



REPUBLIQUE DU BENIN

MINISTRE DE L'ENSEIGNEMENT SUPERIEUR ET
DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE NATIONALE DES SCIENCES, TECHNOLOGIES,
INGENIERIE ET MATHEMATIQUES

INSTITUT NATIONAL SUPERIEUR DE TECHNOLOGIE INDUSTRIELLE
DE LOKOSSA

Filière : Génie Electrique et Informatique

Option : Informatique et Télécommunications

RAPPORT DES TRAVAUX DE FIN D'ETUDES POUR L'OBTENTION DU
DIPLOME DE LICENCE PROFESSIONNELLE

THEME :

REALISATION D'UNE PLATEFORME WEB
D'ARCHIVAGE ET DE MISE A DISPOSITION DES
MEMOIRES ET THESES MUNIE D'UN SYSTEME
DE DETECTION DE PLAGIAT

Rédigé et Soutenu par : Roméo Y. GODONOUKPE & Parménas N. ZINKO

Lieu de Stage : PROTECHNOLOGIES

Tuteur :

Mr KHAKHA Abdou R.
Directeur de PROTECHNOLOGIES

Superviseur :

Dr SAGBO Romaric
Enseignant à l'INSTI

Année Académique : 2020-2021

DEDICACES

Dédicace 1

Je dédie ce présent travail à :

- ✓ Mon père **Edmond S. GODONOUKPE**, qui m'a soutenu dans les moments aussi difficiles qu'agréables grâce aux conseils qu'il ne cesse de me prodiguer ;
- ✓ Ma tendre mère **Béatrice HODONOU**, qui a fait de moi l'homme que je suis et dont l'appui m'a été d'une très grande force.

Je vous suis reconnaissant.

Roméo Y. GODONOUKPE

Dédicace 2

Je dédie ce présent travail à :

A ma très chère mère **Edwige SEGNON** et à mon père **Pierre ZINKO**, qui m'ont toujours soutenu et accompagné dans mes projets à travers leurs sacrifices quotidiens et surtout pour avoir su m'inculquer une très bonne éducation dont je suis énormément fier aujourd'hui.

Je vous suis reconnaissant.

Parménas N. ZINKO

REMERCIEMENTS

Avant de débiter ce rapport, nous tenons à adresser nos sincères remerciements :

- ✓ A Dieu, tout puissant pour sa miséricorde à notre égard et pour l'intelligence dont il nous a fait grâce tout au long de ce travail ;
- ✓ A l'endroit de Monsieur le Directeur Alain ADOMOU, Maître de conférences des universités du CAMES, pour nous avoir permis de suivre une formation de qualité dans de très bonnes conditions ;
- ✓ A l'endroit de Madame la Directrice Adjointe Clotilde GUIDI, Maître de conférences des universités du CAMES, qui n'a ménagé aucun effort quant à la réussite de notre formation ;
- ✓ A l'endroit de Mr Abel M. KONNON, Maître-Assistant des universités du CAMES, notre chef de département pour ses conseils, son suivi et son soutien indéfectible durant toute notre formation ;
- ✓ A l'endroit de notre superviseur Dr SAGBO Romaric, Enseignant à l'INSTI pour ses nombreuses recommandations, ses apports et le suivi lors de la rédaction de ce travail ;
- ✓ A l'endroit de notre tuteur de stage, Mr KHAKHA Abdou R., Directeur de PROTECHNOLOGIES pour nous avoir apporté inspiration et orientation pour l'amélioration de notre travail ;
- ✓ A l'endroit de l'ensemble du personnel de PROTECHNOLOGIES, et aux autres stagiaires pour leurs apports et différentes suggestions concernant notre travail ;
- ✓ Au président, et aux membres du jury qui malgré leurs nombreuses occupations nous ont fait honneur en acceptant d'examiner et de juger notre travail.

SOMMAIRE

DEDICACES.....	i
REMERCIEMENTS	ii
SOMMAIRE	iii
LISTE DES ACRONYMES ET ABREVIATIONS.....	iv
LISTES DES TABLES ET FIGURES	v
CAHIER DES CHARGES.....	vi
RESUME.....	vii
ABSTRACT	vii
INTRODUCTION.....	1
PARTIE N° 1 : PRESENTATION DES STRUCTURES	3
Chapitre n°1 : Présentation de l'INSTI Lokossa	4
Chapitre n°2 : Présentation de la structure d'accueil.....	8
PARTIE N° 2 : DEROULEMENT DU STAGE	10
Chapitre n°3 : Déroulement du stage.....	11
PARTIE N°3 : TRAVAUX DE FIN D'ETUDES	14
Chapitre n°4 : Présentation du projet de fin d'études.....	15
Chapitre n°5 : Conception et réalisation de notre solution.....	32
CONCLUSION	54
BIBLIOGRAPHIE	55
WEBOGRAPHIE.....	56
ANNEXES	57
TABLES DES MATIERES	59

LISTE DES ACRONYMES ET ABREVIATIONS

API : Application Programming Interface

CAMES : Conseil Africain et Malgache pour l'Enseignement Supérieur

EF Core : Entity Framework Core

HTTP : HyperText Transfer Protocol

INSTI : Institut National Supérieur de Technologie Industrielle de Lokossa

MVC : Modèle Vue Contrôleur

MVT : Modèle Vue Template

NDLTD : NetWorked Digital Libray Of Theses and Dissertations

NLP : Natural Language Processing

NLTK : Natural Language Tool Kit

QR : Quick Response

SQL: Structured Query Language

TAL: Traitement Automatique des Langues

UML : Unified Modeling Language

LISTES DES TABLES ET FIGURES

LISTES DES FIGURES

Figure 1: Entrée principale de l'INSTI Lokossa.....	4
Figure 2: Situation géographique de l'INSTI Lokossa.....	5
Figure 3: Organigramme administratif de l'INSTI.....	7
Figure 4: Organigramme de PROTECHNOLOGIES	9
Figure 5 : Page d'accueil et de présentation d'une opération	13
Figure 6: Architecture client-serveur	20
Figure 7: Architecture d'un site web statique	21
Figure 8: Architecture d'un site web dynamique.....	21
Figure 9:Architecture MVT de Django :.....	26
Figure 10: Diagramme de cas d'utilisation du système	39
Figure 11: Diagramme de classe du système	41
Figure 12: Diagramme de séquence du cas « Effectuer un dépôt de mémoire».	42
Figure 13: Diagramme de séquence du cas « Valider un dépôt de mémoire »...	43
Figure 14: Interface Accueil 1.....	50
Figure 15: Interface de dépôt de mémoire	51
Figure 16: Tableau de Bord (pour les enseignants)	51
Figure 17: Interface 2 Tableau de Bord (pour les étudiants)	52
Figure 18: Interface 3 Tableau de Bord (pour les agents).....	52
Figure 19: Interface Administrateur.....	53
Figure 20: Tracés des fonctions $T(n)$ et $f(n)$	58

LISTE DES TABLEAUX

Tableau 1: Différents types de plagiat.....	23
Tableau 2: Différentes classes du SGBD de la plateforme	40
Tableau 3: Différentes classes de complexité temporelle	58

CAHIER DES CHARGES

❖ Objectifs de notre travail de fin d'études

L'objectif général de ce travail est de mettre en place une plateforme web fiable et opérationnelle d'archivage et de mise à disposition des mémoires et thèses munie d'un système de détection de plagiat. Ceci permettra de :

- ✓ Stocker et sauvegarder les thèses et mémoires ;
- ✓ Vérifier l'authenticité des thèses et mémoires par la détection de plagiat à la soumission des documents ;
- ✓ Assurer une mise à jour régulière de la base de données des mémoires ;
- ✓ Faciliter la gestion des tâches pendant la période des soutenances.

❖ Contraintes et moyens mis à disposition

Les contraintes qui pourraient entraver la réalisation de notre projet sont le temps dont nous disposons pour la mise en place de notre plateforme, le système de détection de plagiat et l'obtention de documentation pour certains modules.

Pour atteindre nos objectifs, nous possédons comme atouts nos connaissances théoriques et pratiques acquises lors de notre cursus, notamment en algorithmique et en développement web et les ressources disponibles sur internet.

❖ Le thème du Travail de Fin d'Etude (TFE)

Face à la nécessité d'accéder à la bonne information à partir de n'importe où et de la place de la recherche scientifique dans l'essor de notre nation, nous avons proposé une plateforme facilitant l'accès aux mémoires et thèses à tout usager. L'ensemble des travaux à effectuer nous a donc inspiré le choix du thème suivant :
« REALISATION D'UNE PLATEFORME WEB D'ARCHIVAGE ET DE MISE A DISPOSITION DES MEMOIRES ET THESES MUNIE D'UN SYSTEME DE DETECTION DE PLAGIAT ».

RESUME

Dans le cadre de l'obtention du diplôme de Licence Professionnelle en Génie Electrique et Informatique à l'Institut National Supérieur de Technologie Industrielle (INSTI) de Lokossa, nous avons effectué un stage à PROTECHNOLOGIES pendant une durée de trois (03) mois. Ce stage très enrichissant, nous a permis aussi bien d'approfondir les connaissances reçues au cours de notre formation mais aussi d'en acquérir de nouvelles. Le présent document fait objet de rapport et résumé au mieux le déroulement de ce stage d'une part et d'autre part notre travail de fin d'étude pour l'obtention du dit diplôme. Nous avons donc mis en place une plateforme web d'archivage et de mise à disposition des mémoires et thèses à laquelle nous avons appliqué un système de détection de plagiat pour limiter le phénomène dans nos universités. Ainsi, notre plateforme constituera pour tout utilisateur le meilleur panel à travers lequel il pourra consulter les différents mémoires soutenus par les étudiants. Nous avons également ajouté à notre plateforme une application mobile afin de simplifier au mieux les différentes tâches effectuées par les universités, et les enseignants pendant la période de soutenance.

ABSTRACT

In order to obtain Professional License degree in Electrical and Computer Engineering, we completed an internship at PROTECHNOLOGIES for three (03) months. During this internship period, we deepened our knowledge received during our training and acquire new ones. This document constitutes the report of the various tasks carried and also of the end of studies work for obtaining the said diploma. So, we have set up a web platform for archiving and making available dissertations and theses to which we have applied a plagiarism detection system to limit the harm in our universities. Thus, our platform will constitute for any user the best panel through which he can access to various dissertations supported by students. We have also added a mobile application to our web application in order to simplify as much as possible the different tasks carried out by universities and teachers during the defense period.

INTRODUCTION

Depuis des milliers de décennies et jusqu'à nos jours, de nombreuses recherches scientifiques continuent d'être effectuées au quotidien dans tous les domaines de la vie. Généralement, les résultats de la recherche scientifique servent à faire avancer nos connaissances théoriques ou pratiques et idéalement à résoudre des problèmes de la vie courante, pour une amélioration globale du bien-être de l'humanité. Ces recherches sont effectuées par des chercheurs universitaires, des spécialistes de la recherche ou des intervenants qui produisent chaque année des documents qui soutiennent les résultats obtenus. Par ailleurs, les travaux de fin d'études pour l'obtention des diplômes de Licence, de Master et du doctorat, constituent également une source bien riche de documents scientifiques. Au Bénin, en 2017 et 2018, il a été recensé respectivement treize mille sept cent dix-neuf (13719) et douze mille neuf cent quatre-vingt-neuf (12989) étudiants ayant reçu leur diplôme dans les universités publiques. (*source : Annuaire Statistique du Ministère de l'Enseignement Supérieur et de la Recherche Scientifique 2016-2017 et 2017-2018*). L'obtention de ces diplômes est sanctionnée pour la plupart par la rédaction d'un document qui met en exergue l'étude ou la réalisation d'une solution pour le bien commun. Malgré cette forte capacité que possède nos universités par le biais des étudiants (assez nombreux) à promouvoir la recherche scientifique, et la mise en place de solutions capables de résoudre des problèmes dans notre société, force est de constater que la plupart de ces résultats peinent à être mis en exploitation réelle afin d'atteindre leur objectif initial. La plupart du temps, les solutions proposées abondent les bibliothèques universitaires et demeurent figées dans ce cadre universitaire sans pour autant éclore.

Grâce à notre formation au cours de laquelle nous avons acquis des connaissances théoriques et pratiques, nous avons pensé pouvoir apporter notre contribution au développement des secteurs du numérique et de la recherche scientifique au Bénin

REALISATION D'UNE PLATEFORME WEB D'ARCHIVAGE ET DE MISE A DISPOSITION DES MEMOIRES ET THESES MUNIE D'UN SYSTEME DE DETECTION DE PLAGIAT

en général et dans les universités en particulier, en réalisant une plateforme d'archivage et de mise à disposition des mémoires et thèses munie d'un système de détection de plagiat pour une promotion efficace et durable des solutions développées par les étudiants, chercheurs et autres. Le présent rapport restitue au mieux l'essentiel des différentes activités menées au cours de notre stage de fin de formation pour la mise en place de notre solution et pour l'obtention du diplôme de Licence Professionnelle à l'Institut National Supérieur de Technologie Industrielle (INSTI) de Lokossa. Ce rapport s'articule autour des trois points essentiels suivants :

- ✓ La présentation des structures de formation et d'accueil
- ✓ Le déroulement du stage
- ✓ La présentation du projet de fin d'études intitulé : « **REALISATION D'UNE PLATEFORME WEB D'ARCHIVAGE ET DE MISE A DISPOSITION DES MEMOIRES ET THESES MUNIE D'UN SYSTEME DE DETECTION DE PLAGIAT** ».

PARTIE N° 1 : PRESENTATION DES STRUCTURES

Chapitre n°1 : Présentation de l'INSTI Lokossa

Inscrits depuis trois années académiques en tant qu'étudiant au sein de l'Institut National Supérieur de Technologie Industrielle (INSTI) de Lokossa, nous avons reçu durant cette période, une formation qualifiante capable de nous ouvrir à la vie professionnelle. Dans ce chapitre, nous présenterons donc ce cadre de formation.



Figure 1: Entrée principale de l'INSTI Lokossa

1-1 Situation géographique

L'Institut National Supérieur de Technologie Industrielle de Lokossa est situé au sud-ouest du Bénin, dans le département du Mono précisément dans la ville de Lokossa. Il est situé au quartier Agnivèdji, quittant la mairie de Lokossa, en bordure de la voie bitumée passant par l'ECOBANK, la SITEX et la SBEE sur une superficie d'environ quatre hectares cinq cent (4,5ha). L'emplacement de l'institut est illustré sur le schéma ci- dessous :

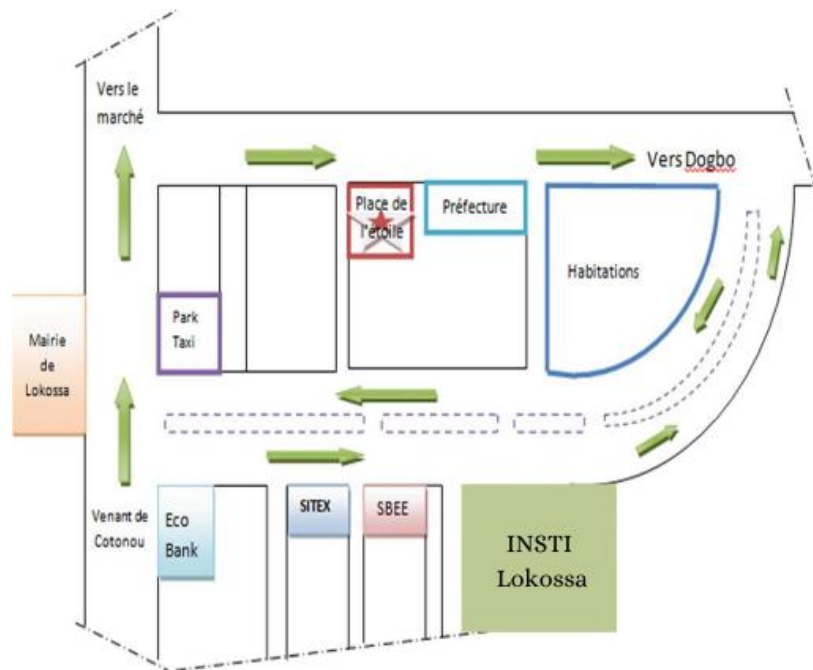


Figure 2: Situation géographique de l'INSTI Lokossa

(Source : Roméo GODONOUKPE & Parménas ZINKO)

1-2 Historique

Créé le 05 novembre 2001, l'Institut National Supérieur de Technologie Industrielle (INSTI) de Lokossa (ex Institut Universitaire de Technologie de Lokossa) est une entité de l'Université Nationale des Sciences, Technologies, Ingénierie et Mathématiques (UNSTIM) d'Abomey qui forme en coopération avec des universités accréditées par le CAMES, des techniciens en licence professionnelle sur une période de trois (03) ans.

1-3 Mission

C'est un établissement d'enseignement supérieur qui forme des techniciens supérieurs uniquement et qui s'occupe du perfectionnement technique et technologique de la formation continue des cadres dans les filières qu'il offre. On distingue dans l'institut cinq (05) différents départements :

- ✓ Génie Electrique et Informatique (GEI) avec ses options : Informatique et Télécommunications (IT), Electricité et Electrotechnique (EE) ;

- ✓ Génie Civil (GC) avec ses options : Bâtiments et Travaux Publics, Ouvrage en bois ;
- ✓ Maintenance des Systèmes (MS) avec ses options : Maintenance Industrielle (MI), Maintenance Automobile (MA) ;
- ✓ Génie Energétique (GE) avec ses options : Energétique, Energies Renouvelables ;
- ✓ Génie Mécanique et Productique (GMP)

L'INSTI de Lokossa, comme tout institut de technologie, a pour vocation de mener la recherche scientifique et technologique, de produire et publier des documents scientifiques tant de ses professeurs que de ses étudiants.

1-4 Administration de l'INSTI

A sa création en 2001, l'institut était dirigé par une administration ayant à sa tête M. ALLOBA Ezéchiel, Docteur Ingénieur en Génie Civil, spécialiste en route, Maître Assistant des Universités/CAMES, Maître-Assistant des universités du CAMES. Il était secondé dans sa tâche par deux adjoints à savoir : M. ADOMOU Alain, Docteur en Physique, Maître-Assistant des universités du CAMES et de feu M. WOUYA K. Emmanuel, Docteur Ingénieur en Résistance des Matériaux, Maître-Assistant des universités du CAMES.

A la rentrée 2009-2010, une nouvelle administration a été mise en place. Elle avait à sa tête M. MOUSSA Aliou Djibril, Docteur en Mathématiques et un adjoint en la personne de M. VISSOH M. Léandre, Docteur en production mécanique.

Ensuite à la rentrée 2012-2013, M. Aristide HOUNGAN, Maître-Assistant des universités du CAMES, prit la tête de l'administration dudit institut qui à son tour sera remplacé par M. ADOMOU Alain Maître de conférences des universités du CAMES avec une adjointe, Mme GUIDI Clotilde TOGNON, Maître de conférences des universités du CAMES.

REALISATION D'UNE PLATEFORME WEB D'ARCHIVAGE ET DE MISE A DISPOSITION DES MEMOIRES ET THESES MUNIE D'UN SYSTEME DE DETECTION DE PLAGIAT

En dehors de ces autorités, il y a au sein de ladite administration un comptable, un chef service chargé des affaires financières, des secrétaires et des chefs de département. Depuis sa création, l'Institut National Supérieur de Technologie Industrielle de Lokossa a toujours eu une administration bien organisée et dirigée par une équipe respectant la hiérarchie en vigueur dans l'administration béninoise. Cette organisation est donc représentée par l'organigramme suivant :

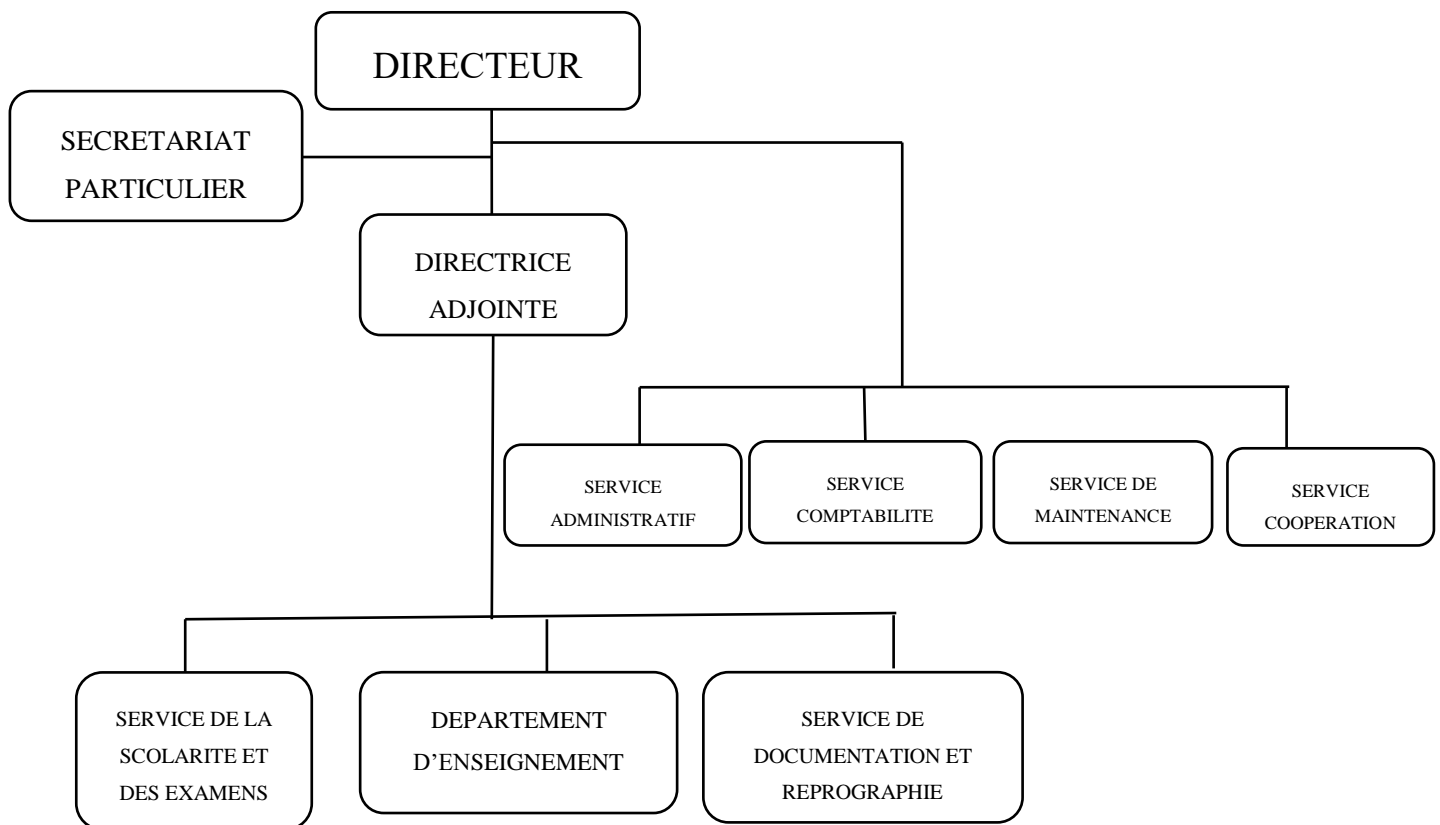


Figure 3: Organigramme administratif de l'INSTI

(Source : Administration de l'INSTI)

Chapitre n°2 : Présentation de la structure d'accueil

Dans le but d'atteindre les objectifs de nos travaux de fin d'études et d'améliorer notre expérience professionnelle, nous avons choisi pour notre stage, une structure dont les domaines d'interventions cadrent avec la conception et la mise en place de solutions web efficace et durable. Dans ce chapitre, nous présenterons l'essentiel sur notre cadre de stage qui est un cabinet.

1-1 Présentation du cabinet PROTECHNOLOGIES

PROTECHNOLOGIES est un cabinet d'ingénierie informatique spécialisé dans la réalisation et l'intégration de solutions digitales appliquées à la gestion, la biométrie et la fin-tech. Son siège est à Cadjehoun sur la route de l'aéroport en République du Bénin. Il offre des prestations de services dans les domaines de :

- ✓ L'ingénierie logicielle : développement de solutions desktop, web et mobile, intégration des solutions et gestion des projets
- ✓ La communication digitale
- ✓ L'audit informatique

Il propose également des modules de formations dans ses domaines d'interventions.

Pour atteindre ses objectifs, PROTECHNOLOGIES dispose de plusieurs partenaires avec lesquels il opère. Au nombre de ceux-ci nous avons :

- ✓ WAOOUH MONDE (cabinet partenaire sur le projet d'élaboration de la politique informatique de BUPDOS-ONG)
- ✓ NEWORLD TECHNOLOGY (cabinet co-fondateur de la solution de paiement en ligne CauriPay)
- ✓ AROBASECOM (cabinet partenaire sur le projet de refonte du portail web des marchés publics du Bénin pour le compte de l'ASSI – PRSCG)

- ✓ SOFT INNOVATION TECHNOLOGIES (cabinet partenaire sur plusieurs projets au profit de MTN Bénin, MTN Cameroun, MOOV Bénin, BOA Bénin, Société Générale Bénin)

1-2 Organigramme de l'administration de PROTECHNOLOGIES

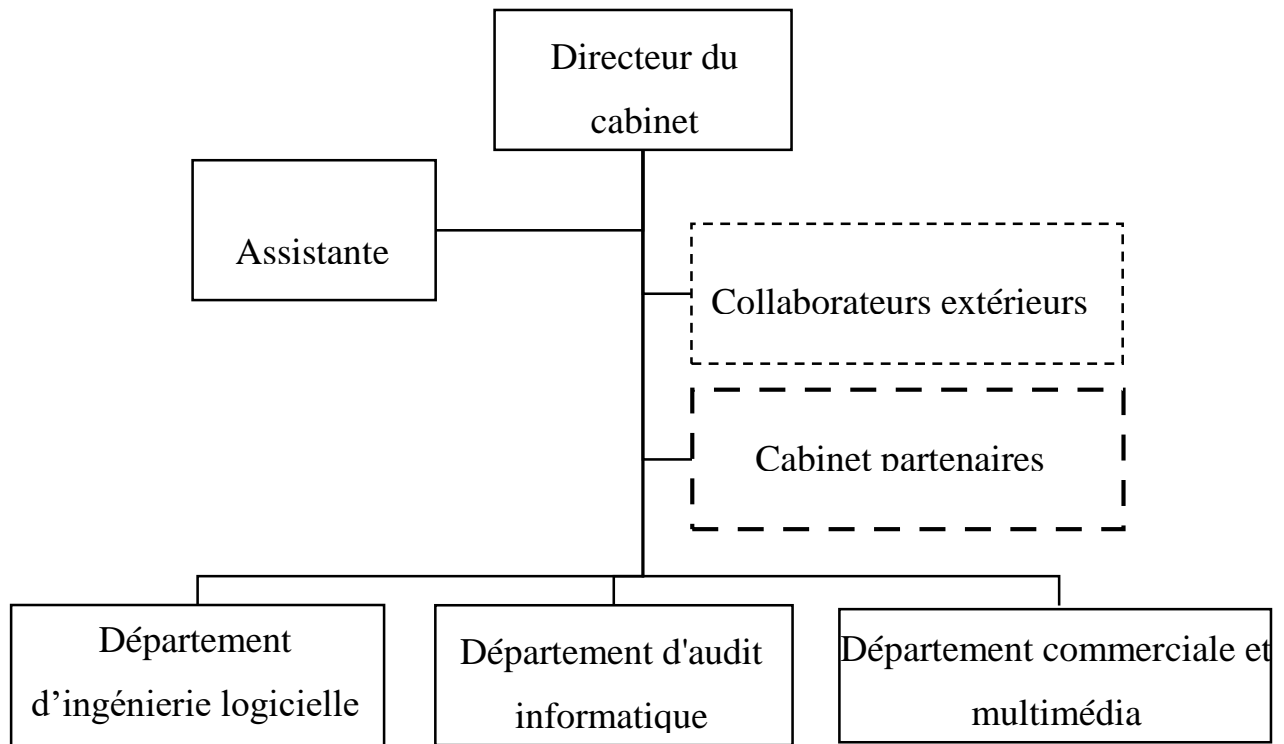


Figure 4: Organigramme de PROTECHNOLOGIES

(Source : Directeur de PROTEHNOLOGIES)

PARTIE N° 2 : DEROULEMENT DU STAGE

Chapitre n°3 : Déroulement du stage

Au cours de notre stage, nous avons mené plusieurs tâches qui se résument à : une initiation à la création d'application Web à partir du Framework ASP .Net Core et la conception d'une application mobile de gestion des dépenses pour des utilisateurs.

3-1 Initiation à la création d'application Web à partir du Framework ASP .Net Core

Au cours de notre stage, nous avons été appelé à prendre en main le Framework Web gratuit et open-source ASP .Net Core. Il a été développé par Microsoft et est basé sur le langage de programmation C#. C'est une nouvelle infrastructure multiplateforme à haute performance permettant de créer des applications pour tous les systèmes d'exploitation notamment Windows, Linux et Mac. Grâce à son modèle **MVC** (Modèle-Vue-Contrôleur), il facilite la création des API et des applications Web avec une interface utilisateur adaptée aux besoins. Il utilise :

- ✓ Razor Pages : modèle de programmation basé sur des pages pour la création d'une interface utilisateur Web plus simple et plus productive.
- ✓ Blazor : infrastructure permettant d'utiliser C# dans le navigateur avec JavaScript, de partager les logiques d'applications côté serveur et côté client écrit avec .Net.
- ✓ Un système de configuration : prêt pour le cloud et basé sur les environnements.
- ✓ Des Outils d'injections de dépendances intégrés.
- ✓ Possibilité d'hébergements simple.

Nos tâches se résument essentiellement à :

- ✓ Comprendre le Modèle MVC du Framework : création des contrôleurs pour la gestion des vues et des modèles, création des vues utilisant Razor pour l'intégration des informations envoyées par le contrôleur.
- ✓ Prendre en main la gestion des modèles avec Entity Framework Core (création de modèle, création d'un contexte de base de données pour coordonner les fonctionnalités EF Core que sont la création, la lecture, la mise à jour, la suppression d'un modèle).

3-2 Conception d'une application de gestion des dépenses quotidiennes

La gestion efficace de l'argent est d'une importante capitale pour tout un chacun. Il serait donc judicieux d'avoir une idée claire et détaillée de ses dépenses et revenus pour ainsi s'orienter vers des possibilités d'épargne et/ou d'investissement. Cette gestion quotidienne est malgré tout une tâche autant difficile au regard des nombreux détails qui échappent à notre attention. Dans ce cadre, nous avons été chargé de concevoir une application mobile de gestion des dépenses et revenus, à partir de laquelle chaque utilisateur pourra renseigner tout au long de sa journée, toutes les opérations au niveau de son portefeuille physique et ceci de manière très simple. L'objectif à travers une telle initiative est d'abord de donner plus de facilité à l'utilisateur pour enregistrer ses dépenses et revenus puis lui en donner une vue globale afin qu'il prenne de meilleures décisions à l'avenir.

A partir de cette application, tout utilisateur pourra :

- ✓ Ajouter une dépense (Coût de la dépense, la catégorie de la dépense, un commentaire servant de détail)
- ✓ Ajouter un revenu (Coût du revenu, la catégorie du revenu, un commentaire servant de détail)

REALISATION D'UNE PLATEFORME WEB D'ARCHIVAGE ET DE MISE A DISPOSITION DES MEMOIRES ET THESES MUNIE D'UN SYSTEME DE DETECTION DE PLAGIAT

- ✓ Définir un seuil journalier de dépenses pour éviter des dépenses inutiles (l'application enverra des notifications lorsque l'utilisateur atteint 50%, 75% et 100% de son seuil)
- ✓ Fournir une statistique mensuelle par catégorie
- ✓ Définir une heure de compte à laquelle l'application notifiera toutes les opérations de la journée
- ✓ Scinder les revenus des dépenses
- ✓ Définir des catégories personnalisées

Les interfaces conçues pour l'application sont les suivantes :

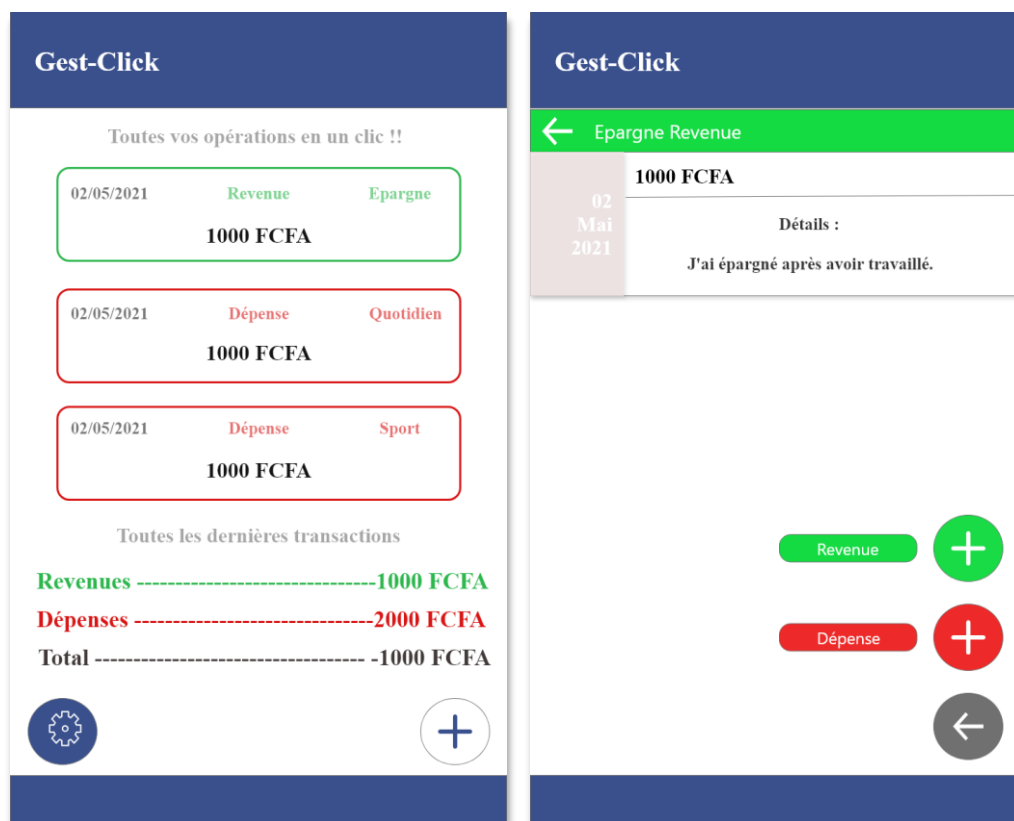


Figure 5 : Page d'accueil et de présentation d'une opération

(Source : Roméo GODONOUKPE & Parménas ZINKO)

PARTIE N°3 : TRAVAUX DE FIN D'ETUDES

Chapitre n°4 : Présentation du projet de fin d'études

INTRODUCTION

Le projet de fin d'étude représente l'objet principal de la production de ce document. Ceci justifie pleinement l'idée d'en faire une présentation pour mieux aborder le sujet. Ainsi, ce chapitre nous permettra d'énoncer les principales raisons qui motivent le choix de cette thématique, l'innovation qu'elle porte et les objectifs fixés tout au long de la mise en place de notre solution, puis évoquer enfin les outils, technologies et langages exploités pour la réalisation de notre projet.

4-1 Problématique

Aujourd'hui personne ne doute que le numérique prend une place de plus en plus illimitée dans notre vie quotidienne, offrant de nouvelles technologies pour limiter au mieux les contraintes de la nature. De par la rapidité et l'ampleur des transformations induites, ces technologies constituent de véritables atouts pour le développement des nations. L'information étant une denrée indispensable à l'essor de l'humanité, les TICs simplifient radicalement l'accès à celle-ci réduisant ainsi les distances entre les populations. Cette révolution numérique motive donc les acteurs du domaine à mettre en place des techniques permettant de numériser beaucoup plus d'informations que possible. Nous assistons donc à une transition quotidienne vers les bibliothèques et archives numériques constituées en général de textes (œuvres scientifiques, littéraires, documents, etc.), de vidéos, d'audios, d'images etc.

Mondialement, les bibliothèques numériques existaient depuis les années 1990, avec la première bibliothèque hybride (associant les deux types d'ouvrages soit physiques et numériques) apparue au Royaume-Uni et à des fins universitaires. Depuis, ce type de bibliothèque ne cesse de se propager offrant de plus en plus de

contenus numériques aux utilisateurs et aux acteurs des universités pour améliorer leurs connaissances. Dans ce cadre universitaire, l'UNESCO à travers une initiative mondiale NetWorked Digital Library Of Theses and Dissertations (NDLTD) est associée à un projet de création d'une Bibliothèque numérique universitaire au Brésil.

Au Bénin, les formations en Licence, Master et Doctorat sont généralement sanctionnées par la rédaction et la soutenance d'un rapport ou mémoire constituant une production scientifique riche et diversifiée. Des milliers de documents de mémoires sont compilés et archivés dans les bibliothèques des centres universitaires et/ou sauvegardés sur des disques en version numérique comme le cas à l'INSTI de Lokossa. En outre, certaines universités, écoles, institutions et sociétés privées proposent des services d'archivage et de diffusion des mémoires en ligne où l'accès au document est payant ou gratuit. Par exemple, **IFRI** (Institut de Formation et Recherche en Informatique) offre sur son site, les mémoires de Licence de chaque année depuis la promotion 2013-2014. En dehors de celui-ci, les plateformes INFRE ou MEMOIRE Online se positionnent. Parallèlement, le **Conseil Africain et Malgache pour l'Enseignement Secondaire (CAMES)** dispose d'une archive numérique institutionnelle nommée DICAMES qui contribue à la mise en valeur du Patrimoine Scientifique africain au service du bien commun. Ce projet trouve l'approbation du recteur Professeur Maxime da Cruz de l'Université d'Abomey-Calavi (UAC) qui à travers sa note de service en date du 16 Avril 2020 demandait à toute la communauté universitaire de l'UAC de mettre en ligne la version corrigée et définitive de leur mémoire de fin de Formation ou de leur thèse de doctorat grâce aux plateformes universitaires <https://bec.uac.bj> et <https://savoirs.cames.online/jspui/re> dans le cadre de la lutte contre le plagiat des Mémoires à l'UAC. Malgré la présence de ces solutions, le taux de disponibilité de bon de projets résolvant des problèmes pertinents dans notre société reste faible aux yeux de la communauté estudiantine,

des structures gouvernementales et des entreprises. La plupart du temps, motivés uniquement par l'obtention de leurs diplômes, les étudiants ne manifestent plus le désir de poursuivre les recherches pour l'amélioration de leurs solutions qui sont désormais rangées.

C'est dans l'optique d'améliorer l'utilisation habituelle et la visibilité des solutions proposées et pour décourager les copies conformes des recherches déjà effectuées dans de nouvelles recherches que nous avons décidé de mettre en place une plateforme numérique d'envergure nationale pour l'archivage, la mise à disposition des mémoires et thèses et d'y intégrer un système de détection de plagiat. Ceci permettra de rendre effective et régulière la disponibilité des travaux effectués chaque année, d'améliorer leur visibilité dans le but d'œuvrer pour leur mise en exploitation. Enfin, notre solution offrira des fonctionnalités afin de simplifier radicalement les tâches effectuées notamment par les universités et les enseignants au cours de la période des soutenances.

4-2 Innovation

En général, les plateformes existantes sur le marché béninois couvrent des mémoires qui datent de plusieurs années (INFRE ou MEMOIRE Online qui n'ont pas pu s'adapter à l'évolution technologique), ou couvrent une seule université (IFRI). De plus, les documents de mémoires sont pour la plupart indisponibles directement en téléchargement. Notre plateforme sera une plateforme complète, avec une étendue nationale (prenant en compte aussi bien les universités publiques que privées) et mettant à disposition de tout utilisateur les mémoires soit pour une simple lecture ou pour des téléchargements. Pour les étudiants, elle constituera l'outil parfait pour mettre en avant leurs travaux aux yeux de toute la communauté. Pour les enseignants, notre plateforme leur permettra d'être en temps réel au courant des différentes soutenances. Par ailleurs, elle permettra de digitaliser les différentes tâches effectuées pendant la période de soutenance

(avant et après les soutenances) pour les rendre plus simples et moins fastidieuses pour les universités. Par exemple, l'administration de chaque université pourra valider les dépôts de mémoire en scannant juste un code QR à partir d'une application mise à leur disposition. De plus, notre plateforme constituera un support de communication efficace pour permettre à l'administration de chaque université d'informer de façon automatisée chaque étudiant et chaque enseignant de la planification des soutenances. Enfin, le plagiat de documents étant un mal récurrent dans nos universités, notre système de détection de plagiat permettra d'identifier au mieux les potentiels cas.

4-3 Objectifs

4-3-1 Objectif général

L'objectif du travail consiste à développer un site web dynamique (plateforme) de dépôt, de mise à disposition et de gestion des tâches effectuées pour la soutenance des mémoires auquel sera appliqué un système de détection de plagiat.

4-3-2 Objectifs spécifiques

Pour atteindre l'objectif général, nous nous sommes fixés les objectifs spécifiques suivants :

- ✓ Implémenter les différentes interfaces pour l'interaction avec les utilisateurs (différentes pages de notre site telles que la page d'accueil, les tableaux de bord des étudiants et enseignants, la page de présentation de mémoires etc...) ;
- ✓ Implémenter le système d'authentification ;
- ✓ Implémenter le module de dépôt de mémoires ;
- ✓ Implémenter les différents modules de modification des mémoires par les étudiants (ajout d'une image de présentation, modification du document, ajout d'un résumé textuel), et par les enseignants (ajout d'une note, d'une mention) ;

- ✓ Implémenter le module d'importation de planification et d'envoi de mail aux enseignants et étudiants ;
- ✓ Implémenter les modules d'exportations de données ;
- ✓ Implémenter les différents modules nécessaires à la mise en place du système de détection de plagiat (construction du dictionnaire de mot, traitement du texte, algorithme de recherche de mots, calculateur de similitude et générateur du pourcentage de plagiat) ;
- ✓ Implémenter les interfaces de l'application mobile ;
- ✓ Implémenter les modules de scannage de code QR et de modification des données dans la base de données depuis l'application ;
- ✓ Intégrer les différents modules aux interfaces.

4-4 Méthodologie

La réalisation de notre projet comporte deux phases :

Phase d'analyse :

- ✓ Etude documentaire : collecte, traitement et stockage d'informations
- ✓ Spécifications : caractéristiques, fonctionnalités et services
- ✓ Conception générale : architecture, ergonomie
- ✓ Conception détaillée : interfaces graphiques, modélisation en UML

Phase de réalisation :

- ✓ Mise en place de l'architecture de la plateforme
- ✓ Codage : implémentation
- ✓ Tests unitaires et tests d'intégration
- ✓ Validation

4-5 Etude globale du système

4-5-1 Le Web

Le World Wide Web (littéralement la « toile mondiale », abrégé **www** ou le **Web**), est un système hypertexte public fonctionnant sur internet. Le Web permet de consulter, avec un navigateur, des pages accessibles sur des sites. Il n'est qu'une des applications d'Internet, distincte d'autres applications comme le courrier électronique, la messagerie instantanée et le partage de fichier pair à pair.

Généralement, pour un site web, l'architecture utilisée est une architecture à deux niveaux entre un client (le navigateur web) et un serveur. On parle d'**architecture client/serveur**. L'architecture s'appuie sur un poste central, qui envoie les données aux machines clientes :

1. L'utilisateur émet une requête HTTP vers le serveur sur lequel est stockée la page HTML
2. Le serveur accède alors au fichier demandé et le retourne au navigateur
3. Le navigateur interprète les balises HTML et affiche la page en résultat

L'architecture est décrite sur la figure suivante :



Figure 6: Architecture client-serveur

(Source : Internet)

Différents sites sur le Web

On distingue sur le Web, deux types de sites : les sites statiques et les sites dynamiques.

L'architecture d'un site web statique est décrite comme suit :

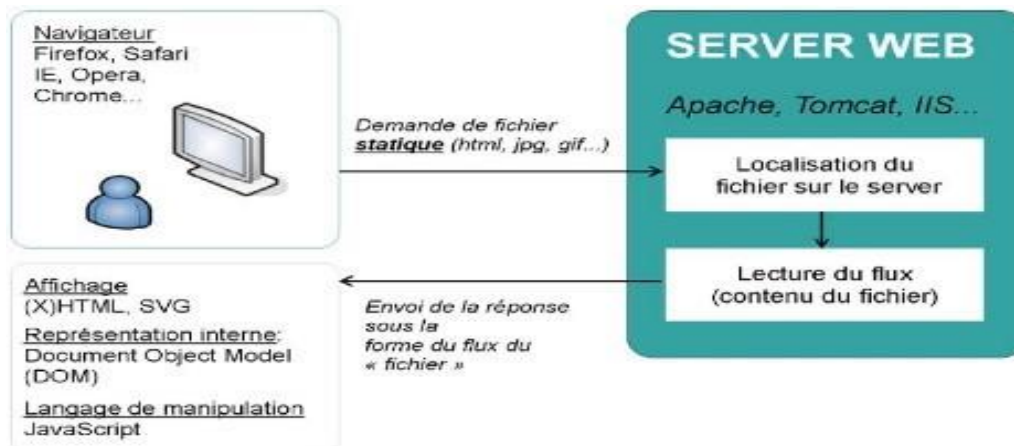


Figure 7: Architecture d'un site web statique

(Source : Internet)

Pour un site web dynamique, nous avons l'architecture ci-après :

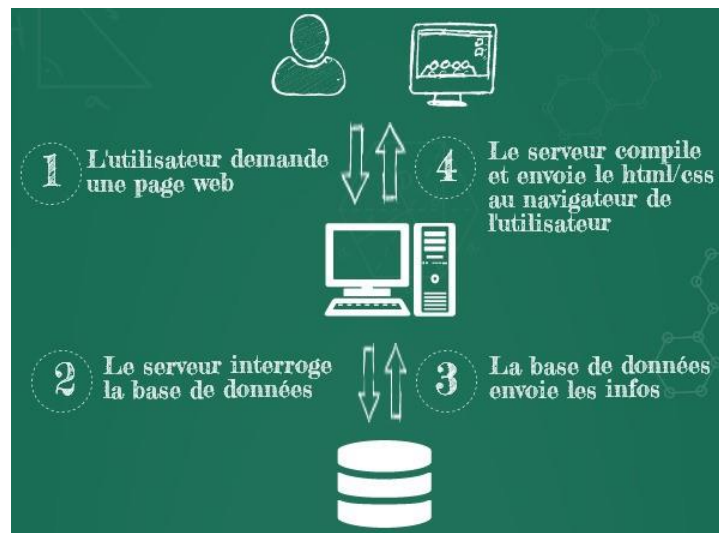


Figure 8: Architecture d'un site web dynamique

(Source : Internet)

La différence au niveau de ces deux types de sites se situe au niveau de la mise à jour de leurs pages. Une page web statique présentera le même contenu à n'importe quel utilisateur quel que soit les caractéristiques de la demande, alors qu'une page web dynamique sera générée à la demande et son contenu varie en fonction des caractéristiques de la demande (heure, formulaire rempli par le demandeur etc...) qui ne sont connues qu'au moment de la consultation.

En fonction des objectifs initiaux et des tâches que doivent effectuer les utilisateurs, notre plateforme sera donc un site web dynamique.

4-5-2 La détection de plagiat

4-5-2-1 Le plagiat

Le plagiat est une forme de contrefaçon consistant à s'approprier le travail d'un autre et le faire passer pour sien ; utiliser des extraits de textes, des images, des données provenant de sources externes sans en mentionner la provenance ou de reformuler ou résumer l'idée originale d'un auteur avec ses propres mots et ne pas mentionner l'auteur. Ce mal récurrent dans le domaine universitaire amène certaines universités à s'exprimer là-dessus. L'Université de Sorbonne stipule que « Le plagiat consiste à copier un modèle, que l'on omet délibérément de désigner. Dans le domaine littéraire, il consiste à recopier entièrement ou partiellement un ouvrage, dont on feint d'être intégralement l'auteur ». Bien que le plagiat soit pour la plupart identifié au « copier-coller », il ne se résume pas entièrement à cela. Il existe également plusieurs autres types de plagiat que nous allons présenter dans le tableau ci-contre.

Types de plagiat	Explication
Faire du copier-coller	C'est la forme la plus connue du plagiat. Le plagiat par copier-coller, également appelé plagiat direct,

	consiste à utiliser un texte provenant d'une autre source sans la citer.
Faire du plagiat en mosaïque	C'est du copier et coller d'un ensemble différents de morceaux de texte pour créer une sorte de « mosaïque » des idées d'autres chercheurs.
Faire de la paraphrase sans citer l'auteur	Consiste à traduire l'idée d'autrui avec ses propres mots, changer la grammaire ou des synonymes.
Faire de l'auto-plagiat	Consiste à utiliser les parties de ses propres travaux antérieurs (article, ensemble documentaire) sans les citer.
Traduire du texte	Consiste à utiliser un outil pour traduire un texte que l'on a copié et penser en être l'auteur.

Tableau 1: Différents types de plagiat

(Source : Internet)

Pour faire face à ces différents types de plagiat bien qu'il en existe qui sont vraiment difficiles à identifier, des techniques de détection sont mises en place. Ces techniques permettent d'identifier un pourcentage de similitude entre les documents.

4-5-2-2 La détection de plagiat

Selon les travaux d'Hermann Maurer, les méthodes de détection de plagiat peuvent être largement classées en trois catégories principales :

- ✓ La première catégorie : elle tente de détecter le style d'écriture et de trouver les changements incompatibles à ce style ;

- ✓ La deuxième catégorie (celle la plus utilisée) : elle se base sur deux ou plusieurs documents pour identifier les chevauchements des parties entre ces documents ;
- ✓ La troisième catégorie : elle reçoit en entrée un document puis recherche les motifs du plagiat sur le Web manuellement ou d'une manière automatisée

Ci-après, nous représentons une classification de quelques approches de détection actuelles utilisées pour identifier le plagiat.

A. Fingerprinting

C'est une méthode représentant le document sous forme d'un ensemble de sous-chaînes (n-gram). Le document suspect est vérifié pour le plagiat par le calcul de son empreinte et la recherche dans cet ensemble d'autres motifs (sous-chaînes) dont les empreintes ont été aussi calculées. Comme algorithme dans ce cas, nous avons l'algorithme de « Rabin-Karp » qui calcule les empreintes à partir d'une fonction de hachage. Il a l'avantage d'être capable de trouver dans un texte une sous-chaîne présente dans un ensemble de k chaînes.

B. String matching

« String matching » est l'approche la plus répandue en informatique. Lorsqu'elle est appliquée au problème de détection de plagiat, les documents sont comparés mot par mot. On trouve plusieurs algorithmes de « String matching » qui existent aujourd'hui, entre autres, les algorithmes « Brute Force » qui vérifie tous les caractères du texte avec le premier caractère du motif et lorsqu'on a une correspondance entre eux, on décale la comparaison entre le deuxième caractère du motif avec le caractère suivant du texte et ainsi de suite.

C. Bag of words

C'est une technique qui consiste à représenter un texte sous forme de vecteur qui est obtenu par fusion des vecteurs des différents mots contenu dans le texte. On recherche donc la similarité entre les documents en comparant les vecteurs de chaque document ; dans la plupart des cas en utilisant la théorie de la similarité par la comparaison de l'écart des angles entre chaque vecteur des documents

D. Citation analysis

C'est une technique d'analyse des citations et est la seule approche de détection de plagiat qui ne se base pas sur la similarité textuelle entre les documents. Elle examine les citations et les références dans les textes pour identifier les patterns similaires dans les séquences de citations. Cette approche est appropriée pour les textes scientifiques, académiques ou d'autres documents qui contiennent des citations. La détection du plagiat par l'analyse des citations est une récente méthode.

E. Stylometry

La stylométrie propose des méthodes statistiques pour quantifier et analyser le style d'écriture d'un auteur, cette méthode emploie la stylométrie pour construire des modèles quantitatifs de style pour les segments d'un texte. L'objectif est d'identifier les segments qui sont stylistiquement différents des autres segments, c'est un indicateur potentiel du plagiat.

4-6 Technologies, outils et langages utilisés

4-6-1 Technologies utilisées

✓ Le Framework Django

Django est un framework de haut niveau, permettant un développement rapide de sites internet, sécurisés et maintenables à partir du langage python. Il est gratuit, open source, a une communauté active et surtout une bonne documentation qui

nous a été très utile dans la mise en place de notre projet. Le framework Django est basé sur l'architecture MVT (Modèle Vue Template) s'inspirant de l'architecture MVC très répandue dans les frameworks web.

Le Modèle interagit avec la base de données et y recherche les items correspondants à une requête et renvoie une réponse facilement exploitable par le programme. La Vue quant à elle est une fonction Python acceptant une requête web et renvoyant une réponse web qui peut contenir le contenu HTML d'une page web, une redirection, une erreur 404 etc. La Vue elle-même contient la logique nécessaire pour renvoyer une réponse. Enfin, le Template est un fichier HTML qui peut recevoir des objets Python et qui est lié à une vue. Il permet également d'afficher des variables et d'utiliser des structures conditionnelles (if/else) ou encore des boucles (for) etc.

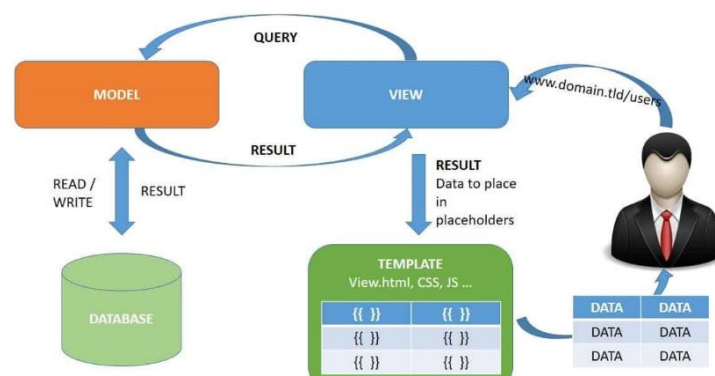


Figure 9:Architecture MVT de Django :

(Source : Internet)

Ce framework nous a servi à définir une architecture de codage plus simple et robuste pour atteindre rapidement nos objectifs.

✓ Framework jQuery

jQuery est une bibliothèque JavaScript libre, rapide, riche en fonctionnalités et multiplateforme créée pour faciliter l'écriture de scripts dans le code HTML des

pages web. Il facilite la manipulation de documents HTML, la gestion des événements, l'animation.

Ce framework nous a été utile pour la création d'animation et pour la présentation de certaines informations sur nos pages.

✓ **Les Frameworks Bootstrap et Uikit**

Bootstrap et Uikit sont des frameworks CSS open source (sous licence MIT) pour le développement web orienté interface graphique. Ils utilisent les langages HTML, CSS et JavaScript pour fournir des outils avec des styles déjà mis en place pour des boutons, des interfaces de navigations et bien d'autres. La particularité entre ces deux frameworks est que Bootstrap est plus efficace dans la mise en place d'un design responsive et Uikit est plus efficace pour la création des animations.

Ces frameworks nous ont donc été très utiles pour la réalisation d'interfaces graphiques simples et agréables aux yeux des utilisateurs.

✓ **MySQL**

MySQL pour My Structured Query Language est un système de gestion de bases de données relationnelles Open Source écrit en C et C++ Il est distribué sous double licence libre GPL (General Public License) et propriétaire et est multiplateforme. Utilisé pour toutes sortes d'applications, il est le plus souvent associé aux applications Web et à la publication de contenus en ligne. MySQL suit un modèle client-serveur avec un serveur intégré qui gère toutes les instructions à la base de données. Il prend également en charge des bases de données volumineuses et de nombreux types de données. L'une des spécificités de MySQL est de pouvoir gérer plusieurs moteurs au sein d'une seule base. Chaque table peut utiliser un moteur différent au sein d'une base. Ceci afin d'optimiser l'utilisation de chaque table.

Notre choix s'est porté sur ce SGBD en raison de ses performances et de la quantité de données à stocker et également grâce à sa facilité d'utilisation.

4-6-2 Outils utilisés

✓ Visual Studio Code

Visual Studio Code ou VS Code est un éditeur développé par Microsoft en 2015. Il est open source, gratuit et surtout disponible sur les systèmes d'exploitation Windows, Linux et Mac. VS Code est développé principalement pour le développement des projets avec JavaScript, Node.js. Toutefois, il peut s'adapter à d'autres types de langages grâce à un système d'extension bien fourni. La capacité de déboguer les applications directement sur l'éditeur sans même avoir recours aux navigateurs, la complétion automatique du code, l'Intégration Native avec Git confèrent un rôle important à Vs Code dans le développement d'application web. Tout ceci justifie pleinement notre choix porté sur cet éditeur.

✓ Android Studio

Android Studio est un environnement de développement pour développer des applications mobiles Android. Il est basé sur IntelliJ IDEA et utilise le moteur de production Gradle. Il peut être téléchargé sous les systèmes d'exploitation Windows, MacOS et Linux. Android Studio permet principalement d'éditer les fichiers Java/Kotlin et les fichiers de configuration XML d'une application Android. Il propose entre autres des outils pour gérer le développement d'applications multilingues et permet de visualiser rapidement la mise en page sur des écrans de résolutions variées simultanément. Il intègre par ailleurs un émulateur permettant de faire tourner un système Android virtuel sur un ordinateur.

Cet outil nous a servi à créer l'application Android de scannage de Code QR pour les tâches de validation de mémoire.

✓ **Diagrams.net**

Diagrams.net est un outil permettant de concevoir des diagrammes (diagrammes de processus, de réseaux, UML) et de les partager. Anciennement Draw.io, Diagrams.net peut être utilisé via un navigateur internet ou hors ligne via une application de bureau. Il propose de nombreux modèles de diagrammes et une large bibliothèque de formes diverses.

Cet outil nous a servi pour la création de nos différents diagrammes UML lors de la phase de modélisation de notre projet.

✓ **L'outil Git**

Git est un système de contrôle de version distribué gratuit et open source créé par Linux Torvalds auteur du noyau Linux et distribué selon les termes de la licence publique générale GNU version 2. Pour prévenir le risque que survienne un problème dans l'implémentation ou de perte de données, cet outil nous a permis d'effectuer des dépôts de notre projet en ligne. Comme Dépôt en ligne, nous avons utilisé GitHub.

✓ **Canva**

Lancé en 2013, Canva est une plate-forme de conception graphique qui permet aux utilisateurs de créer des graphiques, des présentations, des affiches, des documents et d'autres contenus visuels. Son utilisation est très simple et il propose des modèles pour créer des designs.

Cet outil nous a servi à créer ou modifier différentes images pour les utiliser sur nos interfaces graphiques.

4-6-3 Langages utilisés

✓ HTML

Le HyperText Markup Langage, généralement abrégé HTML est un langage de balisage conçu pour la mise en forme des pages web. Il permet de gérer la manière dont le contenu de ses pages va s'afficher sur un écran via le navigateur et également de titrer, sous-titrer, mettre en gras du texte et d'introduire des éléments interactifs comme des images, des liens, des vidéos etc. Enfin, il est plus facilement compris par les robots d'indexation des moteurs de recherche.

Ce langage nous a été incontournable pour la mise en place des différentes pages de notre plateforme.

✓ CSS

CSS (Cascading Style Sheets) pour Feuilles de style en cascade, est un langage informatique qui vient compléter le HTML. Il est utilisé pour gérer la mise en forme d'un site, le style du document HTML à partir des propriétés d'apparence (couleurs, bordures, polices etc.) et de placement (largeur, hauteur etc.)

Ce langage nous a été très utile car il a permis de donner une plus belle apparence aux différentes pages de notre application web.

✓ JavaScript

JavaScript est un langage de programmation qui permet d'implémenter des mécanismes complexes sur une page web. Il est principalement utilisé coté client c'est-à-dire dans le navigateur de l'internaute pour créer du contenu mis à jour de façon dynamique, des animations ou charger des éléments.

✓ Java

Java est un langage de programmation à usage général, évolué et orienté objet dont la syntaxe est proche du C. C'est un langage interprété et est notamment

utilisé pour le développement d'applications d'entreprises et mobiles qui peuvent fonctionner sur différents systèmes d'exploitation.

Ce langage a été utilisé pour le développement de notre application de scannage.

✓ PHP

PHP : HyperText Preprocessor est un langage de scripts multiplateforme et Open Source spécialement conçu pour le développement d'applications web. C'est un langage utilisé le plus souvent coté serveur c'est-à-dire que le serveur interprète le code PHP des pages demandées et génère du code (HTML, CSS par exemple) et des données (images par exemple) pouvant être interprétés et rendus par un navigateur Web. Il est souvent couplé à un serveur pour traiter des informations issues d'une base de données.

Nous nous sommes servis de ce langage pour faire modifier des informations sur des mémoires dans la base de données depuis notre application mobile.

✓ Python

Python est un langage de programmation interprété de haut niveau et open source conçue par Guido Van Rossum. Il est multiplateforme, orienté objet et est utilisé dans plusieurs domaines tels que le développement d'application web, mobile ou desktop. Python est plus répandu dans le monde scientifique et possède de nombreuses bibliothèques optimisées pour l'intelligence artificielle et la science des données.

Chapitre n°5 : Conception et réalisation de notre solution

5-1 Conception

L'étape de conception tient une place centrale dans la mise en place de notre projet. Elle nous permettra de mettre en place une architecture bien claire et détaillée de notre solution. Nous spécifierons à cette étape, les différentes entités logiques et les dépendances logiques entre ces entités (étape de modélisation).

5-1-1 Langage de Modélisation

Pour répondre aux besoins des différents utilisateurs de notre plateforme, il est important pour nous de présenter de manière compréhensible notre système à partir d'une représentation visuelle des objets, des états et des processus dans ce système. Ceci permettra de mieux nous statuer sur le résultat final et la façon dont notre système évoluera. C'est justement ce à quoi sert la modélisation qui consiste à créer une représentation virtuelle d'une réalité en vue d'une utilisation précise. Bien qu'il existe d'autres outils d'analyse tels que la méthode MERISE, notre choix s'est porté sur le langage UML pour ses nombreux avantages. UML, utilise en effet l'approche objet qui est très proche de la réalité permettant donc de mieux illustrer les interactions entre les éléments de notre système.

5-1-2 Modélisation UML

La modélisation UML permettra de définir de manière non ambiguë le fonctionnement futur du système, afin de faciliter sa réalisation. Dans le cadre de notre système, elle s'articulera autour des diagrammes qui suivent : le diagramme de cas d'utilisation, le diagramme de classe (vues statiques) et le diagramme de séquence (vues dynamiques).

5-1-2-1 Vues Statiques

Les vues statiques prennent en compte le diagramme de cas d'utilisation, le diagramme d'objets, le diagramme de classes, le diagramme de composants et le

diagramme de déploiement. Ici, nous ne présenterons que les diagrammes de cas d'utilisation et de classes.

5-1-2-1-1 Diagramme de cas d'utilisation

Les cas d'utilisation sont centrés sur les besoins des utilisateurs et permettent d'y voir clair et mieux les exprimer.

Ici, nous procéderons à :

- ✓ L'identification des acteurs
- ✓ L'identification des cas d'utilisation
- ✓ L'élaboration du diagramme de cas d'utilisation

Identification des acteurs

Un acteur représente un rôle joué par une personne ou une chose qui interagit avec le système (la même personne physique peut donc être représentée par plusieurs acteurs en fonction des rôles qu'elle joue). De plus, un acteur n'est pas nécessairement une personne physique : il peut être un service, une société, un système informatique... Il existe principalement quatre (04) catégories d'acteurs : les acteurs principaux, les acteurs secondaires, les matériels externes et les autres systèmes avec lesquels notre système doit interagir.

Comme acteurs principaux nous avons :

- ✓ Le visiteur : c'est n'importe quel utilisateur qui vient consulter les mémoires mis à disposition sur notre plateforme. Il peut donc rechercher, lire ou télécharger le document d'un mémoire.
- ✓ L'étudiant : c'est un utilisateur du système qui sera appelé à effectuer des dépôts de mémoire après avoir créé un compte, consulter et modifier des informations de ses mémoires, vérifier l'évolution de ses mémoires.

- ✓ L'enseignant : c'est également un utilisateur du système appelé à s'authentifier et consulter les mémoires qui lui sont assignés et à effectuer des modifications nécessaires pour leur validation.

Comme acteur secondaire nous avons :

- ✓ Le secrétariat administratif : Il s'occupe de la gestion des mémoires au niveau de chaque université. Il pourra effectuer les actions nécessaires pour la mise à disposition des mémoires de son établissement sur notre plateforme.

Dans la suite de notre développement nous évoquerons le secrétariat administratif sous le nom d'**Agent**.

- ✓ Administrateur du système : Il s'occupe de la gestion des tables de base, des utilisateurs, des privilèges et de la maintenance sur système.

Comme autre système nous avons :

- ✓ Le système de détection de plagiat : intervient pour identifier la présence de plagiat entre deux documents (le document de mémoire d'un étudiant et des documents disponibles dans notre base des données et qui sont en rapport avec le thème).
- ✓ L'application mobile : Elle servira à scanner le code QR attribué à chaque mémoire pour la validation.

Identification des cas d'utilisation :

Les cas d'utilisation permettent de structurer les besoins des utilisateurs et les objectifs correspondants d'un système. Ils sont un moyen de lister les fonctions que le système doit fournir. Ces fonctions couvrent la période des soutenances allant du dépôt de mémoire (avant soutenance) jusqu'à la validation finale du mémoire (après soutenance).

Selon les différentes tâches assignées à chaque acteur du système nous pourrions définir les cas d'utilisation suivants :

✓ Le visiteur :

- Consulter un mémoire (visiter la page des mémoires et/ou visiter la page d'un mémoire précis)
- Rechercher un mémoire
- Télécharger un mémoire

✓ L'étudiant :

- S'authentifier
- Effectuer un dépôt de mémoire
- Suivre l'évolution d'un mémoire
- Modifier un mémoire
- Supprimer un mémoire
- Modifier son profil

✓ L'enseignant :

- S'authentifier
- Consulter les mémoires assignés
- Ajouter un commentaire à un mémoire
- Attribuer une note à un mémoire après soutenance
- Attribuer une mention à un mémoire après soutenance
- Modifier son profil

✓ L'agent :

- S'authentifier
- Exporter un rapport
- Importer une planification
- Valider un dépôt de mémoire
- Effectuer la validation finale d'un mémoire

- Envoyer une notification de soutenance
- Supprimer un mémoire
- ✓ L'administrateur :
 - S'authentifier
 - Gérer les comptes utilisateurs
 - Maintenir le site
- ✓ Le système de détection de plagiat :
 - Identifier un cas de plagiat

Elaboration du diagramme des cas d'utilisation

Description des cas d'utilisation majeurs de la plateforme.

Cas d'utilisation : Effectuer un dépôt de mémoire

- ✓ Acteur principal : Etudiant ;
- ✓ Autre système : Système de détection de plagiat
- ✓ Objectifs : L'étudiant doit pouvoir accéder au formulaire, renseigner les informations indispensables pour effectuer son dépôt de mémoire.
- ✓ Préconditions : L'étudiant s'est authentifié ;
- ✓ Post conditions : Un mémoire a été enregistré au nom de l'étudiant et affiché sur son tableau de bord ;
- ✓ Scénario nominal
 1. L'étudiant saisie les informations nécessaires au dépôt du mémoire
 - ❖ Les informations relatives à sa formation (Université, écoles, Filière, option)
 - ❖ Les informations relatives au mémoire (thème, superviseur, document)
 - ❖ Les informations relatives au binôme (nom d'utilisateur et mot de passe s'il s'agit d'un couple d'étudiants)

2. Le système vérifie que toutes les informations renseignées sont correctes
3. Le système valide les informations
4. Le système identifie le(s) correspondance(s) avec les mémoires disponibles dans la base de données
5. Le système démarre le test de détection de plagiat
6. Le système n'identifie pas de plagiat, génère un code QR spécifique au mémoire et l'enregistre
7. Le système redirige l'utilisateur vers son tableau de bord et y affiche quelques informations concernant le mémoire enregistré.

✓ Alternatives

3a. Le système identifie une erreur dans les informations renseignées (extension de document non conforme à celle exigées, informations incorrectes pour le binôme dans le cas d'un couple d'étudiants par exemple)

❖ Le système revient au formulaire en renvoyant une erreur

4a. Le système n'identifie aucune correspondance de thématique avec les mémoires déjà stockés.

❖ Le système n'effectue aucun test de plagiat et enregistre le mémoire

6a. Le système identifie un plagiat entre le mémoire voulant être mis en ligne et un mémoire existant

❖ Le cas d'utilisation se termine en erreur et un rapport du test est affiché à l'étudiant

Nous pouvons également ajouter à ce cas d'utilisation une brève description du cas d'utilisation lié à la validation d'un dépôt de mémoire.

Cas d'utilisation : **Valider un dépôt de mémoire**

- ✓ Acteur principal : Etudiant ;
- ✓ Acteur secondaire : Agent ;
- ✓ Autre système : Application Mobile
- ✓ Objectifs : L'étudiant doit pouvoir valider le dépôt de son mémoire auprès de l'administration par un dépôt physique du dit mémoire ;
- ✓ Préconditions : L'étudiant a effectué son dépôt de mémoire en ligne et l'agent et lui sont authentifiés au moment de la validation.
- ✓ Post conditions : Le statut du mémoire avant la soutenance est modifié et le mémoire peut être inséré dans la planification des soutenances
- ✓ Scénario nominal
 1. L'étudiant génère le code QR de son mémoire à partir de son tableau de bord ;
 2. L'agent au secrétariat ouvre son scanner depuis son application mobile ;
 3. L'agent scanne le code QR généré ;
 4. Le système vérifie que les informations inscrites sur le code sont valides ;
 5. Le système valide les informations et modifie le statut avant la soutenance du mémoire.
- ✓ Alternatives
 - 1a. Le système n'arrive pas à générer le code QR
 - ❖ Le cas d'utilisation se termine en erreur
 - 4a. Le système n'arrive pas à valider les informations inscrites sur le code
 - ❖ Le cas d'utilisation se termine en erreur
 - 5a. Le système n'arrive pas à modifier le statut du mémoire

❖ Le cas d'utilisation se termine en erreur

Le diagramme de cas d'utilisation se présente comme suit :

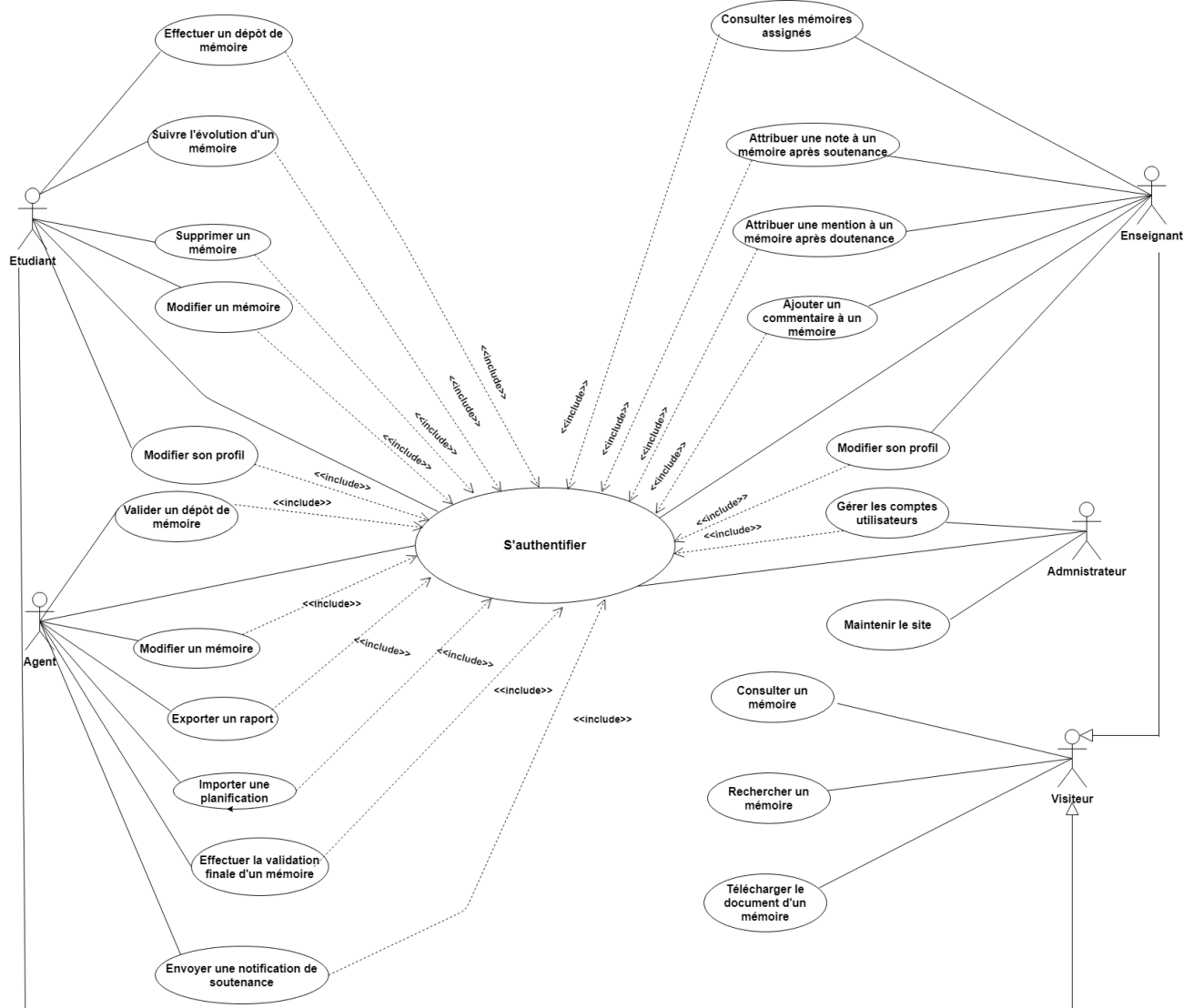


Figure 10: Diagramme de cas d'utilisation du système

(Source : Roméo GODONOUKPE & Parménas ZINKO)

5-1-2-1-2 Diagramme des classes

Le diagramme de classes exprime la structure statique du système en termes de classes et de relations entre ces classes. L'intérêt du diagramme de classe est de modéliser les entités du système. Il s'agira ici donc d'identifier les différentes classes, leurs attributs, leurs méthodes et les relations entre elles.

Identification des classes : une classe est une description abstraite (condensée) d'un ensemble d'objets ayant une sémantique et des caractéristiques communes: elle définit leur structure, leur comportement et leurs relations.

Le tableau ci-après regroupe les différentes classes caractérisant notre base de données :

Classes	Attributs
Universite	id_universite, nom
Entite	id_entite, nom, universite
Filiere	id_filiere, nom, entite
Option	id_option, nom, filiere
Annee_academique	id_annee, intitule
Utilisateur	id_utilisateur, nom_utilisateur, email, mot_de_passe, date_inscription
Enseignant	id_enseignant, nom, prenom, avatar, code_enseignant
Memoire	id_memoire, theme, entite, option, annee, code_memoire, superviseur, moyenne, mention, active, depot_valid, finale_valid, date_depot, resume, image, document, qr_code
Etudiant	id_etudiant, nom, prenom, avatar
Agent	id_agent, ecole
Commentaire	id_commentaire, titre, contenu, memoire, enseignant, date_comment
Soutenance	id_soutenance, memoire, date_soutenance, numero_jury, salle

Tableau 2: Différentes classes du SGBD de la plateforme

(Source : Roméo GODONOUKPE & Parménas ZINKO)

Elaboration du diagramme des classes : Le diagramme des classes de notre système se présente comme suit :

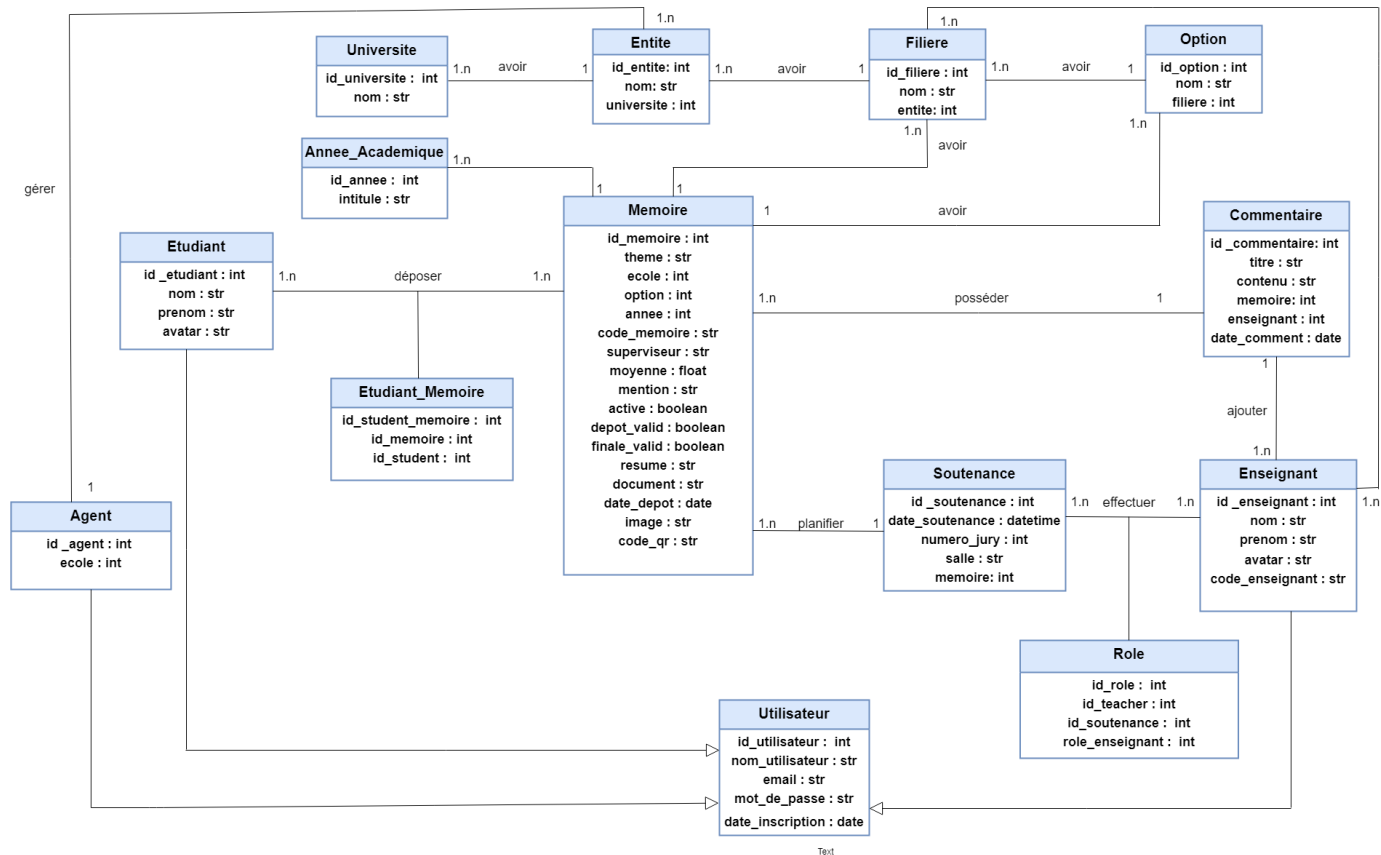


Figure 11: Diagramme de classe du système

(Source : Roméo GODONOUKPE & Parménas ZINKO)

5-1-2-2 Vues dynamiques

Les vues dynamiques permettent de modéliser le comportement dynamique du système et d'indiquer comment ses objets interagissent au moment de l'exécution. Ils prennent en compte les diagrammes de collaboration, de séquence, d'états-transitions et d'activités. Dans le cadre de notre étude, nous ne présenterons que le diagramme de séquence.

Diagramme de séquence du cas « Effectuer un dépôt de mémoire »

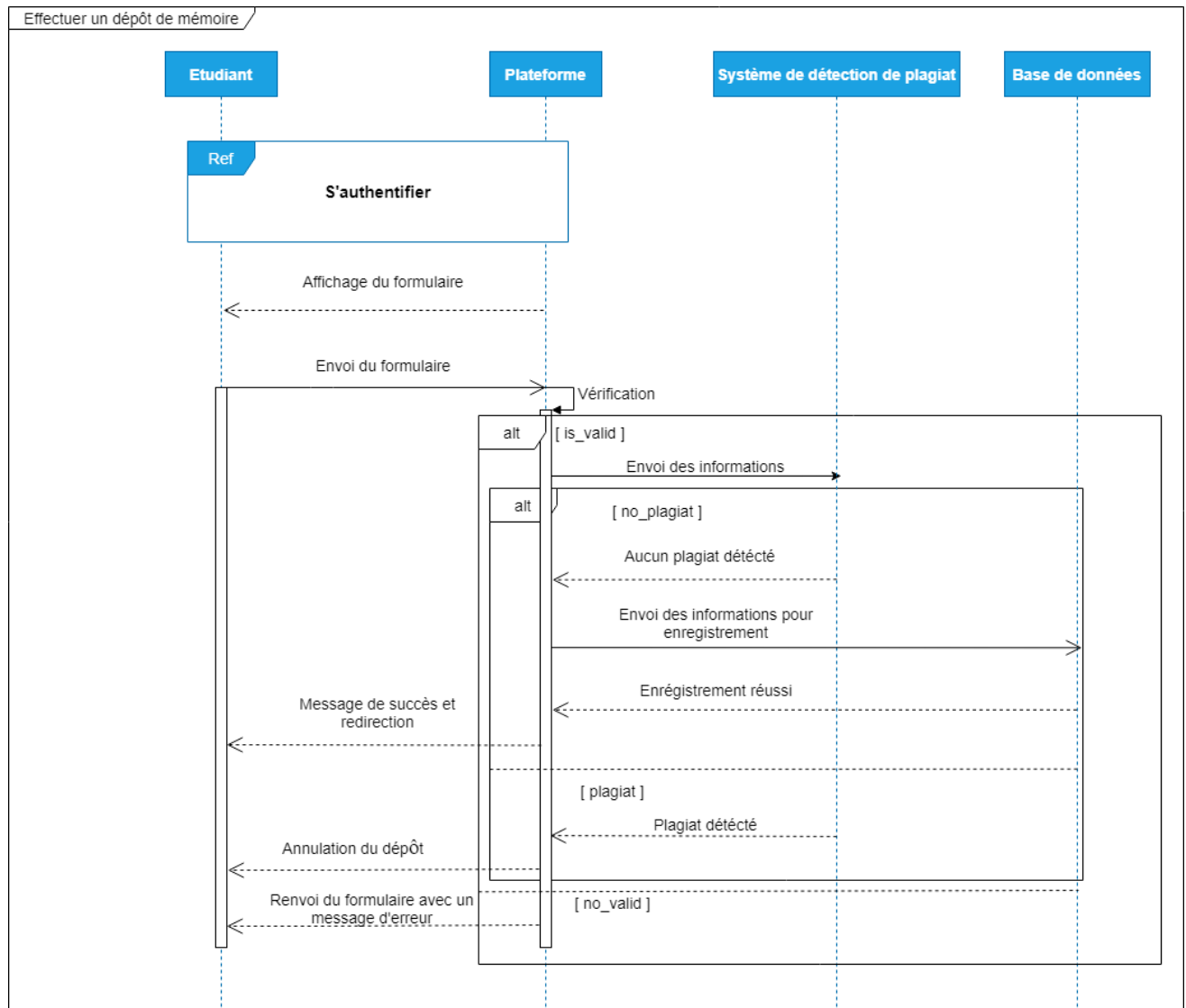


Figure 12: Diagramme de séquence du cas « Effectuer un dépôt de mémoire »

(Source : Roméo GODONOUKPE & Parménas ZINKO)

Diagramme de séquence du cas « Valider un dépôt de mémoire »

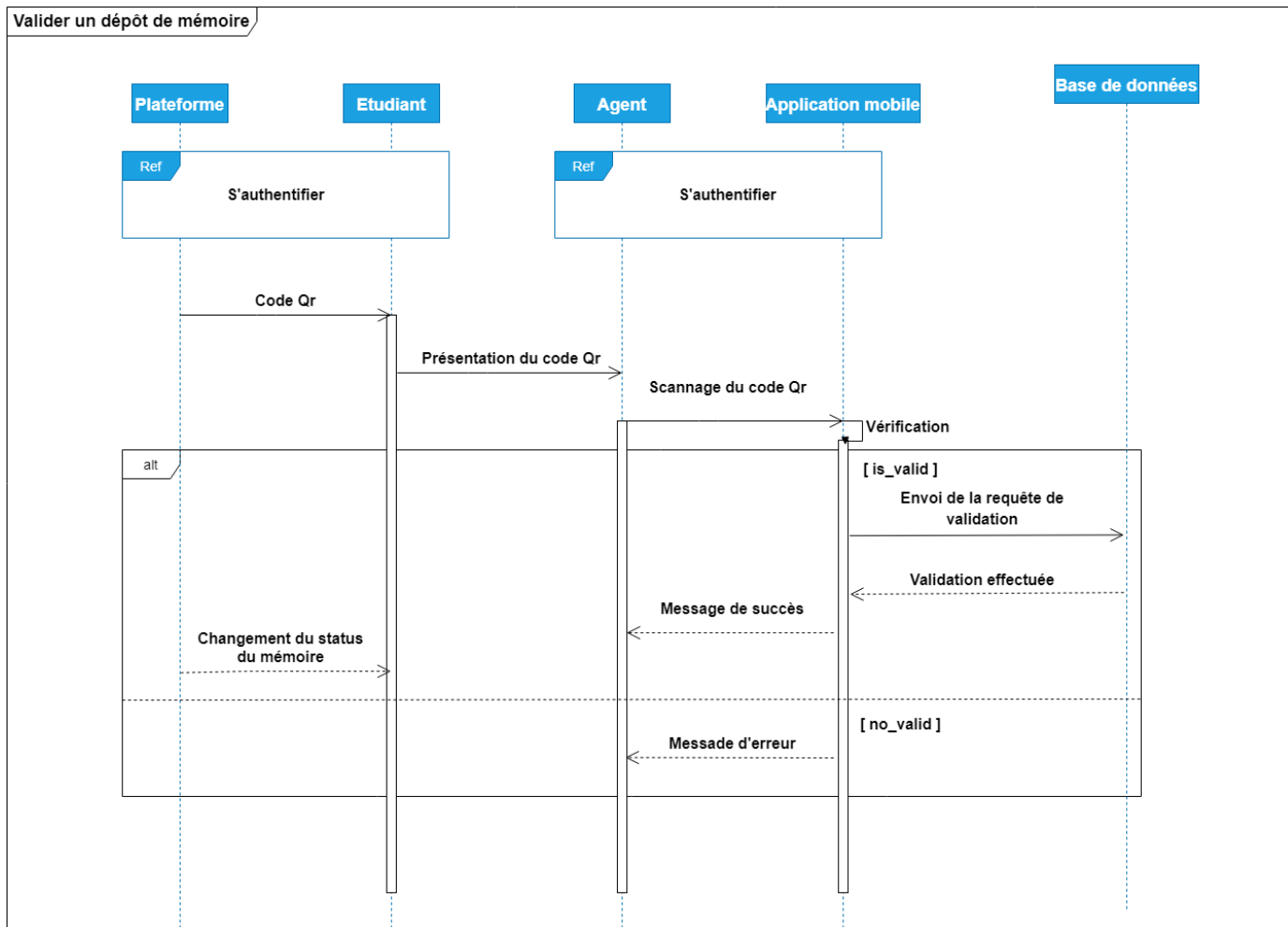


Figure 13: Diagramme de séquence du cas « Valider un dépôt de mémoire »

(Source : Roméo GODONOUKPE & Parménas ZINKO)

5-2 Réalisation

5-2-1 Choix techniques

Dans cette partie, nous indiquerons les approches, algorithmes et outils ayant servi à la mise en place de notre solution tout en justifiant ses différents choix.

5-2-1-1 Choix d'une Approche de détection de plagiat

Les différentes approches de détection de plagiat citées plus haut permettent toutes d'obtenir des résultats moyennement efficaces mais ne sont pas adaptées pour certains cas de plagiat.

Dans un premier lieu, l'approche Fingerprinting se base sur les empreintes calculées pour identifier une similitude. Ceci signifie qu'une légère modification du texte suffit pour rendre inefficace notre détection. Cela montre clairement ses limites dans le cas de la paraphrase où l'intéressé reformule l'idée d'une phrase avec des synonymes.

Par ailleurs, l'analyse des citations (Citation analysis), s'avère efficace mais nous ne pourrions effectuer une analyse si le document ne contient pas des citations.

Concernant la « Stylometry » c'est une méthode puissante au niveau de la détection du plagiat, le seul problème de cette méthode c'est qu'il faut travailler avec un expert linguistique pour avoir une aide linguistique.

Les Bag of Words, comme l'approche Fingerprinting restent limités pour le cas de la paraphrase.

Finalement, nous nous focaliserons sur la dernière méthode (String matching) qui se base sur les concepts sémantiques pour avoir un résultat pertinent. L'objectif en se servant des concepts sémantiques est de traiter au maximum le texte avant d'y appliquer l'algorithme de recherche. Ceci permettra de formater le texte de sorte à ce que tous les mots synonymes à un mot donné soient reconnus comme le même mot ou que les verbes conjugués soient remplacés par leur infinitif etc. Nous pouvons donc réduire au mieux les limites sémantiques.

En définitif, nous utiliserons pour le développement de notre système de détection de plagiat l'approche **String matching**.

5-2-1-2 Choix d'un algorithme de String matching

Il existe plusieurs algorithmes de recherche de caractère. Nous distinguons :

- ✓ L'algorithme naïf ou force brute

Elle est la plus utilisée et consiste à réaliser une comparaison caractère après caractère de la chaîne initiale et de la chaîne recherchée. On ignore les caractères

de la chaîne initiale tant qu'ils sont différents du premier caractère de la chaîne à trouver. Dès qu'on trouve un caractère identique, on parcourt les caractères suivants qui correspondent. Si un caractère diffère alors qu'on n'a pas atteint la fin de la chaîne recherchée, alors on reprend la recherche du premier caractère identique, à partir du caractère suivant dans la chaîne initiale et ceci jusqu'à retrouver toutes les correspondances dans la chaîne initiale.

Sur de petits textes cet algorithme est efficace. Cependant pour des textes assez longs (comme dans notre cas), la complexité temporelle de l'algorithme devient vraiment élevée. On obtient une complexité temporelle de $O(n * m)$.

n représente le nombre de caractère de la chaîne initiale

m représente le nombre de caractère de la chaîne recherche

✓ L'algorithme de Knuth-Morris-Pratt

C'est un algorithme de recherche de caractères bien efficace que l'algorithme de recherche par Force brute. Il recherche les occurrences d'un mot dans une chaîne de texte en utilisant l'observation que lorsqu'une discordance se produit, le mot lui-même contient suffisamment d'informations pour déterminer où la prochaine correspondance pourrait commencer, évitant ainsi le réexamen des caractères précédemment mis en correspondance.

Il a une complexité temporelle de $O(n + m)$.

✓ L'algorithme de Boyer-Moore

L'algorithme de Boyer-Moore est un algorithme de recherche de caractères particulièrement efficace qui ne vérifie pas chacun des caractères dans la chaîne initiale mais peut au contraire sauter certains d'entre eux. L'algorithme de Boyer-Moore est assez surprenant car il effectue sa vérification à l'envers.

Par exemple, s'il commence la recherche de la sous-chaîne « EVOLUTION » au début de la chaîne initiale, il vérifie d'abord la neuvième position en regardant si

elle contient un « N ». Ensuite s'il trouve un « N », il vérifie la huitième position et ainsi de suite jusqu'à ce qu'il ait vérifié la première position du texte pour y trouver un « E ». Si une vérification échoue, on considère qu'aucune correspondance avec la chaîne « EVOLUTION » n'existe au tout début du texte ainsi que les dans les huit positions qui la suivent. Après la vérification d'un seul caractère, l'algorithme est capable de passer plusieurs caractères sans faire de vérifications anodines. Il a une complexité temporelle de $O\left(\frac{n}{m}\right)$.

Pour l'implémentation de notre système de détection de plagiat, notre choix s'est porté sur l'algorithme de Boyer-Moore en raison ses performances.

5-2-1-3 Choix d'une bibliothèque de traitement de Texte

Dans le Traitement Automatique des Langues (TAL) ou le NLP (Natural Language Processing), il existe plusieurs bibliothèques pour le traitement de texte. Ces bibliothèques permettent à la machine de comprendre au mieux le message véhiculé à travers un texte. Généralement, elles proposent des corpus de textes dans plusieurs langues pour faciliter le travail. Comme bibliothèques Python de traitement de texte nous avons :

- ✓ NLTK (Natural Language Tool Kit) est une bibliothèque Python qui permet d'effectuer différentes tâches comme la Tokenization, la Lemmatization, le Stemming. Son corpus est assez riche et supporte le traitement de texte à grande échelle.
- ✓ Spacy : c'est le principal compétiteur de NLTK et propose des fonctionnalités similaires à celui-ci. Son corpus est également riche mais le temps d'exécution de ses scripts est relativement élevé par rapport à celui de NLTK.
- ✓ Scikit-learn : c'est une bibliothèque pour le Machine Learning mais il propose également des outils de traitement de texte. A la différence de

NLTK ou de Spacy, il n'est pas dédié uniquement à la tâche de traitement de texte.

Il en existe bien d'autres en dehors de ceux cités ci-dessus.

Notre choix s'est porté sur NLTK en raison de la robustesse de ses algorithmes, de la richesse de son corpus et de sa rapidité d'exécution.

5-2-1-4 Choix de Code QR

Un code QR est un type de code-barres en deux dimensions (2D) constitués de modules noirs disposés dans un carré à fond blanc. L'agencement de ces points définit l'information que contient le code qui peut être décodée rapidement. A la différence d'un code-barres classique, le code QR peut contenir beaucoup d'informations et surtout des données reconnues directement par des applications permettant ainsi de déclencher facilement des actions comme naviguer sur un site internet, déclencher un appel vers un numéro de téléphone, faire un paiement direct via son téléphone etc.

Notre choix s'est donc porté sur cet outil pour toute tâche de validation de mémoires (validation de dépôt de mémoire et validation finale) en raison de sa facilité et sa rapidité d'utilisation et de création. Pour valider par exemple le dépôt de mémoire d'un étudiant, il suffira juste de lancer l'application de scannage et de viser le code et l'action équivalente sera effectuée.

5-2-2 Implémentation du système de détection de Plagiat

Les principales étapes de détection de plagiat sont :

5-2-2-1 La « Tokenization » :

La Tokenization est une étape dans le processus de normalisation en **TAL** (Traitement Automatique des Langues). C'est un travail de segmentation qui décompose une phrase en ses multiples éléments. Elle segmente le texte en entrée en unités linguistiques manipulables comme les mots, la ponctuation, les données

alphanumériques (désignés ici par «Token») qui sera utile pour l'analyse. Le but de ce procédé au début de notre traitement, est de séparer les unités de base du texte qui se prêteront à une analyse pointue par la suite. On peut penser qu'il suffit de détecter les espaces entre les mots, mais ce n'est pas toujours aussi facile, en particulier en français. La présence des tirets, les apostrophes sont autant de difficultés rencontrées. Les outils disponibles dans la bibliothèque de traitement de texte utilisée nous ont été très utiles pour avoir quand-même des résultats efficaces.

5-2-2-2 L'enlèvement des « Stop Words »

Elle représente la seconde étape dans le processus de traitement de nos différents textes. Elle consiste à supprimer les mots généralistes qui sont fréquents dans la plupart des textes (par exemple : je, étais, avais etc.) et qui n'apportent pas de grandes informations dans la tâche à effectuer. Les « Stop Words » sont établis comme des listes de mots et sont disponibles dans la bibliothèque NLTK.

5-2-2-3 Le « Stemming » et/ou la « Lemmatization »

Cette étape est cruciale pour l'obtention de meilleur résultat. Les Tokens obtenus à la phase de « Tokenization » seront remplacés par leurs concepts associés dans un dictionnaire de mots. Ici, nous tentons de générer le radical de ces Tokens. Le « Stemming » permet de supprimer les préfixes et suffixes d'un mot (le résultat obtenu n'est forcément pas un mot dans la langue) alors que la « Lemmatization » quant à elle identifie le radical d'un mot directement dans une langue donnée à partir du Corpus WordNet de NLTK (le radical obtenu est appelé lemme).

En dehors de notre bibliothèque de traitement de texte, nous avons également conçu un dictionnaire de mots pour l'identification des synonymes.

5-2-2-4 La recherche de mot et le calcul du pourcentage de plagiat

Ici nous implémentons l'algorithme de recherche pour générer ensuite un pourcentage de plagiat. La fonction de recherche prend en entrée les deux textes à comparer (le texte du document suspecté de plagiat et un document référence enregistré déjà dans notre système). Ensuite grâce aux différentes fonctions de traitement de texte, nous formatons les différentes entrées en des listes de mots pour l'étape de recherche.

On note par **listeA** la liste de mots du texte du document suspecté de plagiat et par **listeB** la liste de mots du texte du document référence.

Alors, chaque mot contenu dans la listeA sera recherché dans la listeB grâce à l'algorithme de Boyer-Moore. A chaque fois que le mot est trouvé dans le texte, nous incrémentons (de +1) une variable notée s_h . A la fin de la recherche, nous calculons le pourcentage de plagiat à partir de la formule :

$$P = \frac{2 * s_h}{th_a * th_b} * 100$$

Où :

P représente le pourcentage de plagiat

s_h représente le nombre de mots identiques

th_a représente la taille de la listeA

th_b représente la taille de la listeB

5-2-3 Sécurité du Système

Il est important de mettre en place une bonne sécurité autour de notre plateforme afin de donner une meilleure expérience aux utilisateurs du système. Ainsi, un validateur de mot de passe est associé à tous les champs de mot de passe pour s'assurer de leur robustesse. Par ailleurs ces mots de passe sont bien cryptés avant leur enregistrement. Grâce aux outils fournis par le framework Django nous

protégeons notre plateforme contre les attaques comme le **Cross Site Scripting**, le **Cross Site Request Forgery** (chaque envoi de formulaire est associé à un Token CSRF) et les **Injections SQL**.

5-2-4 Présentation des interfaces

Les interfaces ci-après résument au mieux les différentes fonctionnalités importantes de notre plateforme et de l'application de scannage mis à disposition des universités.

5-2-4-1 Interfaces de la plateforme Web

Accueil :

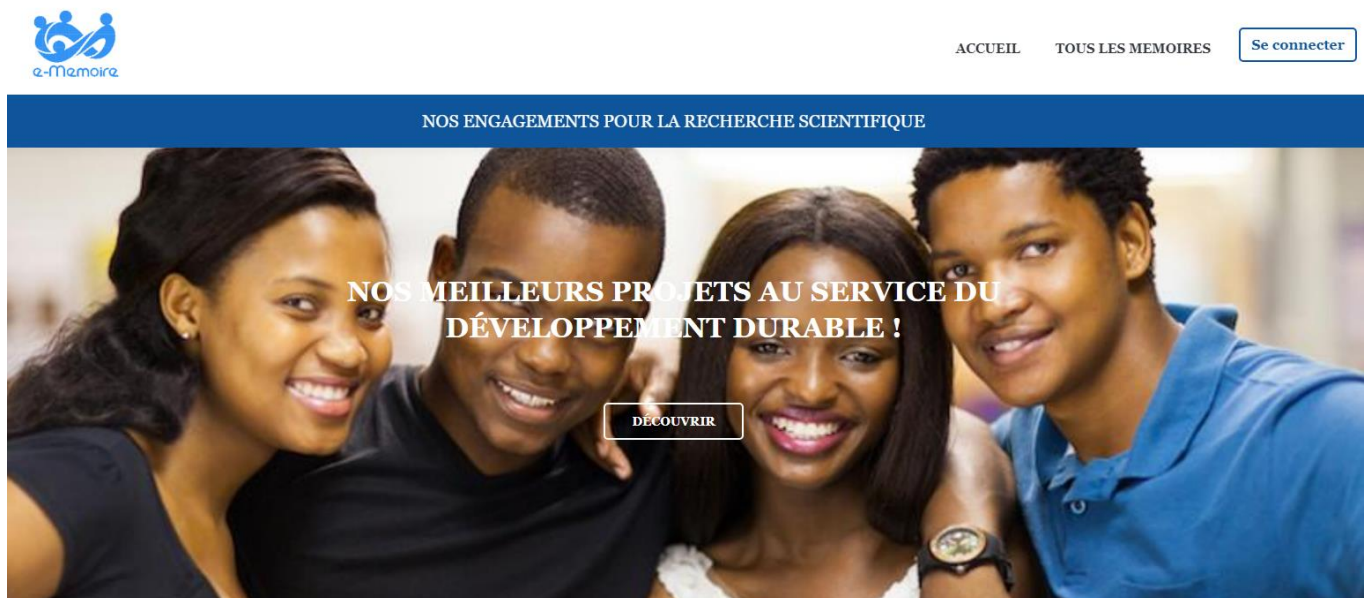


Figure 14: Interface Accueil 1

(Source : Roméo GODONOUKPE & Parménas ZINKO)

REALISATION D'UNE PLATEFORME WEB D'ARCHIVAGE ET DE MISE A DISPOSITION DES MEMOIRES ET THESES MUNIE D'UN SYSTEME DE DETECTION DE PLAGIAT

Dépôt de mémoire :

Accueil

Tableau de Bord

Profile

Dépôt de mémoire

EFFECTUER UN NOUVEAU DÉPÔT

Université *

Ecole/Faculté/Institut *

Filière *

Option

Thème du mémoire *

Superviseur *

Document du Mémoire *

Choisir un fichier Aucun... choisi

ENREGISTRER

Figure 15: Interface de dépôt de mémoire

(Source : Roméo GODONOUKPE & Parménas ZINKO)

Tableau de bord :

E-MEMOIRE

Rechercher ...

Katia

Tableau de Bord

Année Académique 2020-2021

Tous les mémoires

Thème du Mémoire	Etudiants	Filière	Prévue pour le	Lieu
Réalisation d'un système d'enroulement administratif	VIEGBE Patrick	Génie Electrique et Informatique	13 mai 2021	Salle PO4
Traducteur en langue Nationale	GODONOUKPE Roméo & HOUNYOVI Stanislas	Génie Electrique et Informatique	1 juillet 2021	Salle de Lecture
Conception d'un système de détection de plagiat	ZINKO Parménas & HOUDONUGBO Eden	Génie Electrique et Informatique	2 juillet 2021	Salle de Lecture

Copyright © e-Memoire 2021

Figure 16: Tableau de Bord (pour les enseignants)

(Source : Roméo GODONOUKPE & Parménas ZINKO)

REALISATION D'UNE PLATEFORME WEB D'ARCHIVAGE ET DE MISE A DISPOSITION DES MEMOIRES ET THESES MUNIE D'UN SYSTEME DE DETECTION DE PLAGIAT

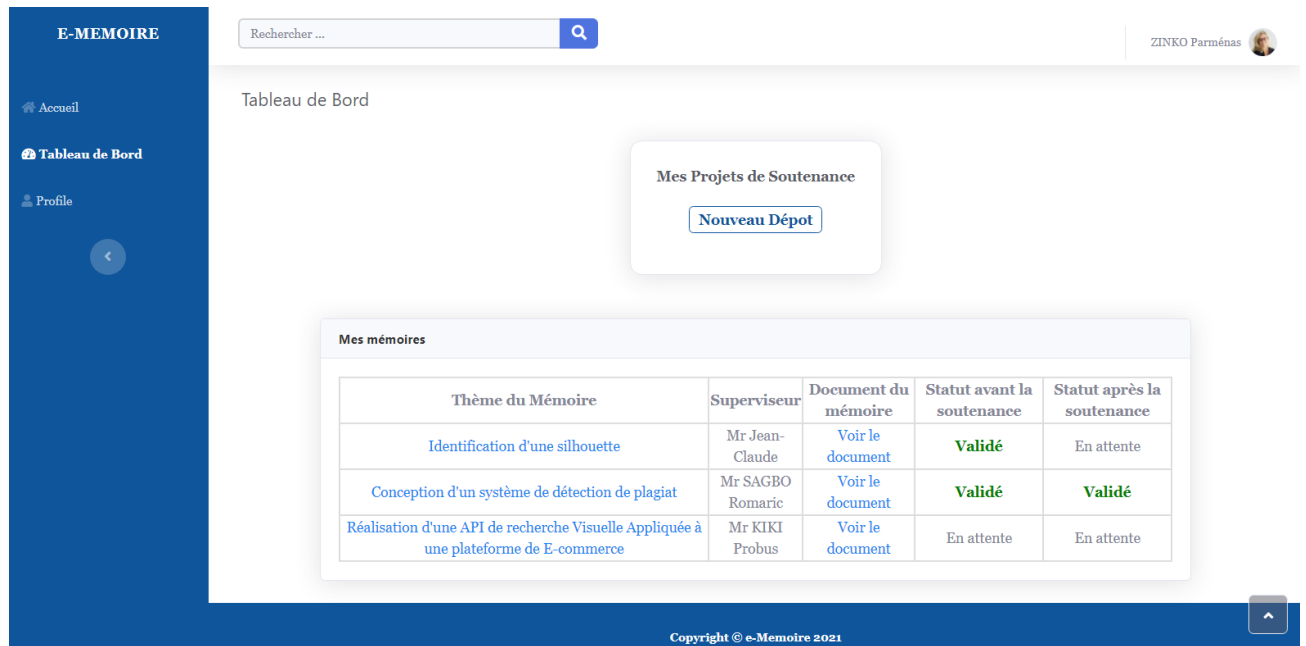


Figure 17: Interface 2 Tableau de Bord (pour les étudiants)

(Source : Roméo GODONOUKPE & Parménas ZINKO)

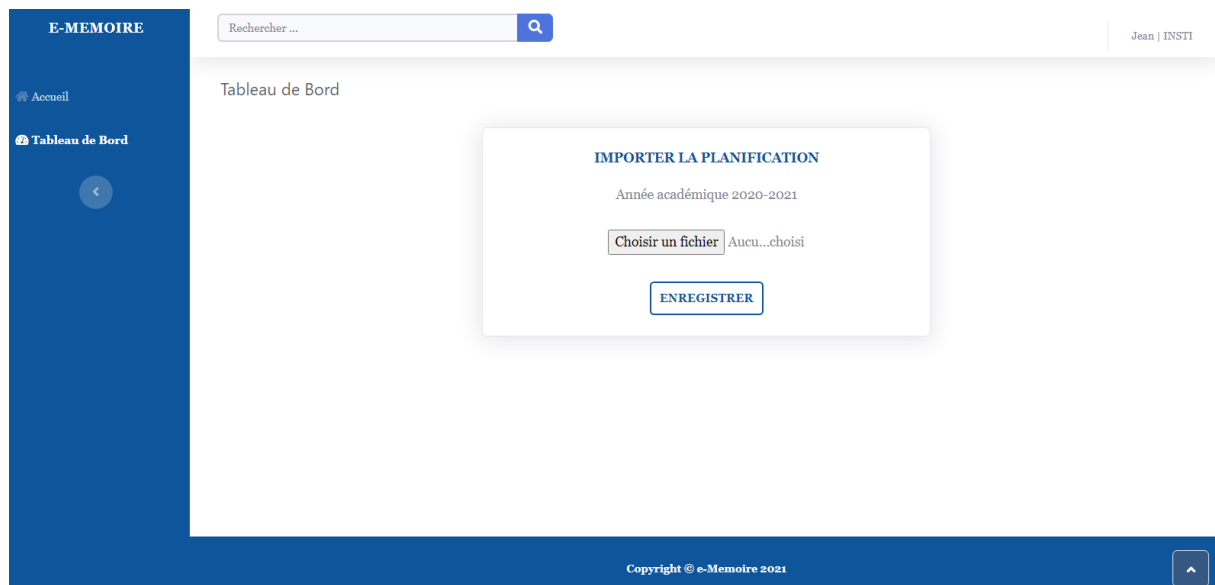


Figure 18: Interface 3 Tableau de Bord (pour les agents)

(Source : Roméo GODONOUKPE & Parménas ZINKO)

Administration

ACCOUNTS		
Commentaires	+ Ajouter	Modification
Ecole/Institut/Facultés	+ Ajouter	Modification
Enseignants	+ Ajouter	Modification
Etudiants	+ Ajouter	Modification
Filières	+ Ajouter	Modification
Mémoires	+ Ajouter	Modification
Options	+ Ajouter	Modification
Planifications	+ Ajouter	Modification
Universités	+ Ajouter	Modification
AUTHENTIFICATION ET AUTORISATION		
Groupes	+ Ajouter	Modification
Utilisateurs	+ Ajouter	Modification

Actions récentes

Mes actions

- Réalisation d'une API de recherche Visuelle Appliquée à une plateforme de E-commerce
Mémoire
- Conception d'un système de détection de plagiat
Mémoire
- Identification d'une silhouette
Mémoire
- Application de Payment Bancaire
Mémoire
- Mise en place d'un chantier Robuste
Mémoire
- Réalisation d'un système d'enroulement administratif
Mémoire
- Traducteur en langue Nationale
Mémoire
- Traducteur en langue Nationale
Mémoire
- Conception d'un système de détection de plagiat
Mémoire

Figure 19: Interface Administrateur

(Source : Roméo GODONOUKPE & Parménas ZINKO)

5-3 Difficultés rencontrées

Pendant la réalisation de notre travail, nous avons rencontré quelques difficultés énumérés ci-après :

- ✓ Difficulté pour le choix d'un algorithme de détection de plagiat
- ✓ Difficulté à traiter efficacement le texte pour le test
- ✓ Difficultés à s'approprier et à maîtriser certains modules qui nous étaient nouveaux

CONCLUSION

Le présent travail constitue notre Travail de Fin d'Etude qui porte sur la réalisation d'une plateforme d'archivage et de mise à disposition des mémoires et thèses au Bénin à laquelle est appliqué un système de détection de plagiat.

Ce travail a été réalisé afin d'apporter notre contribution à la mise à disposition de tout usager les travaux de fin d'études des étudiants inscrits dans les universités béninoises pour faciliter leur mise en exploitation. Les différentes tâches effectuées au cours de notre stage de fin de formation et la réalisation de cette solution, nous ont permis d'approfondir nos connaissances, d'acquérir de nouvelles compétences notamment en terme d'analyse des besoins utilisateurs.

Dans ce rapport, nous avons présenté notre centre de formation, notre structure d'accueil ainsi que les travaux effectués lors de notre stage dans cette structure. Ensuite, dans le cadre de notre projet nous avons présenté les étapes de la conception de notre plateforme d'archivage et notre système de détection de plagiat, puis avons ajouté les résultats obtenus. Nous avons fait face à quelques difficultés lors de notre travail et quelques perspectives sont envisageables afin d'améliorer le fonctionnement de notre plateforme.

Ainsi, des techniques d'intelligence artificielles plus robustes peuvent être utilisées pour optimiser les capacités de notre système de détection de plagiat à faire face aux limites sémantiques imposées dans le cas de la paraphrase. La détection de plagiat avec des documents sur internet, l'amélioration de la présentation des mémoires à partir des vidéos (la réalisation des vidéos sera laissée à la charge de l'étudiant qui peut le faire lui-même ou passer une commande depuis la plateforme), la génération des attestations de fin de formation et la prise en compte des publications scientifiques (effectuées par les chercheurs universitaires), sont autant de possibilités pour améliorer accroître la notoriété de notre système.

BIBLIOGRAPHIE

- ✓ **HAMBI EL MOSTAFA, FAOUZIA BENABBOU, EL HABIB BEN LAHMAR** «*Comparaison Des Techniques De Détection Du Plagiat Académique*» ; Laboratoire Traitement de l'Information et Modélisation
- ✓ **MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE BENIN** «*Annuaire Statistique*»
Année académique 2016-2017 et 2017-2018
- ✓ **SONAWANE KIRAN SHIVAJI, PRABHUDEVA S** «*Plagiarism Detection by using Karp-Rabin and String Matching Algorithm Together*» ; International Journal of Computer Application
- ✓ **CFAURY** «*Complexité d'un algorithme*»,
URL : <https://info.blasepascal.fr/nsi-complexite-dun-algorithme>,
consulté le 16 mai 2021

WEBOGRAPHIE

- ✓ <https://docs.djangoproject.com>
- ✓ <https://openclassrooms.com/fr/courses/1871271-developpez-votre-site-web-avec-le-framework-django>
- ✓ <https://www.stackoverflow.com>
- ✓ <https://www.w3schools.com>
- ✓ <https://www.wikipédia.com>
- ✓ <https://www.github.com>
- ✓ <https://www.google.com>
- ✓ <https://youtube.com>

ANNEXES

Complexité Temporelle d'un algorithme

L'objectif du calcul de complexité temporelle d'un algorithme temporelle est de pouvoir comparer l'efficacité d'algorithmes résolvant le même problème. Elle revient à comparer le nombre d'opérations élémentaires (affectation, calcul arithmétique ou logique, comparaison..) effectuées par l'algorithme.

La complexité en temps d'un algorithme sera exprimée par une fonction, notée T (pour Time) qui dépend de la taille des données passées en paramètres et de la donnée elle-même. On distingue pour un algorithme la complexité dans le meilleur des cas (c'est la situation la plus favorable) et la complexité dans le pire des cas (la situation la plus défavorable).

Ordre de grandeur

Pour comparer des algorithmes, il n'est pas nécessaire d'utiliser la fonction T , mais seulement l'ordre de grandeur asymptotique, noté O .

Une fonction $T(n)$ est en $O(f(n))$ si :

$$\exists n_0 \in \mathbb{N}, \exists c \in \mathbb{R}^+, \forall n \in \mathbb{R}^+, n \geq n_0 \Rightarrow |T(n)| \leq c|f(n)|$$

Autrement dit, $T(n)$ est en $O(f(n))$ s'il existe un seuil n_0 à partir duquel la fonction T est toujours dominée par la fonction f , à une constante multiplicative fixée c près.

Exemples :

$$\checkmark T_1(n) = 7 = O(1)$$

$$\checkmark T_2(n) = 12n + 5 = O(n)$$

$$\checkmark T_3(n) = 4n^2 + 2n + 6 = O(n^2)$$

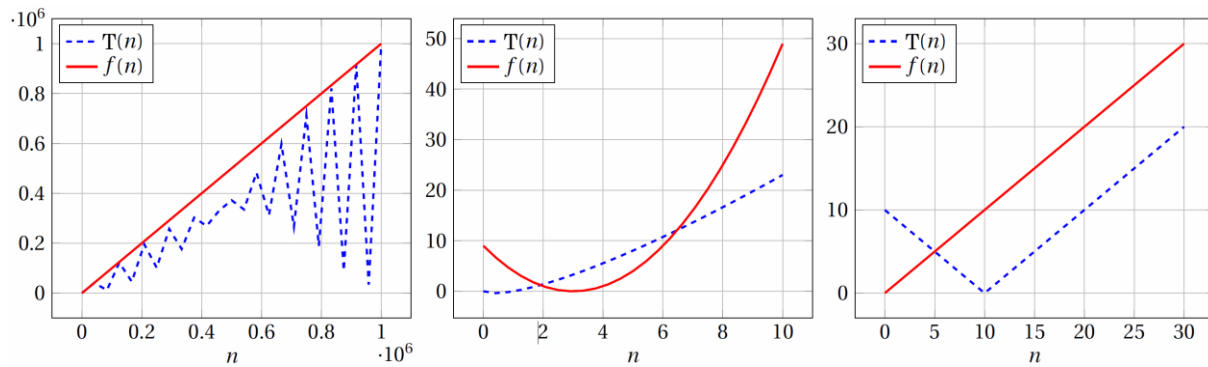


Figure 20: Tracés des fonctions $T(n)$ et $f(n)$

(Source : Internet)

Classes de complexité

O	Type de complexité
$O(1)$	Constante
$O(n)$	Linéaire (utilisé dans le cas de notre étude)
$O(n^2)$	Quadratique
$O(n^3)$	Cubique
$O(\log(n))$	Logarithmique
$O(n * \log(n))$	Quasi-linéaire
$O(n!)$	Factorielle
$O(2^n)$	Exponentielle

Tableau 3: Différentes classes de complexité temporelle

(Source : Internet)

TABLES DES MATIERES

DEDICACES.....	i
REMERCIEMENTS	ii
SOMMAIRE	iii
LISTE DES ACRONYMES ET ABREVIATIONS.....	iv
LISTES DES TABLES ET FIGURES	v
CAHIER DES CHARGES.....	vi
RESUME.....	vii
ABSTRACT	vii
INTRODUCTION.....	1
PARTIE N° 1 : PRESENTATION DES STRUCTURES	3
Chapitre n°1 : Présentation de l'INSTI Lokossa	4
1-1 Situation géographique.....	4
1-2 Historique	5
1-3 Mission	5
1-4 Administration de l'INSTI	6
Chapitre n°2 : Présentation de la structure d'accueil.....	8
1-1 Présentation du cabinet PROTECHNOLOGIES	8
1-2 Organigramme de l'administration de PROTECHNOLOGIES.....	9
PARTIE N° 2 : DEROULEMENT DU STAGE	10
Chapitre n°3 : Déroulement du stage.....	11
3-1 Initiation à la création d'application Web à partir du Framework ASP .Net Core	11
3-2 Conception d'une application de gestion des dépenses quotidiennes ..	12
PARTIE N°3 : TRAVAUX DE FIN D'ETUDES	14
Chapitre n°4 : Présentation du projet de fin d'études.....	15
4-1 Problématique	15
4-2 Innovation	17

4-3 Objectifs	18
4-3-1 Objectif général	18
4-3-2 Objectifs spécifiques.....	18
4-4 Méthodologie.....	19
4-5 Etude globale du système	20
4-5-1 Le Web	20
4-5-2 La détection de plagiat.....	22
4-6 Technologies, outils et langages utilisés	25
4-6-1 Technologies utilisées	25
4-6-2 Outils utilisés	28
4-6-3 Langages utilisés.....	30
Chapitre n°5 : Conception et réalisation de notre solution.....	32
5-1 Conception	32
5-1-1 Langage de Modélisation	32
5-1-2 Modélisation UML	32
5-1-2-2 Vues dynamiques	41
5-2 Réalisation.....	43
5-2-1 Choix techniques	43
5-2-2 Implémentation du système de détection de Plagiat.....	47
5-2-3 Sécurité du Système	49
5-2-4 Présentation des interfaces	50
5-3 Difficultés rencontrées	53
CONCLUSION	54
BIBLIOGRAPHIE	55
WEBOGRAPHIE.....	56
ANNEXES	57
TABLES DES MATIERES	59