorale
Ziguinchor	Ziguinchor	Casamance littorale

Ces zones jouent un rôle crucial dans les activités économiques (pêche, tourisme, transport maritime) et écologiques du pays. Sur le plan touristique, elles regroupent plusieurs pôles majeurs. Voici les principales zones côtières du Sénégal classées du nord au Sud [6] :

- **La Grande Côte** : elle s'étend de la frontière mauritanienne, à hauteur de Saint-Louis, jusqu'à la ville de Rufisque. Cette bande littorale se distingue par sa longue côte sablonneuse encore peu exploitée par le tourisme de masse. Elle offre un fort potentiel pour l'écotourisme et le tourisme culturel.xxxxxx
- Parmi ses points d'intérêt majeurs, on retrouve **Saint-Louis**, ancienne ville coloniale classée au patrimoine mondial de l'UNESCO, **Gandioul**, une zone naturelle riche en biodiversité, **Kayar**, l'un des plus grands centres de pêche artisanale du pays, et **Lompoul**, connu pour sa plage et son désert propice au tourisme de nature et aux séjours en écolodge ;
- **La Presqu'île du Cap-Vert (zone de Dakar)** : située à l'extrémité ouest du continent africain, la presqu'île du Cap-Vert abrite la capitale du Sénégal, **Dakar**, ainsi que plusieurs plages urbaines fréquentées par les populations locales et les touristes internationaux. Cette zone conjugue modernité, dynamisme économique, et richesse culturelle.

Parmi les sites notables figurent **les plages de Yoff, Ngor, Ouakam**, ou encore **l'île de Gorée**. La presqu'île de Dakar joue un rôle stratégique dans le tourisme sénégalais, notamment grâce à son aéroport international, son offre variée en hébergements et son attractivité liée au tourisme d'affaires, culturel et urbain ;

- **La Petite Côte** : elle constitue l'une des zones les plus emblématiques du tourisme balnéaire au Sénégal. Elle s'étend de **Rufisque** à **Joal-Fadiouth**, en traversant des localités comme **Saly-Portudal, La Somone, Ngaparou, Mbour** et **Popenguine**. Cette région se distingue par ses plages aménagées, ses infrastructures touristiques développées, ses stations balnéaires animées, ainsi que ses espaces naturels protégés comme la **réserve de Bandia** ou la **lagune de la Somone**. Très prisée des touristes européens, notamment français et belges, la Petite Côte accueille de nombreux établissements hôteliers, des resorts, et une offre diversifiée en activités de loisirs ;
- **Le Delta du Saloum** : Localisé entre **Joal-Fadiouth** et **Foundiougne**, le delta du Saloum constitue une zone côtière unique, formée de mangroves, d'îles, de bolongs et de forêts maritimes. Il est classé Réserve de biosphère par l'UNESCO et figure parmi les sites les plus riches en biodiversité du pays. Les localités de **Palmarin, Djiffer**, et **Toubacouta** offrent un cadre favorable au développement de l'écotourisme, du tourisme communautaire et de l'observation de la faune. Cette zone attire des voyageurs à la recherche d'authenticité, de calme et de contact avec la nature ;

- **La Casamance littorale** : Située dans le sud du pays, la Casamance littorale s'étend le long de l'océan Atlantique, de **Kafointine** à **Cap Skirring**, en passant par **Diembering** et **Oussouye**, jusqu'à la région de **Ziguinchor**. Dotée de paysages luxuriants, de plages paradisiaques et d'une riche culture Diola, cette région est souvent perçue comme un « joyau caché » du tourisme sénégalais. Cap Skirring est la station balnéaire la plus connue de la zone, disposant d'un aéroport et d'infrastructures touristiques de qualité. Le tourisme y reste toutefois moins intensif que sur la Petite Côte, offrant un environnement plus préservé, adapté au tourisme alternatif, culturel et solidaire.

1.2.3. Les destinations non côtières du Sénégal

Les zones non côtières du Sénégal correspondent à l'ensemble des régions situées à l'intérieur des terres, en dehors du littoral atlantique. En effet, elles couvrent une large partie du territoire national et se distinguent par leur diversité culturelle, écologique et religieuse. Bien que moins fréquentées par les touristes internationaux, ces zones offrent un fort potentiel pour le développement d'un tourisme durable, culturel et religieux. Voici les principales zones non côtières du Sénégal, classées par grandes régions :

- **Le Sénégal oriental** : Composé de la région de Tambacounda et Kédougou, ces régions se distinguent par leur vastes étendues naturelles, leur faune sauvage et leurs paysages montagneux. D'ailleurs le Parc National du Niokolo-Koba constitue une attraction majeure pour les amateurs de la nature. Plus au sud, nous avons Kédougou est réputée pour ses chutes de Dindéfelo, ses villages traditionnels Bassari et Bédik, et ses zones propices à l'écotourisme ;
- **Le Bassin arachidier** : composé des régions de Kaolack, Kaffrine, Fatick et Diourbel, est le centre agricole historique du pays, cette zone possède également un patrimoine religieux et culturel remarquable. À **Kaolack**, la ville sainte de **Médina Baye** attire de nombreux pèlerins musulmans. **Fatick** se distingue par ses traditions sérères, notamment à **Diakhao**, site historique et spirituel ;
- **Le Fouta et la vallée du fleuve Sénégal** : composé de Matam, Podor et Bakel, est située au nord-est, cette région est traversée par le fleuve Sénégal. Le **Fouta Toro**, berceau de la culture Toucouleur, conserve un riche héritage architectural islamique et colonial. De plus, les villes comme **Podor** et **Matam** offrent un cadre propice au tourisme patrimonial, fluvial et solidaire ;
- **Le Centre intérieur** : composé de Thiès, Louga et Tivaouane, la région de **Thiès**, en dehors de sa frange côtière, abrite des lieux religieux majeurs comme **Tivaouane**,

siège de la confrérie **Tidianiya**. Chaque année, des milliers de fidèles s'y rendent pour le **Maouloud**. Ce territoire, bien situé sur les axes routiers stratégiques du pays, constitue un foyer de **tourisme spirituel**.

Le tableau suivant propose une synthèse des principales zones non côtières touristiques, précisant leur département d'appartenance ainsi que leur localisation géographique :

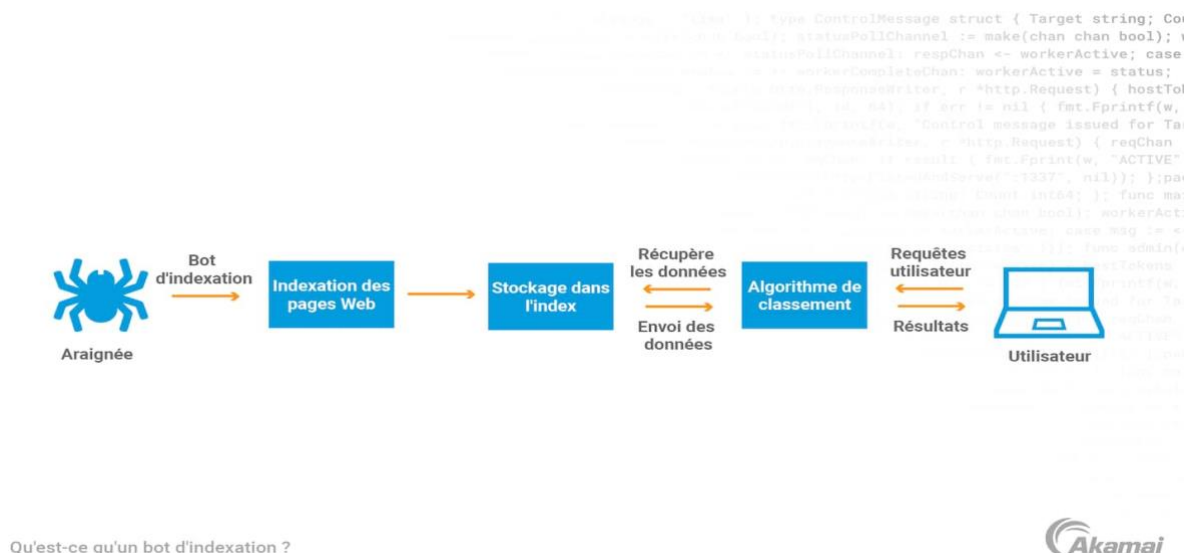
Tableau 2. Tableau des Zones non côtières du Sénégal

Zone touristique	Régions / Départements	Localisation géographique
Sénégal oriental	Tambacounda, Kédougou	Sud-Est
Bassin arachidier	Kaolack, Kaffrine, Fatick, Diourbel	Centre
Fouta et vallée du fleuve Sénégal	Matam, Podor, Bakel	Nord-Est, vallée du fleuve Sénégal
Centre intérieur	Thiès (Tivaouane), Louga	Centre-Nord / intérieur

1.3. Le Web crawling

1.3.1. Définition

Le **web crawling** encore appelé exploration du Web en français est un processus automatisé qui consiste à parcourir systématiquement les pages d'un site web, comme le ferait un internaute, mais à l'aide d'un programme appelé **crawler**. Autrement dit, c'est une **technique d'exploration automatique du web** qui permet de **collecter des liens et des contenus** en naviguant de page en page. De plus, un crawler est un programme de navigation utilisé par les moteurs de recherche. Il parcourt internet pour y analyser les contenus des sites, les archiver et les classer dans les pages de résultats selon leur pertinence [7]. Ces robots informatiques organisent les contenus web dans un grand ensemble appelé l'index.



Qu'est-ce qu'un bot d'indexation ?



Figure 1. **Schéma d'un Crawler (Bot d'indexation)**

Explications : Le **spider** (ou araignée) agit comme un explorateur du web : il parcourt les pages d'un site de manière systématique, un peu comme une araignée se déplaçant sur sa toile. Il repère et localise les informations utiles, puis les transmet à l'extracteur, chargé de les collecter de manière structurée. On peut dire que, pour accéder aux données, le **scraper** (ou grattoir) s'appuie sur l'index créé par le spider, ce qui lui permet de retrouver efficacement les pages pertinentes et d'en extraire les éléments recherchés pour l'utilisateur.

1.3.2. Fonctionnement de base

L'extraction de données sur le Web nécessite deux éléments : le **crawler** et le **scraper**. Le **crawler** est un algorithme d'intelligence artificielle qui parcourt le Web pour rechercher les données particulières requises en suivant les liens sur Internet. Un crawler est constitué d'un ensemble de scripts et d'algorithmes programmés pour accomplir des tâches spécifiques. Ces robots d'indexation parcourent automatiquement les pages web et leurs liens hypertextes, et ce, en continu, 24 heures sur 24 [8].

Le fonctionnement du crawler est relativement simple :

- il commence par visiter une page, l'explore, enregistre son contenu, puis suit les liens qu'il détecte pour accéder à d'autres pages.
- Il sauvegarde ces nouvelles pages, poursuit l'exploration des liens qu'elles contiennent, et répète ce processus indéfiniment. L'exploration débute toujours à partir d'URLs connues.

Lorsqu'un crawler arrive sur un site web, il réalise d'abord une phase appelée *crawl*. Cette étape est primordiale dans le processus de référencement sur les pages de résultats des moteurs de recherche. Durant cette phase, il analyse en détail les différents éléments d'une page web :

texte, images, liens, balises, titres, etc., d'où l'importance de bien structurer ces éléments pour optimiser le référencement naturel (SEO).

Le crawler vérifie également s'il a déjà exploré cette page auparavant. Dans ce cas, il compare la version actuelle avec la précédente pour détecter d'éventuelles modifications. Si la page a été mise à jour, le crawler actualise automatiquement son index avec les nouvelles informations. Après cette étape de crawling, la page est indexée par le moteur de recherche et peut alors être affichée aux internautes en réponse à leurs requêtes.

Cependant, lorsqu'un site ne souhaite pas que certains contenus soient indexés, il peut utiliser une balise meta « noindex » dans le code HTML. Cette balise indique aux robots de ne pas parcourir ni référencer la page dans les résultats de recherche.

1.3.3. Types de Crawling

1.3.3.1. Le crawling standard

Un crawler standard, ou robot d'indexation, est un programme automatisé conçu pour explorer le web de manière systématique en suivant les liens entre les pages. Autrement dit, il parcourt le Web en téléchargeant tous les documents qu'il rencontre. En effet, son objectif principal est de collecter et d'indexer l'ensemble des données disponibles sans filtrage préalable, contrairement aux Crawlers ciblés qui se concentrent sur des thèmes spécifiques. Ces outils, utilisés notamment par les moteurs de recherche comme Google, fonctionnent selon un principe d'exploration exhaustive c'est-à-dire ils parcourent en largeur ou en profondeur et stockent le contenu des pages visitées (HTML, PDF, etc.).

Bien que essentiels pour indexer le web, ils présentent certaines limitations : consommation importante de ressources comme la bande passante, la puissance de calcul, la collecte de données souvent non pertinentes, et nécessite de respecter les règles définies dans les fichiers robots.txt [9,10]. Des projets comme Common Crawl illustrent parfaitement cette approche large et non sélective du web crawling².

1.3.3.2. Le Crawler intelligent

Le crawler pour l'intelligence artificielle est conçu pour collecter massivement des données nécessaires à l'entraînement de modèles d'apprentissage automatique. Il peut récupérer des textes, des images, des vidéos ou des métadonnées pour former des datasets. Ces données sont

² [CS 276 : Recherche d'information et recherche sur le Web](#)

ensuite nettoyées et labellisées avant d'être utilisées pour des tâches comme la reconnaissance d'images, la génération de texte ou la classification automatique. Ce type de crawler joue un rôle fondamental dans les projets de NLP, vision par ordinateur ou assistant virtuel. Il est notamment utilisé par les grandes entreprises technologiques dans leurs projets d'intelligence artificielle. Ces crawlers alimentent la phase d'apprentissage des algorithmes.

1.3.3.3. Usages du Crawling

Le web crawling peut se présenter sous plusieurs formes, selon la méthode employée pour parcourir et collecter les données sur le Web. Parmi celles-ci, on peut citer [11] :

1.3.3.3.1. Le Crawler d'indexation:

Le crawler d'indexation est principalement utilisé par les moteurs de recherche comme Google, Bing ou Yahoo. Il a pour mission de parcourir automatiquement le web afin de découvrir et d'analyser les pages disponibles. En étudiant la structure HTML, les balises et les liens, il permet d'indexer les contenus dans des bases de données pour un classement dans les résultats de recherche. Plus une page est bien structurée et pertinente, plus elle aura de chances d'être bien positionnée. Sans ce type de crawler, une page ne peut pas apparaître sur les moteurs. Googlebot est l'un des exemples les plus connus.

1.3.3.3.2. Le Crawler pour le diagnostic SEO :

Le crawler de diagnostic SEO est un outil spécialisé dans l'analyse technique d'un site web en vue d'optimiser son référencement naturel. Il permet d'identifier les erreurs techniques telles que les liens brisés, les contenus dupliqués ou les problèmes de balisage. Il vérifie aussi la rapidité de chargement des pages, la profondeur du maillage interne, et d'autres aspects structurels. Ces diagnostics sont essentiels pour améliorer la visibilité d'un site dans les résultats des moteurs de recherche.

1.3.3.3.3. Le Crawler pour la veille concurrentielle:

Le crawler de veille concurrentielle est conçu pour surveiller les évolutions d'un marché donné. Il collecte des informations telles que les prix, les promotions, les nouveaux produits ou les changements de contenu chez les concurrents. Ces données sont ensuite utilisées pour réaliser des analyses comparatives et ajuster les stratégies commerciales. Il est très prisé dans des secteurs comme l'e-commerce, le tourisme ou l'hôtellerie. Grâce à ce type de crawler, une

entreprise peut rester compétitive en se basant sur des informations constamment mises à jour. En effet, cette collecte automatisée offre leurs permettra d'avoir une capacité d'anticipation précieuse c'est-à-dire l'identification des tendances émergentes, analyse des stratégies rivales pour une meilleure prise de décision stratégique.

1.3.3.3.4. Le Crawler pour le scraping de données:

Ce type de crawler est destiné à extraire automatiquement des données ciblées depuis des sites web. Il est souvent utilisé pour collecter des informations structurées comme les prix, les avis, les disponibilités ou les caractéristiques techniques de produits. Ces données sont ensuite stockées dans des bases de données pour des analyses ou intégrées dans des services comme des comparateurs. Le scraping est utile dans de nombreux domaines : veille, marketing, recherche ou e-commerce. Contrairement aux crawlers d'indexation, il ne parcourt pas tout le web mais cible des pages précises. Des outils comme Scrapy ou BeautifulSoup sont souvent utilisés à cette fin.

1.3.3.3.5. Le Crawler pour les archives (web archiving):

Le crawler d'archives [12] est utilisé pour conserver une copie des pages web à une date donnée. Il est souvent employé par des institutions ou des bibliothèques numériques dans le but de constituer une mémoire du web. Ces crawlers enregistrent le contenu visible, les images, les scripts et les fichiers intégrés afin de reproduire fidèlement la page originale. Ils sont utiles à des fins de recherche, de documentation ou pour préserver des sites menacés de disparition. L'un des plus célèbres services utilisant ce type de crawler est la Wayback Machine. Ce travail permet de retracer l'évolution d'un site dans le temps.

1.3.3.3.6. Le Crawler pour l'agrégation :

Le crawler d'agrégation permet de centraliser des contenus similaires issus de plusieurs sources. Il est très utilisé dans les plateformes comparatives, notamment dans les domaines de l'hôtellerie, des transports ou de l'e-commerce. Il collecte des informations telles que les prix, les notes ou la disponibilité, et les regroupe pour les présenter de manière synthétique à l'utilisateur. Grâce à ce type de crawler, les internautes peuvent comparer différentes offres rapidement et facilement. Il garantit aussi une actualisation régulière des données. Ce robot est la base technique des comparateurs de services.

1.3.3.3.7. Le Crawler pour le personnel:

Ce type de crawler est développé pour répondre à un besoin spécifique et souvent temporaire. Il peut s'agir par exemple d'extraire tous les articles d'un blog, de récupérer les PDF d'un site institutionnel ou de collecter des données pour un projet de recherche. Il est généralement conçu de manière artisanale avec des outils simples comme Python, BeautifulSoup ou Requests. De plus, ce crawler n'est pas destiné à être utilisé à grande échelle, mais remplit une mission précise. Il offre une grande flexibilité pour des besoins ponctuels. D'ailleurs, les étudiants ou chercheurs en font un usage fréquemment dans leurs travaux.

1.3.3.3.8. Le Crawler pour des tests de performance:

Le crawler de test est utilisé pour tester la capacité d'un site à gérer un grand nombre de requêtes simultanées. Il simule une forte charge pour observer la stabilité du site, les temps de réponse ou les risques de plantage. Ces tests sont souvent réalisés avant une mise en ligne ou lors de mises à jour majeures. Ils permettent de s'assurer que le site reste fonctionnel même en cas de forte affluence.

1.3.3.3.9. Le Crawler pour la veille commerciale

Ces types de crawlers se concentrent sur la collecte et l'analyse de données provenant de diverses sources en ligne comme les forums, les blogs, ou les plateformes d'avis. Ils permettent aussi une entreprise d'identifier les tendances et besoins émergents grâce à l'analyse des retours clients, de surveiller les stratégies et les lançements de produits des concurrents. A cela s'ajoute l'adaptation en conséquence de sa propre stratégie marketing.

Tableau comparatif des types de web crawling

Type de Crawler	Avantages	Limites
Crawling standard	Exploration large du web, utile pour l'indexation générale.	Consomme beaucoup de ressources, collecte parfois des données non pertinentes.
Crawler intelligent	Permet d'entraîner des modèles IA avec de grandes quantités de données.	Nécessite un nettoyage complexe et une labellisation des données.
Crawler pour l'indexation	Optimise le classement des pages sur les moteurs de recherche.	Ne collecte pas toujours les données utiles pour l'analyse métier.
Crawler SEO	Identifie les erreurs techniques et améliore le référencement.	Ne récupère pas de contenu sémantique ou utilisateur.
Crawler veille concurrentielle	Analyse des stratégies des concurrents, aide à l'adaptation marketing.	Peut violer les politiques de certains sites, nécessite une mise à jour fréquente.
Crawler scraping de données	Extrait des données précises pour analyses ou intégration.	Peut être bloqué par les sites ou poser des problèmes juridiques.
Crawler d'archives	Permet de conserver une mémoire du web.	Stockage volumineux, parfois obsolescence rapide des données.
Crawler d'agrégation	Facilite la comparaison d'offres issues de plusieurs sources.	Peut poser problème si les données ne sont pas homogènes ou synchronisées.
Crawler personnel	Flexibilité, adapté aux besoins spécifiques ponctuels.	Peu optimisé pour une grande échelle ou des usages répétés.
Crawler test de performance	Évalue la robustesse des sites sous forte charge.	Peut créer des charges fictives problématiques s'il est mal utilisé.
Crawler veille commerciale	Aide à capter les tendances du marché et surveiller la concurrence.	Traitement complexe de données issues de sources variées.

En somme, ce tableau met en lumière la diversité des types de crawlers, chacun étant conçu pour répondre à des objectifs bien spécifiques, allant de l'indexation générale à la veille concurrentielle, en passant par l'agrégation de données ou l'archivage web. Chaque type présente des avantages mais aussi des limites, notamment en matière de pertinence des données, de ressources nécessaires ou de respect des politiques d'accès.

1.4. Le web scraping

1.4.1. Définition

Le **web scraping**, encore appelé **extraction automatisée de données web**, est une technique qui permet de collecter des informations à partir de sites internet de manière automatisée. Cette méthode utilise des **scripts**³ ou des programmes pour extraire des données structurées, souvent en vue de les analyser ou de les réutiliser dans un autre contexte. D'ailleurs scraping vient du verbe anglais « To Scrap » qui signifie « gratter » en français. Ainsi nous pouvons dire que le web scraping est une méthode efficace pour collecter, structurer et stocker des informations provenant du web. Sa spécificité réside dans l'utilisation exclusive de programmes informatiques pour effectuer cette tâche. En d'autres termes, le web scraping correspond au processus d'extraction automatisée de données à partir de sites web. Les données récupérées sont ensuite transformées et exportées dans un format plus pratique pour l'utilisateur [13]. De plus, le Web Scraping¹ également appelé data scraping est une technique désignée pour l'extraction, la capture et la récolte d'informations provenant de sites web. Cette technique se concentre principalement sur la transformation de données semi-structurées (format HTML) sur le web en données structurées (base de données ou feuille de style) et utilisable. Elle permet ainsi de collecter efficacement des données pertinentes dont nous avons besoin. Ce processus est comparable à un copier-coller automatique offrant ainsi une efficacité renforcée et une diminution des risques d'erreurs.



Figure 2. Schéma du Web Scraping

³ Un **script** est un fichier contenant un ensemble d'instructions ou de commandes écrites dans un langage de programmation.

1.4.2. Fonctionnement de base

Le **scraper** est un outil spécifique créé pour extraire les données du site web. La conception du scraper peut varier considérablement en fonction de la complexité et de l'ampleur du projet, afin qu'il puisse extraire les données rapidement et avec précision [14]. Le web scraping, c'est un programme qui va chercher des pages sur un site web. Ensuite, il lit le code de la page (HTML) pour trouver les informations qui nous intéressent. Enfin, il met ces données en ordre (comme dans un fichier Excel ou une base de données) pour qu'on puisse les utiliser facilement. L'algorithme du web scraping commence par l'identification du site web cible et des données à extraire. Une analyse du code HTML permet de localiser précisément les informations recherchées à travers les URL. Ensuite, une requête est envoyée au site à l'aide d'un outil ce qui permet de récupérer son contenu HTML. Ce code est ensuite parcouru pour extraire les éléments pertinents en s'appuyant sur les balises et classes du site. Les données brutes collectées sont nettoyées afin de supprimer les éléments inutiles et de les organiser dans un format homogène. En effet, une fois prêtes, elles sont enregistrées dans des fichiers comme des CSV ou JSON, ou dans une base de données comme SQL ou MongoDB. Ces étapes permettent d'obtenir des données structurées prêtes à être exploitées.

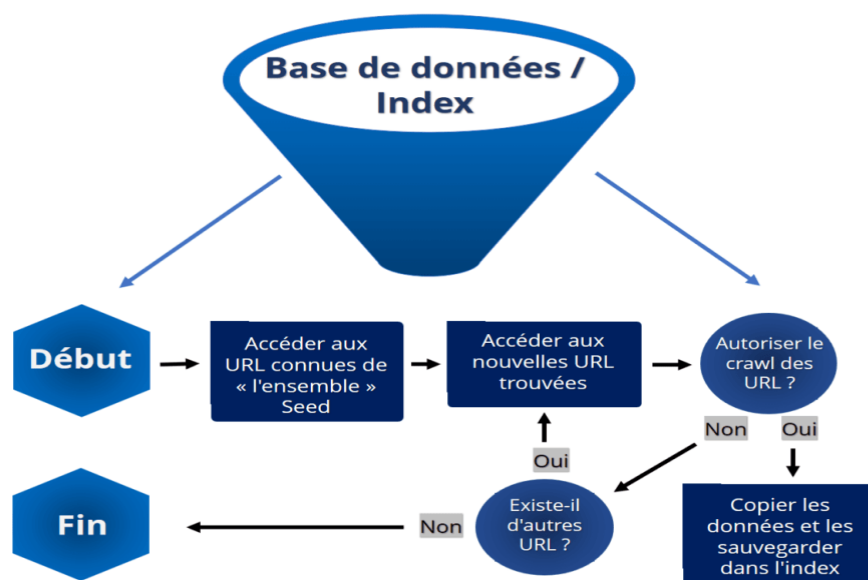


Figure 3. Schéma de fonctionnement du Web Scraping

1.4.3. Types de Scraping

Le **web scraping** peut prendre plusieurs formes, en fonction de la méthode utilisée pour récupérer et traiter les données. Parmi lesquelles on peut citer :

- **Scraping basé sur des requêtes HTTP** : c'est une méthode courante et simple pour extraire des données à partir de sites web statiques. Le programme envoie une requête HTTP (par exemple, une requête GET⁴) pour accéder à une page web. Ensuite, le serveur répond avec le contenu de la page, généralement sous forme de code HTML. Ainsi le programme analyse cet HTML pour trouver les données souhaitées ;
- **Scraping basé sur l'analyse du DOM** : Le DOM est une représentation hiérarchique d'une page web sous forme d'arbre. Chaque élément HTML (balises, attributs, texte, etc.) est un "nœud" dans cet arbre. Le scraping basé sur l'analyse du DOM est une méthode qui consiste à extraire des données directement en explorant et en manipulant la structure HTML d'une page web. Ainsi l'objectif de l'analyse du DOM est de parcourir cette structure pour extraire des éléments spécifiques comme du texte des attributs, des tableaux ou des listes ;
- **Scraping basé sur des APIs** : Une API est un service proposé par un site ou une application pour permettre à des développeurs d'accéder à certaines données ou fonctionnalités. A cet effet, le scraping basé sur les APIs consiste à exploiter les interfaces de programmation (APIs) fournies par certains sites web pour accéder directement aux données de manière structurée et organisée. Contrairement au scraping classique qui analyse le HTML d'une page, cette méthode utilise des requêtes spécifiques à l'API pour récupérer des informations, souvent au format **JSON** ou **XML**. C'est une méthode rapide, efficace et souvent éthique pour récupérer des données. D'ailleurs le programme envoie une requête HTTP (souvent de type GET, POST, etc.) à une URL spécifique de l'API. Ensuite le serveur répond avec les données demandées, généralement dans un format facile à lire comme JSON ou XML ;
- **Scraping basé sur des outils spécialisés** : Il se base sur des frameworks comme *Scrapy* ou des plateformes Octoparse, automatise la collecte de données, adapté aux projets complexes couvrant plusieurs sites. Ces outils peuvent gérer efficacement les tâches ;
- **Le scraping dynamique avec JavaScript** : permet d'extraire des données sur des pages où le contenu est généré par JavaScript, souvent après le chargement initial du HTML. Pour cela, on utilise des outils qui simulent un navigateur web, capables d'exécuter le JavaScript et d'afficher la page comme le ferait un utilisateur. Une fois la

⁴Une **requête GET** est un type de requête HTTP utilisé pour demander des données d'un serveur.

page entièrement chargée, les données peuvent être récupérées comme sur une page statique ;

- **Scraping de fichiers ou flux intégrés** : Le scraping de fichiers ou flux intégrés consiste à extraire des données directement depuis des fichiers disponibles sur un site (comme CSV, PDF, ou Excel) ou des flux structurés tels que RSS ou XML. Cette méthode simplifie l'accès aux informations déjà organisées pour un usage immédiat ;
- **Scraping par machine learning ou reconnaissance visuelle** : utilise des techniques avancées pour extraire des données difficiles à récupérer avec des méthodes classiques, comme celles contenues dans des images (tableaux, graphiques, etc.). Grâce à des outils comme l'OCR⁵ () et des algorithmes d'intelligence artificielle, cette méthode peut analyser des contenus non standards, comme des PDF, captures d'écran ou interfaces complexes. Elle est particulièrement utile quand les données sont cachées dans des formats visuels ou dynamiques [13,14].

⁵C'est une technologie qui permet de convertir des textes présents dans des images, qui lit le texte dans les images.

Tableau 3. Tableau des différentes formes de scraping

Type	Avantages	Limites
Scraping basé sur des requêtes HTTP	Facile à mettre en œuvre pour extraire des données de sites web statiques, accès rapide au contenu d'une page web en envoyant une requête http	Ne fonctionne pas bien avec les sites dynamiques ou ceux nécessitant des sessions authentifiées.
Scraping basé sur l'analyse du DOM	Peut être utilisé pour cibler des données spécifiques sans avoir besoin d'extraire tout le contenu HTML	Peut-être complexe à mettre en œuvre pour des pages avec un DOM très dynamique ou complexe.
Scraping basé sur des APIs	Accède aux données dans un format organisé (JSON, XML), Réduit le risque d'erreurs par rapport aux méthodes de scraping	Dépend de la disponibilité et des limitations de l'API (quotas, restrictions d'accès).
Scraping dynamique avec JavaScript	Permet d'extraire des données générées par JavaScript après le chargement initial de la page, Utilise des outils qui imitent le comportement d'un navigateur	Plus lent et nécessite plus de ressources ; peut être bloqué par certains sites.
Scraping de fichiers ou flux intégrés	Permet d'extraire facilement des données à partir de fichiers disponibles (CSV, PDF), Données déjà organisées pour une utilisation immédiate	Limité aux formats disponibles ; peut ne pas contenir toutes les informations nécessaires.
Scraping basé sur des outils spécialisés	Facilite la collecte de données à grande échelle sans nécessiter de codage complexe	Peut nécessiter une courbe d'apprentissage et une configuration initiale plus complexe.
Scraping par machine learning	Capable d'extraire des données difficiles à récupérer, Utilise l'intelligence artificielle pour analyser et interpréter les données	Nécessite des compétences techniques avancées et peut être coûteux en ressources informatiques.

En somme, ce tableau résume les différentes méthodes de scraping, leurs descriptions et leurs limites respectives, offrant ainsi une vue d'ensemble utile pour choisir la méthode appropriée en fonction des besoins spécifiques d'un projet de collecte de données. Bien que chaque méthode de scraping ait ses propres avantages et inconvénients, nous pouvons dire que le scraping basé sur des APIs se distingue par sa capacité à fournir un accès rapide, structuré et éthique aux données.

1.5. Web Crawling Vs Web Scraping

Les deux processus sont souvent combinés ensemble car le crawling découvre des URLs pertinents et le scraping extrait les données depuis ces URLs. Toutefois, ils présentent quelques différences à savoir:

Tableau 4. Tableau de comparaison entre le web scraping et le web crawling

Critères	Web Scraping	Web crawling
Définition	Extraction ciblée de données spécifiques à partir d'une page web	Exploration automatique et systématique de nombreuses pages web
Objectif principal	Récupérer des informations précises (ex : prix, avis, textes)	Indexer et explorer le web pour découvrir des URLs (ex : moteurs de recherche).
Outils utilisés	<i>Outils</i> comme BeautifulSoup, Scrapy.	Crawlers (robots d'indexation comme Googlebot).
Processus	1.Télécharge le HTML 2.Parse les données 3.Extrait et structure les données	1.Analyse robot.txt 2.Suit les liens 3.Stocke les URLs
Données générées	Données structurées (CSV, JSON, etc)	Liste d'URLs indexées (Métadonnées)

En résumé, le web scraping est utilisé pour extraire des données spécifiques à partir de pages web ciblées, tandis que le web crawling est utilisé pour explorer, indexer des informations globales de plusieurs sites web.

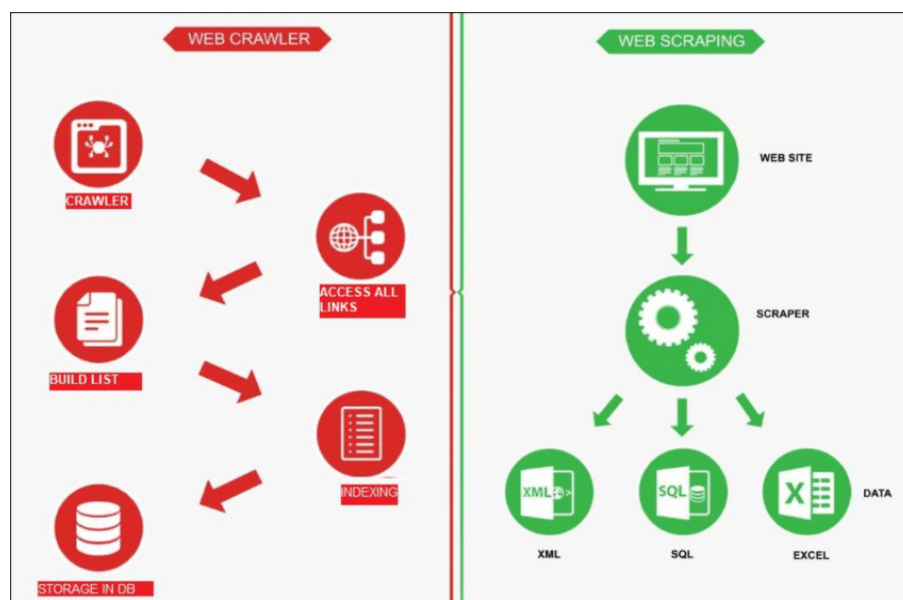


Figure 4. Schéma du Web Crawling & du Web Scraping

En somme, ce premier chapitre nous a permis de mieux comprendre le cadre général du projet en posant les bases théoriques et contextuelles nécessaires. Nous avons exploré la richesse

touristique du Sénégal ainsi que les principes du Web crawling et du Web scraping, deux techniques essentielles pour automatiser la collecte d'informations sur les hébergements. Cette partie nous a également permis de saisir en quoi ces deux approches se complètent et pourquoi elles sont adaptées à la mise en place d'un comparateur d'offres hôtelières dédié aux destinations du Sénégal.

Chapitre 2 : Analyse de l'existant et positionnement

2.1. Analyse de l'existant

Les plateformes de réservations hôtelières jouent un rôle essentiel dans le processus de réservation en ligne, car ils offrent aux utilisateurs la possibilité de comparer facilement les prix et les options disponibles mais aussi leurs limites [16]. D'ailleurs parmi les leaders du secteur nous avons entre autres :

2.1.1. Booking.com :

Booking⁶ est l'une des plus grandes agences de voyages en ligne au monde, offrant aux utilisateurs la possibilité de rechercher et de réserver une large gamme d'hébergements, allant des hôtels aux appartements, en passant par des maisons de vacances et d'autres types de logements. Fondée en 1996 à Amsterdam, elle permet aux partenaires d'hébergement de gérer leurs propres prix et disponibilités. Booking.com est connu pour sa large sélection d'options d'hébergement, son interface conviviale et son support client 24/7 dans plusieurs langues. Le site ne facture pas de frais de réservation et offre une garantie de prix ;

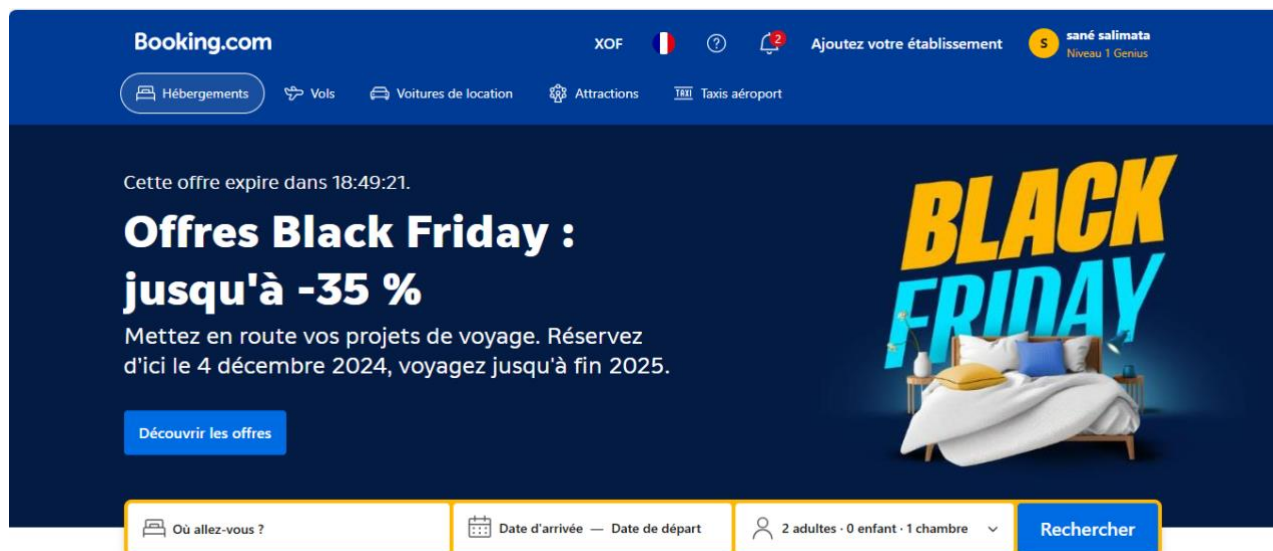


Figure 5. Tableau de bord de Booking. com

2.1.2. HotelsCombined :

HotelsCombined⁷ est un métamoteur de recherche compare les prix de plusieurs sites de réservation, permettant aux utilisateurs de trouver les meilleures offres d'hôtels. Il agrège les résultats pour permettre aux utilisateurs de voir les prix et les options disponibles sur différentes plateformes en un seul endroit. HotelsCombined se concentre sur la transparence des prix et n'ajoute pas de frais supplémentaires lors des réservations ;

⁶ <https://www.booking.com>

⁷ <https://www.hotelscombined.com>

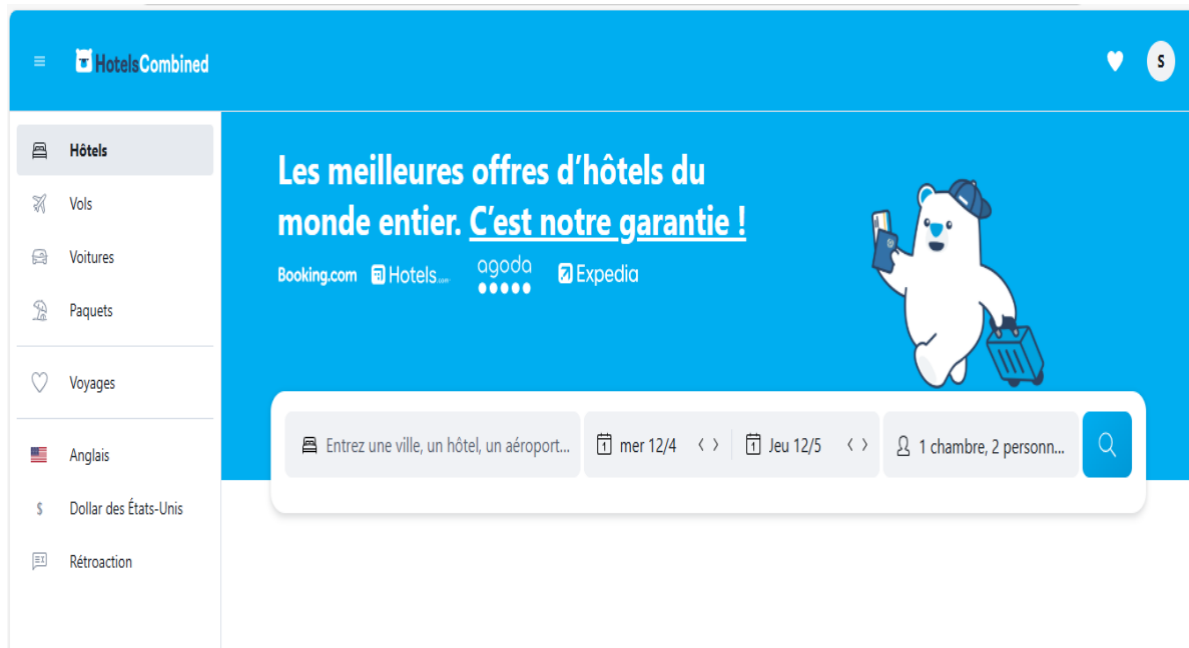


Figure 6. Tableau de bord de HotelsCombined

2.1.3. Trivago :

Trivago⁸ est un métamoteur populaire qui aide les utilisateurs à comparer les prix des chambres d'hôtel sur divers sites. Les utilisateurs peuvent rechercher des hôtels en fonction de divers critères tels que le prix, la localisation et les avis des clients. Trivago ne vend pas directement des chambres d'hôtel, mais redirige les utilisateurs vers d'autres sites pour effectuer leur réservation. Cela permet aux voyageurs de trouver la meilleure offre disponible sur le marché.

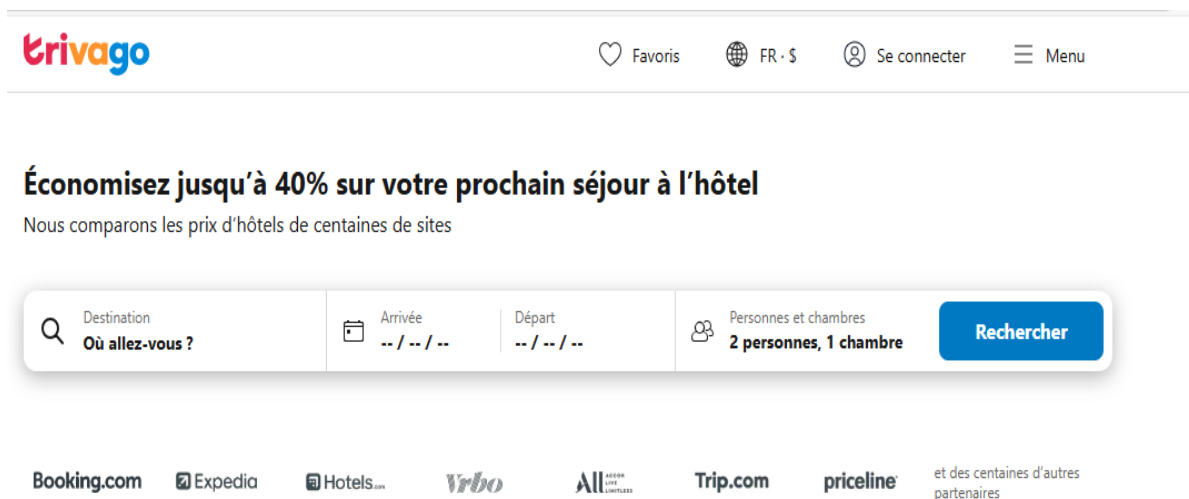


Figure 7. Tableau de bord de Trivago

2.1.4. KAYAK :

⁸ <https://www.trivago.com>

Kayak⁹ permet aux utilisateurs de rechercher et de comparer des prix pour des hôtels, mais aussi pour des vols et des locations de voitures. Il est particulièrement utile pour ceux qui cherchent à combiner plusieurs aspects de leur voyage en une seule recherche. Il offre également des fonctionnalités telles que l'alerte prix, qui informe les utilisateurs lorsque les tarifs changent pour leurs destinations choisies ;



Figure 8. Tableau de bord de Kayak

2.1.5. TripAdvisor :

TripAdvisor¹⁰ est principalement connu comme un site d'avis où les utilisateurs peuvent consulter des commentaires sur des hôtels, restaurants et attractions touristiques. Il permet également aux utilisateurs de comparer les prix des hôtels et de réserver directement via sa plateforme, offrant ainsi une vue d'ensemble des options disponibles ;

⁹ <https://www.trivago.com>

¹⁰ <https://www.tripadvisor.com>

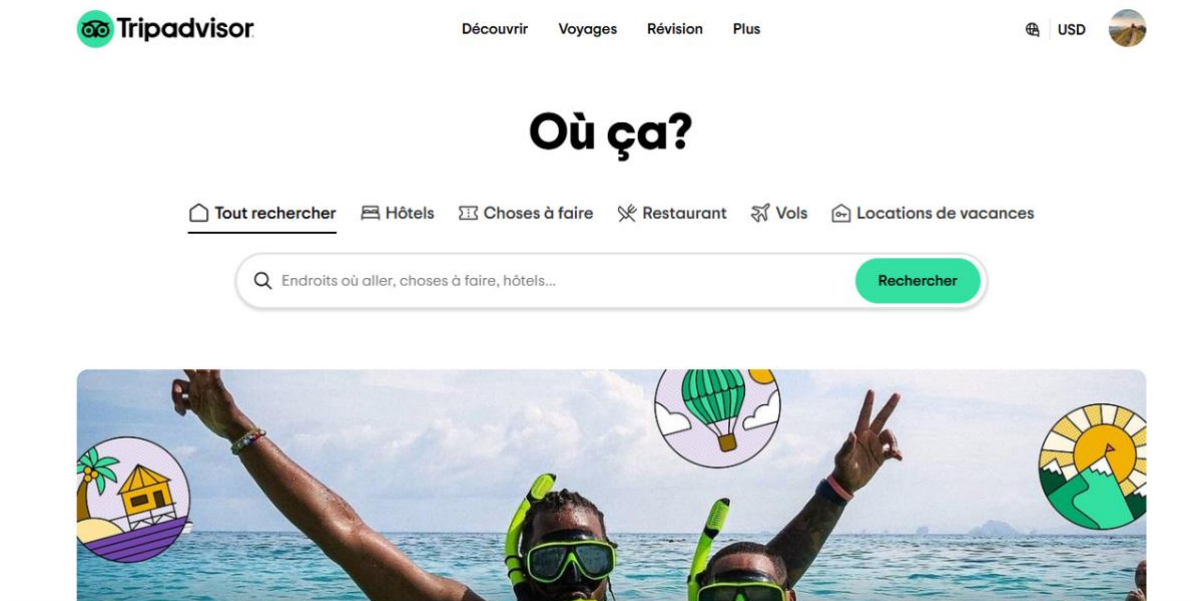


Figure 9. Tableau de bord de TripAdvisor :

2.1.6. Expedia :

Expedia¹¹ est l'une des plus grandes agences de voyages en ligne, offrant une plateforme pour rechercher et réserver des vols, hôtels, locations de voitures et activités. Fondée en 1996, Expedia permet aux utilisateurs de planifier et réserver l'ensemble de leur voyage sur une seule plateforme. Elle offre également des réductions pour les réservations groupées et un programme de fidélité pour récompenser les clients réguliers ;

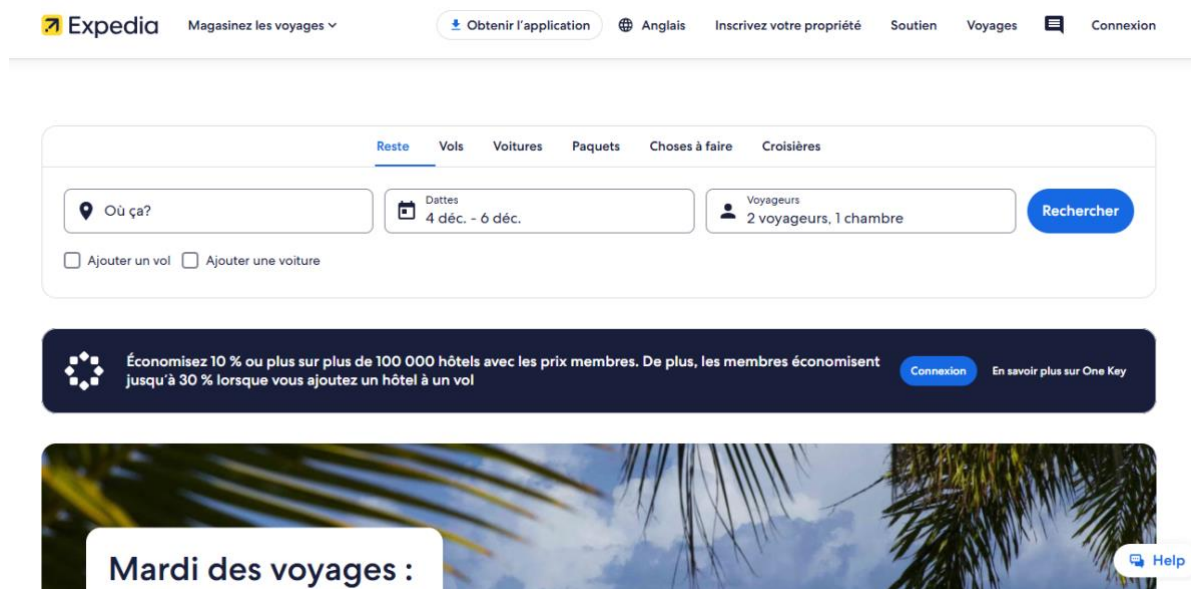


Figure 10. Tableau de bord de Expedia :

2.1.7. Agoda :

¹¹ <https://www.expedia.com>

Agoda¹² est une plateforme de réservation en ligne spécialisée dans les hôtels et logements, particulièrement en Asie. Fondée en 2005, elle propose également des locations de vacances et des activités. Agoda se distingue par ses offres exclusives et ses tarifs compétitifs, ainsi que par son programme Genius qui offre des réductions supplémentaires aux membres ;

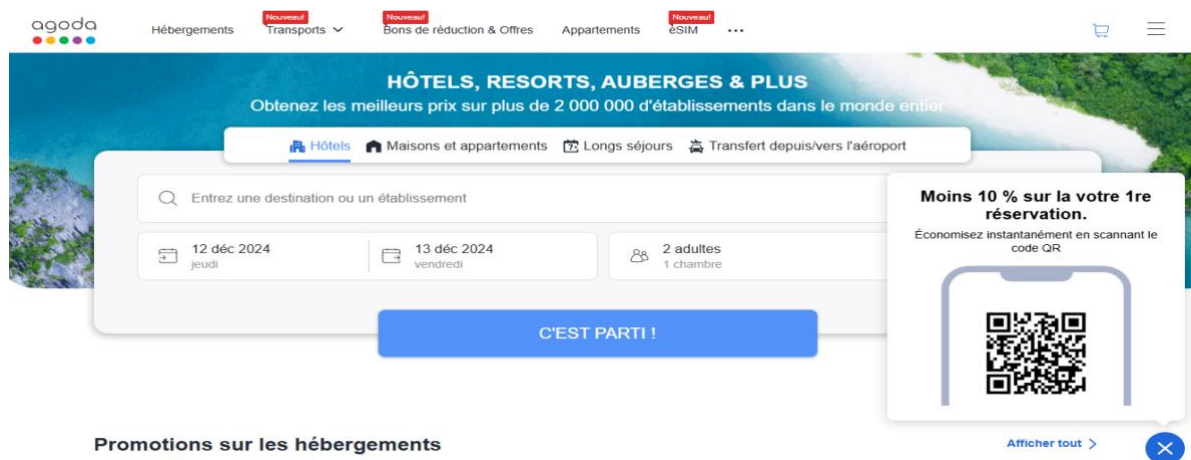


Figure 11. Tableau de bord de Agoda

2.1.8. Amadeus Hotel Plus :

Amadeus Hotel Plus¹³ est un système de distribution global (GDS) permettant aux agences de voyages d'accéder à un large inventaire d'hôtels. C'est une solution technologique destinée aux professionnels du secteur hôtelier, permettant la gestion des réservations et l'optimisation des ventes. Amadeus fournit également des outils d'analyse pour aider les hôtels à comprendre le comportement des clients et à améliorer leur stratégie commerciale ;

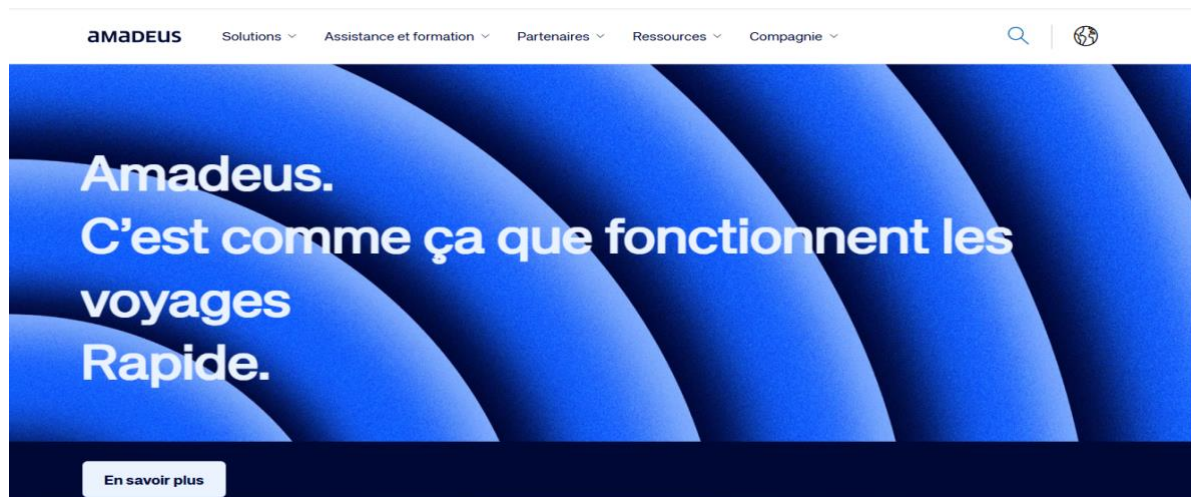


Figure 12. Tableau de bord de Amadeus Hotel Plus

2.1.9. Etraviax :

¹² <https://www.agoda.com>

¹³ <https://amadeus.com>

Etraviax¹⁴ est une plateforme spécialisée qui offre une API pour permettre aux agences de voyages d'accéder à un large éventail d'hébergements. Elle est moins connue que certains autres agrégateurs, mais il se concentre sur la fourniture d'offres compétitives pour les réservations d'hôtels en Europe. Il propose une interface simple pour comparer les prix et trouver rapidement des options d'hébergement adaptées aux besoins des voyageurs ;



Figure 13. Tableau de bord de Etraviax

2.1.10. HRS (Hotel Reservation Service) :

HRS (Hotel Reservation Service)¹⁵ est une plateforme dédiée à la réservation d'hôtels principalement pour les voyageurs d'affaires, offrant des tarifs négociés. Il propose une large gamme d'options d'hébergement avec un accent particulier sur le service client et l'optimisation des coûts pour les entreprises ;

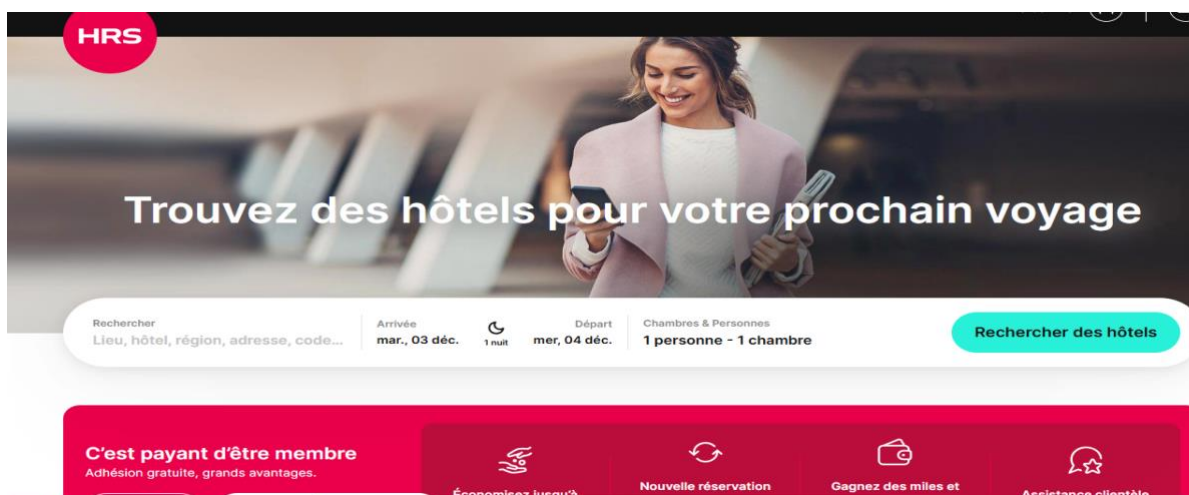


Figure 14. Tableau de bord de Hotel Reservation Service

2.1.11. Priceline :

¹⁴ <https://www.etraviax.com>

¹⁵ <https://www.hrs.com>

Priceline¹⁶ est une agence de voyages en ligne qui permet aux utilisateurs de réserver des hôtels, vols et voitures, se distinguant par ses offres "Express Deals". Priceline est une plateforme qui permet aux utilisateurs de réserver des hôtels, vols et voitures à des tarifs réduits grâce à son modèle "Name Your Own Price" (Nommez votre propre prix). Fondée en 1997, elle se concentre sur l'obtention du meilleur rapport qualité-prix pour ses clients ;

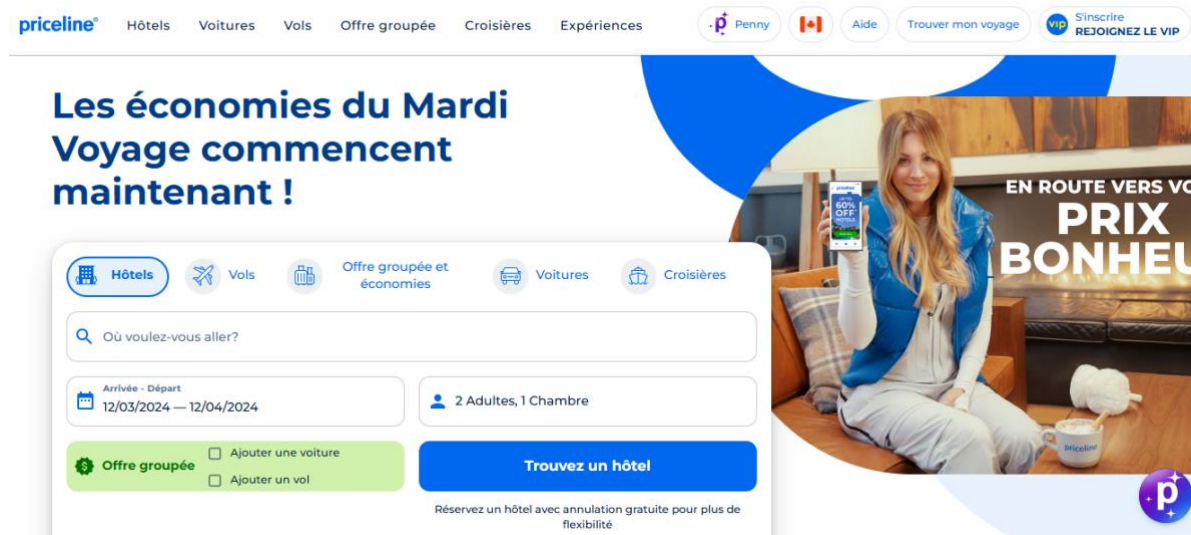


Figure 15. Tableau de bord de Priceline

2.1.12. Amadeus Value Hotels :

Amadeus Value Hotels¹⁷ permet aux agences de voyages d'accéder à un large inventaire hôtelier via le système GDS Amadeus. Elle offre une interface facile à utiliser pour rechercher et réserver des hôtels tout en fournissant des informations détaillées sur chaque propriété ;

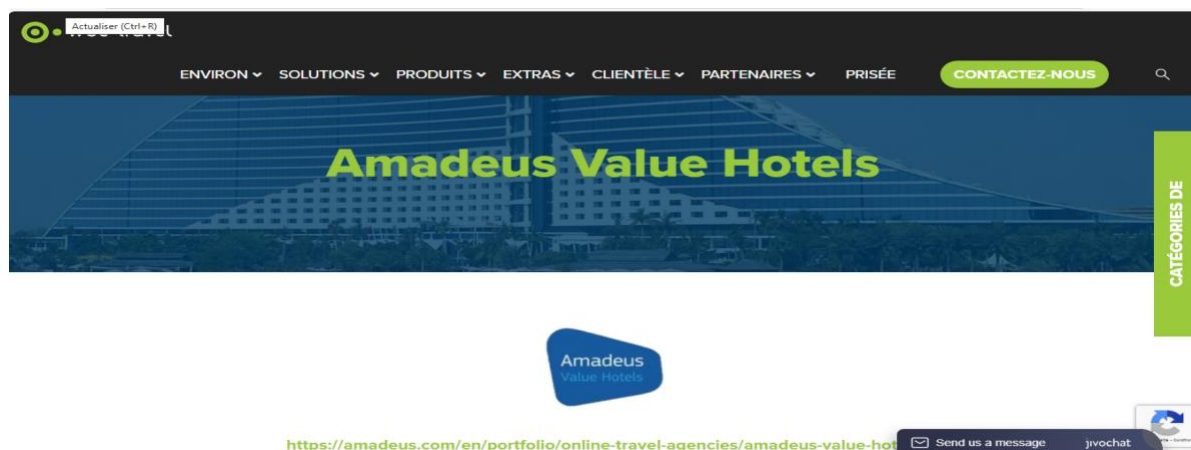


Figure 16. Tableau de bord de Amadeus Value Hotels

2.1.13. Airbnb :

¹⁶ <https://www.priceline.com>

¹⁷ <https://amadeus.com>

AirBnb¹⁸ est une plateforme en ligne qui permet aux particuliers de louer des logements (maisons, appartements, chambres) à court terme. Elle met en relation des hôtes, qui proposent leurs biens, et des voyageurs, à la recherche d'hébergements souvent plus personnalisés et économiques que les hôtels. Airbnb offre une large gamme d'options, allant des logements modestes aux propriétés de luxe. La plateforme est connue pour sa simplicité d'utilisation, son système de réservation sécurisé et ses évaluations mutuelles entre hôtes et voyageurs ;

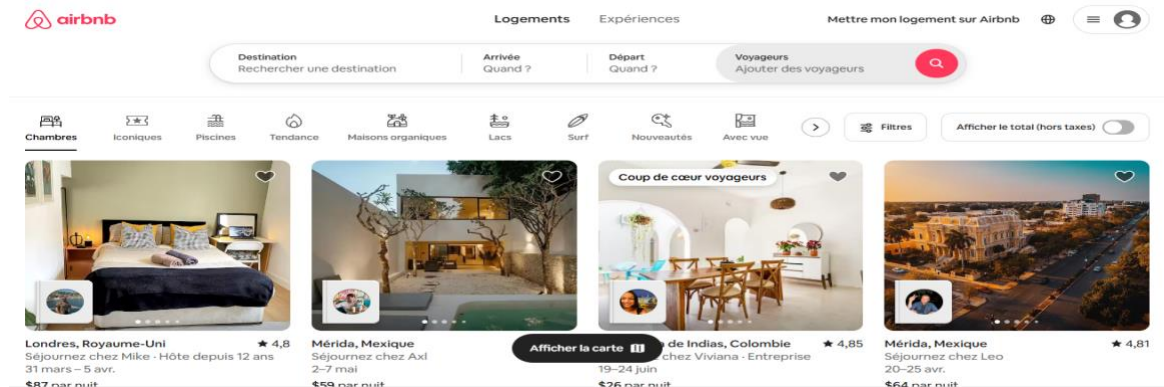


Figure 17. Tableau de bord de AirBnb

2.1.14. Cozycozy :

Cozycozy¹⁹ est un métamoteur de voyage spécialisé dans la comparaison d'hébergements, offrant une vision exhaustive du marché avec plus de 15 millions de solutions d'hébergement répertoriées comme les hôtels, les appartements dans 200 pays. En effet, la plateforme se distingue par sa capacité à scanner simultanément l'ensemble des sites de réservation en ligne, permettant aux utilisateurs d'accéder en un seul point à toutes les offres disponibles pour leurs dates de séjour ;



Figure 18. Tableau de bord de Cozycozy

2.1.15. Skyscanner :

¹⁸ <https://www.airbnb.com>

¹⁹ [Location vacances, Airbnb, Hôtel en France & ailleurs | cozycozy](#), (consulté le 11Mai 2025)

Skyscanner²⁰ est un métamoteur de voyage initialement spécialisé dans la comparaison de vols, avant d'élargir son offre aux hébergements et locations de voiture. D'ailleurs, la plateforme permet aux utilisateurs de comparer diverses options en fonction de critères précis comme les dates, la destination, le nombre de voyageurs et d'affiner leur recherche grâce à des filtres avancés (type d'hébergement, équipements, politiques d'annulation, etc.) ;

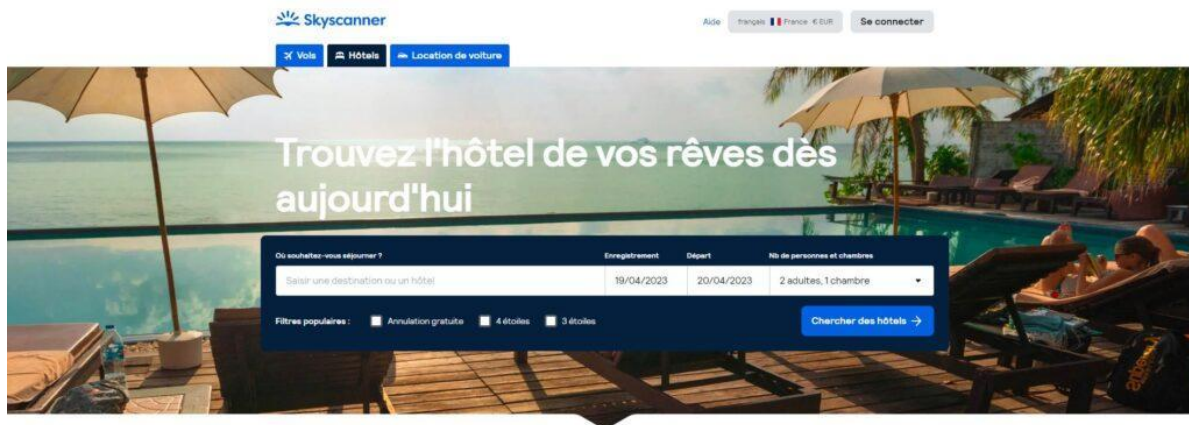


Figure 19. Tableau de bord de Skyscanner

2.1.16. Liligo :

Liligo²¹ est un métamoteur de voyage français lancé en 2006, spécialisé dans la comparaison de vols, hébergements et locations de voiture. En effet, la plateforme permet aux utilisateurs d'effectuer des recherches ciblées parmi une large gamme d'options d'hébergement comme des hôtels, des appartements etc. Il s'applique sur divers filtres à savoir : le prix, les notations clients et la localisation géographique ;



Figure 20. Tableau de bord de Liligo

2.1.17. eDreams :

²⁰ [Skyscanner | Trouvez le vol pas cher, économisez temps et argent !](#), (consulté le 1& Mai 2025)

²¹ [Comparateur de vol et Billets d'avion pas cher | LILIGO](#)

L'agence de voyages eDreams²² est une plateforme de voyage en ligne d'origine américaine, actuellement intégrée à un groupe espagnol, spécialisée dans la réservation de vols, séjours et services complémentaires à tarifs compétitifs. L'interface propose une recherche modulaire permettant de combiner différents services comme des vols seuls, des hôtels, ou des packages vol+hébergement, avec un système de filtres avancés pour sélectionner le type d'hébergement souhaité (hôtel, appartement, villa, etc.). On peut aussi préciser, si on le souhaite, réserver un appartement, un hôtel, une auberge de jeunesse, une maison d'hôtes, un bateau hôtel, une villa, un camping, une résidence étudiante ;

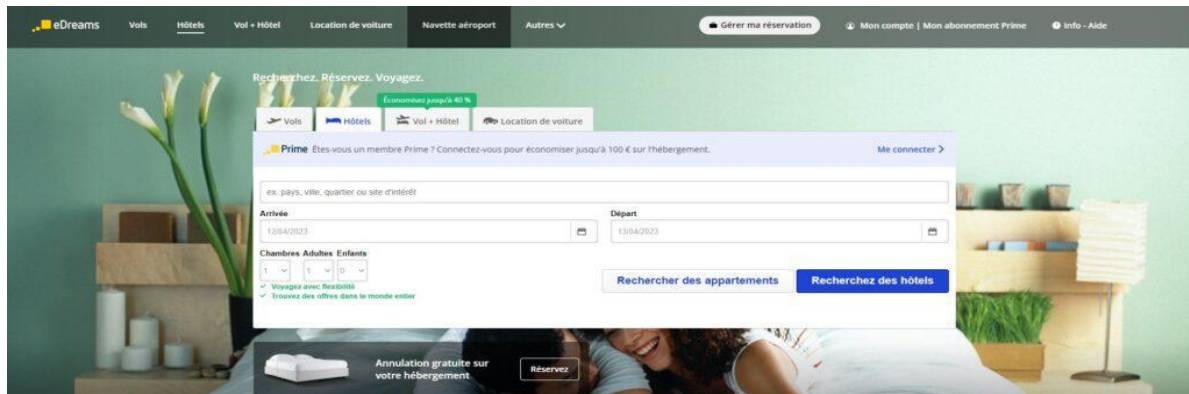


Figure 21. Tableau de bord de eDreams

2.1.18. Momondo :

Initialement conçu pour la comparaison de vols, Momondo²³ s'est progressivement imposé comme une plateforme polyvalente couvrant l'ensemble des besoins voyage comme hôtels, locations privées, transports terrestres etc. De plus, la plateforme se distingue par trois atouts principaux à savoir : une interface ergonomique permettant une comparaison prix immédiate, une politique de flexibilité des réservations, et un accès gratuit à l'ensemble des services ;

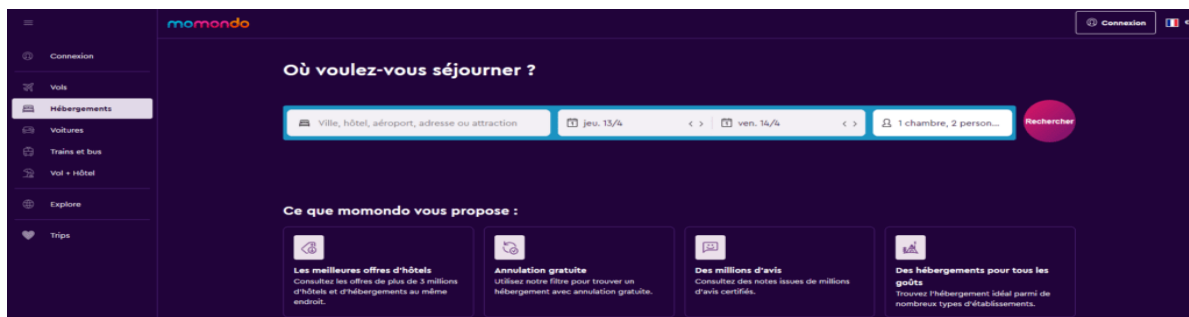


Figure 22. Tableau de bord de Momondo

2.1.19. Hotels.com :

²² [eDreams - Réservation de vols, voyage pas cher & séjour dernière minute](#)

²³ [Comparateur de vol. Billets d'avion & vols pas chers | momondo](#)

Hotels.com²⁴ se positionne comme une plateforme internationale de réservation hôtelière, basée à Dallas (États-Unis), offrant un catalogue étendu d'établissements à l'échelle mondiale. Son interface utilisateur intuitive permet une recherche simplifiée : l'utilisateur renseigne simplement la destination, les dates de séjour et le nombre de voyageurs pour accéder aux options disponibles. La plateforme se distingue par son programme de fidélité avantageux, particulièrement adapté aux voyageurs fréquents. Cette combinaison de choix diversifié en fait un acteur majeur du secteur des OTA (Online Travel Agencies).

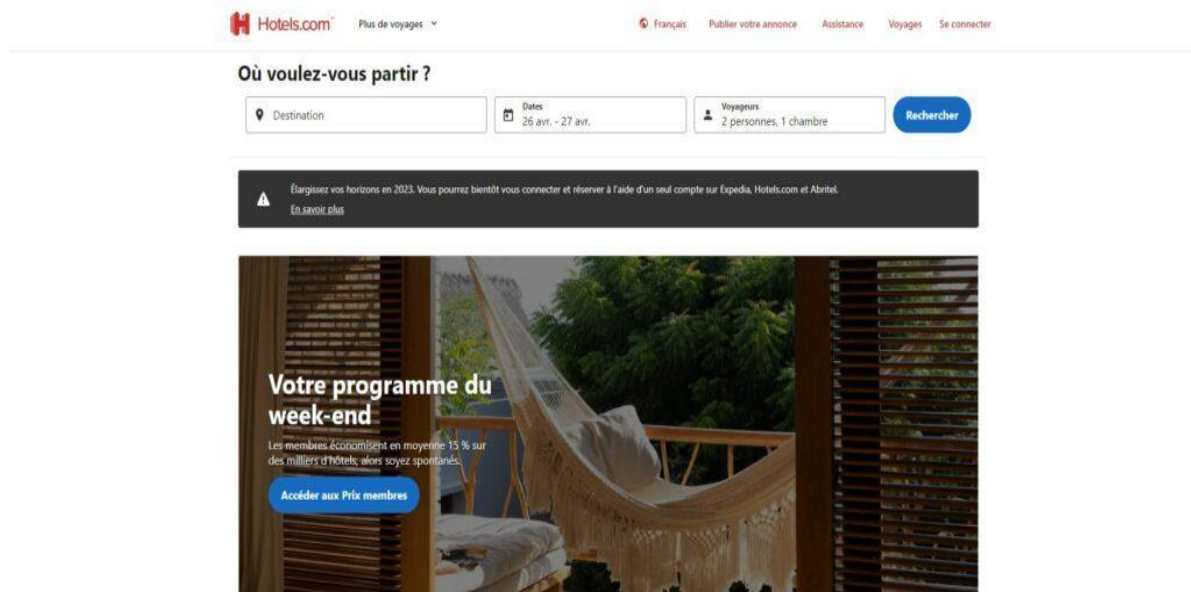


Figure 23. Tableau de bord de Hotels.com

D'ailleurs le tableau ci-dessous montre les avantages et les limites de certaines plateformes leaders prêtes à l'emploi.

²⁴ [Hotels.com - Deals & Discounts for Hotel Reservations from Luxury Hotels to Budget Accommodations](#)

Tableau comparatif des différentes plateformes leaders :

Plateforme	Avantages	Limites
Booking.com	Large sélection, pas de frais de réservation, interface conviviale	Parfois prix plus élevés que les concurrents
HotelsCombined	Compare plusieurs sites, pas de frais cachés	Ne permet pas de réserver directement
Trivago	Comparaison efficace, nombreux filtres	Redirection vers d'autres sites
KAYAK	Recherche multi-produits, alertes de prix	Redirige souvent, interface parfois chargée
TripAdvisor	Commentaires d'utilisateurs, comparateur intégré	Moins fiable pour la réservation directe
Expedia	Réservations groupées, programme de fidélité	Peut facturer des frais supplémentaires
Agoda	Spécialisé en Asie, bons tarifs, programme Genius	Service client parfois critiqué
Amadeus Hotel Plus	Accès large inventaire, outils d'analyse	Réservé aux professionnels
Etraviatx	Offres compétitives, simple d'utilisation	Moins connu, moins de choix
HRS	Tarifs négociés, optimisé pour entreprises	Moins orienté grand public
Priceline	Offres Express, modèle « Name Your Price »	Pas de détails complets avant achat
Amadeus Value Hotels	Interface intuitive, bon support agences	Moins accessible au grand public
Airbnb	Grande variété, options personnalisées	Qualité inégale, frais parfois élevés
Cozycozy	Vue exhaustive, plus de 15M offres	Plateforme relativement récente
Skyscanner	Filtres puissants, multi-produits	Peut manquer de clarté sur certains tarifs
Liligo	Interface claire, bon tri des résultats	Moins de partenaires comparés
eDreams	Offres combinées, tarifs compétitifs	Frais cachés possibles
Momondo	Interface ergonomique, services gratuits	Informations parfois redondantes
Hotels.com	Programme fidélité avantageux, interface simple	Certaines offres pas toujours mises à jour

En somme, choisir le bon type de Comparateurs dépendra des besoins spécifiques des utilisateurs, qu'ils soient voyageurs individuels ou professionnels du secteur.

2.2. Positionnement

L'essor des technologies numériques a transformé le secteur du tourisme, notamment à travers la création de comparateurs de destinations hôtelières. Ces outils permettent aux utilisateurs de comparer les offres d'hébergement, d'analyser les prix et d'accéder à des avis clients, facilitant ainsi la prise de décision lors de la planification de voyages. D'ailleurs, on note la transformation numérique bouleverse profondément le secteur du tourisme et de l'hôtellerie, offrant des opportunités inédites tout en imposant de nouveaux défis. Des efforts sont nécessaires pour soutenir les hôtels dans leur transition numérique, tirer parti des TIC pour moderniser les pratiques touristiques, et exploiter des innovations telles que les destinations intelligentes. C'est dans ce sens que plusieurs études sont menées sur différents comparateurs de destinations hôtelières. Nous avons entre autres :

- **Les Avis sur les réservations en ligne :** Certains [17] auteurs mettent l'accent sur les avis des clients sur les réservations en ligne. D'ailleurs, les hôteliers et restaurateurs subissent une forte pression due aux avis laissés par les clients sur des plateformes comme TripAdvisor ou Booking.com. Ces avis sont devenus essentiels pour les tests des services hôteliers, contrairement à un produit physique. Les clients ne se fient plus uniquement à la communication des établissements, qu'ils perçoivent comme de la publicité, mais préfèrent consulter les expériences partagées par d'autres consommateurs. Ces commentaires les rassurent et influencent directement leur choix final. Des études montrent que lorsque les avis en ligne mentionnent la qualité de service, ils ont un impact plus fort sur les décisions²⁵. D'autres recherches [18] montrent que la majorité des voyageurs évitent systématiquement les hôtels ayant reçu des commentaires négatifs, et ce même avec leur classement en étoiles. Les avis clients constituent en effet un critère décisif dans le processus de réservation, jouant un rôle clé dans la construction de la réputation numérique des établissements. D'ailleurs un avis positif est perçu par les consommateurs comme une garantie de qualité, renforçant leur confiance et leur intention d'achat. À l'inverse, des évaluations négatives peuvent conduire à un manque de confiance des touristes et entacher l'image de l'étoile. De plus la figure ci-dessous montre la répartition des avis en ligne sur les hôtels par site d'évaluation dans le monde en 2017.

²⁵ [Les auteurs des avis en ligne, leurs profils et leurs influences sur les choix d'un hôtel - DUMAS - Dépôt Universitaire de Mémoires Après Soutenance](#), (consulté le 10 Mai 2025)

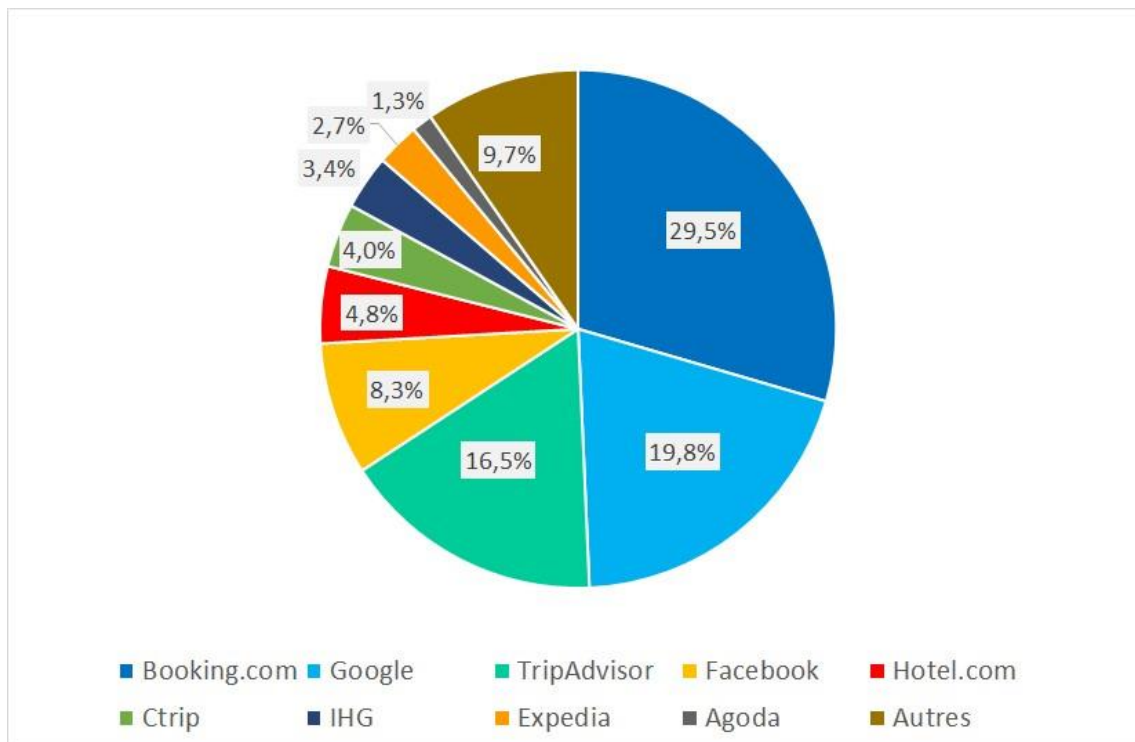


Figure 24. Répartition des avis en ligne des hôtels par site d'évaluation dans le monde en 2017²⁶

Nous notons que cette figure illustre la répartition des avis en ligne que l'on pouvait trouver sur les hôtels parmi les principaux sites d'évaluation dans le monde en 2017. Booking.com regroupait, cette année-là, près de 30 % du total des commentaires présents en ligne ayant pour but d'évaluer un établissement hôtelier.

- **La Révolution digitale du secteur hôtelier** : Une autre étude [19,20] se concentre sur la révolution digitale du secteur hôtelier. En effet, l'arrivée des OTA (comme Booking ou Expedia) a révolutionné le secteur de l'hôtellerie en modifiant les habitudes de réservation et les canaux de distribution. Autrement dit, ces plateformes, apparues entre le XXe et XXIe siècle, offrent une meilleure visibilité en ligne grâce à leur référencement optimal et leurs campagnes publicitaires. Pour les hôteliers, leur utilisation est devenue quasi indispensable pour apparaître dans les premiers résultats de recherche et vendre des nuitées. D'ailleurs, « en 2013, 70 % des internautes ayant recherché un hôtel en ligne ont consulté le site Booking.com » selon un exemple tiré du 8eme baromètre publié par le JDN / Kantar Media Compete²⁷. A cela s'ajoute que la chambre d'hôtel se prête bien à la vente en ligne puisqu'elle s'achète à distance sans

²⁶ Source : Statista, *Distribution des avis présents sur Internet à propos des hôtels dans le monde en 2017, selon les principaux sites d'évaluation*, 15/07/2019 [En ligne]. Disponible sur <https://tinyurl.com/mxck2ez2>, (consulté le 10 Mai 2025)

²⁷ [Trafic des sites : Booking en tête sur les hôtels et eDreams sur les vols](#), (consulté le 10 Mai 2025)

inspection préalable. Les hôteliers ont d'ailleurs été parmi les premiers à digitaliser leur offre, publiant dès les années 1990 leurs catalogues sur Internet. L'émergence des plateformes comme Expedia et Booking (créées en 1996) a marqué un tournant, elles se sont progressivement imposées comme des acteurs dominants. Aujourd'hui, ces comparateurs agrègent une offre sans commune mesure avec les chaînes hôtelières (2 à 2,5 millions de chambres contre 600 000 à 1 million pour les plus grandes chaînes), démontrant leur puissance sur le marché²⁸.

- La Comparaison par critères de choix :** La comparaison des différents comparateurs permettra aux touristes de faire un meilleur choix selon les critères du site. Certaines études [21] montrent que si les méthodes de recherche d'hôtels varient selon les individus (consultation de photos, lecture d'avis, recommandations), les attentes au moment de la réservation convergent vers quatre critères essentiels. D'abord, les clients recherchent systématiquement les tarifs les plus avantageux pour un établissement donné. Ensuite, ils exigent une transparence totale sur le prix final, sans frais cachés lors du paiement. Puis, ils rejettent les techniques commerciales agressives (messages urgents, incitations à réserver immédiatement). Enfin, ils attendent une parfaite adéquation entre la prestation promise lors de la réservation (description, photos) et la réalité sur place. Ces exigences communes, qui relèvent à la fois de la rationalité économique et de la confiance dans la transaction, constituent aujourd'hui le socle minimal de qualité attendu par les consommateurs dans le processus de réservation en ligne. D'ailleurs une étude comparative des meilleurs tarifs a été faite sur quatre sites de réservation (Booking, Expedia, Hotels.com, et TripAdvisor) sur 10 réservations de séjour en hôtel en France et en Europe²⁹ [22] :

Tableau 5. Tableau Comparatif des meilleurs tarifs trouvés par les 4 sites sur 6 Séjours :

Séjour	Expedia	Hotels.com	Booking	Direct hôtelier
Paris - Mercure Notre Dame St Germain des Prés - 2 pers. - 1 ch. - 1 nuit du 10 au 11 déc.	175,00 €	175,00 €	175,00 €	172,00 €
Lyon - Hotel Victoria centre perrache confluence - 2 pers. - 2 ch. - 2 nuits du 12 au 14 déc.	244,00 €	244,00 €	244,00 €	220,00 €
Grenoble - Novotel Centre - 2 pers. - 1 ch. - 1 nuit du 10 au 11 déc.	90,00 €	90,00 €	90,00 €	86,00 €
Arreau (Midi-Pyr.) - Résid. Les Balcons de la Neste - 4 pers. - 2 ch. - 7 nuits du 11 au 18 mars	333,00 €	333,00 €	373,00 €	470,00 €
Bruxelles - Hotel Neufchatel - 1 pers. - 1 ch. - 2 nuits du 4 au 6 janv.	129,00 €	136,00 €	137,00 €	138,00 €
Budapest - Bohem Art Hotel 2 pers. 1 ch - 5 nuits du 24 au 29 déc.	597,00 €	597,00 €	597,00 €	505,12 €
Total	971,00 €	978,00 €	1 019,00 €	1 086,00 €

²⁸ [L'hôtellerie réinventée à l'heure du digital : le monde selon Accor | Cairn.info](#), (consulté le 10 Mai 2025)

²⁹ [Quels sont les meilleurs comparateurs d'hôtels ? \[Tests\] - Le Même en Mieux](#), (consulté le 10 Mai 2025)

Ici, nous constatons que, pour un séjour de 6 jours, le comparateur Expedia, avec un tarif de (971euro), présente des tarifs beaucoup plus moins chers que les autres sites de réservation.

- **Les évolutions et Perspectives Futures :** Certains auteurs [23] soulignent que les comparateurs de voyage sont devenus des outils incontournables pour les voyageurs, offrant une comparaison rapide des prix (vols, hôtels, locations) et transformant ainsi les habitudes de réservation. En effet, leur développement technologique, marqué par l'Intelligence Artificielle, permet désormais des recommandations personnalisées basées sur les préférences et l'historique des utilisateurs. Leur accessibilité mobile renforce leur utilité dans un contexte de digitalisation accrue. À l'avenir, ces plateformes pourraient élargir leur offre en incluant des services complémentaires comme des conseils, des itinéraires, ou des assurances, répondant ainsi aux attentes diversifiées des voyageurs (budget, style de voyage). De plus, les comparateurs, par leur capacité d'innovation et d'adaptation, sont appelés à renforcer leur rôle clé dans l'industrie du voyage, en combinant ainsi l'optimisation des coûts³⁰.

Dans un contexte de projet orienté vers la collecte ciblée et structurée de données pour la mise en place d'un comparateur d'hébergements, il devient nécessaire de se positionner sur une approche parmi les différents comparateurs existants. En effet, la recherche actuelle sur les comparateurs d'hébergements touristiques présente plusieurs limites, notamment l'absence de solutions adaptées aux contextes locaux comme celui des destinations côtières du Sénégal. A cela s'ajoute le manque d'intégration de techniques automatisées telles que le web crawling et le web scraping. Pour répondre à ces défis, notre étude vise à concevoir un système de comparaison d'hébergements inspiré des pratiques de plateformes de référence qui utilisent le web crawling et le scraping pour collecter, structurer et actualiser leurs données. Notre objectif est de développer une solution locale, performante et accessible, capable d'agréger automatiquement les données d'hébergement disponibles sur le web pour les destinations côtières sénégalaises. D'où notre choix de **Booking**, **TripAdvisor** et **Airbnb** comme comparateurs de destinations hôtelières. Nous avons choisi cette approche afin que les plateformes **Booking.com**, **TripAdvisor** et **Airbnb** utilisent le web crawling et le web scraping pour récupérer et gérer les informations sur les hébergements. En effet, le système sera capable de fournir des informations actualisées, structurées, facilitant ainsi la prise de décision pour les touristes et les acteurs du secteur hôtelier.

³⁰ [L'Utilité des Comparateurs de Voyage : Économies et Astuces – Explore Le Monde](#), (consulté le 11Mai 2025)

En conclusion, ce chapitre a examiné les travaux et solutions existants liés au web scraping, crawling, et aux plateformes de comparaison d'hébergements. La revue de littérature a permis d'identifier les outils et méthodes couramment employés, ainsi que les limites des systèmes actuels. Nous avons ensuite défini notre positionnement en insistant sur les spécificités de notre projet, la cible et l'objectif qui est d'automatiser la collecte et la comparaison des offres hôtelières.

Chapitre 3 : Cahier des charges

3.1. Contexte et justificatifs du projet

Certaines zones du Sénégal, riches en atouts touristiques, souffrent d'un manque de plateformes centralisées permettant aux voyageurs de comparer efficacement les offres d'hébergement disponibles. En effet, les informations concernant les hôtels, telles que les prix, les disponibilités, les types de services proposés ou encore les avis des clients, sont souvent dispersées sur plusieurs sites, ce qui rend la recherche longue, confuse et peu pratique pour les touristes. Cette dispersion complique la planification d'un séjour et peut décourager certains visiteurs, en particulier ceux qui ne sont pas familiers avec la région ou qui ne parlent pas couramment la langue locale.

De plus, cette fragmentation des données limite considérablement la visibilité des établissements hôteliers locaux, qui peinent à se démarquer face aux grandes chaînes internationales déjà bien référencées sur les plateformes mondiales. Les petites structures d'hébergement, souvent gérées de manière familiale, ne disposent pas toujours des compétences ou des ressources pour se rendre visibles en ligne ou optimiser leur présence numérique. Cela freine leur développement économique, réduit leur taux d'occupation et diminue leur potentiel de revenus [24].

Dans ce contexte, la mise en place d'un comparateur d'hébergements spécifiquement adapté aux réalités locales apparaît comme une solution pertinente. Un tel outil, basé sur les technologies de web scraping et d'agrégation de données, permettrait de collecter automatiquement des informations issues de multiples sites, de les structurer de façon cohérente, et de les présenter sur une interface unique, claire et conviviale. Ce système faciliterait la comparaison des offres, permettant aux utilisateurs de trouver rapidement un hébergement adapté à leur budget, à leurs besoins et à leurs préférences.

D'ailleurs, une telle solution numérique offrirait non seulement un gain de temps significatif aux voyageurs, mais leur permettrait également de prendre des décisions plus éclairées grâce à une meilleure lisibilité des offres disponibles. Par ailleurs, elle pourra renforcer la compétitivité des hôtels locaux, en leur donnant une meilleure visibilité sur Internet, ce qui est aujourd'hui un facteur essentiel pour attirer une clientèle internationale. En effet, en favorisant l'accès équitable à l'information, ce type de comparateur peut jouer un rôle crucial dans la valorisation du tourisme local et contribuer au développement durable du secteur hôtelier dans les régions côtières du Sénégal. Il représente donc une opportunité concrète pour dynamiser le secteur touristique, tout en soutenant les économies locales.

3.2. Mission du projet

Dans une optique d'amélioration de leur taux de réservation, nous souhaitons offrir aux établissements hôteliers des destinations du Sénégal un outil de gestion plus efficace et moderne. En effet, ces établissements, souvent de taille modeste, rencontrent de nombreuses difficultés pour atteindre une clientèle large, notamment à cause du manque de visibilité sur les plateformes internationales et de l'absence d'outils numériques adaptés à leurs réalités. La mission de ce projet est donc de concevoir une solution pratique et accessible, qui facilite la comparaison des hébergements dans ces zones, en centralisant les informations essentielles telles que les tarifs, les disponibilités, les types de chambres, les services offerts, les avis clients et la localisation.

L'objectif principal est de simplifier la prise de décision pour les touristes, en leur fournissant des données fiables, claires, actualisées en temps réel, et présentées dans une interface facile à utiliser, même pour ceux qui découvrent le pays pour la première fois. Autrement dit, en regroupant ces informations sur une seule plateforme, on permet aux voyageurs de gagner du temps, d'éviter les mauvaises surprises et de choisir leur hébergement en toute confiance, en fonction de leurs besoins, préférences et aussi de leur budget.

C'est dans cette perspective que nous proposons la création d'un comparateur de destinations hôtelières dédié aux zones touristiques du Sénégal. En effet, ce comparateur, reposant sur des technologies de web scraping et d'agrégation de données, jouera un rôle d'intermédiaire entre l'offre locale et la demande touristique nationale et internationale. A cet effet, il contribuera à améliorer la performance commerciale des hôtels, en augmentant leur visibilité en ligne et en leur donnant un accès plus direct à leur clientèle cible. Par ailleurs, cette solution encouragera les hôtels à maintenir à jour leurs informations et à améliorer la qualité de leurs services, dans un monde de compétitivité. Enfin, le comparateur pourra également servir de support à la promotion du tourisme local, en valorisant les destinations côtières souvent méconnues, mais riches en potentiel.

Ce projet constitue donc une vraie opportunité pour développer l'économie locale, en aidant les professionnels du tourisme à mieux utiliser les outils numériques. En même temps, il permet d'améliorer l'expérience des voyageurs, en leur donnant un accès facile et rapide à toutes les informations utiles. Il s'inscrit aussi dans les objectifs actuels de modernisation du secteur touristique, soutenus par plusieurs organisations nationales et internationales, qui encouragent la digitalisation pour rendre le tourisme plus efficace et plus accessible.

3.3. Périmètre du projet

Le projet se concentre spécifiquement sur les destinations du Sénégal, surtout sur les zones côtières du Sénégal en prenant en compte leurs principales destinations touristiques telles que Dakar, Saly, Joal-Fadiouth, Saint-Louis, Cap Skirring ou encore Mbour. Ces localités attirent chaque année de nombreux visiteurs, en raison de leurs plages, leur patrimoine culturel, leur gastronomie et leur ambiance chaleureuse. Cependant, malgré leur potentiel touristique élevé, ces zones souffrent encore d'un manque de structuration numérique, en particulier en ce qui concerne la centralisation des offres d'hébergement. Ainsi, la plupart des touristes doivent visiter plusieurs sites différents pour trouver une chambre, comparer les prix ou lire les avis, ce qui peut décourager certains voyageurs, notamment ceux qui viennent pour la première fois. Ce manque de visibilité numérique pénalise également les petits établissements hôteliers, qui n'ont pas toujours les moyens techniques ou financiers pour figurer sur les grandes plateformes internationales.

De plus, le projet vise à réduire ces inégalités, en proposant une solution qui regroupe toutes les informations utiles sur une seule interface claire et facile à utiliser. Ce comparateur d'hôtels pourrait ainsi devenir un outil de valorisation du tourisme côtier sénégalais, en rendant les destinations locales plus visibles et plus accessibles. Il répond aussi aux recommandations des plans de développement touristique du Sénégal, qui mettent l'accent sur la promotion du tourisme durable, régionalisé et digitalisé.

3.4. Cibles du projet

Dans ce projet, les cibles de ce système peuvent inclure :

- Les touristes à la recherche d'hébergements adaptés ;
- Les établissements hôteliers souhaitant accroître leur visibilité ;
- Les agences de voyage et opérateurs touristiques ;
- Les institutions publiques chargées de la promotion touristique ;
- Les chercheurs et développeurs en technologies du tourisme.

3.5. Besoins / Exigences

Pour mettre en place un comparateur d'hébergements touristiques destiné aux zones touristiques du Sénégal, il est très important d'identifier tous les besoins et exigences du projet. Cela permet de mieux organiser le travail, de choisir les bons outils et de s'assurer que la solution sera utile et facile à utiliser pour les touristes. Nous allons donc commencer par présenter les besoins fonctionnels, c'est-à-dire ce que le système doit pouvoir faire (comme afficher les prix, les avis ou les disponibilités des hôtels). Ensuite, nous parlerons des besoins

technologiques, qui concernent les logiciels, langages de programmation et plateformes nécessaires pour construire la plateforme. Nous verrons aussi les besoins en données, qui sont indispensables pour que le comparateur fonctionne correctement, en récupérant des informations fiables depuis différents sites. Enfin, nous aborderons les besoins non fonctionnels, comme la rapidité, la sécurité ou encore la facilité d'utilisation. La partie suivante (3.6) expliquera les contraintes techniques, c'est-à-dire les limites ou difficultés que nous pourrions rencontrer lors du développement de notre plateforme.

3.5.1. Besoins fonctionnels

A long terme, ce comparateur de destinations hôtelières permettra aux Touristes et Voyageurs :

- De trouver des hébergements selon le prix, la localisation, les avis ;
- De comparer les offres de différentes plateformes pour une destination ;
- De voir des informations complètes sur les hôtels ;
- De filtrer et trier les résultats en fonction de leurs préférences ;
- De recevoir des recommandations personnalisées ;
- D'accéder à des données actualisées comme la disponibilité des chambres et les promotions en cours.

Il permettra aux hôtels de :

- Mettre à jour leurs informations sur les tarifs, les disponibilités ;
- Suivre les consultations et les réservations sur la plateforme ;
- Recevoir des notifications pour les réservations ;
- Gérer leurs offres et répondre aux avis clients ;
- Bénéficier d'une visibilité accrue et des promotions ;
- Bénéficier d'outils pour voir leur performance.

Pour l'administrateur de la plateforme, il permettra :

- Superviser la collecte de données pour garantir leur fiabilité ;
- Gérer les utilisateurs et les comptes des propriétaires d'hôtels ;
- Automatiser la détection des problèmes techniques ;
- Paramétrer les algorithmes de recommandation pour la satisfaction des utilisateurs.

3.5.2. Besoins technologiques

La mise en œuvre de ce dispositif nécessitera l'utilisation de certains outils adaptés à notre approche et à nos objectifs. Ainsi, il est demandé d'utiliser :

- Un langage de programmation fluide et moderne ;

- Une base de données relationnelles ;
- Un Framework de développement rapide et efficace ;

3.5.3. Besoins en données

Dans le cadre de ce projet, nous aurons besoins des données suivantes [31] :

- **Du Type du logement :** Il s'agit de la catégorie de l'hébergement proposé par l'établissement. Chaque type de logement répond à des besoins différents en termes de confort, de capacité, ou de budget. Nous pouvons avoir plusieurs types de logement dont :
 - **Une Chambre Single :** C'est une chambre conçue pour accueillir une seule personne, avec un lit simple. Elle est idéale pour les voyageurs en solo ou les professionnels en déplacement ;
 - **Une Chambre Double :** Chambre pouvant accueillir deux personnes. Elle est généralement équipée d'un lit double ou de deux lits simples. Parfaite pour les couples ou les amis ;
 - **Une Suite :** C'est un logement plus spacieux et haut de gamme qu'une chambre standard. Une suite comporte souvent un coin salon, une chambre séparée, voire une petite cuisine. Elle est destinée aux clients recherchant plus de confort ou de luxe ;
 - **Une Villa :** C'est un hébergement indépendant de type maison, généralement avec plusieurs chambres, un salon, une cuisine, parfois un jardin ou une piscine privée. Adaptée aux familles ou aux groupes ;
 - **Un Appartement :** C'est des logements équipés pour des séjours de courte ou longue durée. Ils comprennent une ou plusieurs pièces, une salle de bain, une cuisine et offrent plus d'autonomie aux clients ;
- **Nom de la structure:** C'est le nom commercial de l'établissement qui propose l'hébergement (exemple : Hôtel Teranga, Résidence du Soleil, Auberge Keur Coumba). Il permet d'identifier clairement l'hébergement sur une plateforme de comparaison ou lors d'une réservation ;
- **Type de la structure :** Il s'agit ici de la forme ou du statut de l'établissement hôtelier. Cela peut influencer le niveau de service, le confort, ou le public ciblé. Parmi les types de structure, nous pouvons citer :
 - **Les Hôtels :** C'est des structures classiques d'hébergements proposant différents niveaux de confort (étoiles), souvent avec des services complémentaires (restauration, piscine, room service...) ;

- **Les Auberges :** Hébergements simples et convivial, souvent de petite taille, destinés à une clientèle à la recherche d'un cadre authentique ou plus économique ;
- **Les Campements :** Établissements touristiques en pleine nature, souvent utilisés dans les zones rurales ou écotouristiques. Ils peuvent proposer des tentes, cases traditionnelles ou bungalows ;
- **Le Prix en FCFA :** C'est le tarif affiché pour une nuit ou un séjour dans l'hébergement, exprimé en Francs CFA (XOF). Ce prix varie selon le type de logement, la saison, les services inclus ou encore la localisation de l'établissement ;
- **Les Photos :** C'est les images fournies par l'établissement pour illustrer ses offres. Elles permettent de visualiser les chambres, les équipements, l'environnement extérieur ou les services proposés. Les photos jouent un rôle important dans la décision du client et dans la mise en valeur de la structure sur une plateforme de réservation.

3.5.4. Besoins non fonctionnels

Notre plateforme permettra en termes de **Performance** :

- D'offrir des recherches et comparaisons rapides ;
- De gérer un grand volume de données en temps réel.

En termes de **Fiabilité** elle permettra :

- De garantir la disponibilité de la plateforme avec peu d'interruptions ;
- De mettre en place des mécanismes de reprise après incident.

En ce qui concerne la **Sécurité**, la plateforme permettra :

- De protéger les données des utilisateurs et des hôtels ;
- De respecter les normes de protection des données comme le RGPD.

Concernant la **Compatibilité**, elle pourra :

- Fonctionner sur divers appareils et navigateurs ;
- S'intégrer aux systèmes de réservation des hôtels.

Par rapport à la **Facilité d'utilisation**, la plateforme peut :

- Offrir une interface intuitive et simple à utiliser ;
- Proposer des guides pour une prise en main facile.

En matière de **Maintenabilité**, ce comparateur permettra :

- De faciliter les mises à jour et corrections de bugs ;
- De fournir une documentation complète pour les développeurs.

En termes de **Conformité légale et éthique**, elle doit :

- Respecter les directives des sites scrappés (robots.txt) ;
- Éviter toute violation des conditions d'utilisation ou de droits.

3.6. Contraintes techniques

Nous pouvons être confrontés aux :

- Respect des politiques et des conditions d'utilisation des sites web ciblés ;
- Respect des réglementations en matière de protection des données et de confidentialité ;
- Contraintes de performances.

En somme, ce chapitre nous a permis de bien poser les bases du projet en expliquant le contexte, les objectifs, ainsi que les besoins et les contraintes. En prenant en compte les attentes fonctionnelles, technologiques entre autres. Cela nous permettra de construire une solution solide qui correspondra aux besoins des utilisateurs tout en respectant les limites techniques.

Chapitre 4 : Conception et réalisation

4.1. Conception

4.1.1. Architecture générale

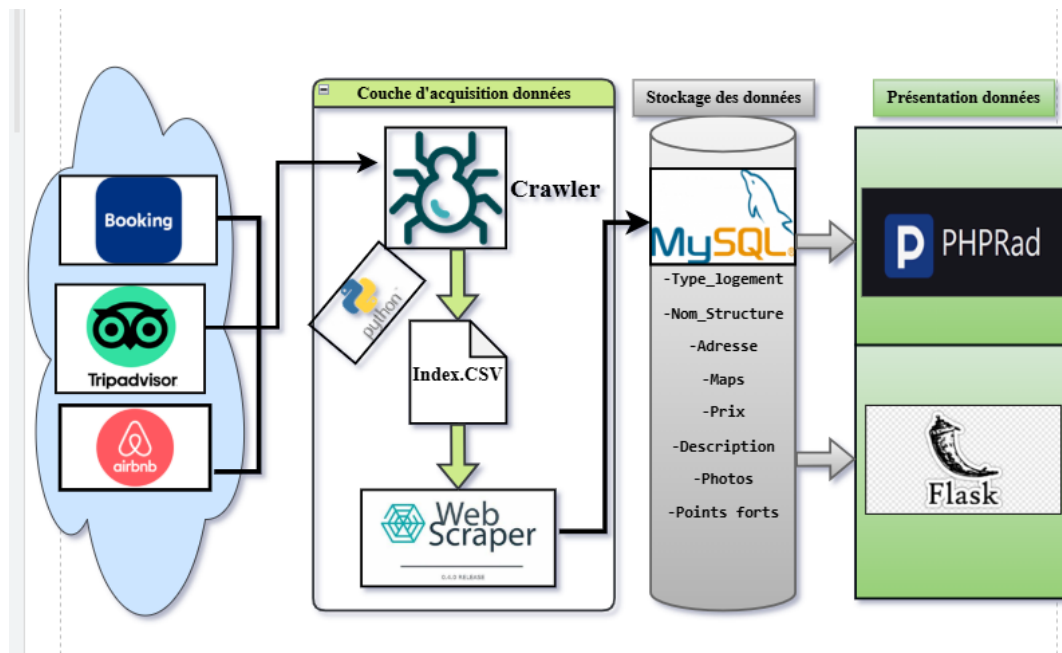


Figure 25. Modélisation Architecture générale du système

En arrière-plan, l'architecture proposée permet d'automatiser la collecte, le traitement et la visualisation des données touristiques issues de plusieurs plateformes comme **Booking**, **TripAdvisor** et **Airbnb**. Le système est divisé en trois grandes couches. D'abord, la **couche d'acquisition des données** utilise des scripts Python (**crawler**) et l'outil **Web Scraper** pour extraire automatiquement les informations ciblées, qui sont ensuite indexées dans un fichier **CSV**. Ces données sont ensuite stockées dans une base **MySQL**, structurées selon plusieurs champs tels que le type de logement, l'adresse, le prix ou encore les photos. Enfin, dans la **couche de présentation**, deux outils sont mobilisés : **PHPRad** pour la création rapide d'interfaces web de gestion, et **Flask**, un micro-framework Python, pour afficher dynamiquement les données collectées. Ce processus permet d'avoir une plateforme à jour, flexible et fonctionnelle pour analyser les données touristiques.

4.1.2. Couche d'acquisitions de données

Dans le cadre de notre projet, nous appelons Sources les plateformes de réservation en ligne. En effet, ces sites représentent des moyens essentiels pour collecter des informations sur les offres d'hébergement disponibles. Aujourd'hui, plusieurs plateformes sont utilisées par les voyageurs pour rechercher, comparer et réserver des logements. Pour notre étude, nous avons choisi trois sites comme échantillon représentatif : **Booking.com**, **TripAdvisor** et **Airbnb**. Ces plateformes sont parmi les plus populaires à l'échelle mondiale et sont aussi largement

utilisées par les touristes au Sénégal. Elles proposent des données variées et utiles telles que les prix, les disponibilités, les avis clients, les notes, les photos et les descriptions détaillées. Ce choix nous aidera à mettre en place un système de web scraping pour récupérer automatiquement les informations depuis ces sites. Il nous permettra aussi de mieux comprendre comment les données sont organisées sur chaque plateforme [32].

- Booking est un comparateur de destination touristique accessible via l'url suivant : www.booking.com . Elle permet aux voyageurs de comparer facilement différentes offres dans plusieurs pays. Sur ce site, on trouve plusieurs types d'hébergements comme les hôtels, les campements-auberges, les villas, et les maisons d'hôtes. Ces catégories permettent aux utilisateurs de choisir l'option qui correspond le mieux à leurs besoins et à leur budget. D'ailleurs, le tableau suivant montre en détails la présentation de Booking avec ses catégories d'hébergements. On y observe la disponibilité ou non de chaque type d'hébergement, ce qui permet d'avoir un aperçu de l'offre disponible pour les utilisateurs.

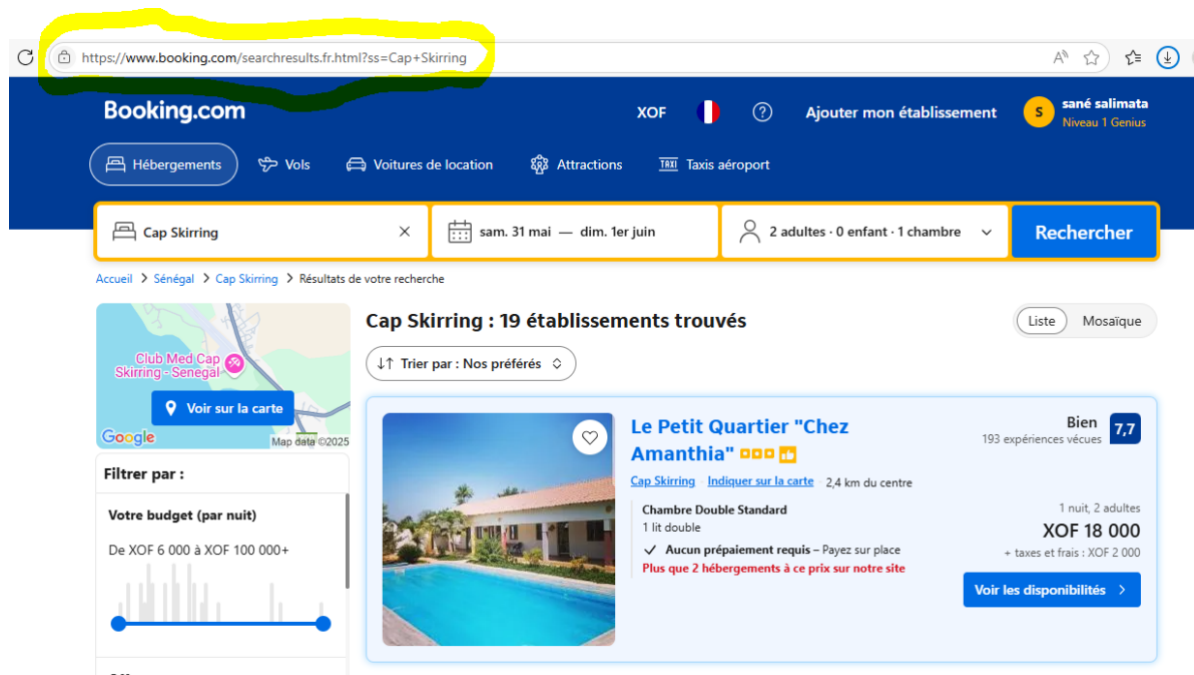
Tableau 6. Présentation des types d'hébergements dans Booking :

Catégories Hébergement	Hotel	OUI
	Campement-Auberge	OUI
	Motels	NON
	Villa	OUI
	Maison d'hôtes	OUI
	Sites culturels	NON
	Lodge	OUI
	Appartements	OUI
	Camping	OUI

Explications :

Pour faire une recherche des hébergements dans une ville du Sénégal, dans Booking, on peut procéder de la manière suivante :

Exemple : <https://www.booking.com/searchresults.fr.html?ss=Cap+Skirring>, cette URL recherche et affiche les types de logements de Cap Skirring disponibles dans Booking.



- <https://www.booking.com> : Il s'agit du site web principal de la plateforme de Booking ;
- /searchresults.fr.html : Cela signifie que la page affichera les résultats de recherche en **langue française** (.fr.html) ;
- ?ss=Nom_de_la_Ville: C'est la requête de recherche. Le paramètre (pour "search string") indique que l'utilisateur cherche des hébergements à telle ville ou zone.

Page de présentation de la structure :

Une fois à l'intérieur, on peut y trouver des informations essentielles comme : le type de logement, le nom de la structure, l'adresse, l'emplacement, la description, ses points forts, le prix, les photos etc.

Figure 26. Exemple de page Booking

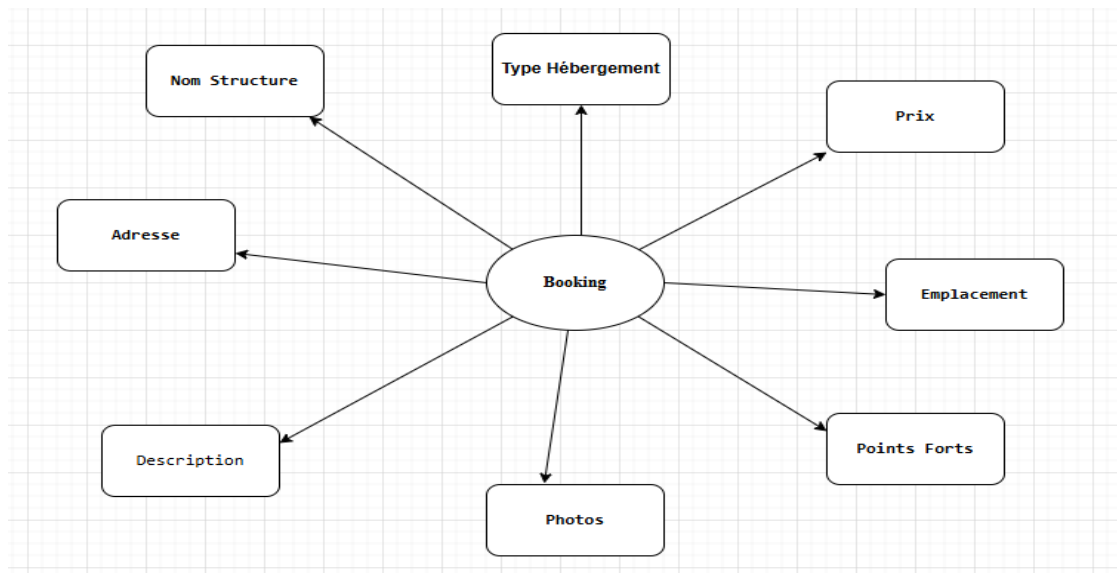


Figure 27. Figure en étoile pour l'organisation de la source de Booking

- **TripAdvisor** est une plateforme populaire qui aide les voyageurs à planifier leurs séjours en leur proposant une grande variété d'hébergements touristiques. Accessible via l'URL suivante : www.tripadvisor.com, elle permet de rechercher et de comparer différentes options comme les hôtels, les auberges, les maisons d'hôtes, les appartements, les villas ou encore les lodges. En plus des hébergements, TripAdvisor présente aussi des activités, des restaurants et des lieux culturels à visiter, ce qui en fait une plateforme plus complète que d'autres. Les utilisateurs peuvent lire des avis, voir des photos partagées par d'autres voyageurs, et choisir l'option qui correspond le mieux à leurs préférences et à leur budget.

Tableau 7. Présentation des types d'hébergements dans TripAdvisor :

Types de propriétés	Hotel	OUI
	Campement-Auberge	OUI
	Villa	OUI
	Maison d'hôtes	OUI
	Sites culturels	NON
	Lodge	OUI
	Appartements	OUI
	Motels	OUI
	Camping	OUI

Explications :

Ici dans TripAdvisor, pour faire une recherche des hébergements dans une ville du Sénégal, on peut procéder de la manière suivante :

Exemple : <https://www.tripadvisor.com/Hotels-g293831.html#SPLITVIEWMAP>, cette URL recherche et affiche les types de logements disponibles dans TripAdvisor.

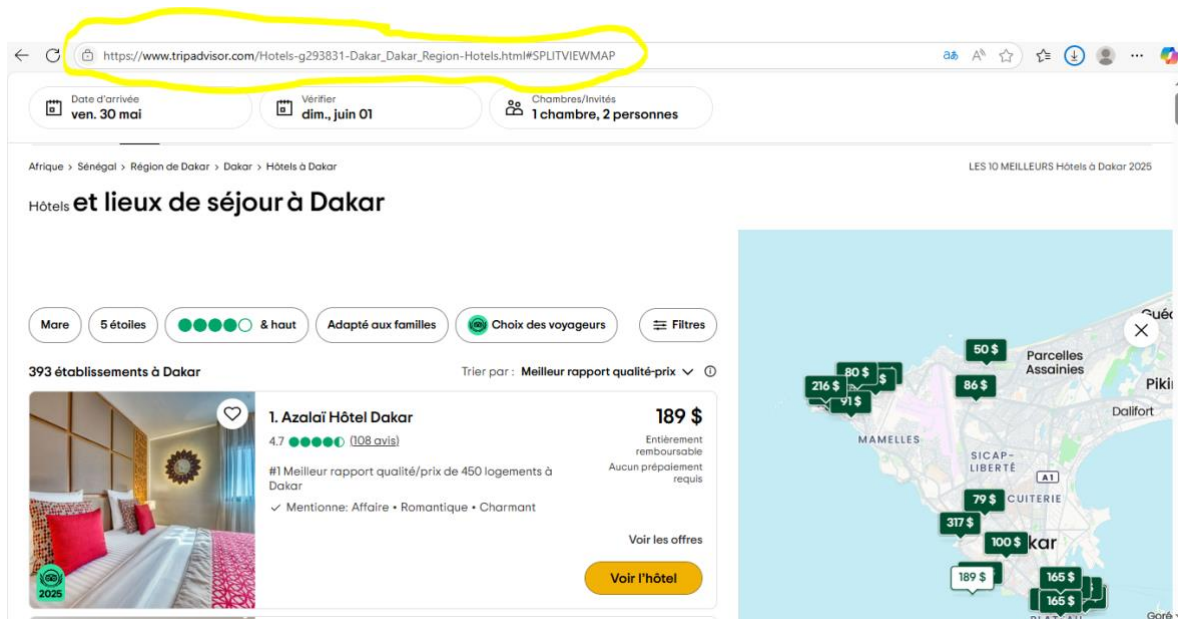


Figure 28. Exemple de page Tripadvisor

- www.tripadvisor.com : C'est le site principal de la plateforme TripAdvisor, qui aide les voyageurs à trouver et comparer des hébergements.
- `/Hotels-g293831.html` : Ce segment du lien indique que la page affiche la liste des hôtels dans une région spécifique, ici **Dakar (code de zone : g293831)**.
- `#SPLITVIEWMAP` : Cette partie active un affichage divisé sur la page, avec la liste des hôtels à gauche et la carte interactive de leur emplacement à droite.

Page de présentation de la structure :

Une fois à l'intérieur, on peut y trouver des informations essentielles comme : le type de logement, le nom de la structure, l'adresse, l'emplacement, la description, ses points forts, le prix, les photos etc.

Remarque : On constate que dans TripAdvisor, pour avoir plus d'informations sur un type de logement, la recherche nous redirige vers le site de Booking.com.

Exemple : Si on clique sur Voir l'offre de la figure suivante, on sera redirigé vers Booking pour plus d'informations sur l'hôtel Azalai de Dakar.

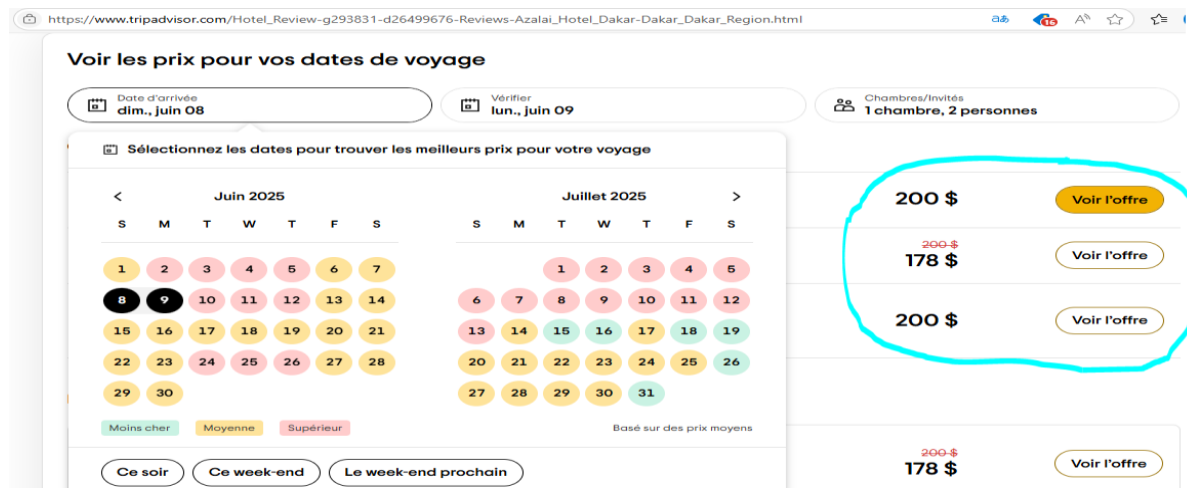


Figure 29. Page detail tripadvisor

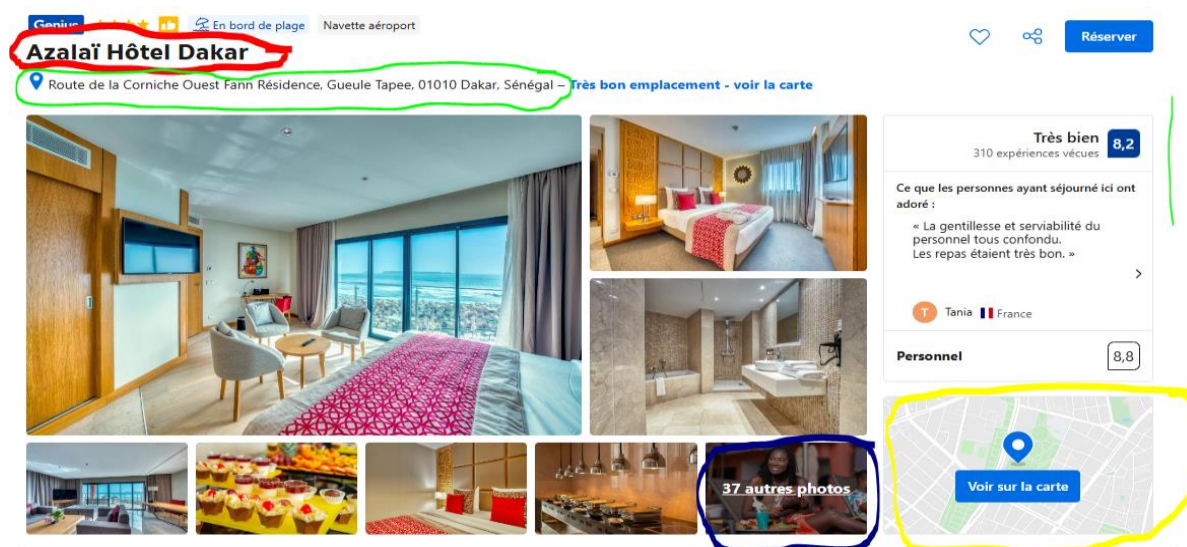


Figure 30. Page detail redirection tripadviso vers booking

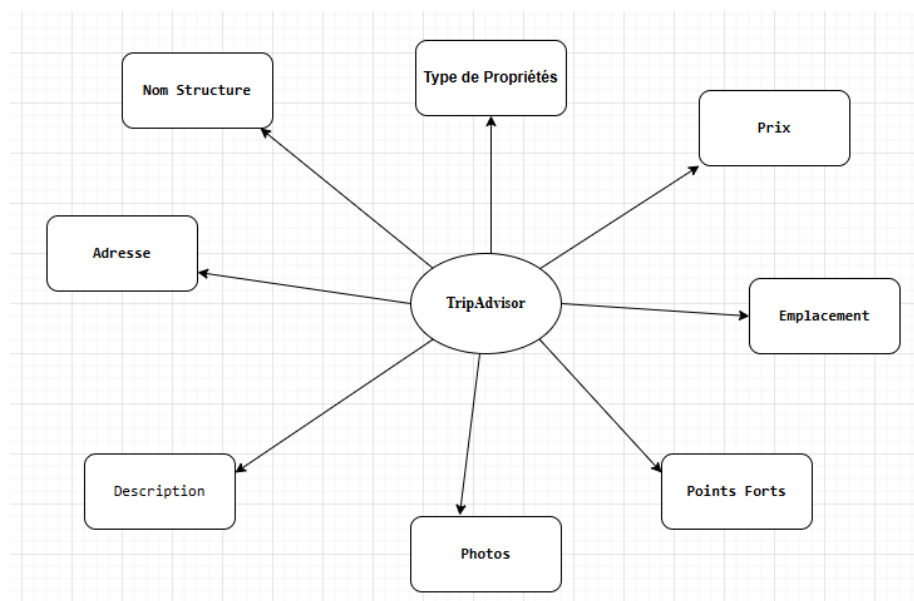


Figure 31. Figure en étoile pour l'organisation de la source de TripAdvisor :

- **Airbnb** est une plateforme en ligne très utilisée par les voyageurs pour réserver des logements dans plusieurs pays à travers le monde. Accessible via le site www.airbnb.com, elle propose une grande diversité de types d'hébergements. On y trouve des appartements, des maisons, des studios, des chambres privées, des villas, des maisons d'hôtes, des bungalows, ou même des logements insolites comme des cabanes ou des yourtes. Airbnb permet aux particuliers de mettre leur logement en location, ce qui offre plus de flexibilité et souvent des prix plus abordables que les hôtels.

Tableau 8. Présentation des types d'hébergements dans Airbnb

Types de propriétés	Hotel	NON
	Appartements	OUI
	Villa	OUI
	Maison d'hôtes	OUI
	Sites culturels	NON
	Bungalows	OUI
	Chambres privées	OUI
	Maison	OUI
	Camping	NON

Explications :

La recherche des hébergements qui existent dans une ville du Sénégal, dans Airbnb, on peut se faire de la manière suivante :

Exemple : <https://fr.airbnb.com/rooms/1334865753946079072?>, cette URL recherche et affiche les types de logements disponibles dans Airbnb.

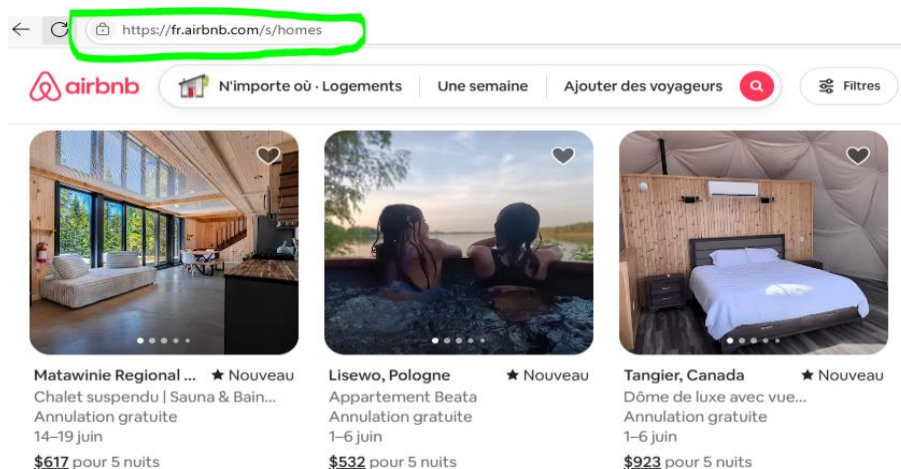


Figure 32. Page AirBNB

- **fr.airbnb.com** : Il s'agit du site principal d'Airbnb en version française. C'est une plateforme mondiale qui permet aux particuliers de proposer et de réserver des hébergements, allant d'une simple chambre à une maison entière, pour de courts ou longs séjours.
- **/rooms/** : Ce segment de l'URL indique qu'il s'agit d'une page dédiée à une **annonce spécifique** d'un logement mis en location sur Airbnb.
- **1334865753946079072** : Il s'agit de l'**identifiant unique du logement**. Chaque hébergement sur Airbnb possède un numéro d'identification propre, ce qui permet d'y accéder directement via une URL personnalisée. Ce lien mène donc à la fiche détaillée d'un seul logement disponible à la réservation (par exemple, un appartement à Dakar Nommé : **La Palme des Mamelles**).

Page de présentation de la structure : Une fois à l'intérieur, on y trouve différentes informations essentielles comme : le type de logement, le nom de la structure, l'adresse, l'emplacement, la description, les équipements, le prix, les photos etc.



 N'importe où

 Une semaine

 Ajouter des voyageurs
 

Devenir hôte





La Palme des Mamelles

 Partager
  Enregistrer





Afficher toutes les photos

Logement entier : appartement - Ndakhar, Sénégal
 3 voyageurs · 1 chambre · 1 lit · 1 salle de bain
 ★ 5,0 · 5 commentaires

 Perle rare ! Les réservations pour ce logement sont fréquentes.

https://fr.airbnb.com/rooms/1334865753946079072?source_impression_id=p3_1748302722_P3AG7bnD2Oifvrg&check_in=2025-06-30&guests=1&adult...

Photos

Équipements

Commentaires

Emplacement

Le logement
 À la recherche d'un confort absolu et d'un logement unique proche des commerces et restaurants, cet appartement spacieux est pour vous. Il est niché au cœur des mamelles, à 5 minutes de la corniche et du monument de la Renaissance. Ce logement offre toute...

Lire la suite

Ce que propose ce logement

 Cuisine

 Wifi

 Espace de travail dédié

 Parking gratuit sur place

 Télévision

 Lave-linge

 Climatisation

 Privé : patio ou balcon

 Détecteur de monoxyde de carbone

 Détecteur de fumée

 Perle rare ! Les réservations pour ce logement sont fréquentes.

\$67 pour 1 nuit

ARRIVÉE

30/06/2025

DÉPART

01/07/2025

VOYAGEURS

1 voyageur

Réserver

Aucun montant ne vous sera débité pour le moment

 Signaler cette annonce

https://fr.airbnb.com/rooms/1334865753946079072?source_impression_id=p3_1748302722_P3AG7bnD2Oifvrg&check_in=2025-06-30&guests=1&adult...

Photos

Équipements

Commentaires

Emplacement

\$67 pour 1 nuit
 ★ 5,0 · 5 commentaires

Réserver

Où se situe le logement
 Ndakhar, Dakar Region, Sénégal

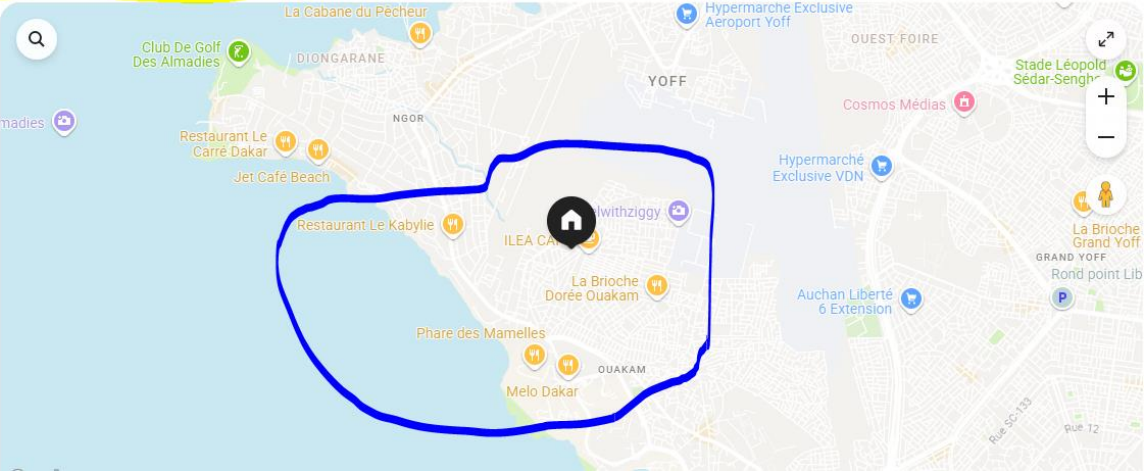


Figure 33. Page detail AirBNB

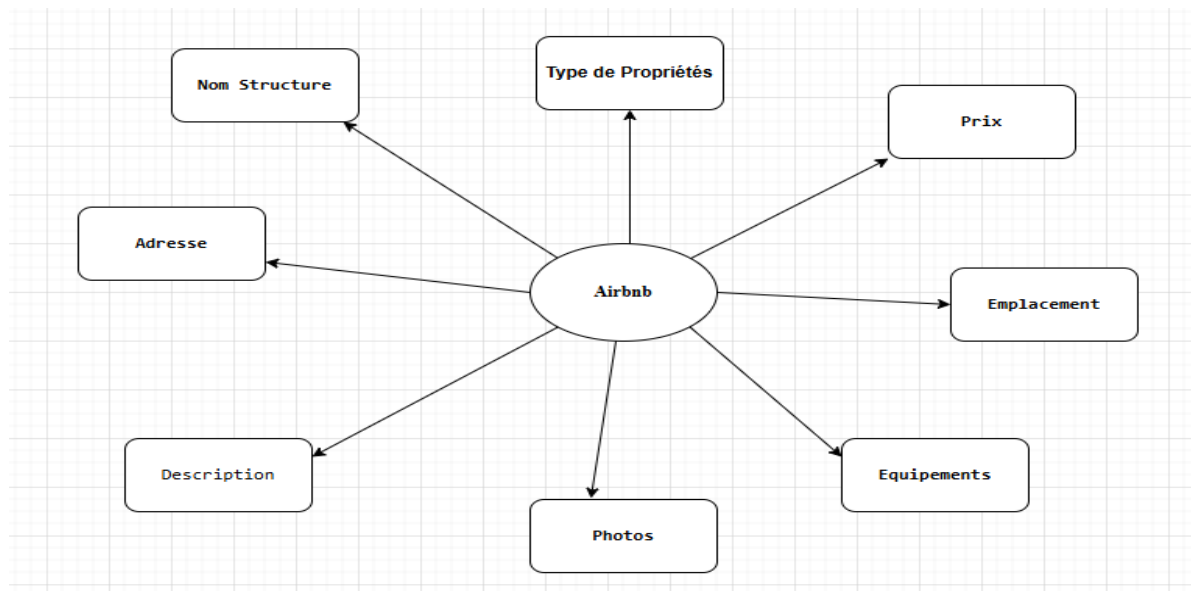


Figure 34. Figure en étoile pour l'organisation de la source de Airbnb :

Illustration : Pour crawling et scraping sur booking

Nous avons utilisé Python pour automatiser la **collecte, l'analyse et le stockage de données d'hébergements touristiques** disponibles sur **Booking.com**, notamment pour les destinations situées au Sénégal. Il est composé de deux modules principaux : un **crawler** et un **scraper**, chacun ayant un rôle bien défini dans le processus global :

- Le module **CRAWLER_MODULE** lit une liste de destinations à partir du fichier input/ville.csv. Pour chaque ville, il construit dynamiquement une URL de recherche sur Booking.com et envoie une requête HTTP. À partir de la page de résultats obtenue, il extrait plusieurs informations essentielles comme le nom de l'hébergement, son lien vers la fiche complète, son adresse, ainsi qu'une image principale. Ces données sont ensuite stockées dans un fichier CSV (output/index.csv), tout en supprimant les doublons pour éviter les redondances.
- Le **SCRAPER_MODULE** lit chaque ligne de ce fichier index. Pour chaque lien d'hébergement, il accède à la page détaillée et en extrait plusieurs informations : la description, le prix par nuit, les points forts, les coordonnées géographiques (latitude, longitude), le type de logement, ainsi que jusqu'à cinq images supplémentaires. Les images sont téléchargées localement dans un dossier structuré. Si le prix n'est pas trouvé dans la page, une méthode de récupération alternative (Get_price_new) est tentée.
- Le module de sauvegarde : Les données finales, une fois nettoyées et standardisées, sont insérées dans une base de données **MySQL** nommée DESTINATIONSSENEGAL,

dans la table hébergements. Le système enregistre également les hébergements traités avec succès dans un fichier output/succes.txt, et ceux pour lesquels une erreur survient (comme l'absence de prix) dans output/erreurs.txt.

Le système utilise plusieurs bibliothèques Python comme requests, BeautifulSoup, pandas, concurrent.futures, mysql.connector et logging. Il est robuste grâce à des mécanismes de gestion des erreurs, des tentatives multiples de récupération, et des suppressions automatiques de doublons dans les fichiers CSV. Il permet ainsi de construire une base de données locale complète et structurée d'offres touristiques disponibles sur Booking.com, ce qui peut être très utile pour alimenter un portail touristique local ou un projet de valorisation de destinations régionales.

4.1.3. Couche de presentation

La couche de présentation est développée grâce à PHPRAD et est constituée de plusieurs parties :

- **Un moteur de recherche :** Il permet aux utilisateurs de retrouver facilement les informations dont ils ont besoin. Elle utilise des mots-clés saisis par l'utilisateur pour rechercher dans la base de données les hébergements ou informations touristiques correspondantes. Cette recherche peut se faire par nom, adresse, type de logement ou encore par prix. Le moteur de recherche peut aussi classer les résultats selon différents critères comme la pertinence ou le prix [33].
- **Un module de réservation :** La couche de réservation permet à l'utilisateur de réserver un hébergement directement depuis la plateforme. L'utilisateur peut choisir une offre, consulter les détails (prix, adresse, photos, disponibilités), puis cliquer pour envoyer une demande de réservation. Cette couche utilise une base de données (comme **MySQL**) pour stocker les réservations et vérifier les disponibilités [34].
- **Une Couche pour l'Analyse et le Monitoring :** La couche d'analyse et monitoring est utilisée pour suivre le fonctionnement du système et analyser les données récoltées. Elle permet de générer des statistiques utiles comme le nombre de recherches effectuées, les types d'hébergements les plus demandés ou encore les zones géographiques les plus visitées. Grâce à cette couche, les responsables du système peuvent aussi détecter des erreurs ou des problèmes techniques (comme une panne de connexion à une source de données) et intervenir rapidement [35].

Explication : Dans ce diagramme de cas d'utilisation pour la gestion des réservations hébergements, nous avons cinq acteurs principaux : le Client, le Gestionnaire de Réservation, le Gérant, le Comptable et l'Opérateur Financier. Le Client peut rechercher et filtrer les hébergements, consulter les détails, faire une réservation, mais aussi payer par divers moyens (carte, virement, mobile money), et imprimer son reçu. Il doit obligatoirement s'authentifier avant d'effectuer ces actions. Le Gestionnaire de Réservation peut ajouter et consulter les hébergements. Le Gérant est chargé du reporting et du monitoring, tandis que le Comptable intervient dans la gestion des paiements. L'Opérateur Financier, quant à lui, participe au processus de paiement. Toutes les opérations principales passent par une authentification préalable pour garantir la sécurité et la traçabilité du système.

❖ **Diagramme de Cas d'utilisation Gestion de la Base de Données :**

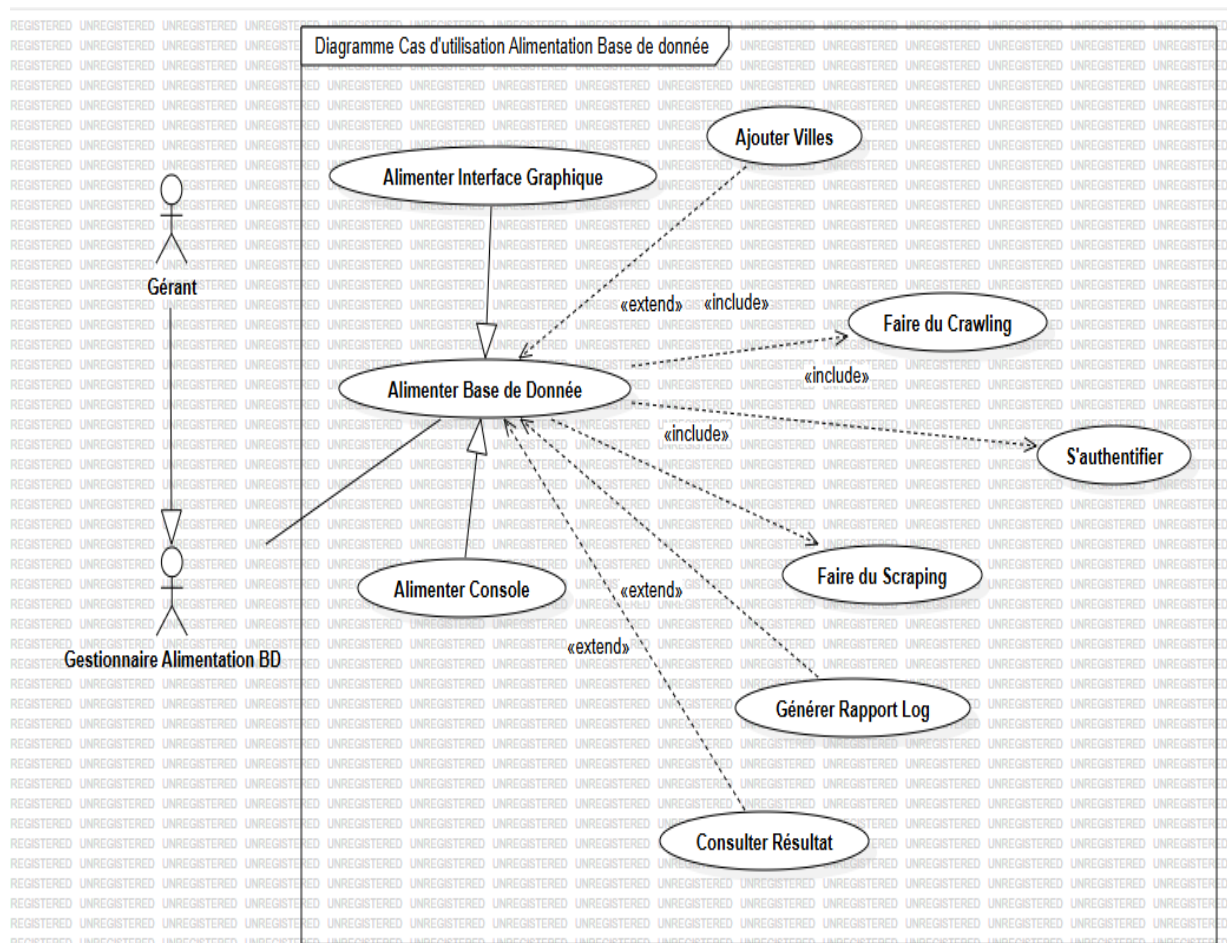


Figure 36. Diagramme de Cas d'utilisation Alimentation Base de Données

Explication : Pour ce diagramme, nous avons définis deux acteurs : le Gérant et le Gestionnaire de l'Alimentation de la Base de Données. Leur rôle principal consiste à alimenter la base de données par différents moyens, notamment via une interface graphique ou via la

console. Ce processus inclut plusieurs fonctionnalités telles qu'ajouter des villes, faire du crawling, faire du scraping, générer un rapport log et consulter les résultats. Avant toute opération, l'utilisateur doit obligatoirement s'authentifier. Ce diagramme met ainsi en évidence les différentes manières d'intégrer des données dans le système et les outils nécessaires pour assurer une alimentation efficace et sécurisée de la base de données.

4.1.4.2. Diagrammes de séquences

Les diagrammes de séquence sont des outils de modélisation utilisés en génie logiciel pour représenter le déroulement temporel des interactions entre les différents acteurs (utilisateurs ou systèmes) et les composants d'un système. Ils servent à montrer comment les différentes personnes ou systèmes communiquent entre eux dans un ordre précis. Autrement dit, ils permettent de visualiser les échanges de messages entre les acteurs et les composants d'une application. C'est un bon moyen pour mieux comprendre comment fonctionne un processus, repérer les étapes importantes, et voir s'il y a des erreurs ou des oublis dans les échanges [38]. D'ailleurs, ces diagrammes sont très utiles pendant la phase de conception d'un système³¹.

❖ Diagramme de Séquence d'Authentification

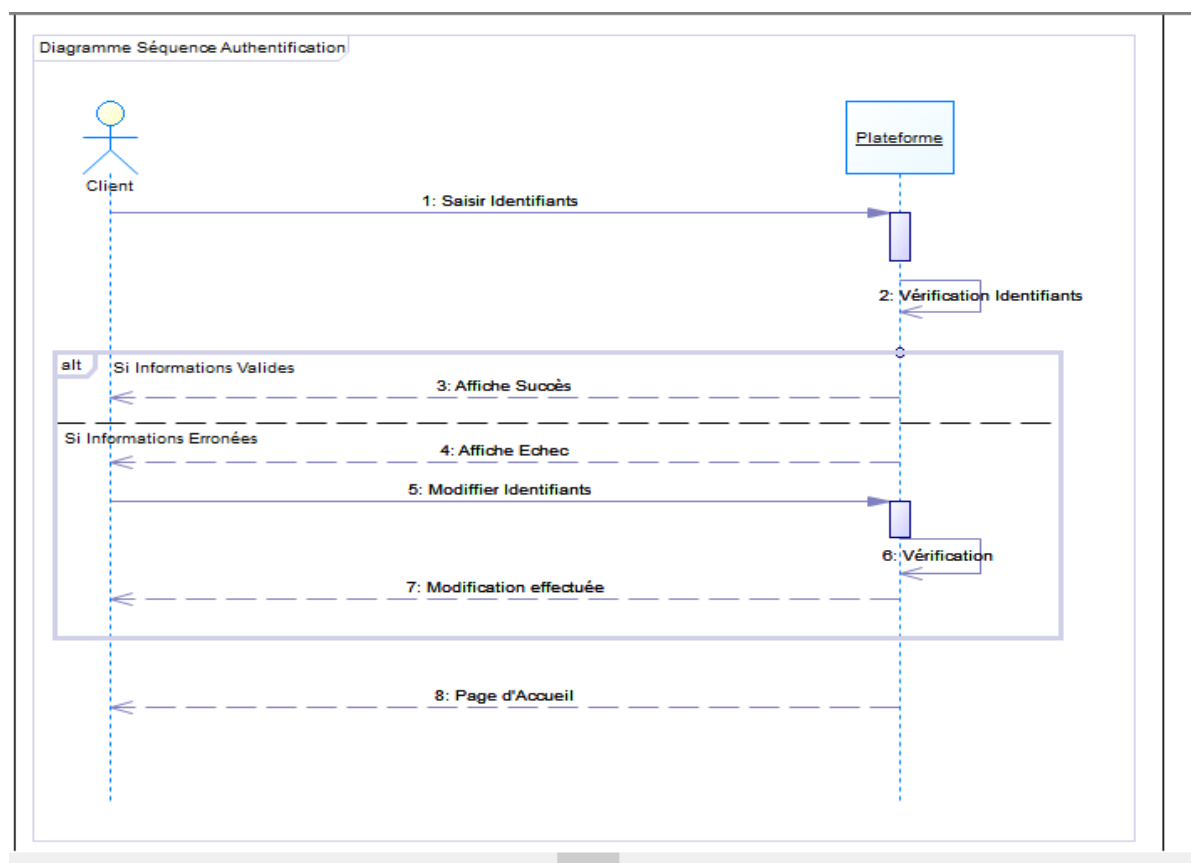


Figure 37. Diagramme de Séquence d'Authentification

³¹ [UML Cours 5 : Diagramme de séquences](#)

Explication : Dans ce diagramme de séquence d'authentification, le client commence par saisir ses identifiants, qui sont ensuite transmis à la plateforme pour vérification. Si les informations sont valides, la plateforme affiche un message de succès et redirige le client vers la page d'accueil. En cas d'erreur dans les identifiants, un message d'échec s'affiche. Le client peut alors modifier ses identifiants et les renvoyer pour une nouvelle vérification. Une fois les informations corrigées et validées, le système affiche le message « modification effectuée » et redirige également vers la page d'accueil.

❖ Diagramme de Séquence gestion de réservation

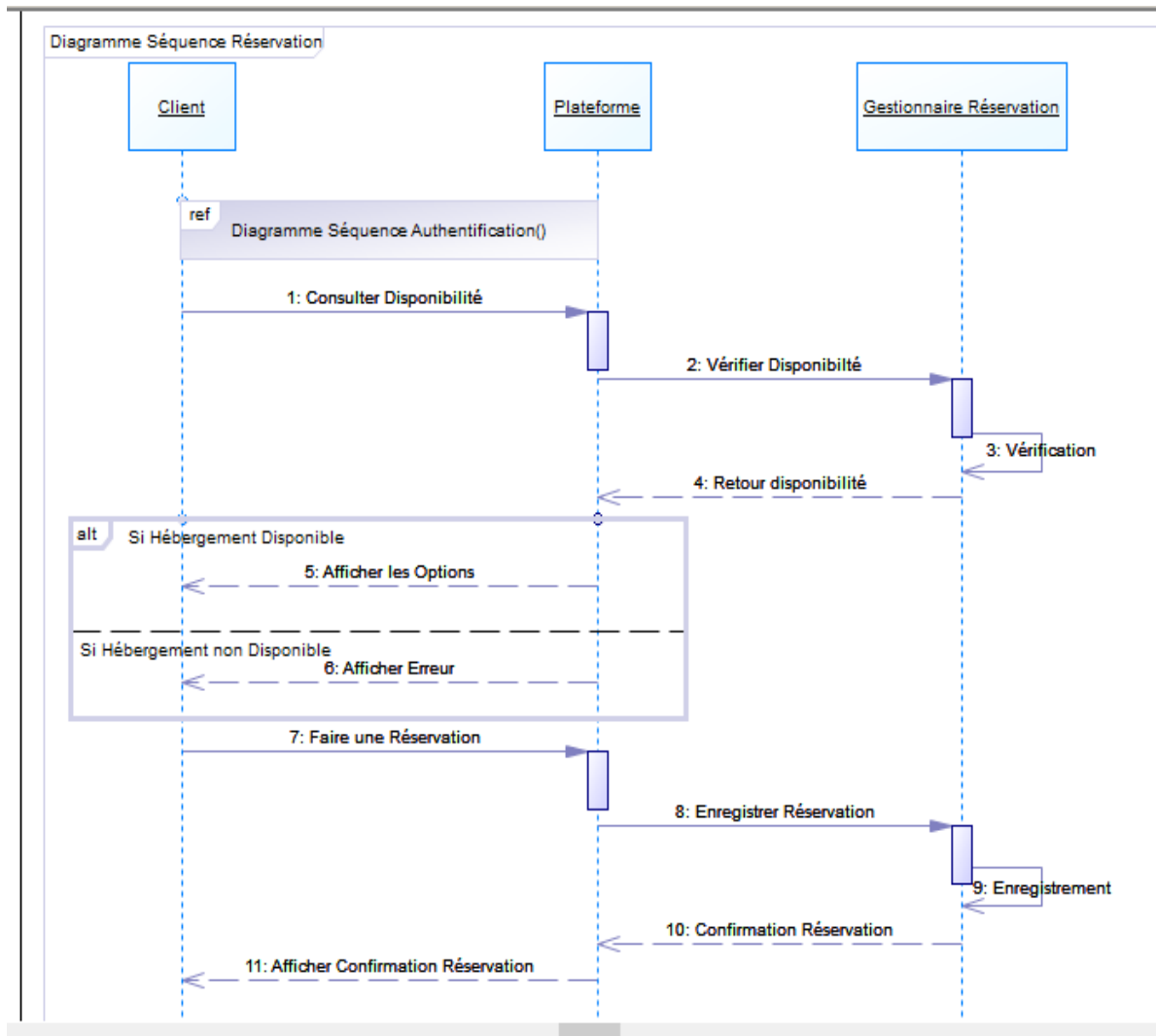


Figure 38. Diagramme de Séquence gestion de réservation

Explication : Pour ce diagramme de séquence de réservation, le client, après s'être authentifié, commence par consulter la disponibilité d'un hébergement. La plateforme transmet cette demande au gestionnaire de réservation qui effectue la vérification. Une fois la réponse obtenue, si l'hébergement est disponible, la plateforme affiche les options possibles. Dans le cas contraire, un message d'erreur est affiché. Si l'utilisateur décide de poursuivre, il

peut effectuer une réservation. Celle-ci est alors transmise au gestionnaire pour enregistrement. Une fois la réservation enregistrée, la plateforme renvoie une confirmation de la réservation au client.

❖ Diagramme de Séquence des Paiements

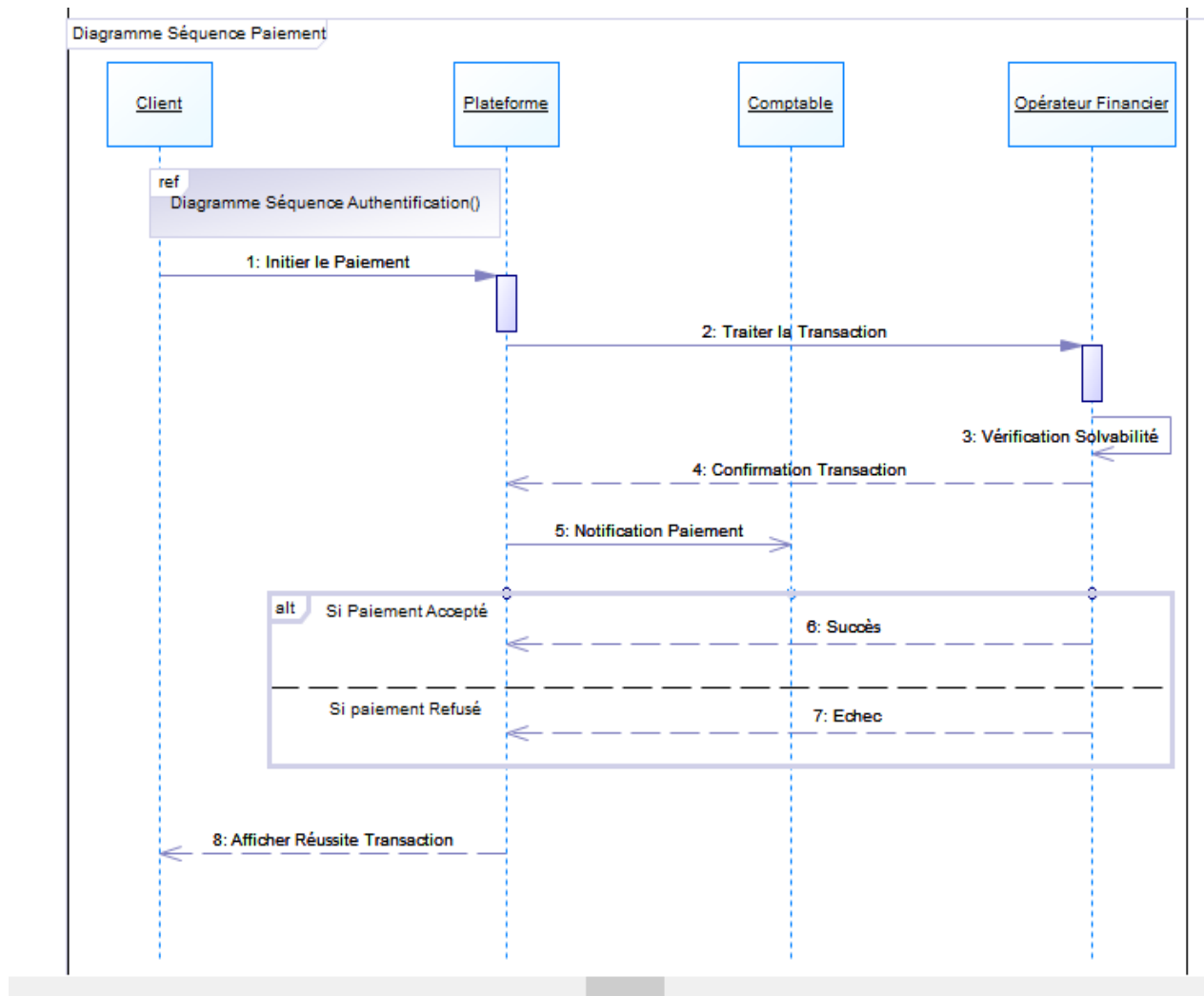


Figure 39. Diagramme de Séquence des Paiements

Explication : Dans ce diagramme de séquence relatif au paiement, le client initie la procédure depuis la plateforme. Celle-ci transmet la transaction au service comptable, qui effectue une vérification de la solvabilité via l'opérateur financier. Une fois la transaction confirmée, la plateforme est notifiée. Si le paiement est accepté, un message de succès est retourné au client ; dans le cas contraire, un échec est signalé. Enfin, une réussite de la transaction est affichée au client.

4.1.4.3. Diagramme de classes

Les diagrammes de classes sont des outils de modélisation utilisés pour représenter la structure statique d'un système. Ils montrent les différentes classes du système, leurs attributs, leurs

méthodes (ou opérations), ainsi que les relations entre elles comme l'héritage, l'association ou l'agrégation. Ces diagrammes permettent de comprendre comment les données sont organisées et comment les différentes parties du système interagissent entre elles [39].

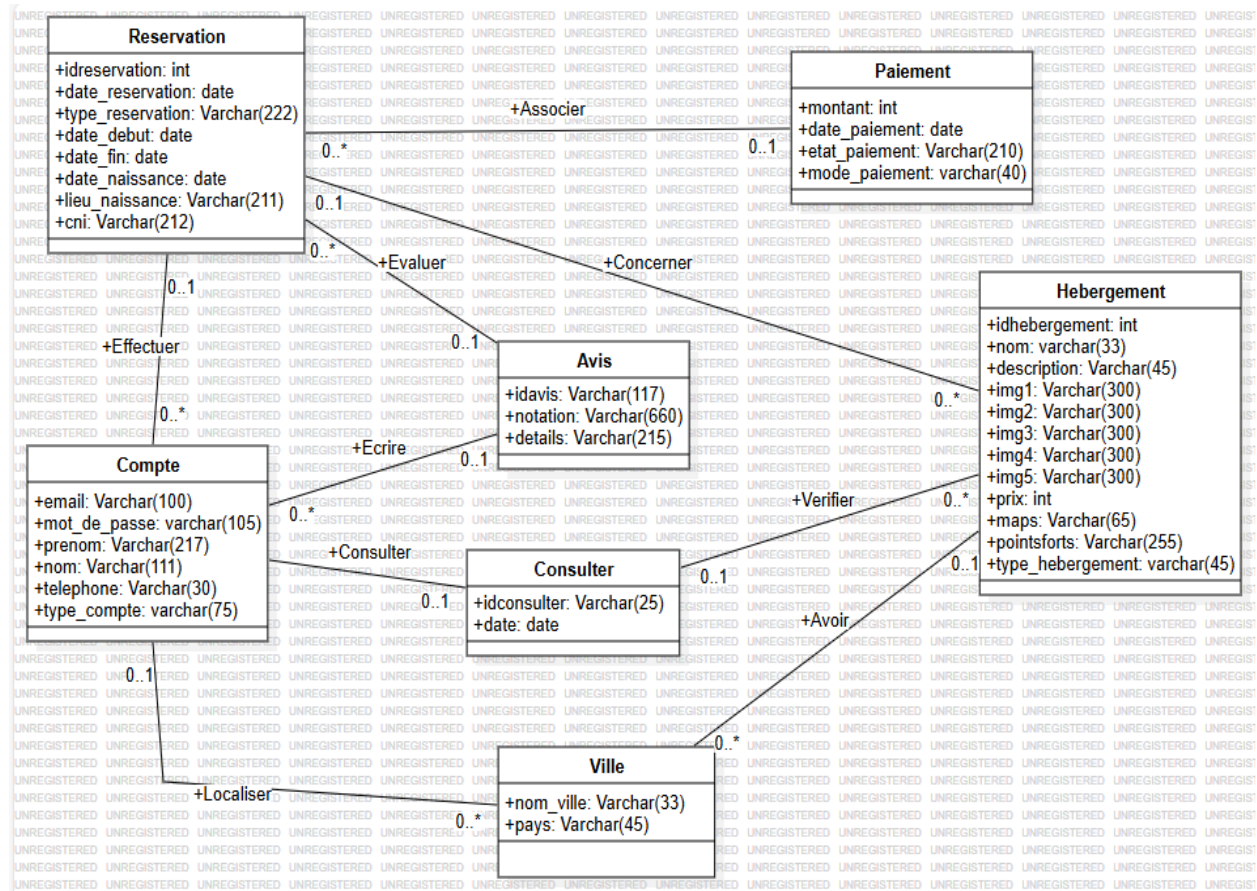


Figure 40. Diagramme de Classe Gestion Hébergements

Explication : Le diagramme représente les différentes classes du système de gestion des réservations d'hébergements. Nous avons les classes suivantes :

- **Compte** : contient les informations personnelles de l'utilisateur (email, mot de passe, prénom, etc.) et est lié à une réservation ;
- **Réservation** : enregistre les données relatives à une réservation, incluant les dates, la pièce d'identité, etc. ;
- **Paiement** : regroupe les informations sur les paiements effectués, tels que le montant, la date, l'état du paiement et le mode de paiement ;
- **Hébergement** : contient les informations des hébergements disponibles, incluant les images, le prix, les points forts, les types d'hébergements et les liens vers Google Maps ;
- **Ville** : indique la localisation géographique des hébergements ;
- **Avis** : permet aux utilisateurs de noter et de laisser des commentaires sur les hébergements ;

- **Consulter** : enregistre les consultations de fiches d'hébergement par les utilisateurs.

Ce diagramme permet ainsi de modéliser les entités principales de l'application et leurs relations pour une gestion efficace des réservations.

4.1.4.4. Modèle physique de donnée (SQL)

❖ Modèle Physique de Données

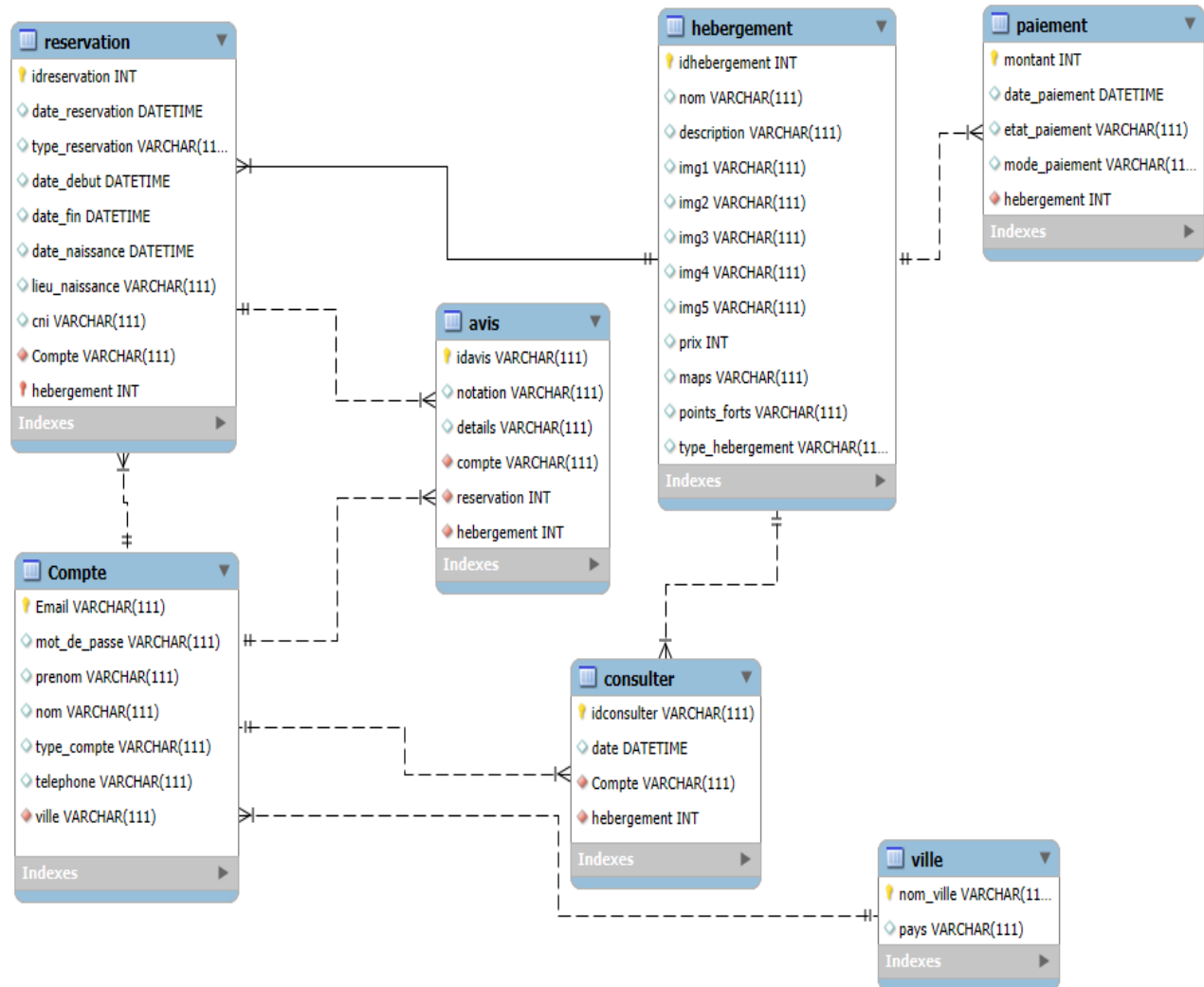


Figure 41. Modèle Physique de Données gestion hébergement

Explication : Ce diagramme de classe représente le modèle physique de données, autrement dit le **modèle relationnel** de notre système. Il illustre les relations entre les différentes tables de la base de données. En effet, chaque entité est associée à une ou plusieurs clés primaires et étrangères. Voici les principales tables et leurs liens.

- **Reservation** : liée aux tables Compte, Hébergement, Paiement et Avis ;
- **Paiement** : connectée à Reservation ;
- **Compte** : reliée à Ville, Reservation, Consulter et Avis ;
- **Avis** : dépend à la fois de Reservation et Compte ;

- **Consulter** : fait le lien entre Compte et Hébergement ;
- **Hébergement** : dépend de la table Ville, et est connectée à Reservation, Consulter et Avis ;
- **Ville** : reliée à Compte et hébergement ;

4.2. Réalisation

4.2.1. Création de la base de données

Pour la création de notre base de données, nous allons d'abord nous connecter sur notre interface phpMyAdmin [40,41]. Une fois connecté, il suffit juste de cliquer sur « nom de base de données » et y taper le nom de notre base de données nommé « **destination_senegal** ». Il suffira juste ensuite de cliquer sur « **Créer** » et la base de données est créée.



Figure 42. Création de la base de Données

❖ Importation du Fichier SQL

Après la création de notre Base de Données, nous passons à l'importation du fichier SQL. Pour cela, il faut cliquer sur « importer », sélectionner le fichier SQL à importer, puis lancer son importation en cliquant sur « Exécuter ».

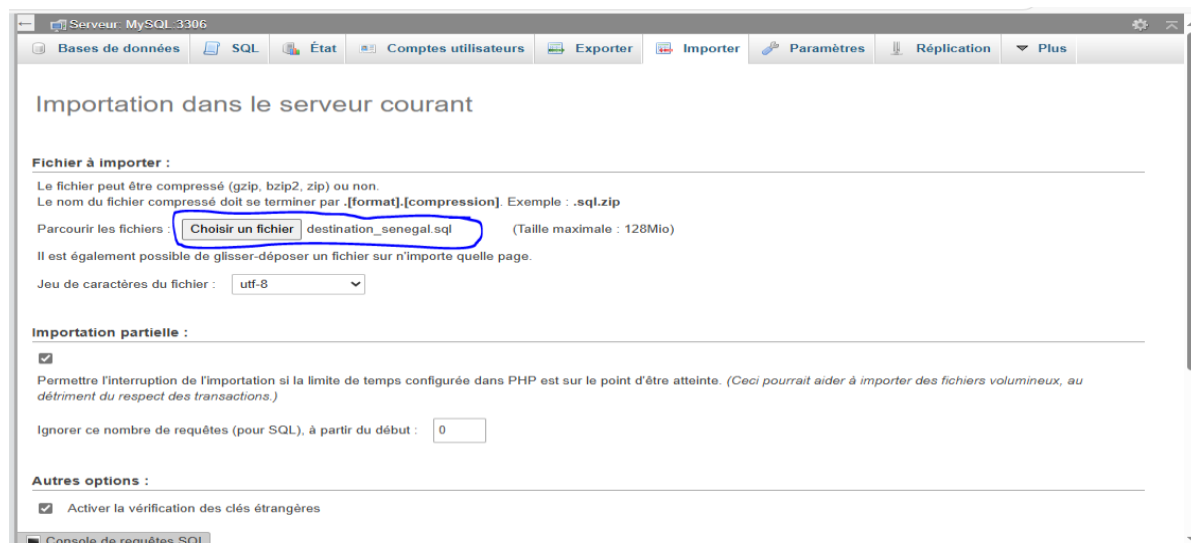


Figure 43. Importation des données du fichier SQL

Après l'exécution du fichier SQL, on note que l'importation des données a réussi. Les données ont ensuite été importé dans la base de données « destination_senegal ». Ainsi la base de données, avec ses différentes tables, se présentera comme suit :

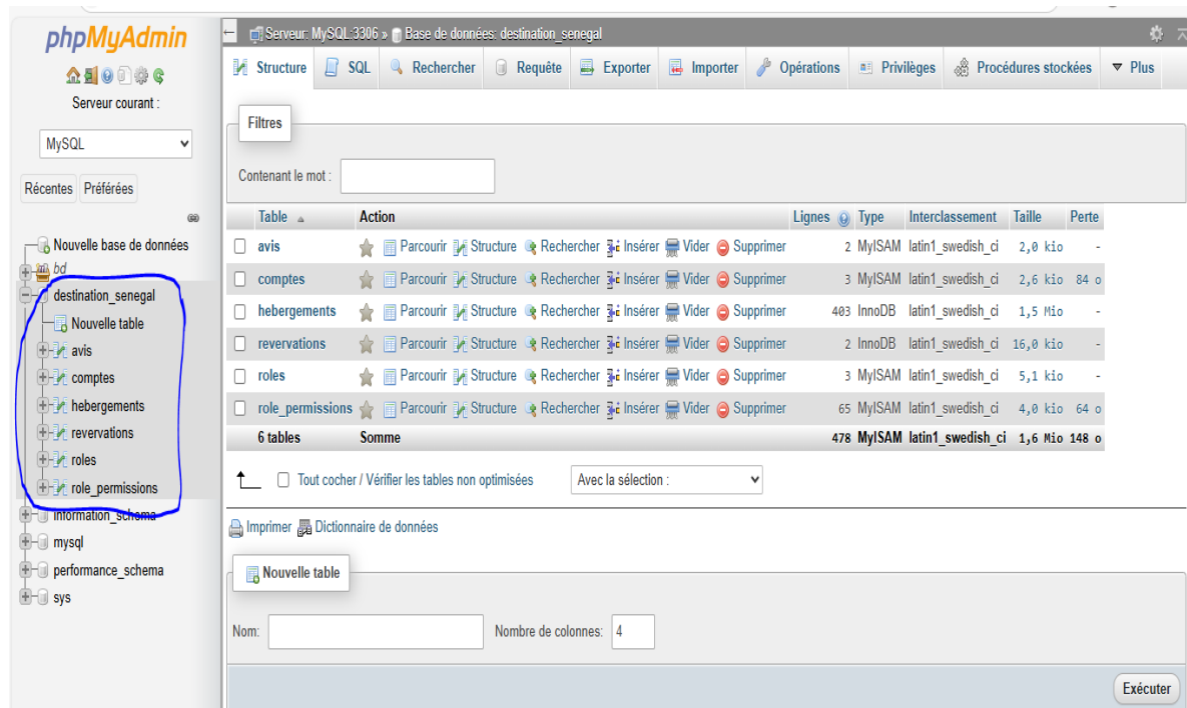


Figure 44. Base de données destination_senegal

4.2.2. Réalisation du Back END ou acquisition des données

4.2.2.1. Le crawler

Le Crawler ou bot d'indexation est un script ou programme automatisé qui parcourt les pages web afin d'en extraire des informations. Dans notre projet, on lui donne un fichier csv contenant la liste des villes côtières du Sénégal. Il interroge les plateformes comme Booking, TripAdvisor ou Airbnb pour récupérer tour à tour les noms et les url des hébergements. Ensuite, il stocke tout dans un fichier nommé **index.csv** et garde l'historique des hébergements trouvés afin d'optimiser la recherche la prochaine fois.

4.2.2.2. Le scraper

Le scraper est un programme qui visite des pages web de manière automatique pour récupérer certaines informations précises. En effet, dans notre cas, il charge le fichier index.csv stocké pour le crawler et pour chaque hébergement, il parcourt la page et récupère un certain nombre de données des hébergements de chaque ville comme : le nom, la description, le google Maps, le prix, les points forts, l'url, les photos etc. Ces données sont ensuite stockées dans un fichier nommé **hébergement.csv** et dans la base de données mysql.

4.2.2.3. Prétraitement des données

Pour le traitement des données, nous pouvons constater que le scraper a stocké les données trouvées dans deux fichiers outputs différents. En effet, un fichier nommé **reussites.csv** contenant toutes les données complètes citées dans la section 4.2.2.2 et un fichier nommé **echec.csv** contenant des données incomplètes.

4.2.2.4. Stockage des données

Le stockage des données se fait directement dans la base de données mysql. En effet, nous pouvons constater que les données issues du scraper sont stockées dans notre base de données DestinationSenegal dans la table hébergement.

The screenshot shows the phpMyAdmin interface. On the left, the database 'destination_senegal' is selected, and the table 'hebergements' is highlighted. The main area displays the table structure and data. The columns are: type_hebergement, ville, nom, url, img1, im, description, and maps. The data rows include hotels like 'Casa Mara Dakar', 'International Hotel Dakar', 'Terrou-Bi', 'Hotel Lagon 2', 'Azalai Hôtel', and 'AXIL HÔTEL'.

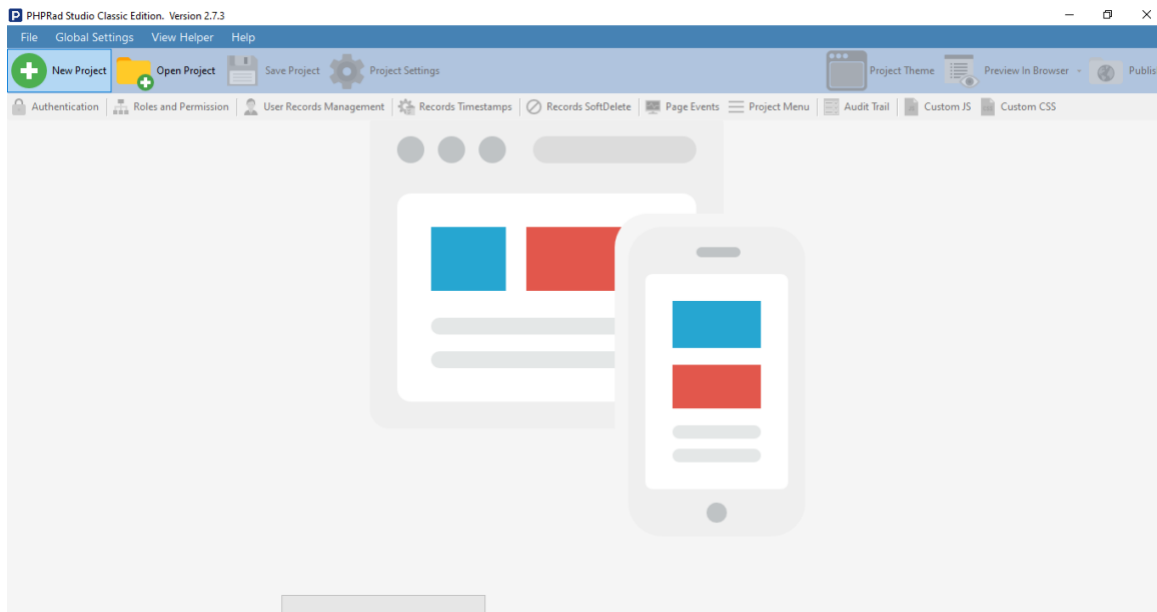
type_hebergement	ville	nom	url	img1	im	description	maps
Hôtel	Mermoz Sacre-Coeur Dakar	Casa Mara Dakar	https://www.booking.com/mara-dakar.f...	output/images/casa_mara_da		Située à Dakar à 10 minutes à pied de la plage L...	https://www.google.com/maps/place/14.7027939829923
Hôtel	Dakar	International Hotel Dakar	https://www.booking.com/vdn...	output/images/international_hc		L'établissement International Hotel Dakar vous acc...	https://www.google.com/maps/place/14.7431812+-17.4...
Hôtel	Gueule Tapee Dakar	Terrou-Bi	https://www.booking.com/bi.fr.html...	output/images/terrou_bi_img1		Face à la mer le Terrou-Bi propose un hébergement...	https://www.google.com/maps/place/14.6771833462740
Hôtel	Dakar	Hotel Lagon 2	https://www.booking.com/2.fr.html?c...	output/images/hotel_lagon_2		Situé juste en face de la mer l'Hotel Lagon 2 se ...	https://www.google.com/maps/place/14.6676554550468
Hôtel	Gueule Tapee Dakar	Azalai Hôtel Dakar	https://www.booking.com/dakar.fr.h...	output/images/azalai_hôtel_de		L'établissement Azalai Hôtel Dakar vous accueille ...	https://www.google.com/maps/place/14.678981+-17.45...
		AXIL HÔTEL	https://www.booking.com/dakar.fr.htm...	output/images/axil_hôtel_img1		L'établissement AXIL HÔTEL se trouve à	https://www.google.com/maps/place/14.675097+-17.43...

Figure 45. Chargement des données dans la base de données

4.2.3. Réalisation du moteur de recherche

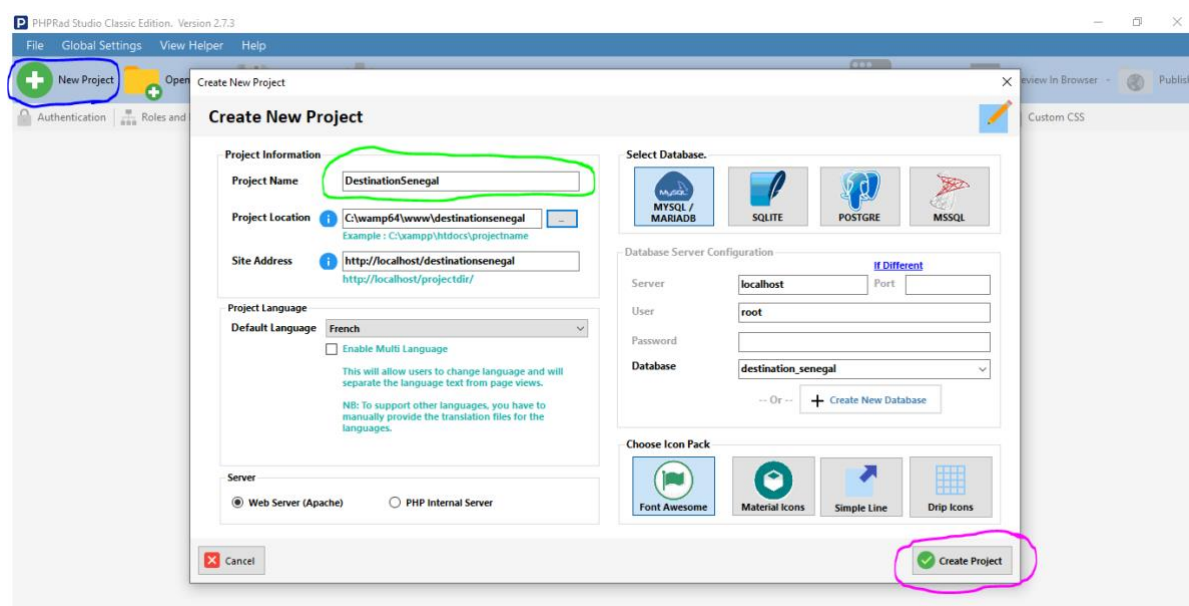
4.2.3.1. Installation du framework

Pour démarrer notre projet, nous allons installer PHHRAD comme framework de développement.

Figure 46. Interface PHPRAD

4.2.3.2. Création du projet

Pour démarrer la création du projet, nous lançons PHPRad en mode administrateur. Une fois l'interface ouverte, nous cliquons sur « **New Project** » afin d'initier un nouveau projet. Après avoir cliqué sur Nouveau projet, une interface de configuration s'ouvre. Dans celle-ci, nous indiquons le nom du projet, définir l'emplacement de sauvegarde, choisir la langue du projet, puis sélectionner le type de serveur utilisé. Si la base de données existe déjà, on doit renseigner son nom ainsi que les informations de connexion nécessaires. Sinon, il est également possible de créer une nouvelle base de données directement depuis cette interface. Nous allons enfin appuyer sur « Create projet ».

Figure 47. Création du Projet

Nous avons ainsi créé notre projet DestinationSenegal dans notre dossier notre wampServer dans le fichier www :

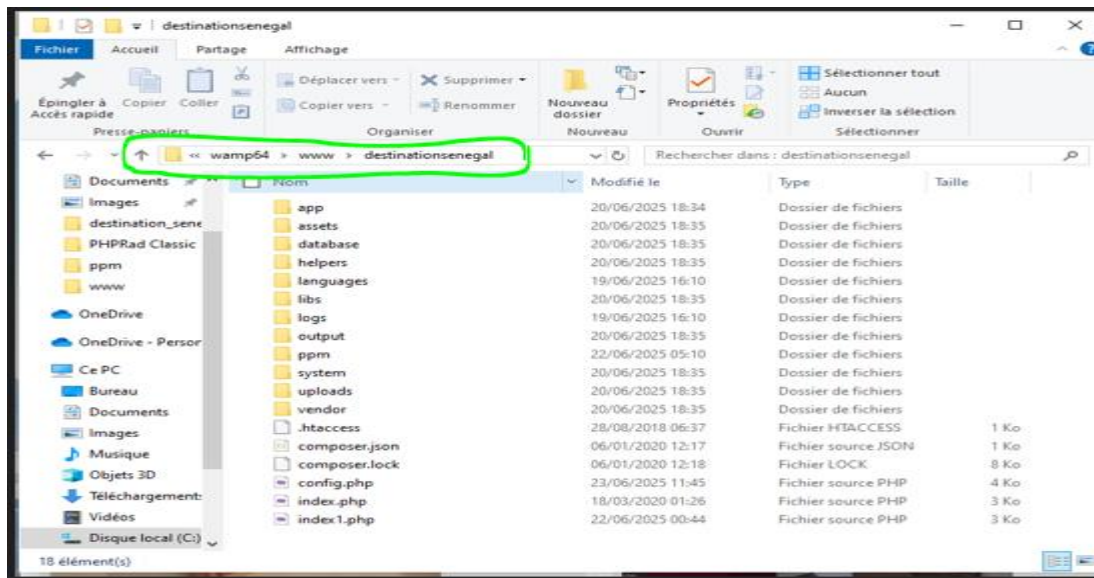


Figure 48. Enregistrement du projet

Dans notre projet destinationsenegal on y retrouve différents dossiers qui jouent chacun un rôle spécifique. Nous avons entre autres :

- App : contient la logique du backend (contrôleurs, modèles, gestion des requêtes) ;
- Assets : contient les fichiers statiques (CSS, JS, images, polices personnalisées) ;
- Database : contient notre base de données pour modification ou mise à jour des données ;
- Helpers : contient des fonctions utilitaires partagées dans tout le projet ;
- Languages : gère les traductions des langues si le projet prend en charge plusieurs langues ;
- Libs : bibliothèques PHP externes ou personnalisées utilisées dans le projet ;
- Uploads : dossier où sont stockés les fichiers envoyés par les utilisateurs (ex : photos) ;
- System : contient les fichiers système internes de PHPRad : sécurité, requêtes SQL ;
- Templates : gère les modèles de pages HTML/PHP utilisés pour générer les vues ;
- Vendor : contient les bibliothèques PHP installées automatiquement générées par Composer ;
- Output : contient les images et les topimages générées ;
- Ppm : Contient le fichier .ppm utilisé pour ouvrir/modifier le projet dans PHPRad. Ainsi, en cliquant sur ce dossier ppm, on y retrouve le fichier « destinationsenegal.ppm » propre à l'application PHPRad, qui contient la

configuration, la structure de la base de données, les modules, les pages et les options choisies lors de la création du projet.

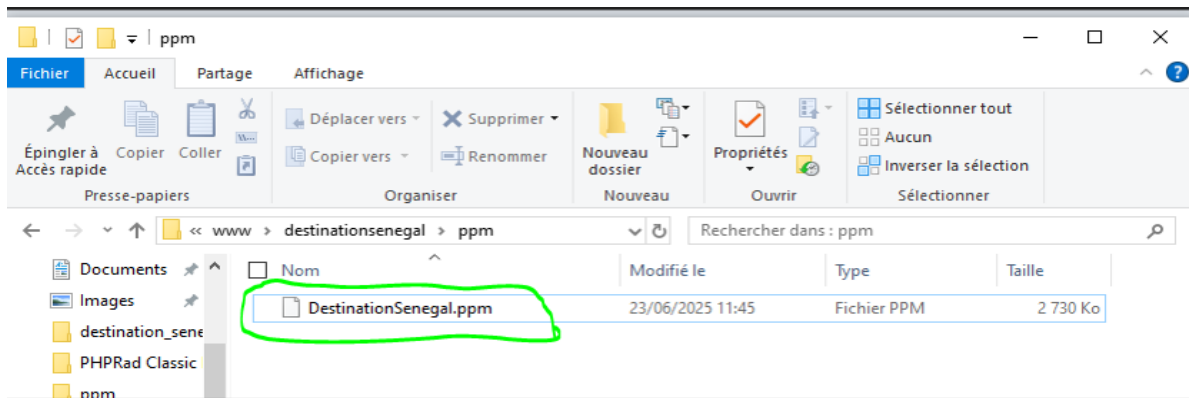


Figure 49. Enregistrement du fichier ppm

Nous notons aussi des fichiers générés tels que :

- .htaccess : Fichier Apache pour gérer les URL propres, les redirections ;
- Composer.json : Déclare les dépendances PHP du projet gérées par Composer ;
- Composer.lock : Verrouille les versions exactes des dépendances PHP utilisées ;
- Config.php : Fichier de configuration (base de données, chemins) ;
- Index.php : Point d'entrée principal du projet. Il redirige vers les pages appropriées selon l'URL.

4.2.3.3. Choix du thème

Dans cette interface, il y'a plusieurs types de Templates où on a la possibilité de choisir un thème que nous voulons en cliquant sur « Select Template ». Nous pouvons même choisir la couleur de fond et la police des écritures de la barre de navigation principale en cliquant sur « Navbar Variant ». Pour le choix du thème, nous allons cliquer sur « Project Theme ».

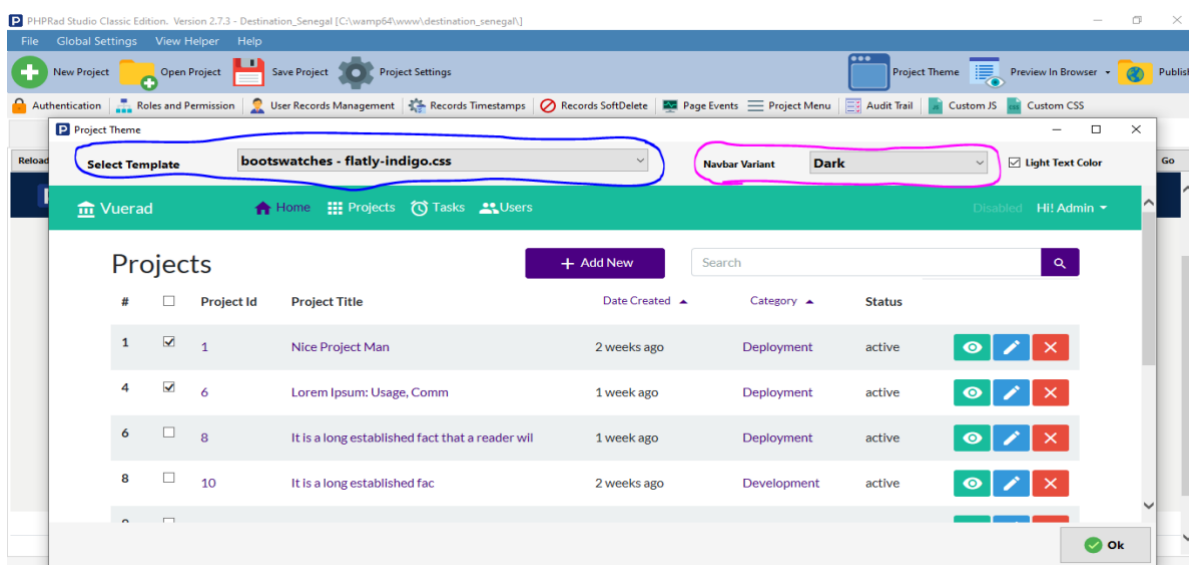


Figure 50. Choix du thème

4.2.3.4. Gestion des pages

Pour la gestion des pages, nous avons une section nommée « **Database Tables** » où toutes les tables de notre base de données sont listées. Il y'a aussi un menu composé de plusieurs modules qui sont :

- List Page : qui affiche la liste complète des données ;
- View Page : qui affiche les détails d'un élément ;
- Add Page : qui modifie un élément existant ;
- Edit Page : qui ajoute un nouvel élément ;
- Delete Page : qui supprime un élément.

A cela s'ajoute qu'on peut gérer le choix design de la page dans l'onglet « **Design** ». Dans l'onglet « **Custom** » aussi, on peut manuellement faire des modifications du contenu de la page générée comme ajouter du code HTML ou Javascript ou encore une mise en forme spécifique. Il existe aussi une section nommée « **Pages Fields** » qui correspond à la configuration des champs qui seront affichés dans la page. Il permet de voir les types de données de la table (FieldsName), quels champs de la table seront visible (Output), ou encore quel champ peut être utilisé dans la barre de recherche (Search).

Dans la section « **Fields Properties** », nous pouvons personnaliser les propriétés d'un champ sélectionné, comme son affichage, son alignement, sa relation avec d'autres tables ou son rôle dans la base de données.

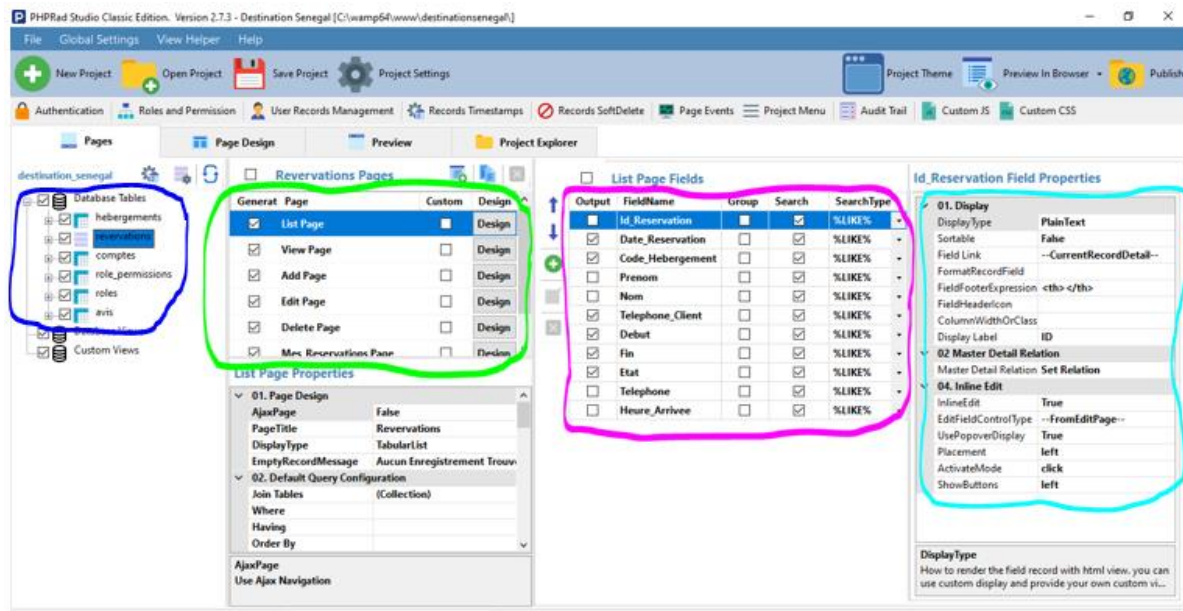


Figure 51. Gestion des pages

4.2.3.5. Gestion du design

La gestion du design permet de personnaliser l'apparence des pages. Pour ce faire, nous allons cliquer sur « **Page Design** » ou nous pourrions ajouter une nouvelle section ou une nouvelle colonne. On peut aussi y gérer dans l'onglet « **Pages, Components and Widgets** » les composants comme « Header », « Button », « Search Field » où on peut ajouter ou modifier des données des tables. Il existe aussi en bas des composants tels que :

- Charts : qui consiste à faire des graphiques ;
- Sub page : qui permet de créer des sous pages ;
- Records Count : qui permet de voir le nombre total d'enregistrement dans une table ;
- View Component : qui permet de voir le mode d'affichage d'un champ ;
- Page Filter Components : qui consiste à sélectionner les filtres visibles en haut d'une page.

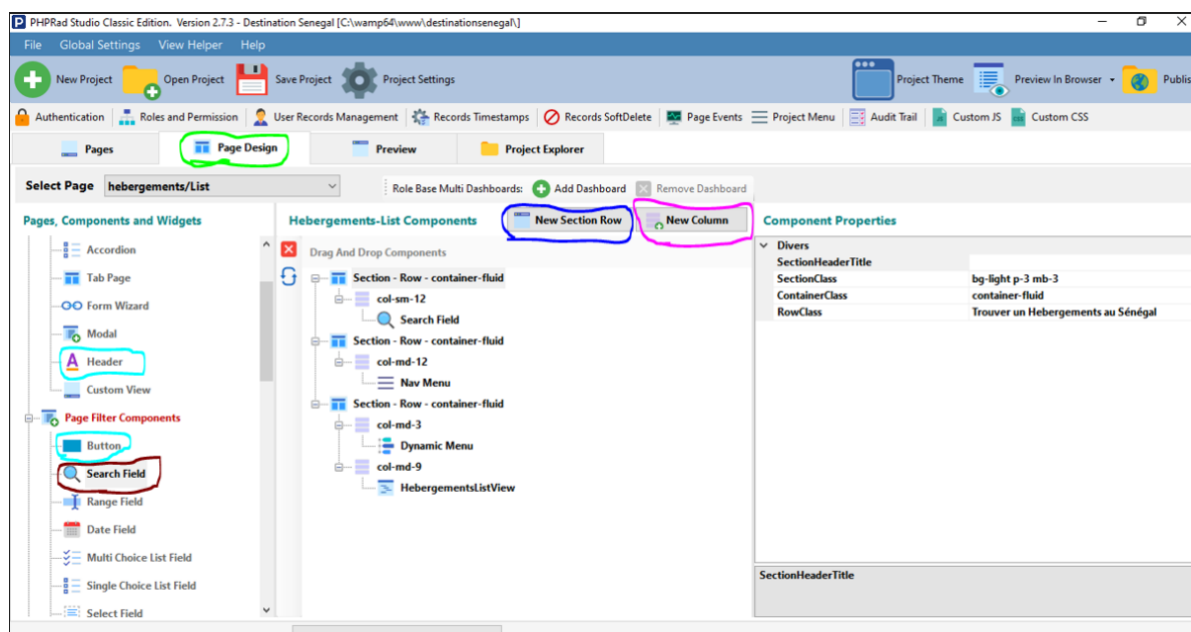


Figure 52. Gestion du design

4.2.3.6. Gestion du menu

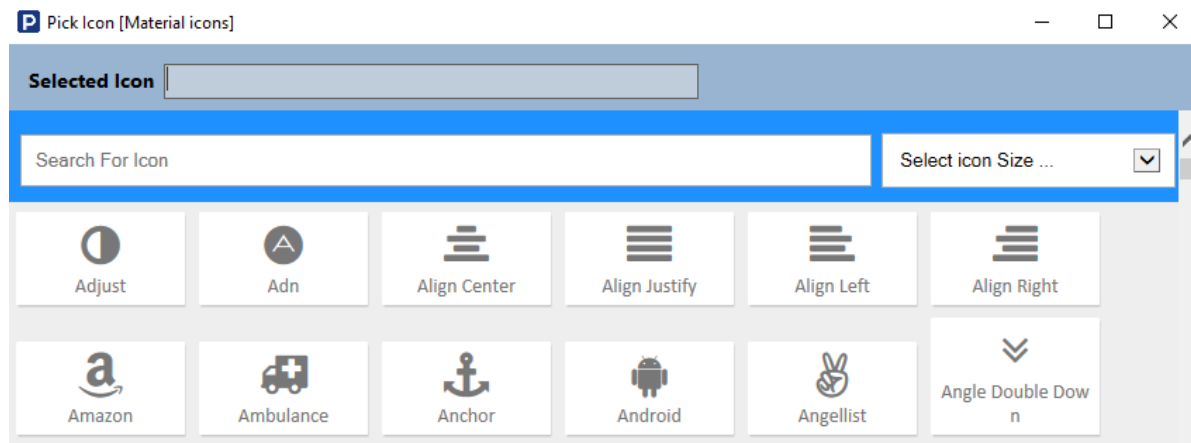
La gestion du menu permet d'organiser la navigation de l'application en ajoutant, supprimant ou réorganisant les éléments du menu. Chaque menu a les propriétés suivantes :

Path : Contient le lien vers la page à ouvrir ;

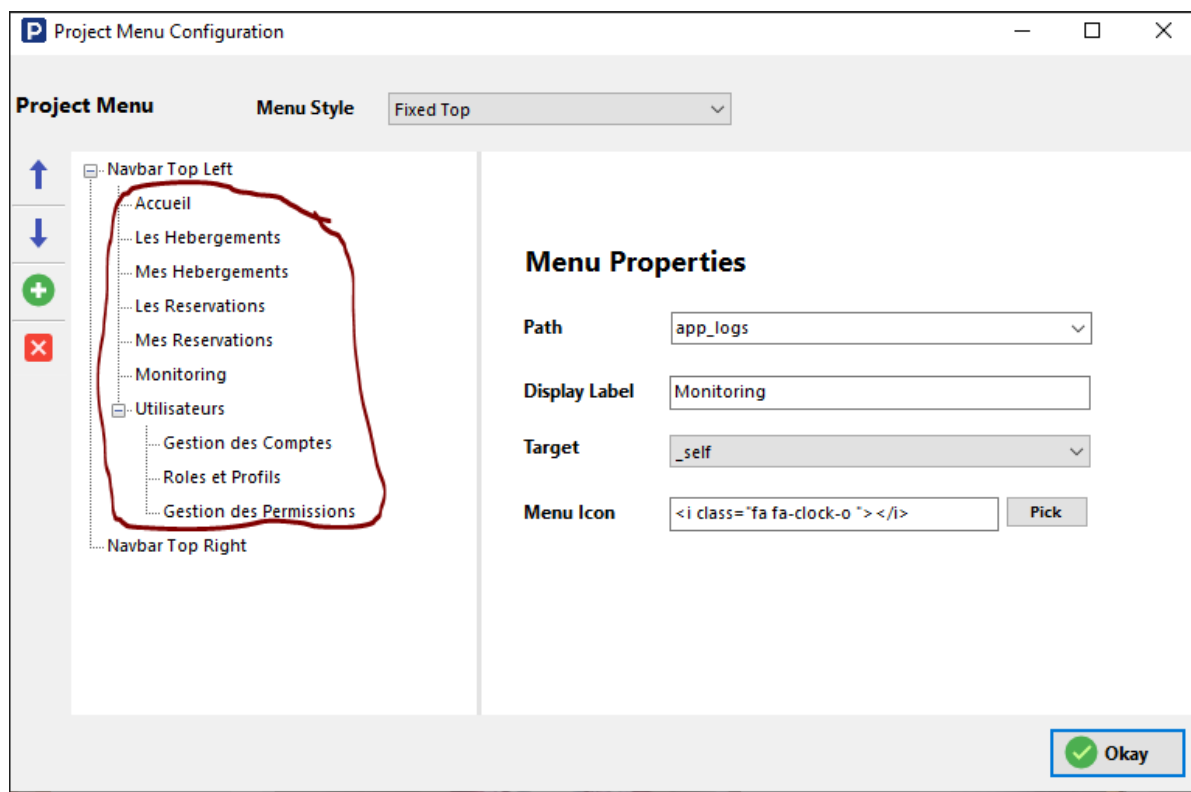
Display Label : Contient le nom du menu visible à l'utilisateur ;

Target : contient l'onglet où s'ouvre la page ;

Menu Icon : Constitue l'icône affichée à gauche du libellé du menu (Voir figure ci-dessous) ;

Figure 53. Liste Menu Icon

Pour ce faire, nous allons cliquer sur « Project Menu ».

Figure 54. Gestion du menu

4.2.3.7. Gestion des utilisateurs

La gestion des utilisateurs permet de gérer les authentifications de chaque utilisateur. Pour ce faire, nous allons cliquer sur « Authentification » et ensuite choisir la table qui va gérer le processus comme notre table **Compte**. Cela veut dire que les utilisateurs devront s'authentifier à partir des données contenues dans la table **Compte**. Nous avons entre autres :

- User Name Field : champ qui contient le nom de l'utilisateur pour se connecter ;
- Password Field : champ qui contient le mot de passe ;

- User ID Field : champ qui contient l'identifiant unique de l'utilisateur ;
- Email Field : champ qui contient l'adresse email ;
- Profile Photo : Champ qui contient le fichier de la photo de l'utilisateur ;

Pour les options d'authentications, nous avons choisi de cocher le champ « **UserName or Email** », c'est-à-dire l'utilisateur peut se connecter soit avec son nom d'utilisateur, soit avec son email. Nous avons aussi cocher le champ « **Allow User to Register** » qui permettra à n'importe quel utilisateur de créer un compte à partir d'un formulaire de connexion et le champ « **Auto Login After Registration** » qui Connecte automatiquement l'utilisateur après son inscription dans la plateforme. Il existe aussi d'autres champs chacun jouant un rôle spécifique. Nous avons entre autres :

- Force Email Verification : qui exige une confirmation par email après l'inscription de l'utilisateur ;
- Enable Password Reset : qui active la fonction du « Mot de Passe Oublié » ;
- Password Reset Interval : qui donne à l'utilisateur un délai entre deux tentatives de réinitialisation de mot de passe ;
- Administrator Approve Account : l'inscription n'est valide que si l'administrateur l'approuve.

Nous avons aussi choisi de sélectionner la table « hebergement/Accueil », cela veut dire que l'utilisateur ne peut pas voir toutes les pages sauf la page « hebergement/Accueil » qui ne nécessite pas une authentification.

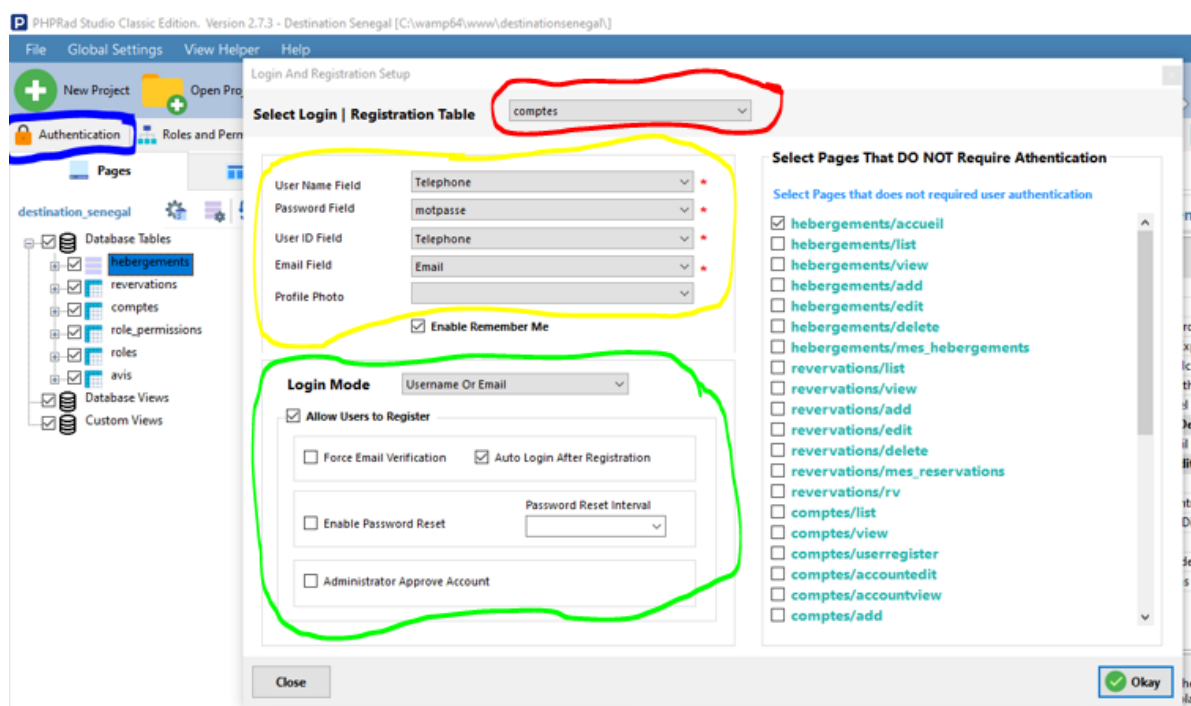


Figure 55. Gestion des utilisateurs

4.2.3.8. Gestion des rôles et privilèges

Pour la gestion des rôles, chaque utilisateur a ses propres privilèges et permissions dans le système. Dans notre projet nous avons trois utilisateurs qui sont : l'administrateur, le gérant et le client. L'administrateur a toutes les permissions (Ajouter, modifier, lister, consulter et supprimer des données) dans la plateforme. Il est chargé aussi de donner des permissions au gérant ou au client en cochant les actions qu'ils peuvent effectuer dans chaque page de la plateforme. Afin de gérer cela, nous devons cliquer sur « Rôles et Permissions ».

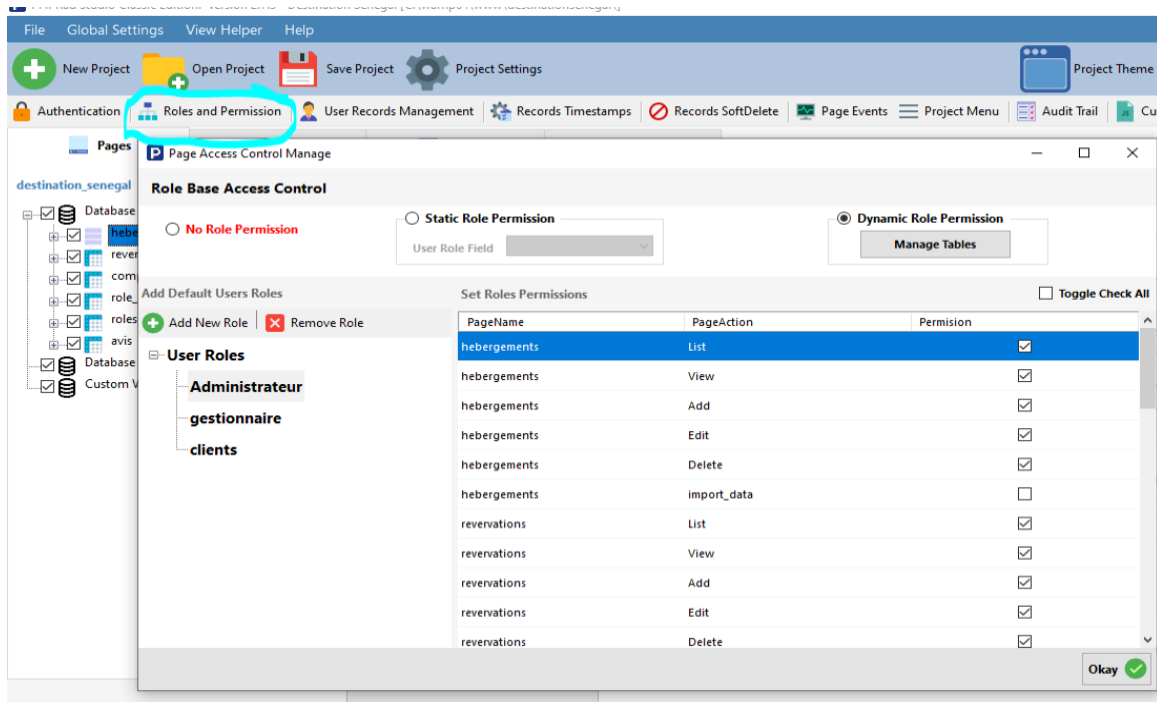


Figure 56. Gestion des rôles et privilèges

4.2.3.9. Gestion du monitoring ou audits

La gestion du monitoring ou de l'audit permet de suivre les actions effectuées par les utilisateurs dans l'application (ajout, modification, suppression de données). Elle offre un historique détaillé des opérations, utile pour la traçabilité, la sécurité et le contrôle des accès. Pour ce faire, nous allons cliquer sur « Audit Trail » et voir si un utilisateur a consulté, modifié, ajouté ou supprimé un hébergement. On peut voir aussi si un utilisateur a modifié, ajouté ou supprimé une réservation ou encore s'il a ajouté ou supprimé un compte.

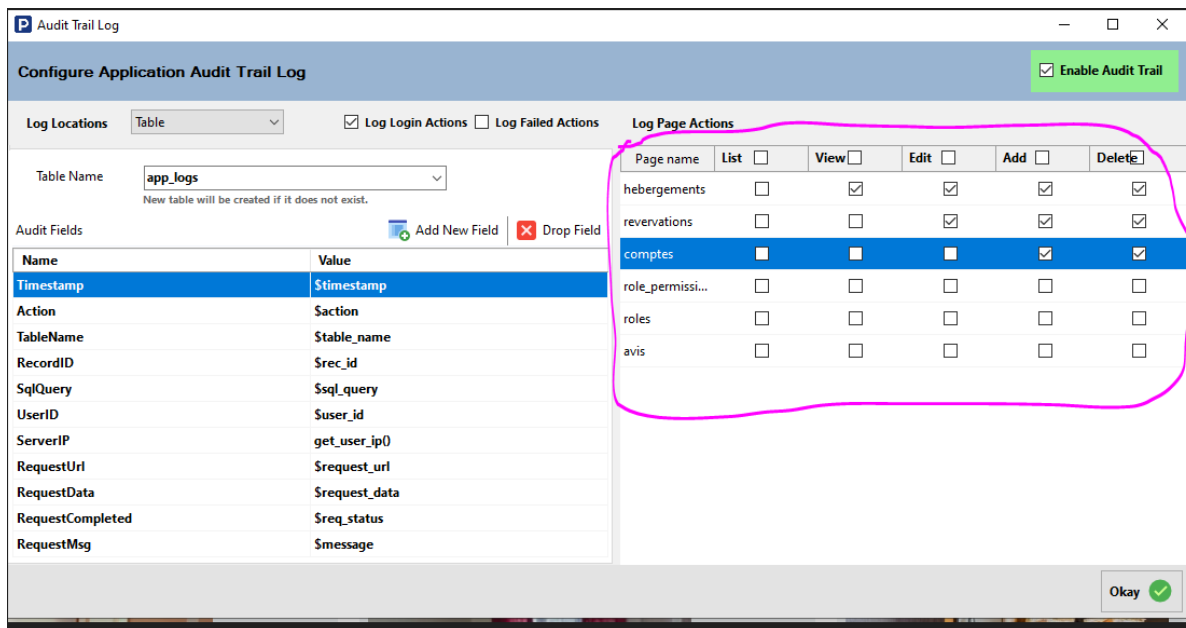


Figure 57. Gestion du monitoring ou des audits

Dans notre projet, seul l'administrateur peut voir la liste des utilisateurs qui ont consulté la plateforme avec leurs emails, l'heure, la date entre autres avec la page « app_logs ».

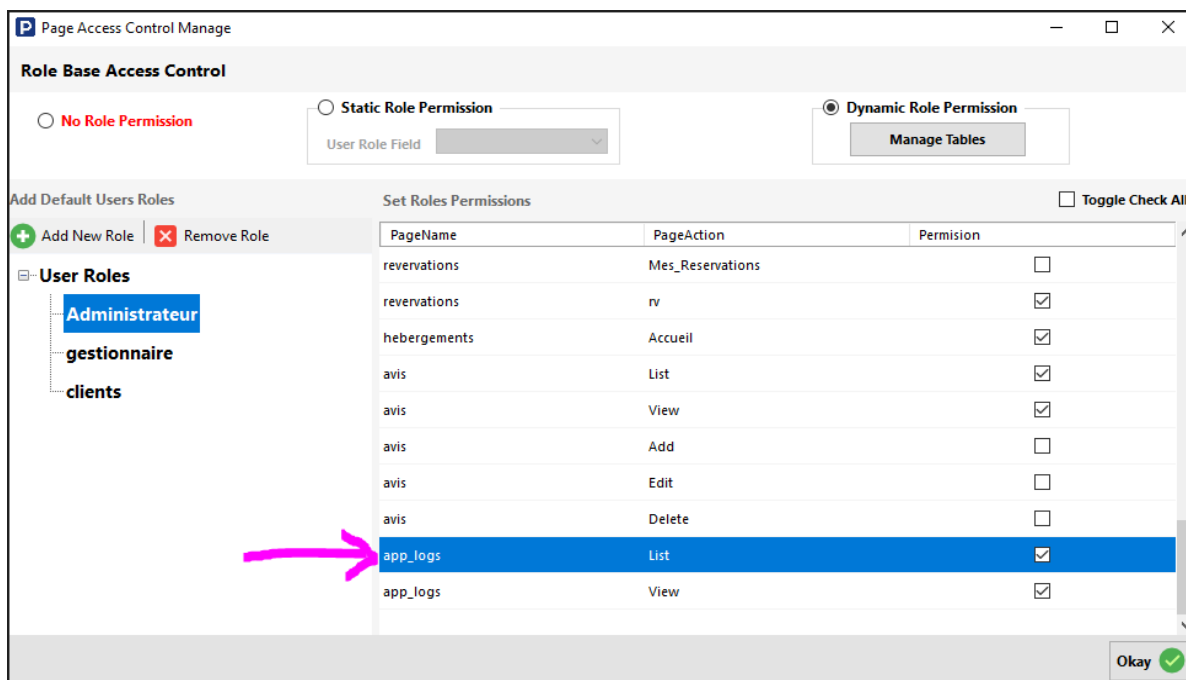


Figure 58. Affichage du monitoring

4.2.3.10. Déploiement de la plateforme

Dans cette section, on peut voir le nom du site contenant notre projet (SiteName), l'adresse du site (SiteAdress), le répertoire de notre projet (Project Directory), l'adresse du serveur de base de données où MySQL est installé (DatabaseServer), le nom de l'utilisateur MySQL utilisé pour la connexion (DatabaseUserName), le nom de la base de données

(DatabaseName). Pour le déploiement de notre plateforme, nous allons appuyer sur « **Publish** ».

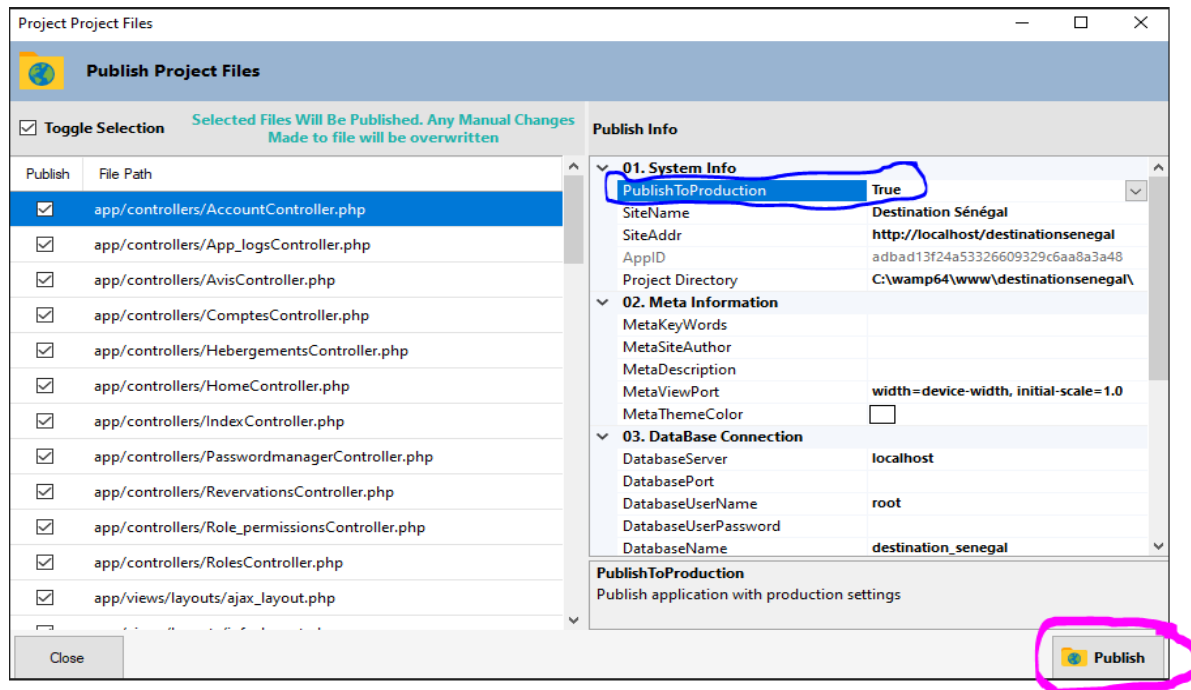


Figure 59. Publication du projet Destination Sénégal

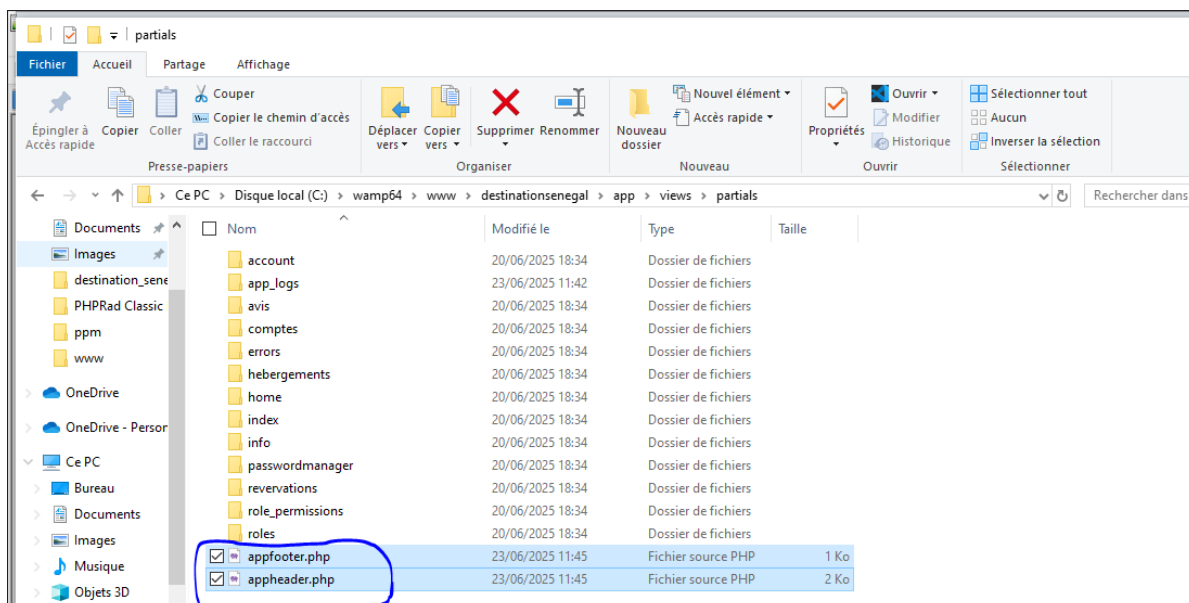
Ce fichier de configuration nous permet de gérer les paramètres de connexion de notre projet à la base de données afin de faciliter la maintenance et la modification sans toucher au code principal de l'application. Il est composé des éléments suivants :

- `define("DB_HOST", "localhost")` : qui définit le nom d'hôte du serveur de base de données (ici, localhost, c'est-à-dire le serveur local) ;
- `define("DB_USERNAME", "root")` : qui définit le nom d'utilisateur pour accéder à la base de données (ici, root, est l'utilisateur par défaut de MySQL que nous avons utilisé) ;
- `define("DB_PASSWORD", "")` : qui définit le mot de passe de la base de données (ici, c'est vide car nous travaillons en local) ;
- `define("DB_NAME", "destination_senegal")` : qui définit le nom de la base de données à utiliser (ici, destination_senegal est le nom de notre base de données) ;
- `define("DB_TYPE", "mysql")` : qui définit le type de base de données (ici, nous avons utilisé MySQL) ;
- `define("DB_PORT", "")` : qui définit le port utilisé pour la connexion (ici, le port par défaut est utilisé) ;
- `define("DB_CHARSET", "utf8")` : qui définit le jeu de caractères utilisé lors de la connexion (ici, nous avons utilisé utf8 comme jeu de caractères).

```

85
86 // Database Configuration Settings
87 define("DB_HOST", "localhost");
88 define("DB_USERNAME", "root");
89 define("DB_PASSWORD", "");
90 define("DB_NAME", "destination_senegal");
91 define("DB_TYPE", "mysql");
92 define("DB_PORT", "");
93 define("DB_CHARSET", "utf8");
94

```

Figure 60. Fichier configuration du projetFigure 61. Affichage du fichier app

Explication : Dans le dossier app de notre projet se trouve un dossier views contenant deux sous dossiers : **layouts** et **partials**. Dans ce dossiers **partials** se trouvent deux fichiers sources PHP qui sont :

- **appHeader.php** : qui contient généralement le code html et php qui doit apparaître en haut de chaque page du site. Il est utilisé dans notre projet pour gérer l'en-tête commun de toutes les pages de notre plateforme ;
- **appFooter.php** : qui contient le code html et php qui doit apparaître en bas de chaque page. Il sert à afficher le pied de page commun à toutes les pages de notre plateforme.

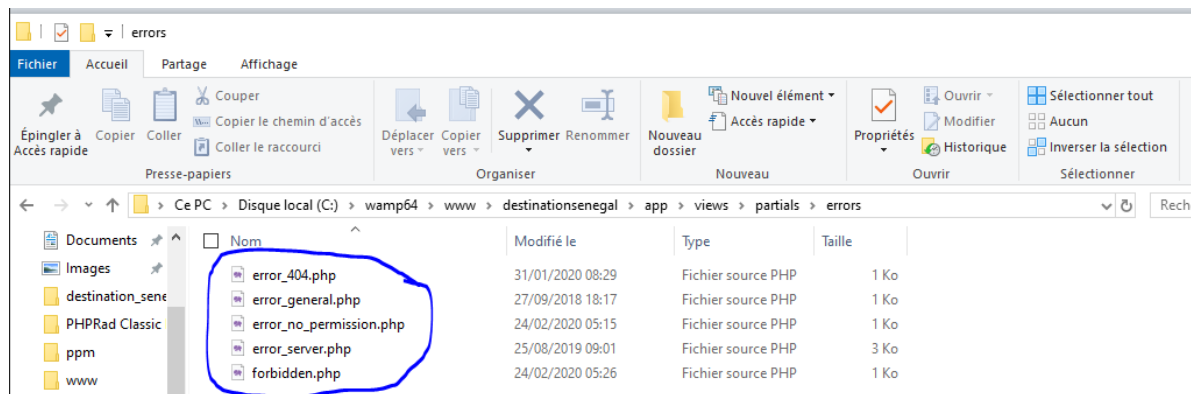


Figure 62. Affichage du fichier Errors

Explication : Notre fichier Errors contient les erreurs liées à la page d'accueil, à la page de connexion, au serveur etc. Ces fichiers PHP sont utilisés pour afficher des pages d'erreur personnalisées selon le type de problème rencontré. Ces messages d'erreurs sont souvent affichés en anglais que nous pouvons modifier en français. Chaque message d'erreur a un rôle spécifique. Nous avons :

- error_404.php : qui affiche une page d'erreur lorsque l'utilisateur tente d'accéder à une page qui n'existe pas (erreur HTTP 404) ;
- error_general.php : Sert de page d'erreur générique pour afficher un message en cas d'erreur inattendue ;
- error_no_permission.php : qui affiche une page d'erreur lorsqu'un utilisateur tente d'accéder à une page sans avoir les droits nécessaires ;
- error_server.php : Présente une page d'erreur en cas de problème interne du serveur (erreur HTTP 500) ;
- forbidden.php : affiche une page d'erreur lorsque l'accès à une ressource est strictement interdit (erreur HTTP 403).

En conclusion, la conception du système présentée dans ce chapitre a permis d'organiser les différents éléments du projet au sein d'une architecture robuste et efficace. Chaque composant a été clairement défini et intégré afin de garantir une cohésion parfaite et une interconnexion parfaite.

Chapitre 5 : Présentation de la solution

5.1. Accès à la plateforme

Pour accéder à la plateforme Destination Sénégal il faut taper l'adresse suivante : <http://localhost/destinationsenegal/>. Dans cette page d'accueil, si l'utilisateur veut visiter la plateforme, s'il a un compte, il doit fournir son nom d'utilisateur ou email et son mot de passe et cliquer sur « **S'identifier** ».

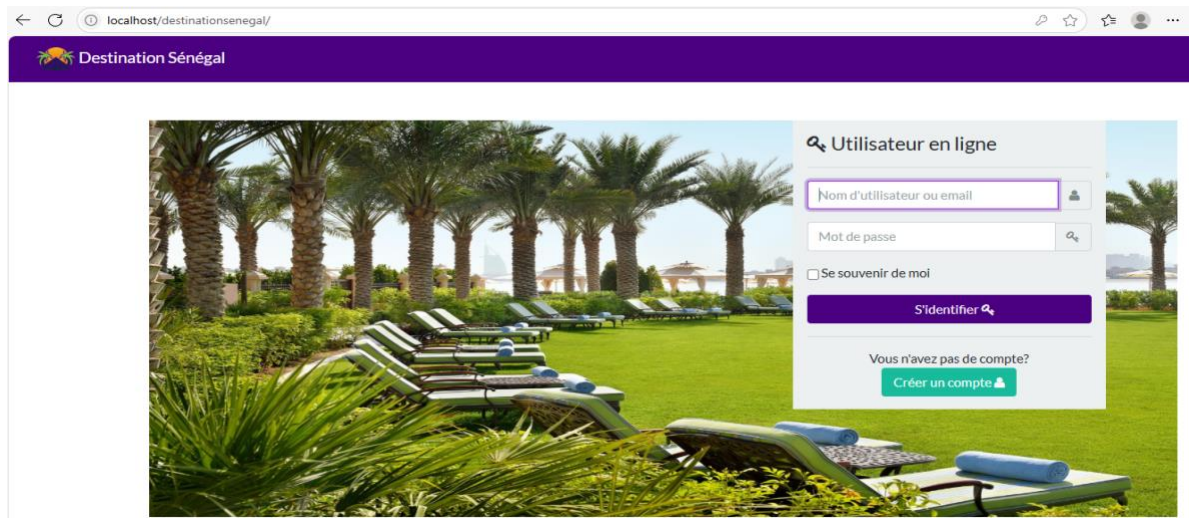


Figure 63. Page d'accueil de la plateforme

5.2. Page de connexion et déconnexion

La page de connexion permet aux utilisateurs de se connecter à leur compte en saisissant leur email et leur mot de passe. Ils doivent s'assurer que les informations fournies sont exactes pour pouvoir accéder aux fonctionnalités de la plateforme. Depuis cette page, si l'utilisateur ne possède pas de compte, alors il a la possibilité de créer un nouveau compte en appuyant sur « **Créer un compte** ».

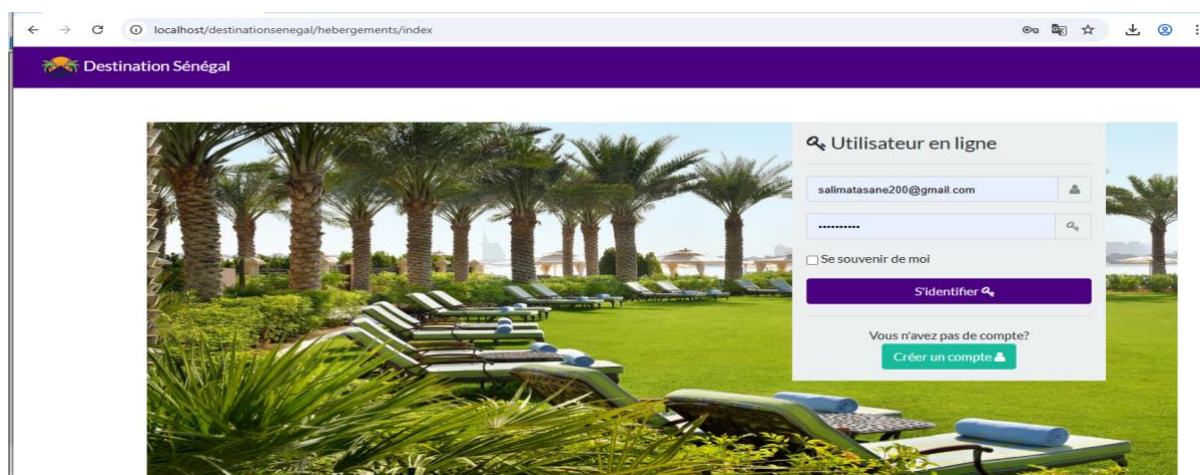


Figure 64. Page de connexion

5.3. Page création de compte

La page de création de compte permettra aux utilisateurs de créer un compte pour pouvoir accéder à notre plateforme. Pour cela, il suffira de renseigner les informations suivantes : l'email, le téléphone, le mot de passe, confirmer le mot de passe, le prénom, le nom, le sexe, la nationalité.

Figure 65. Création de compte

5.4. Espace Administrateur

Dans l'espace administrateur, il apparaît les onglets suivants : Accueil, les hébergements, les réservations, le monitoring et la gestion des utilisateurs. L'administrateur gère les fonctionnalités suivantes dans la plateforme :

- Gestion des hébergements : il peut consulter, modifier ou supprimer un hébergement.

Figure 66. Gestion hébergements

Il peut aussi à partir de là, voir la fiche du logement.

Destination Sénégal Accueil Les Hebergements Les Reservations Monitoring Utilisateurs 781812471!

Fiche du logement

Code Logement:	7100
Type Logement:	Appartement
Ville:	Dakar
Nom:	Appartement sympathique Jules Ferry.
Telephone:	
Email:	
Url:	appartement-sympathique-jules-ferry.fr.html?checkin=2025-08-08&checkout=2025-08-09&no_rooms=1
Image Principale:	

Figure 67. Fiche de logement

- Gestion des réservations : il a la possibilité de voir la liste des réservations déjà émises mais aussi de voir, modifier ou de supprimer une réservation.

Destination Sénégal Accueil Les Hebergements Les Reservations Monitoring Utilisateurs 781812471!

Gestion des Reservations + Nouvelle Reservation Chercher

#	ID	Logement	Arrivée	Départ	Prenom	Nom	Etat	
1	7	7100	2025-08-19	2025-06-29	Asta	Seye	EMISE	View Edit Delete
2	6	7027	2025-07-05	2025-08-08	Fatou	Sow	EMISE	View Edit Delete
3	5	7100	2025-06-28	2025-07-04	Salimata	Sané	EMISE	View Edit Delete
4	4	6425	2025-06-21	2025-06-22	Awa	SOW	EMISE	View Edit Delete
5	3	7101	2025-06-13	2025-06-11	Edouard Ngor	SARR	EMISE	View Edit Delete

EXCEL Importer Records: 5 sur 5

Figure 68. Liste des réservations

Il peut aussi ajouter une nouvelle réservation en cliquant sur « **Nouvelle Reservation** », renseigner les champs et cliquer sur « **Réserver** ».

Destination Sénégal Accueil Les Hebergements Les Reservations Monitoring Utilisateurs 781812471!

Nouvelle Reservation

Logement *	Entrer Logement
Prénom *	Entrer Prénom
Nom *	Entrer Nom
Telephone *	Entrer Telephone
Date Arrivée *	Entrer Date Arrivée
Date Départ *	Entrer Date Départ
Heure Arrivée	Entrer Heure Arrivée

RESERVER

Figure 69. Ajout d'une nouvelle réservation

- Gestion du monitoring : il a la possibilité de voir les utilisateurs qui ont consulté la plateforme avec leurs emails, leur mot de passe, la date et l'heure de la consultation.

#	Log Id	Timestamp	Action	Tablename	Recordid	Sqlquery	Userid	Serverip	Requesturl	Requestdata
1	12	2025-06-23 19:24:40	userlogin	comptes		SELECT * FROM comptes WHERE Telephone = ? OR Email = ? LIMIT 1	781812471	::1	index/login/	{"username":"salimatasane200@gmail.com","password":"781812471S","motpasse":"\$2y\$10"
2	11	2025-06-23 19:23:51	userlogout	comptes			781812471	::1	index/logout	{}
3	10	2025-06-23 19:02:49	userlogin	comptes		SELECT * FROM comptes WHERE Telephone = ? OR Email = ? LIMIT 1	781812471	::1	index/login/	{"username":"salimatasane200@gmail.com","password":"781812471S","motpasse":"\$2y\$10"

Figure 70. Gestion du monitoring

- Gestion des comptes : Il peut créer, voir, modifier ou supprimer un compte.

#	Email	Téléphone	User Role Id	Prenoms	Nom	Sexe
1	salimatasane200@gmail.com	781812471	1	Salimata	Sané	Femme
2	s.s2@zig.univ.sn	760179878	2	Moussa	Diatta	Homme
3	esarr@sarrigroupe.com	772568987	3	Edou	Sarr	Homme

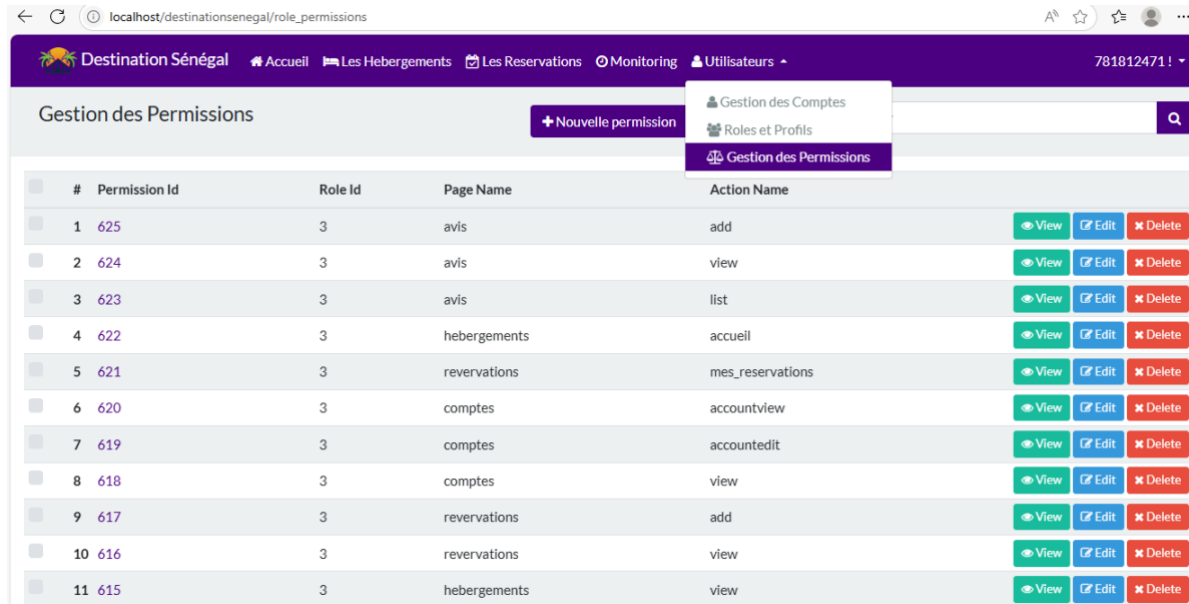
Figure 71. Gestion des comptes

- Gestion des rôles : il a la possibilité de créer, de modifier ou de supprimer un rôle mais aussi voir la liste des rôles.

#	Role Id	Role Name
1	3	clients
2	2	gestionnaire
3	1	Administrateur

Figure 72. Gestion des rôles

- Gestion des permissions : il a la responsabilité d'affecter à chaque utilisateur une permission en cliquant sur « **Nouvelle permission** » selon le rôle que chacun joue dans la plateforme. Il peut aussi créer, modifier ou supprimer une permission. Il peut aussi voir la liste des permissions qu'il a affectées.

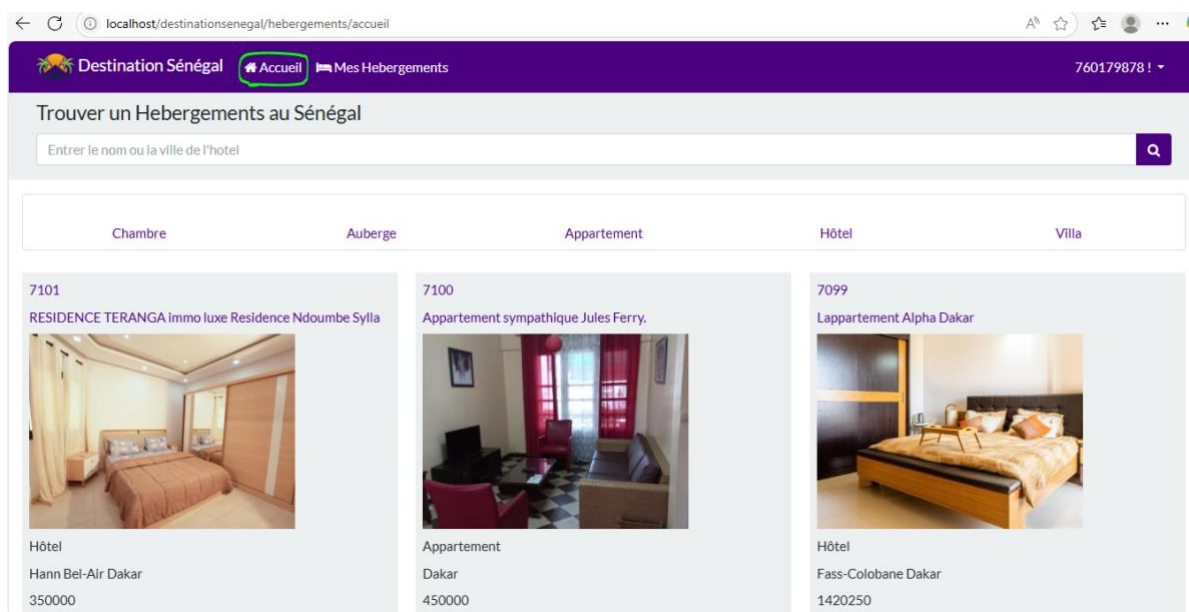


#	Permission Id	Role Id	Page Name	Action Name	
1	625	3	avis	add	View Edit Delete
2	624	3	avis	view	View Edit Delete
3	623	3	avis	list	View Edit Delete
4	622	3	hebergements	accueil	View Edit Delete
5	621	3	reervations	mes_reservations	View Edit Delete
6	620	3	comptes	accountview	View Edit Delete
7	619	3	comptes	accountedit	View Edit Delete
8	618	3	comptes	view	View Edit Delete
9	617	3	reervations	add	View Edit Delete
10	616	3	reervations	view	View Edit Delete
11	615	3	hebergements	view	View Edit Delete

Figure 73. Gestion des permissions

5.5.Espace Gérant

Dans cette espace, le gérant a deux onglets qui sont l'accueil et mes hébergements. Il peut accéder à différents éléments de la plateforme. Dans l'onglet Accueil, il peut rechercher un type d'hébergement en entrant le nom ou la ville de l'hôtel dans le champ « Trouver un hébergement au Sénégal ». A partir de là, il peut voir les informations de chaque hébergement.



Chambre	Auberge	Appartement	Hôtel	Villa
7101 RESIDENCE TERANGA Immo luxe Residence Ndoumbe Sylla  Hôtel Hann Bel-Air Dakar 350000	7100 Appartement sympathique Jules Ferry.  Appartement Dakar 450000	7099 L'appartement Alpha Dakar  Hôtel Fass-Colobane Dakar 1420250		

Figure 74. Page d'accueil gérant

Le gérant peut aussi effectuer les tâches suivantes :

- Gestion des hébergements : il a la possibilité de voir ou d'ajouter un type d'hébergement, de consulter, de modifier ou de supprimer un hébergement.

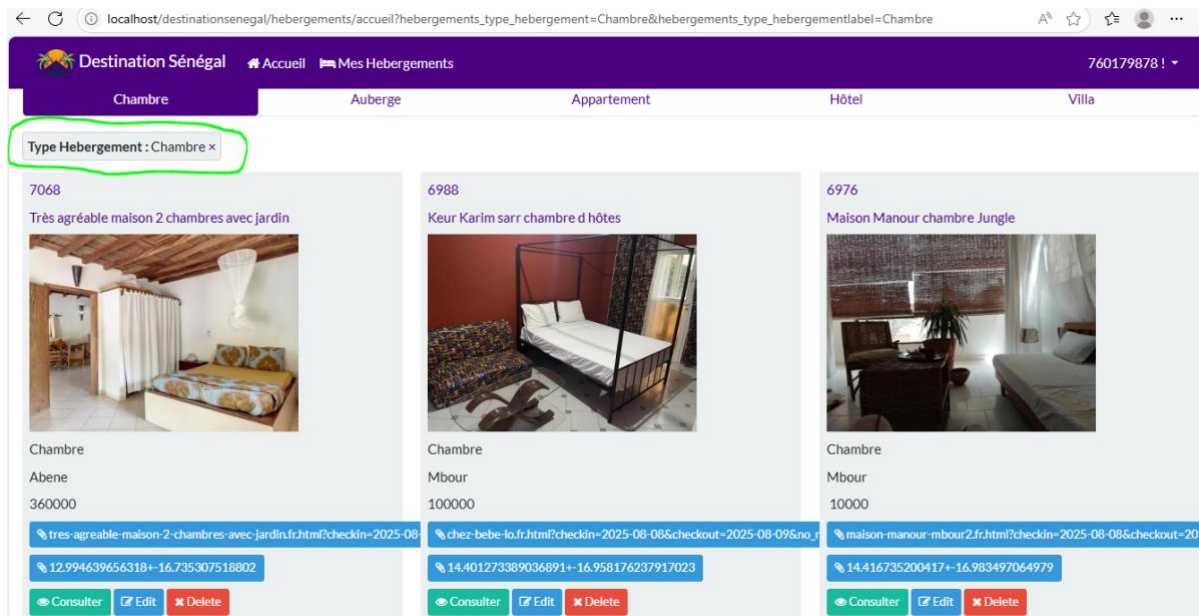


Figure 75. Gestion hébergements

Il peut aussi voir la liste de ses logements en cliquant sur mes hébergements.

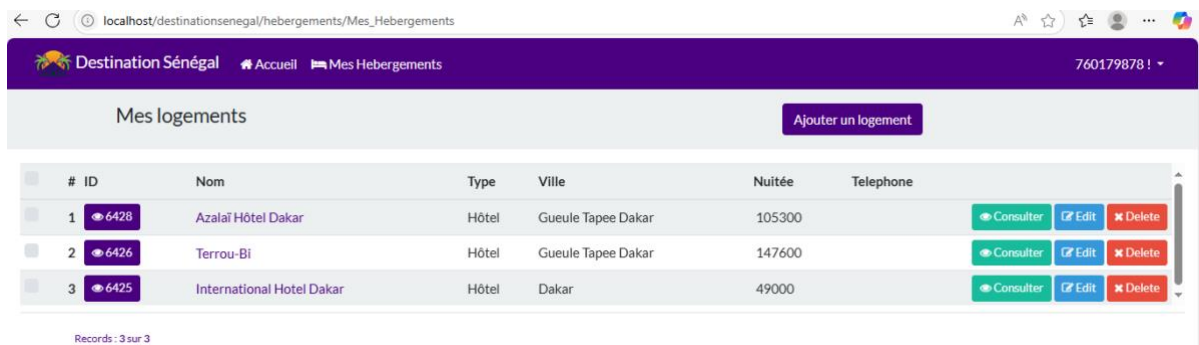


Figure 76. Liste des logements

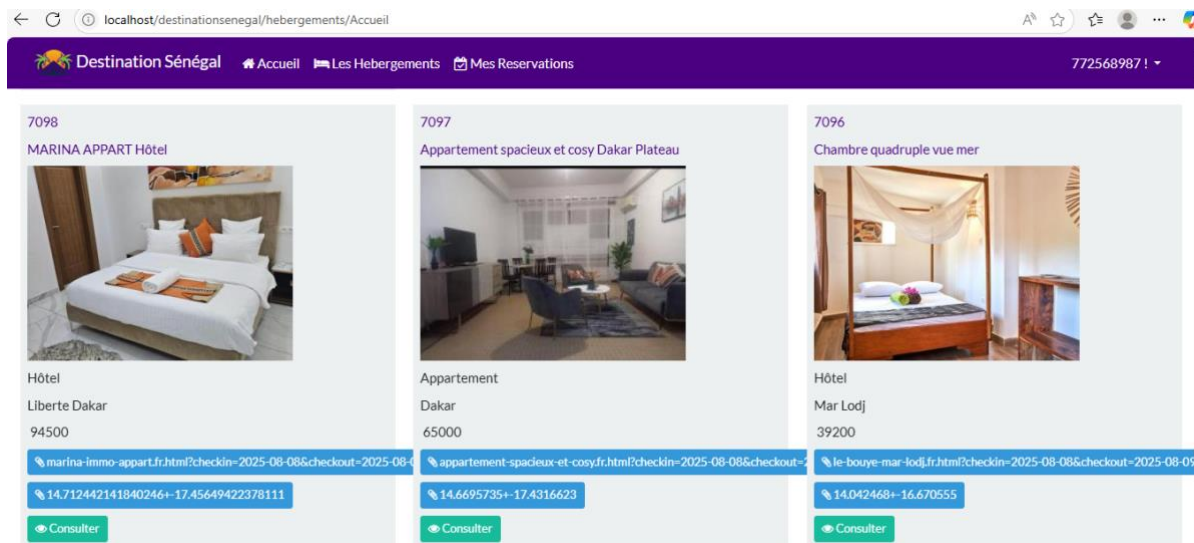
Il peut aussi ajouter un nouvel hébergement en cliquant sur « Ajouter un logement ». Ainsi un formulaire s'affichera où renseignera les champs et cliquer sur « Terminer ».

Figure 77. Ajout d'un logement

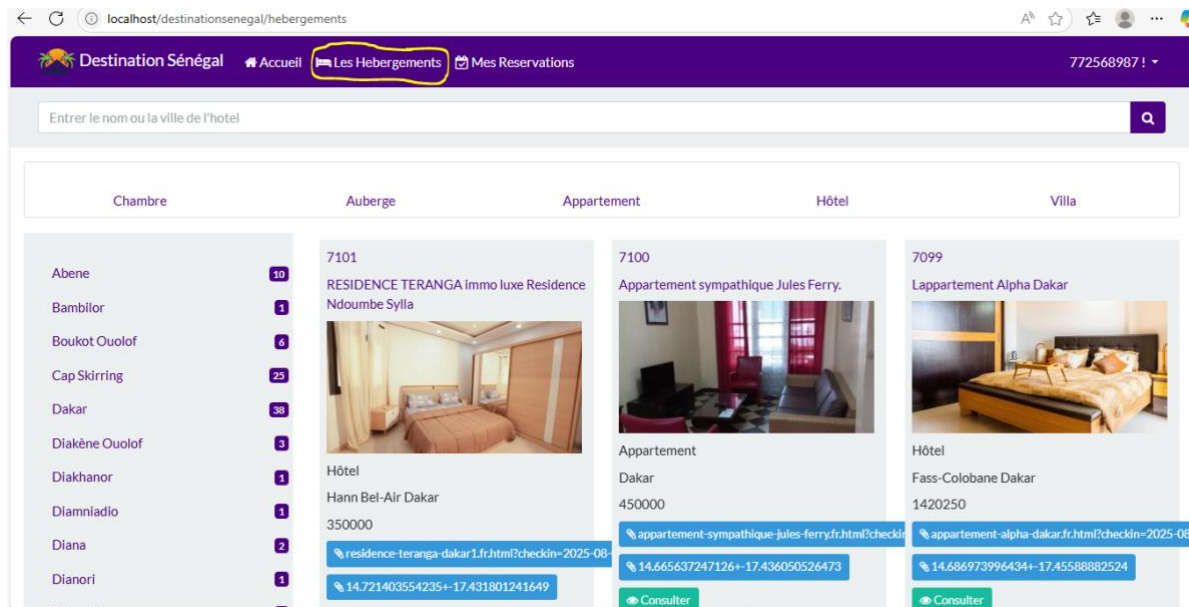
5.6.Espace Voyageur

Le voyageur dispose d'un espace qui lui est attribué, il est composé des onglets **Accueil**, **Les Hébergements** et **Mes Réservations**. Il a accès aux fonctionnalités suivantes :

Dans l'accueil, il peut juste consulter les informations de chaque type d'hébergement.

Figure 78. Page d'accueil voyageur

- Gestion des hébergements : il peut voir les différents types d'hébergements qui se trouvent dans la plateforme. En cliquant sur « les hébergements », il peut aussi sélectionner une ville dans la liste des villes se trouvant à gauche et voir les hébergements qui s'y trouvent.

Figure 79. Affichage des hébergements

Il peut faire une réservation en fournissant son prénom, nom, son numéro de téléphone, la date de départ et d'arrivée et son heure d'arrivée à l'hébergement et cliquer sur « **Réserver** ».

Figure 80. Page de Réservation

Il peut voir sa fiche de réservation avec les informations qu'il a fourni.

Id réservation:	7
Prenom:	Asta
Nom:	Seye
Téléphone:	782134562
Compte Operateur:	772568987
Date réservation:	2025-06-23 23:18:13
Date Arrivée:	2025-08-19
Date Départ:	2025-06-29
Code Logement:	7100
Nom du logement:	Appartement sympathique Jules Ferry.
Ville:	Dakar
Type Logement:	Appartement

Figure 81. Fiche de réservation

- Gestion des avis : Après son séjour, le voyageur peut donner un avis sur le champ « Recommandez-vous cet établissement ? » en cliquant sur Oui ou Non.

Code Logement:	7100
Nom du logement:	Appartement sympathique Jules Ferry.
Ville:	Dakar
Type Logement:	Appartement
Prix Nuitée:	450000
Heure Arrivée:	14:10:00
Etat réservation:	EMISE

Reservation * 7

Recommandez-vous cet établissement ? *

☒ Oui ☐ Non

TERMINER

Figure 82. Gestion des avis

Il peut aussi voir la liste contenant ses différentes réservations émises dans la plateforme en cliquant sur « **Mes Réservations** ».

#	ID	Date Réservation	Code Logement	Arrivée	Départ	Comptes	Etat	
1	7	2025-06-23 23:18:13	7100	2025-08-19	2025-06-29	772568987	EMISE	View
2	6	2025-06-23 22:45:04	7027	2025-07-05	2025-08-08	772568987	EMISE	View
3	5	2025-06-23 22:42:55	7100	2025-06-28	2025-07-04	772568987	EMISE	View

EXCEL Importer Records : 3 sur 3

Figure 83. Affichage des réservations émises

En résumé, ce chapitre a présenté en détail la solution développée, en décrivant les différentes interfaces utilisateur et leurs fonctionnalités spécifiques. De l'accès à la plateforme aux espaces dédiés à l'administrateur, au gérant et au voyageur, chaque composant a été conçu pour offrir une expérience fluide et adaptée aux besoins de chaque profil. Cette organisation modulaire garantit une gestion efficace et une bonne utilisation de la plateforme.

Conclusion et perspectives

Dans ce mémoire, nous avons étudié comment concevoir un comparateur d'hébergements hôteliers en utilisant des techniques de **crawling** et de **scraping web**, en nous concentrant sur les destinations touristiques du Sénégal. L'objectif était de créer une plateforme qui permet de récupérer automatiquement les informations sur les hôtels, appartements ou villas à partir de plusieurs sites web, afin d'aider les utilisateurs à comparer facilement les différentes offres disponibles. Dans un premier temps, nous avons analysé les besoins des voyageurs : ils cherchent souvent un logement qui correspond à leur budget, leur emplacement préféré et leurs critères de confort. Grâce au scraping web, notre système peut collecter ces informations automatiquement, les organiser et les afficher de façon claire, ce qui simplifie la recherche et la prise de décision. Ensuite, nous avons mis en place une plateforme de réservation d'hébergement capable de gérer les besoins de ces voyageurs. A cela s'ajoute la création d'une base de données bien structurée pour stocker toutes les données récupérées, ainsi qu'une interface qui affiche les résultats de manière claire. Cette étape a montré l'importance d'avoir une plateforme, capable de s'adapter aux mises à jour fréquentes des sites sources, tout en diminuant les doublons et les erreurs.

Enfin, nous avons remarqué les défis rencontrés, comme le respect des règles d'utilisation des sites web, la protection des données, ou encore la nécessité de faire évoluer la plateforme en fonction des changements sur internet. Pour que ce type de projet réussisse, il faut non seulement une bonne solution technique, mais aussi penser à l'utilisateur final et à la qualité des informations fournies. En effet, ce projet de mise en place de notre plateforme d'hébergements pour le Sénégal est une initiative utile pour valoriser le tourisme local et faciliter l'accès aux informations pour les voyageurs. Il montre qu'avec les bonnes méthodes et outils, il est possible de créer un service efficace, innovant mis au service du développement touristique.

L'amélioration d'une telle plateforme d'hébergements permettrait de maintenir un équilibre entre innovation technologique, respect de l'éthique numérique et satisfaction des besoins évolutifs des utilisateurs. Il faudra continuer à l'améliorer en tenant compte des besoins des utilisateurs, des évolutions technologiques et du respect des pratiques éthiques du web.

Références

1. Ouellet, A. (2024). Tourisme rural intégré en Basse-Casamance (Sénégal). Survie d'un projet de développement touristique communautaire en contexte de crise (s)? *Belgeo. Revue belge de géographie*, (1).
2. Bocoum, S. (2024). Tourisme rural intégré, facteur de résilience des territoires enclavés: étude du campement villageois dans la localité d'Affiniam (Basse Casamance/Sénégal). *Sciences Eaux & Territoires*, (46), 8222.
3. BALDE, C. O., GUEYE, T. N., & NDOYE, P. S. (2020). Tourisme et croissance économique inclusive au Sénégal. *Repères et Perspectives Economiques*, 4(2).
4. Schlechten, M. (1988). Tourisme balnéaire ou tourisme rural intégré? Deux modèles de développement sénégalais.
5. Niang, N. A. (2009). *Dynamique socio-environnementale et développement local des régions côtières du Sénégal: l'exemple de la pêche artisanale* (Doctoral dissertation, Université de Rouen).
6. Noblet, M. (2015). *L'adaptation au changement climatique en zone côtière au Canada et au Sénégal, une comparaison Nord-Sud* (Doctoral dissertation, Université de Picardie Jules Verne).
7. Karthikeyan, T., Sekaran, K., Ranjith, D. et Balajee, JM (2019). Extraction de contenu personnalisé et classification de texte à l'aide de techniques de grattage Web efficaces. *Journal international des portails Web (IJWP)* , 11 (2), 41-52.
8. Delalandre, M., & DI, A. P. (2016). Développement d'un web crawler pour la protection des mangas sous copyright.
9. **Desai, K., Devulapalli, V., Agrawal, S., & Patel, A.** (2017). *Review of Different Types of Web Crawler, Its Issues, Applications and Research Opportunities. International Journal of Advanced Research in Computer Science*, 8(3).
10. Diagne, O. (2023). Web Scraping et veille concurrentielle: généralité, état de l'art et étude de cas sur les boutiques en ligne. Visiter le 12/03/2024.
11. Hayet, S., & Souhila, Y. (2019). *Etude comparative de web scraping: Etude des outils scrapy et Apache-nutch* (Doctoral dissertation, Université Mouloud Mammeri). Visiter le 28/06/2024
12. Faheem, M., & Senellart, P. (2014). Crawl intelligent et adaptatif d'applications Web pour l'archivage du Web. *ISI*, 19(4).
13. Isaac, F., Hamon, T., Fouqueré, C., Bouchard, L., & Emirikian, L. (2001). Extraction informatique de données sur le web. *Revue DistanceS*, 5(2), 195-210.
14. Karthikeyan, T., Sekaran, K., Ranjith, D. et Balajee, JM (2019). Extraction de contenu personnalisé et classification de texte à l'aide de techniques de grattage Web efficaces. *Journal international des portails Web (IJWP)* , 11 (2), 41-52.

15. Brinkmann, M., & van der Merwe, B. (2021). *Web Crawling and Data Mining: Concepts and Techniques* (2nd ed.). Springer International Publishing. <https://doi.org/10.1007/978-3-030-75015-2> *(Chapitre 3: "Fundamentals of Web Crawlers", pp. 45-68) *
16. Aguida, A. A., & Mebarkia, M. (2021). *Conception et réalisation d'une application mobile pour la réservation d'hôtel* (Doctoral dissertation, Université Mohamed el-Bachir el-Ibrahimi Bordj Bou Arréridj Faculté de Mathématique et Informatique). Visiter le 18/04/2024
17. Carelli, J. (2020). Industrie hôtelière et avis en ligne : Etude des usages dans le but de comprendre les attentes des consommateurs à travers leurs parcours expérientiels.
18. Idrissi, F. C. AVIS EN LIGNE: UNE SOURCE DE DONNEES UTILE POUR L'ELABORATION «DU PARCOURS CLIENT» EN HOTELLERIE: LE CAS DES ETABLISSEMENTS D'HEBERGEMENT DE MARRAKECH. *Revue Ivoirienne des Arts, des Sciences de l'Information et du Patrimoine*, 167.
19. FRIEL, M. O., & KATIA, M. H. La transformation digitale dans le secteur touristique et de l'hôtellerie en Algérie « Cas plateformeNamlatic Hotel Booking ».
20. Ndiaye, M. (2020). *Digitalisation et tourisme au Sénégal : opportunités et défis*. Revue Économique Africaine.
21. Beauvisage, T., Beuscart, J. S., Cardon, V., Mellet, K., & Trespeuch, M. (2013). Notes et avis des consommateurs sur le web : Les marchés à l'épreuve de l'évaluation profane. *Réseaux*, (1), 131-161.
22. Tamba, K. (2021). *Comparateurs de prix et visibilité des PME hôtelières en Afrique de l'Ouest*. Revue Africaine du Tourisme et de l'Innovation.
23. Fasquelle, D., & Got, P. (2015). *Rapport d'information sur l'impact du numérique sur le secteur touristique français*. Assemblée nationale.
24. Petit, S. (2010). *Une analyse du tourisme international: fragmentation de la production, flux croisés, effets redistributifs* (Doctoral dissertation, Université de Lille 1–Sciences et Technologies).
25. PORNIN, Thomas, Comment choisir un langage de programmation, 2eme édition, Paris, édition H et K, 2005, 124 pages;
26. Muller, P. A. (2006). *De la modélisation objet des logiciels à la metamodélisation des langages informatiques* (Doctoral dissertation, Université Rennes 1).
27. Muller, P. A., & Gaertner, N. (2000). *Modélisation objet avec UML* (Vol. 514, p. 235). Paris: Eyrolles.
28. Roy, G. (2007). *Conception de bases de données avec UML*. PUQ.
29. Peltier, M. (2011, November). Développement d'applications Web avec le framework PHP Symfony. In *JRES (Journées réseaux de l'enseignement et de la recherche) 2011*.

30. Fuchs, P. F., & Poulain, P. (2024). *Introduction à la programmation Python pour la biologie* (Doctoral dissertation, Université Paris Cité).
31. Idrissi, F. C. AVIS EN LIGNE: UNE SOURCE DE DONNEES UTILE POUR L'ELABORATION «DU PARCOURS CLIENT» EN HOTELLERIE: LE CAS DES ETABLISSEMENTS D'HEBERGEMENT DE MARRAKECH. *Revue Ivoirienne des Arts, des Sciences de l'Information et du Patrimoine*, 167.
32. Marinval, P. (1986). Recherches expérimentales sur l'acquisition des données en Paleocarpologie. *ArchéoSciences, revue d'Archéométrie*, 10(1), 57-68.
33. Pouyllau, S. (2011). Les moteurs de recherche profitent aussi de la sémantique. *Documentaliste-Sciences de l'information*, 48(4), 36-37.
34. Zibani, K., & Medani, O. (2024). *Le Yield management: Outil de gestion dans l'hôtellerie (cas de l'hôtel: "THE BEST TITANIC")* (Doctoral dissertation, Université Mouloud Mammeri).
35. Zibani, K., & Medani, O. (2024). *Le Yield management: Outil de gestion dans l'hôtellerie (cas de l'hôtel: "THE BEST TITANIC")* (Doctoral dissertation, Université Mouloud Mammeri).
36. Pascal Roques. Les cahiers des programmeurs UML2 : Modéliser une application web 3ème édition. Edition Eyrolles 2007.
37. Goncalves, G., & Hémerly, F. (2000, May). Des cas d'utilisation en UML à la gestion de rôles dans un système d'information. In *INFORSID* (pp. 367-379).
38. Apvrille, L., de Saqui-Sannes, P., & Khendek, F. (2005, March). Synthèse d'une conception UML temps-réel à partir de diagrammes de séquences. In *Colloque Francophone sur l'ingénierie des protocoles (CFIP'05), Bordeaux, France* (Vol. 112).
39. Rigou, Y. (2022). *Une approche de découverte des diagrammes de classes UML par l'apprentissage profond* (Doctoral dissertation, Université du Québec à Rimouski).
40. Dehainsala, H., Pierra, G., Bellatreche, L., & Aït-Ameur, Y. (2007). Conception de bases de données à partir d'ontologies de domaine: Application aux bases de données du domaine technique. *Actes des 1ère Journées Francophones sur les Ontologies (JFO'07)*, 215-230.
41. Heurtel, O. (2008). *PHP et MySQL: maîtrisez le développement d'un site web dynamique et interactif*. Editions ENI.

Tables des matières

Dédicace	i
Remerciements	ii
Résumé	iii
Abstract.....	iv
Sommaire	v
Liste des figures	vi
Liste des tableaux	ix
Sigle et abréviations.....	x
Introduction générale	1
Chapitre 1 : Généralités	4
1.1. Tourisme au Sénégal.....	5
1.2. Les destinations du Sénégal.....	6
1.2.1. Destinations côtières	6
1.2.2. Les zones côtières du Sénégal	7
1.2.3. Les destinations non côtières du Sénégal	9
1.3. Le Web crawling.....	10
1.3.1. Définition.....	10
1.3.2. Fonctionnement de base	11
1.3.3. Types de Crawling	12
1.4. Le web scraping.....	17
1.4.1. Définition	17
1.4.2. Fonctionnement de base	18
1.4.3. Types de Scraping.....	18
1.5. Web Crawling Vs Web Scraping.....	22
Chapitre 2 : Analyse de l'existant et positionnement	24
2.1. Analyse de l'existant.....	25
2.2. Positionnement	37
Chapitre 3 : Cahier des charges	42
3.1. Contexte et justificatifs du projet.....	43
3.2. Mission du projet	44
3.3. Périmètre du projet	45
3.4. Cibles du projet.....	45
3.5. Besoins / Exigences	45

3.5.1.	Besoins fonctionnels	46
3.5.2.	Besoins technologiques	46
3.5.3.	Besoins en données	47
3.5.4.	Besoins non fonctionnels	48
3.6.	Contraintes techniques	49
Chapitre 4 : Conception et réalisation		50
4.1.	Conception	51
4.1.1.	Architecture générale	51
4.1.2.	Couche d'acquisitions de données	51
4.1.3.	Couche de presentation	62
4.1.4.	Conception du SI avec UML	63
4.1.4.1.	Diagramme User case	63
4.1.4.2.	Diagrammes de séquences	65
4.1.4.3.	Diagramme de classes	67
4.1.4.4.	Modèle physique de donnée (SQL)	69
4.2.	Réalisation	70
4.2.1.	Création de la base de données	70
4.2.2.	Réalisation du Back END ou acquisition des données	71
4.2.2.1.	Le crawler	71
4.2.2.2.	Le scraper	71
4.2.2.3.	Prétraitement des données	72
4.2.2.4.	Stockage des données	72
4.2.3.	Réalisation du moteur de recherche	72
4.2.3.1.	Installation du framework	72
4.2.3.2.	Création du projet	73
4.2.3.3.	Choix du thème	75
4.2.3.4.	Gestion des pages	76
4.2.3.5.	Gestion du design	77
4.2.3.6.	Gestion du menu	77
4.2.3.7.	Gestion des utilisateurs	78
4.2.3.8.	Gestion des rôles et privilèges	80
4.2.3.9.	Gestion du monitoring ou audits	80
4.2.3.10.	Déploiement de la plateforme	81
Chapitre 5 : Présentation de la solution		85

5.1.	Accès à la plateforme.....	86
5.2.	Page de connexion et déconnexion.....	86
5.3.	Page création de compte	87
5.4.	Espace Administrateur.....	87
5.5.	Espace Gérant	90
5.6.	Espace Voyageur	92
	Conclusion et perspectives	96
	Références.....	97
	Tables des matières.....	100
	Annexe 1 : Scraper Booking.....	103
	Annexe 2 : Crawler Booking	105

Annexe 1 : Scraper Booking

```
#!/usr/bin/env python3
# -*- coding: utf-8 -*-
"""
Created on Wed Jun 18 20:54:48 2025
@author: ufrses
"""
import requests
from bs4 import BeautifulSoup
import re
import csv
import os
import time
import concurrent.futures
import logging
import mysql.connector
import pandas as pd
logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - %(message)s')
SELECTORS = {
    'coordinates': {'id': 'map_trigger_header_pin', 'attr': 'data-atlas-latlng'},
    'description': {'name': 'p', 'attrs': {'data-testid': 'property-description'}},
    'prices': ['.fcab3ed991', '.prco-valign-middle-helper', '.bui-price-display__value']
}
from urllib.parse import urlparse, parse_qs, urlencode, urlunparse
from datetime import datetime, timedelta
def modifier_dates(url):
    q = parse_qs(urlparse(url).query)
    ci = datetime.strptime(q['checkin'][0], "%Y-%m-%d") + timedelta(days=30)
    co = datetime.strptime(q['checkout'][0], "%Y-%m-%d") + timedelta(days=34)
    q['checkin'], q['checkout'] = [ci.strftime("%Y-%m-%d")], [co.strftime("%Y-%m-%d")]
    return urlunparse(urlparse(url)._replace(query=urlencode(q, doseq=True)))
def extract_coordinates(soup):
    geo_element = soup.find('a', {'id': SELECTORS['coordinates']['id']})
    if geo_element and geo_element.has_attr(SELECTORS['coordinates']['attr']):
        latlng = geo_element[SELECTORS['coordinates']['attr']].split(',')
        if len(latlng) == 2:
            try:
                return {
                    'latitude': float(latlng[0]),
                    'longitude': float(latlng[1])
                }
            except ValueError:
                return None
    return None
def extract_description(soup):
    desc_element = soup.find(SELECTORS['description']['name'],
SELECTORS['description']['attrs'])
    if desc_element:
        return desc_element.get_text(separator=' ', strip=True).replace(", ", " ")
```

```
    return None
def extract_price(url):
    headers = {
        "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36
(KHTML, like Gecko)"
        " Chrome/114.0.0.0 Safari/537.36"
    }
    response = requests.get(url, headers=headers)
    soup = BeautifulSoup(response.text, 'html.parser')
    # Essai d'extraction d'un prix sur la page (exemple : élément avec classe contenant 'price')
    price = None
    # Exemple sélecteur générique
    price_candidates = soup.select('div.bui-price-display__value, span[data-testid="price-
and-discounted-price"]')
    if price_candidates:
        price = price_candidates[0].get_text(strip=True)

    return price
.....
```

Annexe 2 : Crawler Booking

```

Import requests
from bs4 import BeautifulSoup
from datetime import datetime, timedelta
import csv
import os
import pandas as pd
import time
import random
import re
import logging
from concurrent.futures import ThreadPoolExecutor, as_completed

# === Logging configuration ===
logging.basicConfig(
    filename='crawler.log',
    level=logging.INFO,
    format='%(asctime)s - %(levelname)s - %(message)s'
)

USER_AGENTS = [
    "Mozilla/5.0 (Windows NT 10.0; Win64; x64)...",
    "Mozilla/5.0 (Macintosh; Intel Mac OS X 13_2)...",
    "Mozilla/5.0 (X11; Linux x86_64)...",
]

def safe_find_text(parent, tag, attrs):
    el = parent.find(tag, attrs)
    return el.text.strip() if el else ""

def safe_find_attr(parent, tag, attrs, attr_name):
    el = parent.find(tag, attrs)
    return el[attr_name] if el and el.has_attr(attr_name) else ""

def clean_filename(name):
    return re.sub(r"[^\w]", "", name).lower()

def telecharge_image(url, nom):
    dossier = "output/topimages"
    os.makedirs(dossier, exist_ok=True)
    try:
        response = requests.get(url, timeout=10)
        response.raise_for_status()
        chemin = os.path.join(dossier, nom + ".jpg")
        with open(chemin, "wb") as f:
            f.write(response.content)
        logging.info(f"Téléchargement réussi de l'image : {nom}")
    except Exception as e:
        logging.warning(f"Échec du téléchargement de {url} : {e}")

```

```

def Get_URL(url, inclure_expression=False):
    pos = url.find("?")
    pos = pos if pos != -1 else len(url)
    fin = pos + 1 if inclure_expression else pos
    checkin = str(datetime.now().date() + timedelta(days=50))
    checkout = str(datetime.now().date() + timedelta(days=51))
    return url[:fin] + f"?checkin={checkin}&checkout={checkout}" + "&no_rooms=1"

def supprimer_doublons_par_nom(fichier_csv):
    lignes_uniques = {}
    try:
        with open(fichier_csv, newline="", encoding='utf-8') as f:
            reader = csv.reader(f)
            for ligne in reader:
                if len(ligne) < 2:
                    continue
                nom = ligne[1].strip()
                if nom not in lignes_uniques:
                    lignes_uniques[nom] = ligne

        with open(fichier_csv, 'w', newline="", encoding='utf-8') as f:
            writer = csv.writer(f)
            for ligne in lignes_uniques.values():
                writer.writerow(ligne)
        logging.info(f"Suppression des doublons terminée dans {fichier_csv}")
    except Exception as e:
        logging.error(f"Erreur lors de la suppression des doublons : {e}")

def supprimer_doublons_par_2e_colonne(chemin_fichier):
    """
    Supprime les doublons d'un fichier CSV en fonction de la 2e colonne
    et réécrit le fichier original avec les données nettoyées.

    :param chemin_fichier: Chemin vers le fichier CSV.
    """
    df = pd.read_csv(chemin_fichier)
    df_unique = df.drop_duplicates(subset=df.columns[1], keep='first')
    df_unique.to_csv(chemin_fichier, index=False)

def Spider(url, ville):
    headers = {"User-Agent": random.choice(USER_AGENTS)}
    try:
        response = requests.get(url, headers=headers, timeout=10)
        response.raise_for_status()
    except requests.RequestException as e:
        logging.error(f"Erreur HTTP sur {ville} : {e}")
        return
    soup = BeautifulSoup(response.content, 'html.parser')
    hotel_results = []
    for el in soup.find_all("div", {"data-testid": "property-card"}):
        nom = safe_find_text(el, "div", {"data-testid": "title"})

```

```

link_tag = el.find("a", {"data-testid": "title-link"})
lien = link_tag["href"] if link_tag and link_tag.has_attr("href") else ""
image_principale = safe_find_attr(el, "img", {"data-testid": "image"}, "src")

location = safe_find_text(el, "span", {"data-testid": "address"})
location = location.replace(", ", " ").replace(", ", " ").replace(" ", "")
nom = nom.replace(", ", " ").replace(", ", " ").replace(" ", "")

if not nom:
    continue

nom_image = clean_filename(nom)
telecharge_image(image_principale, nom_image)
lien = Get_URL(lien)

hotel_results.append({
    "ville": location,
    "nom": nom,
    "lien": lien,
    "image_principale": nom_image + ".jpg"
})

logging.info(f"Hôtel trouvé : {nom}")
print(nom)

file_path = "output/index.csv"
os.makedirs("output", exist_ok=True)
file_exists = os.path.isfile(file_path)

with open(file_path, mode='a', newline="", encoding='utf-8') as f:
    fieldnames = ["ville", "nom", "lien", "image_principale"]
    writer = csv.DictWriter(f, fieldnames=fieldnames)
    if not file_exists:
        writer.writeheader()
    for hotel in hotel_results:
        writer.writerow(hotel)

```

....