**RÉPUBLIQUE DEMOCRATIQUE DU CONGO**

**PROVINCE DU NORD-KIVU**

**ENSEIGNEMENT SUPÉRIEUR ET UNIVERSITAIRE**



[**www.ci-ucbc.edu.cd**](http://www.ci-ucbc.edu.cd) **|** [**www.congoinitiative.org**](http://www.congoinitiative.org)

**Faculté de Technologie et sciences de l’ingénieur**

**Département de Génie Informatique**

**Système de recommandation basé sur les contenus pour recommander les sujets de mémoire aux étudiants finalistes**

*Mémoire présenté et soutenu en vue de l’obtention du diplôme de licence en faculté de technologie et sciences de l’ingénieur*

Mention : Génie informatique

**Par Kambale Muteya Cedric**

Dirigé par : Dr. Nsenge Mpia Hértitier, PhD

Professeur Associé

**2023-2024**

1. **SIGNATURES D’ÉQUIPE DE DIRECTION ET DE LA FACULTÉ**

**SUJET :** « ***Système de recommandation basé sur les contenus pour recommander les sujets de mémoire aux étudiants finalistes*** »

Par : Kambale Muteya Cedric 5520

(+243 974 891820, [cedricarnoldmbulamuteya@gmail.com](mailto:cedricarnoldmbulamuteya@gmail.com))

Conformément aux politiques de l’UCBC, ce mémoire est accepté en vue de satisfaire partiellement aux exigences du diplôme de licence

Date

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Dr. Nsenge Mpia Hértitier, PhD. Professeur Associé

Directeur

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Erick Kalwana, MSc.

Secrétaire de la Faculté de technologie et sciences de l’ingénieur

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Félix Katimika, CT, Coordonnateur de la faculté des Sciences Appliquées et Secrétaire Général académique chargé de la recherche

1. **EPIGRAPHE**

*"La recherche de la vérité est plus précieuse que sa possession."*

**Albert Einstein**

*"L'innovation est le moteur du progrès, et la diversité des idées est le carburant de la recherche académique."*

**Anonyme**

1. **DEDICACE**

A ma mère !

1. **REMERCIEMENTS**

Avant tout, nous remercions le bon Dieu qui nous a donné santé, courage et volonté pour réaliser ce présent travail de mémoire.

1. **RESUME**
2. **ABSTRACT**
3. **TABLES DE MATIERES**
4. **SIGLES, ABBREVIATIONS ET ACRONYMES**
5. **LISTE DES TABLEAUX**
6. **LISTES DE FIGURES**
7. **PRINCIPALES DÉCOUVERTES DE LA RECHERCHE**
8. **CONCLUSIONS**
9. **LES RECOMMANDATIONS CLÉS**
10. INTRODUCTION GÉNÉRALE
    1. **Contexte de l’étude**

Dans le domaine académique, il est essentiel de gérer efficacement les sujets de mémoire afin d'éviter les répétitions et de promouvoir l'innovation. De nombreuses universités se retrouvent fréquemment confrontées à des sujets de mémoire déjà abordés, ce qui entraîne un gaspillage de ressources et entrave les progrès académiques. Il existe nombreux défis rencontrés par les universités dans la gestion des sujets de mémoire. Historiquement, l'absence de systèmes automatisés a conduit à une répétition non productive des thèmes de mémoire, ce qui entrave l'innovation académique et l'utilisation efficace des ressources. Dans nombreuses institutions, ce problème persiste, rendant urgente la mise en place d'une solution intelligente et collaborative pour la gestion des sujets de mémoire. Ceci est source de plagiat dans les travaux de mémoire. Le plagiat constitue une menace sérieuse pour l'intégrité académique et la qualité des recherches produites (McIntire et al., 2024).

Plusieurs travaux ont tenté de proposer des systèmes pouvant recommander des cours et des pratiques efficaces de supervision des mémoires afin d’optimiser le choix des étudiants dans leurs cursus académiques et de les accompagner de façon efficiente. Pendant COVID-19 par exemple, plusieurs plateformes d'apprentissage en ligne, de nombreux cours ont été disponibles et permettaient aux utilisateurs de sélectionner la meilleure option. (Jena et al., 2022) ont développé un système de recommandation en utilisant le mécanisme de filtrage collaboratif pour la recommandation de cours d'apprentissage en ligne. Leur travail se concentre sur les modèles basés sur l'approche de K-voisin le plus proche (KNN), la décomposition en valeurs singulières et les modèles de filtrage collaboratif basés sur les réseaux neuronaux. Pour leur part, (Liang et al., 2023) ont proposé un modèle d'apprentissage par renforcement hiérarchique avec critique des acteurs pour la recommandation de cours. Ils ont soutenu l'idée qu'il faut appliquer le Machine Learning dans le processus académique. Leur système de recommandation de cours a été jugé efficace et joue un rôle clé en aidant les apprenants à sélectionner les cours appropriés et en améliorant l'efficacité de l'apprentissage en ligne (Liang et al., 2023).

Toutefois, dans l’aspect des recommandations des sujets de mémoires, la revue de littérature a révélé une absence de travaux scientifiques. Ainsi, l’objectif de cette étude a été de créer une plateforme collaborative qui utilise des algorithmes de Machine Learning permettant aux finalistes de sélectionner des topics, de suivre et de gérer les sujets de mémoire à l'Université Chrétienne Bilingue du Congo (UCBC).

## Énoncé du problème

Cette étude traite de deux problèmes distincts. Premièrement, le choix des sujets de mémoire par les étudiants, qui peut impliquer une duplication à UCBC, ce qui entraine la non validation du sujet et peut limiter l'innovation si validé et peut entraîner une utilisation inefficace des ressources académiques. En plus, il y a le plagiat dans les mémoires, ce qui compromet la réputation académique et la qualité de la recherche. Actuellement, l'université ne dispose pas d'un système intelligent capable d'analyser les sujets déjà réalisés pour éviter la répétition et recommander des domaines de recherche peu explorés tout en offrant une solution solide pour la détection du plagiat. La littérature existante met l'accent sur l'importance d'un tel système pour améliorer la gestion académique, encourager de nouvelles recherches et garantir l'authenticité des travaux des étudiants (Dávila Morán, 2022).

## Objectif principal de l'étude

Cette étude a eu pour objectif global le développement d’une application web de gestion des travaux de mémoires au sein de l’UCBC avec fonctionnalité de recommandation des sujets de mémoire universitaires basés sur les mémoires déjà réalisés à l’UCBC.

## Objectifs spécifiques

Partant de l’objectif principal, cette recherche a pour objectifs spécifiques ce qui suit :

1. Collecter les données relatives aux sujets de mémoire précédemment réalisés dans la base de données de l'UCBC.
2. Implémenter des algorithmes d'apprentissage automatique pour l'analyse sémantique de ces sujets
3. Déployer le modèle efficace de recommandation de mémoire dans une plateforme de gestion de travaux de mémoire à l’UCBC
4. Évaluer l'efficacité du système proposé.

## Questions de recherche

## A la lumière des objectifs spécifiques sus-évoqués, cette étude a eu pour questions de recherche ce qui suit :

1. Comment collecter efficacement les données des sujets de mémoire précédemment réalisés ?
2. Quels algorithmes d'apprentissage automatique sont les plus appropriés pour l'analyse sémantique des sujets de mémoire et l'identification de similarités ?
3. Comment concevoir une interface utilisateur qui facilite la soumission et la recherche de sujets de mémoire tout en étant intuitive et conviviale ?
4. Le système proposé détecte-t-il efficacement le plagiat ?

## Justification

Le choix de ce sujet se justifie par la nécessité croissante de promouvoir l'innovation dans les universités, d'optimiser l'utilisation des ressources académiques, et de maintenir l'intégrité académique en luttant contre des recherches répétitives au niveau des sujets de mémoire. En développant ce système intelligent de gestion des sujets de mémoire avec des fonctionnalités anti-plagiat, cette recherche vise à répondre à un besoin critique identifié dans le milieu universitaire, contribuant ainsi à l'amélioration des pratiques académiques et à la bourse de connaissances en intelligence artificielle et en gestion académique.

## Importance de l'étude

Les universités, les chercheurs et les étudiants bénéficieront principalement de cette étude. Cette recherche permettra de réduire la duplication des sujets de mémoire, de les réorienter selon les problématiques et les domaines de recherches non-exploités, d'optimiser les ressources académiques, de promouvoir l'innovation et d'assurer l'authenticité des recherches en détectant les travaux répétés et assurant l’intégration de la technologie dans des universités de façon effective (McManus et al., 2024). Les administrateurs universitaires, les enseignants et les étudiants sont les principales parties prenantes.

* 1. **Limitations**

Les limites de cette étude comprennent plusieurs facteurs, notamment la disponibilité et la qualité des données des sujets de mémoire collectés. Nombreuses données ont été manquantes, incomplètes et de mauvaise qualité, ce qui peut avoir affecté l'analyse et les résultats du système proposé. De plus, l'intégration et le traitement de grandes quantités de données présentent souvent des défis techniques et logistiques. Ainsi, les résultats de cette recherche doivent être compris à la lumière de ces limites.

* 1. **Délimitation du sujet**

Cette recherche s’est concentrée sur les données de l'UCBC, en particulier les départements qui gèrent les sujets de mémoire dans différentes facultés, le service académique chargé de la recherche. La recherche s’est ainsi déroulée dans le cadre géographique de l'UCBC, avec une attention particulière aux bases de données académiques et aux archives des sujets de mémoire déjà réalisés par les étudiants de l’UCBC. Pour réduire l’impact des limitations susmentionnées, cette étude s’est focalisée sur l'utilisation de techniques spécifiques de traitement du langage naturel (NLP) afin d’optimiser le prétraitement des données (Pirnau et al., 2024).

Il est important de noter que cette étude s’est limitée à analyser les sujets de mémoire déjà réalisés à l'UCBC, à implémenter des modèles Machine Learning et à déployer le modèle validé dans un environnement web. En plus, bien que le domaine de la génération de contenu par IA se développe rapidement (Kalota, 2024), cette étude s’est arrêtée exclusivement sur la détection de similitudes et de plagiat dans les recherches existantes afin d’identifier les similitudes, recommander des domaines de recherche peu explorés et d’intégrer des fonctionnalités anti-plagiat pour assurer l'authenticité des recherches.

## Subdivision du travail

Cette recherche comprend trois chapitres. Le premier chapitre fait la revue de littérature. Dans ce chapitre, il s’agit d’aborder d’abord des concepts clés relatifs à cette étude dans la section nommée revue de littérature théorique et ensuite de présenter quelques travaux scientifiques ayant abordé les recherches sur les modèles de recommandation des sujets de mémoire. Cette section a été intitulée revue de littérature empirique. Le deuxième chapitre présente la méthodologie utilisée dans cette recherche afin d’atteindre les objectifs assignés. Le dernier chapitre illustre les résultats et la discussion relative à ces résultats obtenus.

PREMIER CHAPITRE *:* REVUE DE LITTÉRATURE

## Introduction

Dans cette section on passe en revue les concepts, théories et travaux antérieurs pertinents pour le développement d'une application web intelligente de recommandation des sujets de mémoire. Cette revue fournit un cadre théorique et empirique pour l'étude, mettant en évidence les lacunes et les opportunités d'innovation dans ce domaine.

# 1.1. Revue de littérature théorique

## 1.1.1. Systèmes de recommandation

(Béchet, 2012) part de la question « Pourquoi des tels systèmes » pour affirmer qu’un système de recommandation a pour objectif de fournir à un utilisateur des ressources pertinentes en fonction de ses préférences. Ce dernier voit ainsi réduit son temps de recherche mais reçoit également des suggestions de la part du système auxquelles il n’aurait pas spontanément prêté attention. Différentes définitions ont été faites des systèmes de recommandation ; cependant comprenons que les systèmes de recommandation sont des outils logiciels et techniques d'intelligence artificielle qui aident les utilisateurs à identifier les éléments les plus pertinents dans une grande variété d'options. Jena et al., (2022) ont proposé un système de recommandation basé sur des techniques d'intelligence machine, utilisant notamment les modèles KNN, SVD et NCF, pour recommander des cours sur la plateforme Coursera. L'évaluation montre que KNN offre de meilleures performances en termes de taux de réussite et de précision par rapport aux autres modèles.

### A. Classification des Systèmes de Recommandation

Les systèmes de recommandation sont généralement classés en trois grandes catégories :

Figure 1.1 Classification des systèmes de recommandation (Isinkaye et al., 2015)

#### A.1. Systèmes de recommandation basée sur le contenu (Content-Based Filtering)

Ce type de système recommande des éléments similaires à ceux que l'utilisateur a déjà consultés ou appréciés, en se basant sur les caractéristiques intrinsèques des éléments (Aggarwal, 2016). Par exemple, pour notre cas des sujets de mémoires, le système analyse les thèmes, méthodes et domaines d'étude des mémoires précédents pour proposer des sujets qui n'ont pas encore été explorés ou qui sont innovants. Dans le cadre de notre travail, le système de recommandation basé sur le contenu utilisera les caractéristiques (intitulé du sujet, problématique, hypothèses, questions de recherches etc.) des mémoires déjà réalisés pour proposer des sujets de mémoire aux étudiants finalistes.

#### A.2. Systèmes de recommandation collaboratifs (Collaborative Filtering)

Les systèmes de filtrage collaboratif recommandent des éléments en se basant sur les préférences de plusieurs utilisateurs similaires. Ils identifient des modèles communs parmi les utilisateurs ayant des comportements ou des préférences similaires (Aggarwal, 2016). Ce type de système pourrait involontairement encourager des sujets similaires s'il se base uniquement sur les préférences communes des étudiants, ce qui pourrait être contre-productif dans la lutte contre la duplication des sujets. Les méthodes du filtrage collaboratif peuvent être regroupées en deux catégories générales : à base de mémoire (heuristique, voisinage proche) et à base de modèles (figure 1.1). Notons que dans le filtrage collaboratif à base de mémoire, les notes des utilisateurs pour les items qui sont stockés dans le système sont directement utilisées pour prédire les notes des items non notés. Le filtrage collaboratif basé sur les utilisateurs, introduit pour la première fois dans le système GroupLens (Resnick & Varian, 1997), a pour principe de fonctionnement très simple, celui de déterminer les utilisateurs qui sont similaires à l’utilisateur courant, puis calculer une valeur de prédiction pour chaque item candidat à la recommandation en analysant les notes que les voisins de l’utilisateur courant ont exprimées sur cet item. Nous ne nous sommes pas concentrés ici vu que, pour notre travail nous avons utilisé un SR basé sur le contenu.

#### A.3. Systèmes de recommandation hybrides (Hybrid Recommender Systems)

Un système hybride combine plusieurs méthodes de recommandation, par exemple, en utilisant à la fois des méthodes basées sur le contenu et des méthodes collaboratives. Cela permet de tirer parti des avantages de chaque méthode tout en minimisant leurs inconvénients (Mpia et al., 2023). Dans ce cas, un système hybride pourrait être configuré pour d'abord éliminer les sujets déjà traités (basé sur le contenu), puis pour analyser les préférences des étudiants afin de proposer des sujets innovants et non redondants.

### A.4. Limites des systèmes de recommandation

Avec l'arrivée du web et la quantité énorme d'informations qu'il contient, les systèmes de recommandation sont devenus des outils très utiles pour gérer la surcharge d'informations et aider les utilisateurs à trouver ce dont ils ont besoin, par exemple, quand on cherche un film à regarder sur Netflix, une musique à écouter Spotify ou toute autre plateforme de streaming, le système te propose des films en fonction de tes goûts. Cependant, ces systèmes ont aussi quelques limites, dont (i) le *démarrage à froid (Cold Start Problem)*. C’est un problème qui se pose avec l’arrivée d’un nouvel élément dans le système de recommandation soit un utilisateur soit un item. Le système doit le gérer comme un cas spécial, le problème de démarrage à froid c’est qu’on n’a pas assez d’informations pour émettre des recommandations (Yuan & Hernandez, 2023) et (ii) *le biais de filtrage collaboratif*. Les systèmes de recommandation collaboratifs peuvent avoir tendance à recommander des éléments populaires ou courants, ce qui peut créer un biais. Ce phénomène est appelé "l'effet de popularité", où les items moins connus ou de niche sont rarement recommandés, limitant ainsi la diversité des suggestions (Gonzalez et al., 2022).

## 1.1.2. Recommandation des sujets de mémoire

La recommandation des sujets de mémoire est le processus par lequel des algorithmes Machine Learning suggèrent des sujets de recherche en se basant sur des données historiques et des intérêts individuels (Algarni & Sheldon, 2023). Ce processus vise à éviter la duplication des travaux et à promouvoir l'innovation académique. **Pour le cas échéant, comme signalé à l’introduction, cette étude a utilisé les données des travaux de mémoires déjà réalisés à l’UCBC, pour toutes les facultés. En effet, un mémoire** est un document rédigé par l’étudiant pour obtenir son diplôme de fin de cycle, ce document peut-être un sujet d’actualité ou un thème pour trouver une solution à un problème (Man & Zhan, 2023). Le sujet de mémoire doit généralement répondre au domaine de formation de l’étudiant.

## 1.1.3. Suivi et archivage

Le suivi et archivage est un processus essentiel dans la gestion de l'information et des documents. Le suivi consiste à surveiller, contrôler et gérer les flux de documents et d'informations à travers leurs cycles de vie, depuis leur création jusqu'à leur disposition finale. L'archivage, quant à lui, implique la conservation et la gestion sécurisée des documents importants et pertinents à long terme(Fitriyani et al., 2022). Ce processus permet de garantir l'intégrité, l'accessibilité et la disponibilité des documents pour des besoins futurs, tout en assurant leur protection contre les pertes ou dommages. Le suivi et l'archivage des documents sont des processus cruciaux dans la gestion de l'information. Le suivi implique la surveillance continue des documents, tandis que l'archivage garantit leur conservation à long terme (Alpasan et al., 2021). En outre, la curation numérique joue un rôle essentiel dans la gestion des documents électroniques, assurant leur intégrité et leur accessibilité (Yakel, 2007).

Nous pouvons citer ici quelques sites web de dépôts et d’archivage des mémoires. Certains sont des dépôts indépendants et d’autres sont mise dans des universités pour l’archivage :

1. *MemoireOnline :* Une plateforme en ligne qui donne accès à des milliers de mémoires, thèses, articles et autres publications académiques. Ces documents sont généralement soumis par des étudiants, des chercheurs ou des professionnels de divers domaines d'étude. En fournissant un accès centralisé à une large gamme de travaux universitaires, la plateforme vise à faciliter la recherche académique.
2. SGLT ULPGL : est un outil permettant la recherche du sujet de mémoire, l'enregistrement et le suivi de l'évolution des travaux de fin d'étude des étudiants à l’ULPGL GOMA.(*SGT ULPGL*, s. d.)
3. La plateforme DANTE (Dépôt et Archivage Numérique des Travaux Étudiants) : la plateforme institutionnelle en ligne de l'ensemble des travaux étudiants soutenus à l'Université Toulouse-Jean Jaurès. (*Accueil · Dante · Dante*, s. d.)

## 1.1.4. Plagiat

Le plagiat est l'acte de présenter le travail ou les idées d'autrui comme les siens sans reconnaissance adéquate, ce qui compromet l'intégrité académique. Ce comportement va au-delà de la simple violation des règles institutionnelles ; il remet en question les principes fondamentaux de la confiance et de l'intégrité qui sont essentiels à toute communauté académique (Rumanovská et al., 2024). Le plagiat peut prendre plusieurs formes :

1. **Plagiat direct** : Copier mot à mot le texte d'une source sans citation ni guillemets. C'est l'une des formes les plus flagrantes de plagiat.
2. **Plagiat par paraphrase** : Reformuler les idées d'un auteur sans lui attribuer le crédit approprié, ce qui peut donner l'impression que les idées sont originales.
3. **Plagiat mosaïque** : Emprunter des phrases ou des idées de plusieurs sources, en les modifiant légèrement, mais sans donner crédit aux auteurs originaux.
4. **Plagiat d'auto-plagiat** : Réutiliser ses propres travaux précédents sans le signaler, ce qui est fréquent dans les publications académiques.
5. **Plagiat involontaire** : Manquer de citer correctement une source, soit par négligence, soit par manque de connaissance des règles de citation.
6. **Plagiat par omission** : Oublier de citer des sources importantes ou omettre délibérément certaines références afin de s'approprier des idées sans donner crédit aux auteurs.

En effet, aujourd’hui l'émergence des technologies d'intelligence artificielle générative (IA), telles que les grands modèles de langage (LLM) comme ChatGPT, a précipité un changement de paradigme dans les domaines de l'écriture académique, du plagiat et de la propriété intellectuelle. Voilà pourquoi l'intégration croissante de l'IA dans la sphère universitaire nécessite une réévaluation de l'originalité de l'écriture, de l'objectif de l'apprentissage de la recherche et de l'écriture, ainsi que des cadres régissant la propriété intellectuelle (PI) et le plagiat (Hutson, 2024) .

## 1.1.5. Machine Learning

L'apprentissage automatique, en anglais Machine Learning (ML), ou apprentissage artificiel ou apprentissage statistique est un champ d'étude de l'intelligence artificielle qui se fonde sur des approches mathématiques et statistiques pour donner aux ordinateurs la capacité d’apprendre à partir de données, c'est-à-dire d'améliorer leurs performances à résoudre des tâches sans être explicitement programmés pour chacune. Le ML est devenu indispensable dans les activités quotidiennes. Elle offre une variété d'applications allant de simples applications web à des systèmes complexes tels que la reconnaissance vocale, la vision par ordinateur, le traitement du langage naturel et la reconnaissance faciale (Mpia et al., 2023).

## 1.1.6. Analyse sémantique

L'analyse sémantique est une technique utilisée en traitement automatique du langage naturel (TALN) pour comprendre le sens des mots, phrases, ou documents en se basant sur leurs contextes et relations(Suleman & Korkontzelos, 2021). Pour un système de recommandation basé sur les contenus pour les sujets de mémoire, l'analyse sémantique joue un rôle clé dans deux aspects : la recommandation de sujets pertinents et la détection de plagiat. Grâce à l'analyse sémantique, il est possible de mesurer la similitude sémantique entre les sujets déjà présents (contenus des mémoires précédents) et de nouveaux sujets suggérés.

### A.1. Analyse sémantique pour la recommandation

L'analyse sémantique permet d'évaluer la similarité sémantique entre les sujets déjà existants (contenus des mémoires antérieurs) et de nouveaux sujets proposés. Cela fonctionne en trois temps. Premièrement, il y a la représentation vectorielle des documents. Ici, chaque sujet de mémoire peut être représenté par un vecteur dans un espace de caractéristiques. Des techniques comme le TF-IDF (Term Frequency-Inverse Document Frequency), Word2Vec, ou BERT (Bidirectional Encoder Representations from Transformers) peuvent être utilisées pour convertir des documents en vecteurs.

(1)

Où TF est la fréquence du terme t dans le document d, et IDF est l'inverse de la fréquence des documents contenant le terme t dans le corpus D. Deuxièmement, il y a l’usage de la mesure de similarité. En fait, une fois que les sujets sont représentés sous forme de vecteurs, on peut calculer la similarité entre les vecteurs à l'aide de mesures telles que la similarité cosinus, la distance euclidienne, ou la distance de Jaccard (Han et al., 2021). La similarité cosinus est particulièrement courante pour des applications de TALN :

(2)

Où A et B sont les vecteurs des deux sujets. Enfin, il y a le processus de recommandation des sujets. En comparant un nouveau sujet proposé avec l'ensemble des sujets existants, vous pouvez recommander les sujets qui sont sémantiquement similaires ou complémentaires. L'objectif est de suggérer des sujets innovants tout en évitant les doublons.

### A.2. Analyse sémantique pour la détection de plagiat

La détection de plagiat vise à identifier les contenus très similaires à ceux existants, et l'analyse sémantique permet de détecter non seulement les copies exactes, mais aussi les reformulations ou paraphrases(Alvi et al., 2021). Les mémoires sont indexés à l'aide de modèles tels que BERT, qui capturent les subtilités sémantiques au-delà des simples correspondances de mots. Lorsqu'un nouveau mémoire est soumis, il est comparé à l'index des mémoires existants pour identifier les passages ayant une forte similarité sémantique. Des techniques comme l'Analyse Sémantique Latente (LSA) (Cheddak et al., 2024) ou l'Allocation de Dirichlet Latente (LDA) peuvent être utilisées pour détecter des similitudes sémantiques cachées. Un seuil de similarité peut être défini pour marquer les textes qui pourraient être plagiés si leur score de similarité dépasse un certain pourcentage, nécessitant ainsi une révision supplémentaire. On peut ainsi calculer la similarité Cosinus avec cette formule :

(3)

Où A et B sont les vecteurs des documents à comparer.

## Revue de littérature empirique

Pimau et al. (2024) ont examiné, à travers leur recherche : « Content Analysis Using Specific Natural Language Processing Methods for Big Data », les 1000 articles les plus cités sur le COVID-19 en utilisant des outils de traitement du langage pour comprendre les tendances et les sentiments dans les recherches scientifiques. Ils ont analysé les titres, les résumés et les mots-clés des articles pour voir quels mots sont les plus utilisés et si les titres sont similaires aux résumés. Ils ont utilisé des outils comme NLTK, TextBlob, VADER, et Azure Machine Learning pour mesurer les sentiments des titres et des résumés. L'objectif a été de créer une liste des termes les plus fréquents avec des représentations visuelles et d'évaluer les opinions des auteurs sur les sujets traités, mais les détails spécifiques sur la performance ne sont pas mentionnés.

Jena et al. (2023) ont développé un système de recommandation de cours en ligne utilisant des techniques de filtrage collaboratif telles que KNN et la décomposition en valeurs singulières. Liang et al. (2023) ont proposé un modèle d'apprentissage par renforcement hiérarchique pour la recommandation de cours, mettant en évidence l'importance de l'apprentissage automatique dans le processus académique. Ces études montrent l'efficacité des systèmes de recommandation dans l'optimisation des choix académiques.

Morán (2022) dans sa recherche intitulée « The Perception of Academic Plagiarism in Industrial Engineering Students at a Public University in Lima » a évalué comment les étudiants en ingénierie industrielle d'une université publique à Lima perçoivent le plagiat académique. L'étude s'est concentrée sur trois aspects : les connaissances, les motivations, et les pratiques liées au plagiat. Il a utilisé une méthode quantitative sans expérimentation. Un questionnaire de 35 questions, utilisant une échelle de Likert à cinq points, a été administré à un échantillon de 155 étudiants parmi une population de 2145, ciblant ceux des 8e et 9e cycles. Les résultats révèlent que 56,1% des étudiants voient le plagiat de manière négative, tandis que 43,9% ne le trouvent pas important.

## Conclusion partielle

Nous avons présenté ici une revue détaillée des concepts, théories et études empiriques liés à la recommandation des sujets de mémoire et à la détection de plagiat. Les points clés incluent l'importance des systèmes de recommandation et des techniques de ML dans l'optimisation des choix académiques et la maintenance de l'intégrité académique.

DEUXIEME CHAPITRE *:* METHODOLOGIE DE RECHERCHE

# 2.0. Introduction

La présente recherche propose de concevoir et déployer un système de recommandation basé sur le contenu pour suggérer des sujets de mémoire innovants aux étudiants finalistes de l'UCBC, dans le but de favoriser l'originalité et l'innovation académique. À travers une analyse sémantique des mémoires existants, effectuée à l'aide d'algorithmes d'apprentissage automatique, le système identifiera les sujets redondants et essayera de suggérer un autre sujet. Le modèle sera intégré dans une plateforme web développée avec Django, offrant ainsi un outil interactif de gestion des travaux académiques et de recommandations de sujets.

Cette section détaillera la conception de la recherche, les matériels et méthodes utilisés, ainsi que le processus d'analyse et de traitement des données, en vue de fournir une compréhension claire de la démarche scientifique adoptée pour atteindre les objectifs de ce projet.

# 2.1. Conception de la recherche

Le processus de notre recherche commence par la collecte de données des mémoires de l'UCBC, où sont extraits les titres, résumés, le nom de l’auteur et des infos sur lui. Ces données servent à construire un modèle de filtrage par contenu qui sera utilisé pour comparer les nouveaux sujets proposés par les étudiants avec les sujets déjà existants. L'objectif est d'identifier les similarités et les redondances afin de recommander des sujets novateurs et éviter la duplication. Pour concevoir le modèle de recommandation, des algorithmes d'apprentissage automatique comme le TF-IDF seront utilisés pour analyser les sujets de mémoire et créer un profil sémantique détaillé. Ce modèle, intégré dans une plateforme web développée avec Django, permettra aux étudiants de soumettre leurs propositions de sujets, qui seront ensuite analysées en temps réel pour fournir des recommandations personnalisées et éviter les redondances. Notons que cette plateforme web sera aussi un système de gestion de suivi d’évaluation des travaux de mémoires des étudiants par leurs superviseurs.

Selon défini dans le manuel de rédaction de mémoire à l’UCBC, Le choix du sujet de mémoire doit se baser sur un problème ou un défi pertinent pour la communauté, et l'étudiant est libre de sélectionner un sujet qui l'intéresse, à condition qu'il soit en rapport avec ses compétences et connaissances acquises. Le sujet doit être précis et concis, et l'influence des enseignants sur le choix de sujet doit être évitée pour ne pas perturber la réflexion individuelle. Si l’étudiant ne parvient pas à choisir un sujet, le conseil facultaire pourra lui en proposer plusieurs.

Les conditions de rétention d’un sujet incluent l’originalité et une longueur maximale des mots, avec la validation par les membres de la faculté qui organisent une séance de défense des propositions de ces sujets de mémoire. Pour ce faire, chaque étudiant doit présenter 1 à 3 sujets avec description de la problématique, l’hypothèse de travail, l’objectif du travail, la méthodologie de travail, le plan provisoire de la recherche.

Après validation d’un sujet, les changements de sujet peuvent être demandés si le sujet est déjà traité, non authentique, si le directeur ou l’encadreur manque, ou si les ressources sont inaccessibles.

Tout cela se fait manuellement, ce qui présente plusieurs inconvénients, tels que le risque de perte de temps pour rechercher le sujet de mémoire lui-même, les travaux et les informations associées, et la possibilité de redondances dans les sujets à traiter. De plus, ce système est encombré par le nombre croissant de finalistes et souffre d'un manque de suivi des travaux par les facultés.

Pour arriver à bout, tant soit peu, à tout cela au niveau des facultés, nous avons conçu le système SGT (système de gestion des travaux), un système à 5 types d’utilisateurs principaux offrant plusieurs fonctionnalités.

## 2.1.1. Présentation des utilisateurs et acteurs du SGT

1. **La faculté**

La faculté ici représente un membre du conseil facultaire qui a la capacité de réaliser plusieurs actions essentielles. Elle peut enregistrer les enseignants (encadreur et directeurs), les étudiants. Elle est également responsable de l'affectation des directeurs, encadreurs et des décisions aux différents sujets proposés. De plus, elle peut consulter les travaux déposés et valider ou invalider les corrections apportées, tout en planifiant les délais de dépôt des sujets et des travaux. La faculté notifie les directeurs et encadreurs par email concernant l'évolution des travaux. Elle peut également imprimer les fiches d’évaluation des travaux de recherche par département ou année académique, et envoyer des messages de diffusion à tous les étudiants finalistes. Enfin, la faculté permet de rechercher, par mot-clé et par département, les sujets des étudiants, de consulter les détails y relatifs, ou encore de télécharger les travaux déposés.

1. **L’étudiant**

L’étudiant peut se connecter au système pour proposer et/ou déposer son sujet et le reformuler si nécessaire après que le système lui aura présenté des travaux se reprochant de ce qu’il aura proposé et lui recommander d’autres travaux. Il pourra aussi changer son mot de passe, communiquer avec son directeur et encadreur, échanger des documents, recevoir des notifications par mail, déposer son travail ou les corrections des lecteurs, et rechercher et télécharger les travaux

1. **Le directeur**

Le directeur ou encadreur doit être un enseignant permanent ou visiteur de l’UCBC, compétent dans le domaine de recherche de l’étudiant, et posséder la disponibilité et les qualifications nécessaires pour diriger un travail scientifique. Il doit également avoir au minimum un grade de chef de travaux ou un diplôme de Master, DEA, ou DES (Bunduki Phd Honoré et al., 2017). Il dispose de fonctionnalités pour retrouver tous les travaux qu’il dirige, communiquer et envoyer des corrections aux étudiants, recevoir des notifications par email des messages reçus, envoyer un message de diffusion à tous les étudiants ou par promotion, consulter l’historique des messages du groupe directeur-encadreur-étudiant, envoyer ou annuler le feu vert, et rechercher et télécharger les travaux.

1. **L’encadreur**

L’encadreur peut, de la même manière, retrouver tous les travaux qu’il encadre, recevoir des notifications par email, envoyer des messages de diffusion à tous les étudiants ou par promotion, communiquer et envoyer des corrections aux étudiants, consulter l’historique des messages du groupe directeur-encadreur-étudiant, et rechercher et télécharger les travaux.

1. **L’administrateur**

L’administrateur du système a la possibilité de gérer tous les utilisateurs, c'est-à-dire créer, éditer, rechercher ou supprimer un utilisateur, envoyer les paramètres de connexion à un utilisateur, gérer les facultés et les départements, et rechercher et télécharger les travaux. Le secrétariat général à la recherche jouissant des privilèges administrateur, peut suivre tous les travaux des étudiants, accéder aux fonctions de directeur, aux fonctions de la faculté, ainsi qu’aux fonctions de l’administrateur.

Mais au-delà de ces 5 utilisateurs principaux du système nous ne manquerons pas de citer deux acteurs aussi importants : un Mail API qui permet d’envoyer les mails pour la notification et un Plagiat API un service externe qui permet la détection de plagiat.

## 2.1.1. Diagrammes de cas d’utilisation du SGT

Après avoir identifié les acteurs qui régissent notre champ d’étude, nous vous exposons les diverses situations d'utilisation au niveau des acteurs. Pour ce faire nous trouvons interessant d’introduire UML (Unified***M****odeling****L****anguage*) qui est un standard largement utilisé pour la conception et la visualisation de systèmes logiciels. Il permet de représenter les différents aspects d'un système grâce à une série de diagrammes, parmi lesquels on retrouve des diagrammes structurels (comme les diagrammes de classes ou d'objets) et comportementaux (comme les diagrammes de cas d'utilisation ou de séquence). Ces diagrammes offrent une vue abstraite qui aide à communiquer efficacement la conception entre développeurs, analystes et autres parties prenantes (Koç et al., 2021). L'utilisation d'UML facilite non seulement la documentation du système, mais aussi l'identification des besoins et des potentiels problèmes dans la phase de développement logiciel.

1. **Diagramme de cas d’utilisation d’un étudiant**

Ici nous représentons les besoins fonctionnels d’un étudiant :

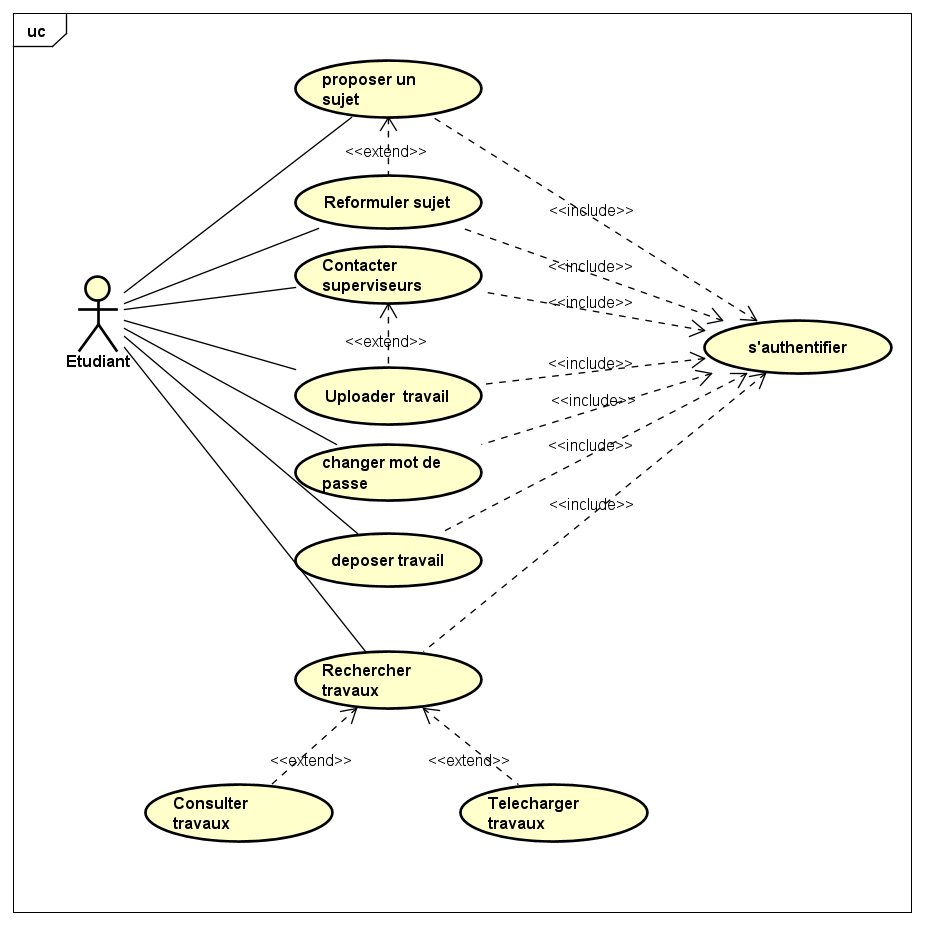
****

Figure 2.1. Diagramme de cas d'utilisation d'un étudiant

Il est démontré que tout utilisateur authentifié avec les droits d'un étudiant peut proposer un sujet, le reformuler après recommandation du système ; il peut contacter ses superviseurs, changer son mot de passe et rechercher les travaux. Certaines fonctionnalités en impliquent d'autres, ce qui explique l'utilisation des termes « extend » et « include » ; par exemple, la recherche de travaux inclut leur consultation ou leur téléchargement. De plus, l'utilisateur externe Mail API permet d'envoyer un email au moment où l'étudiant soumet son sujet.

1. **Diagramme de cas d’utilisation de la faculté, directeur et encadreur**

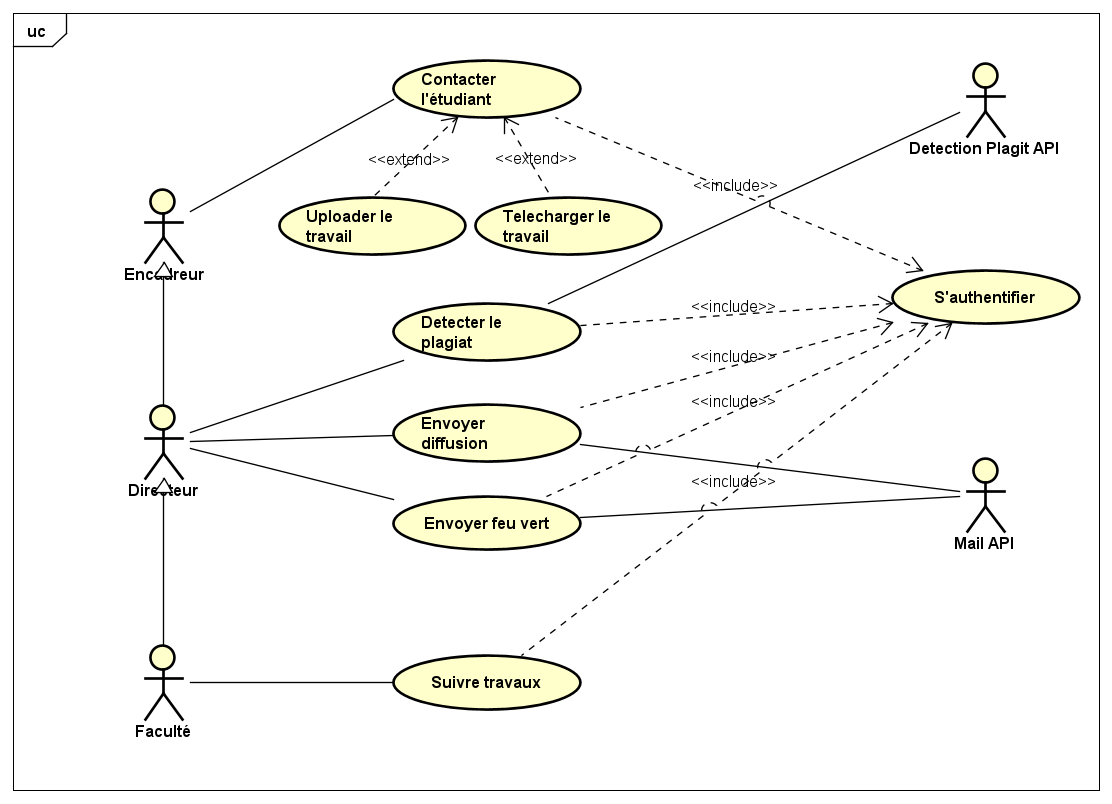
****

Figure 2.2. Diagramme de cas d’utilisation du doyen, directeur, encadreur et lecteur

Un encadreur peut contacter un étudiant, envoyer un message de diffusion, changer de mot de passe et rechercher des travaux. Un directeur étant un encadreur, en plus de fonctionnalités liées à celui-ci, il peut envoyer un feu vert à l’étudiant (sauf dans le cas de la où le finaliste n’a qu’un encadreur). Ils peuvent aussi détecter le plagiat.

1. **Diagramme de cas d’utilisation de la faculté**

Tout utilisateur authentifié avec les droits d’un membre du conseil de la faculté peut gérer les sujets proposés, gérer les travaux déposés, planifier le délai de dépôt de sujets ou travaux et envoyer les notifications.

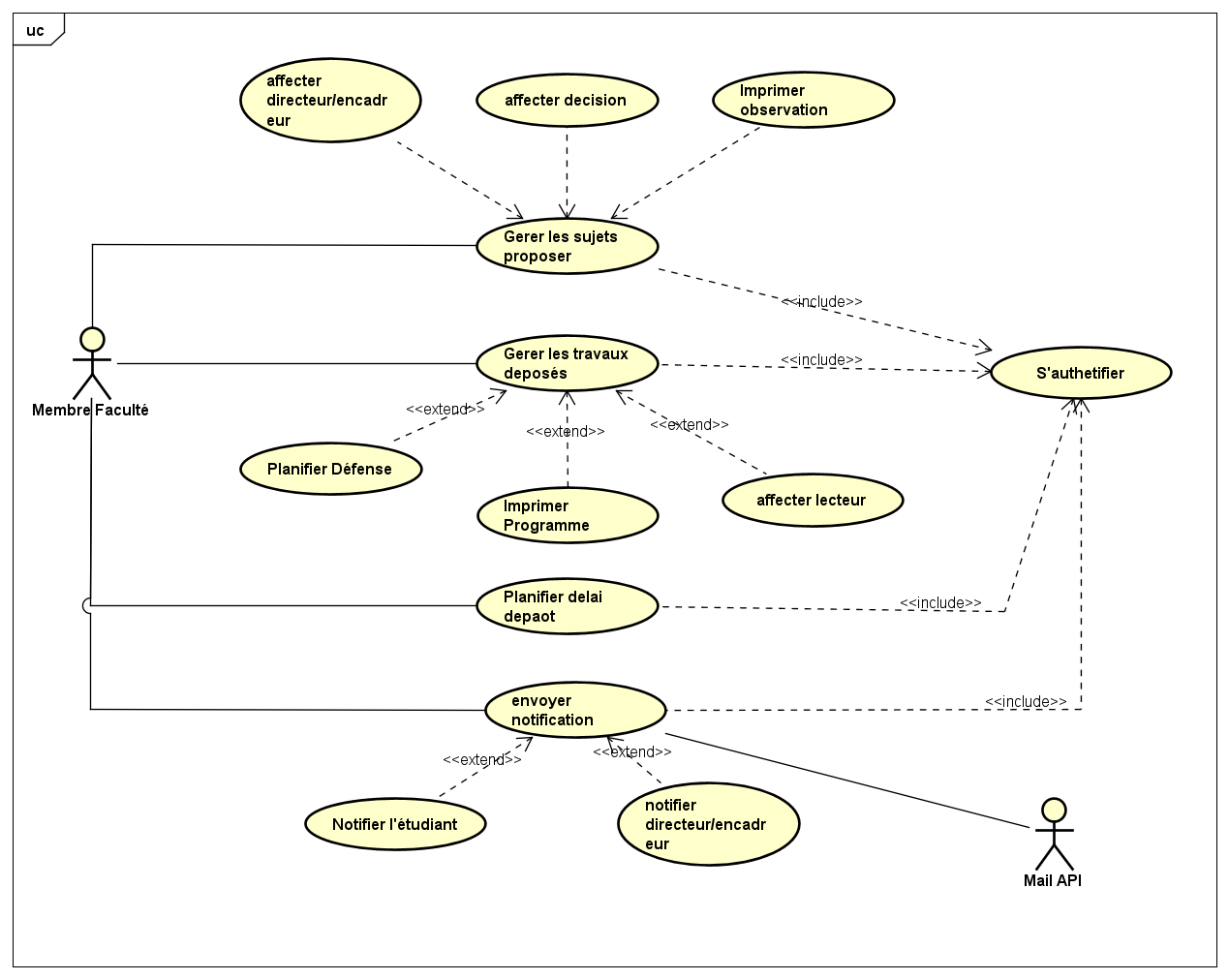
****

Figure 2.3. Diagramme de cas d'utilisation de la faculté

1. **Diagramme de cas d’utilisation de l’administrateur**

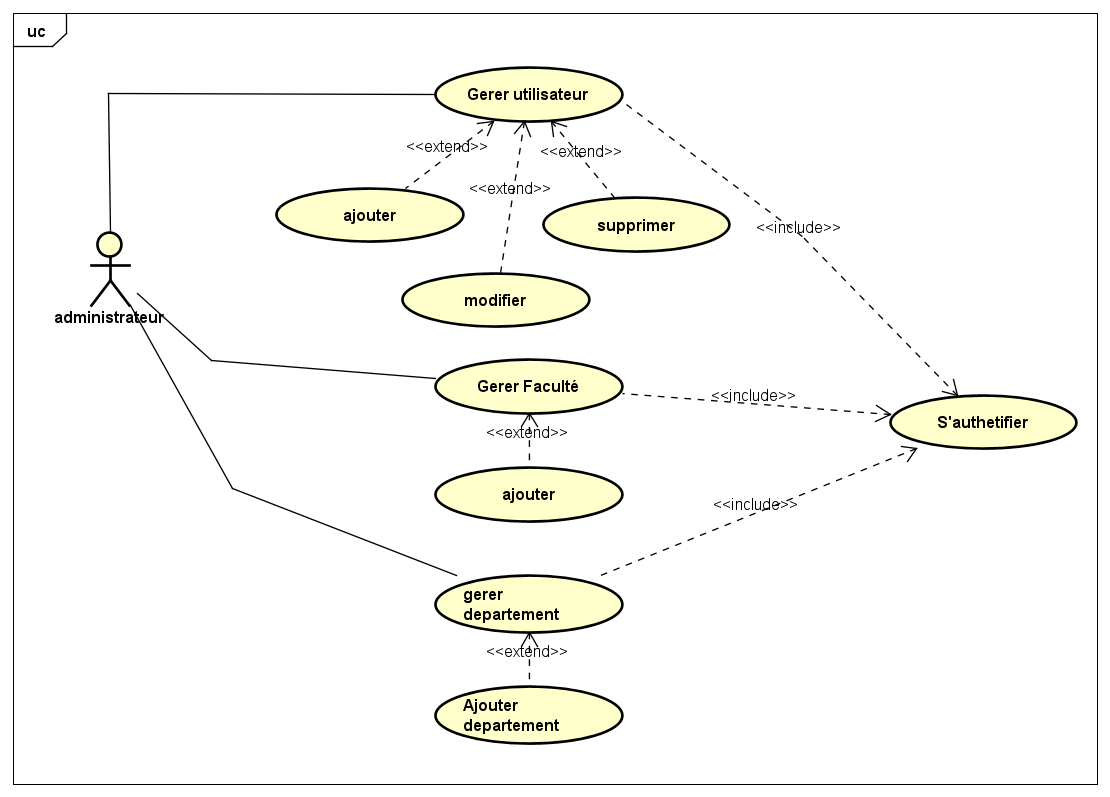
****

Figure 2.4 Diagramme de cas d'utilisation d'un administrateur

Les utilisateurs, les facultés et les départements sont gérés par l'administrateur.

## 2.1.2. Diagrammes de séquences

Notre système a une page d’accueil qui présenter les mémoires déjà réalisés au sein de l’université. Il ne demande pas d’autorisation pour etre vue. En effet, pour profiter des fonctionnalités recherche de sujet de mémoire il faut se connecter voilà pourquoi l'utilisateur sera invité à remplir le formulaire, et si les conditions sont remplies, le système des privilèges au système. Les figures suivantes illustrent le déroulement chronologique des interactions entre les utilisateurs et le système.

### Diagramme de Séquence s’authentifier

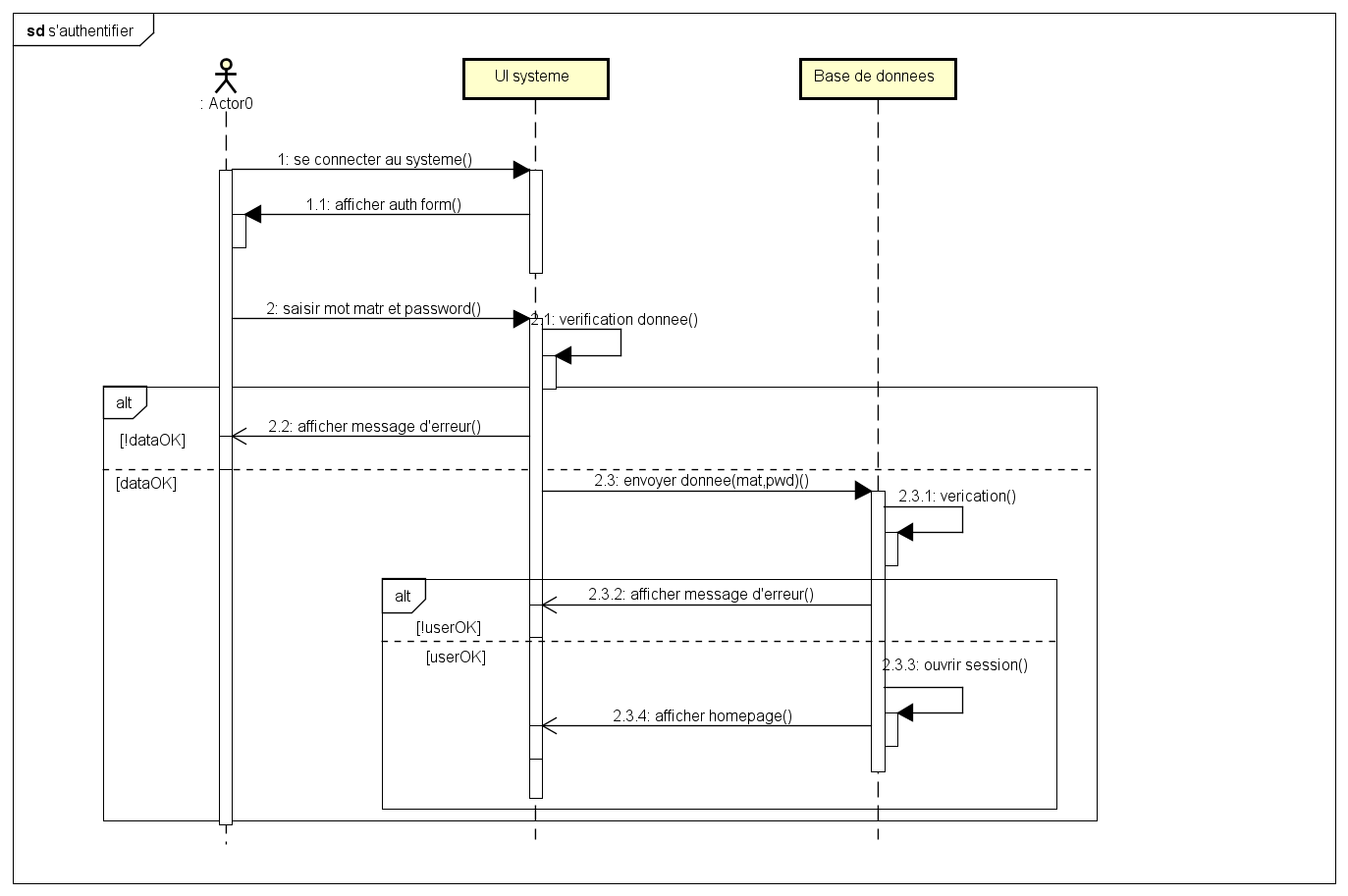


Figure 2.5 Diagramme de séquence S'authentifier

### Diagramme de séquence pour étudiant

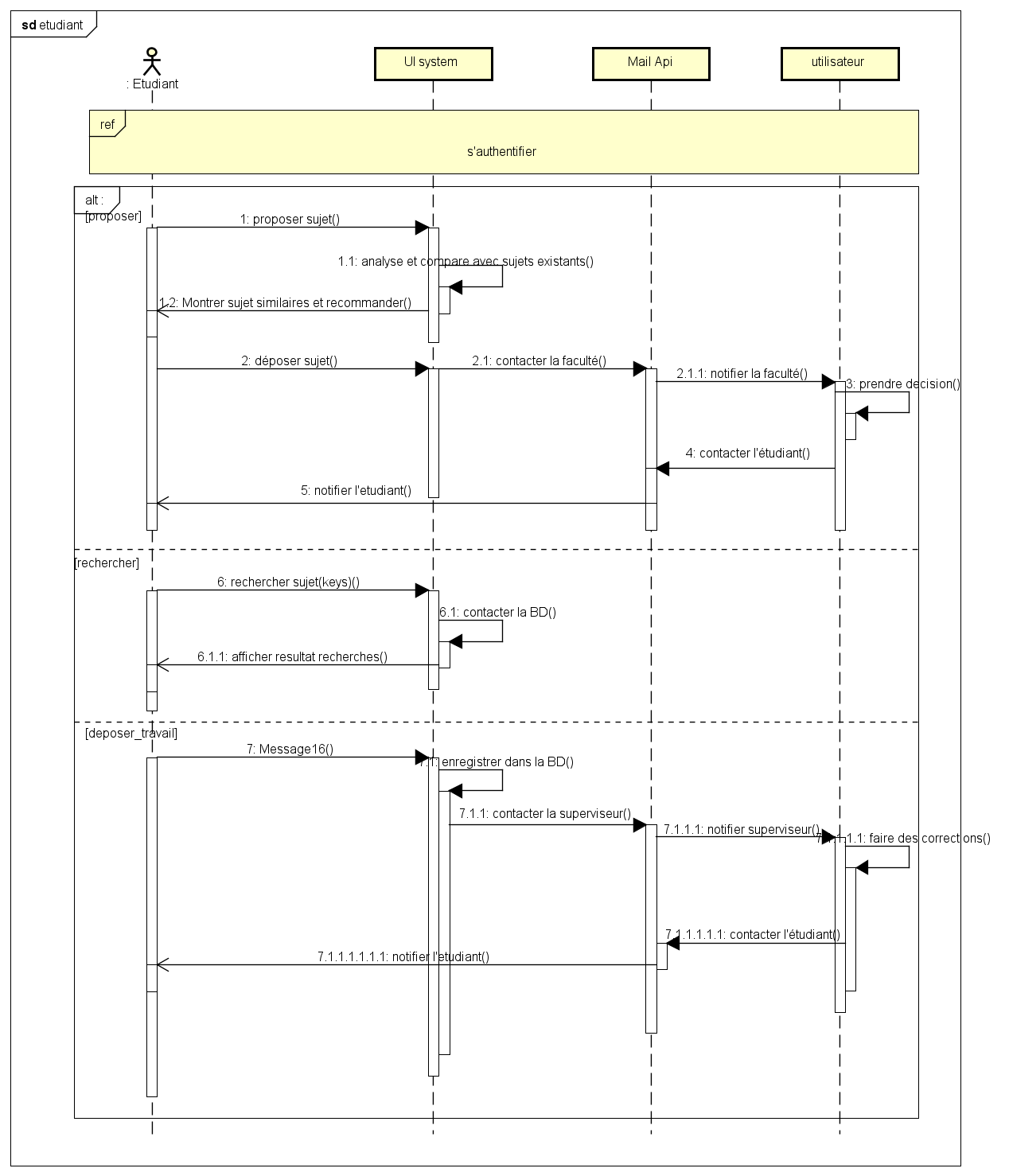
Ici nous avons les cas d’utilisateur de l’étudiant selon l’ordre des séquences

Figure 2.6 Diagramme de séquence Etudiant

### Diagrammes de séquence superviseurs

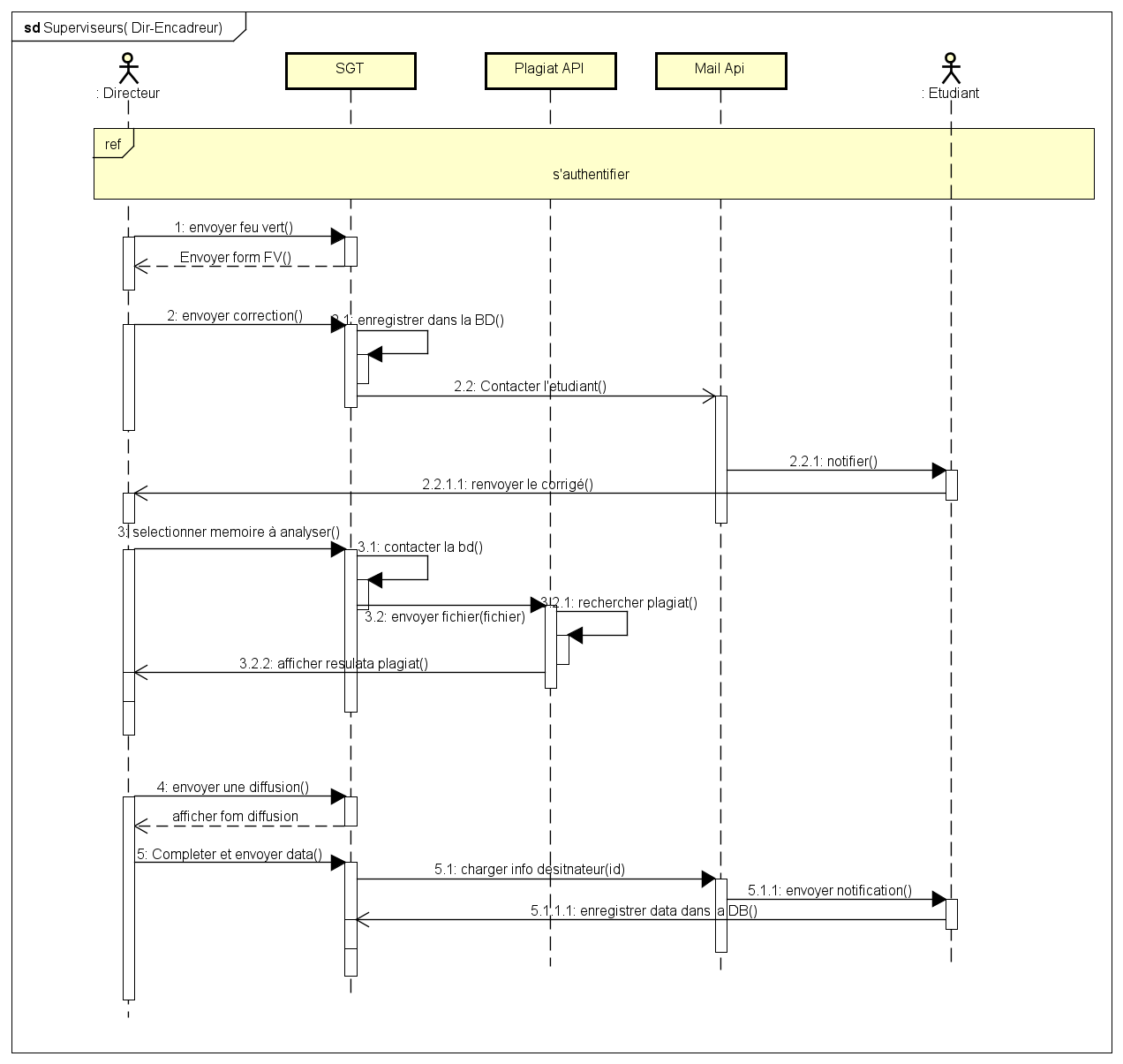


Figure 2.7 Diagramme de séquence directeur-encadreur

### Diagramme de séquence faculté

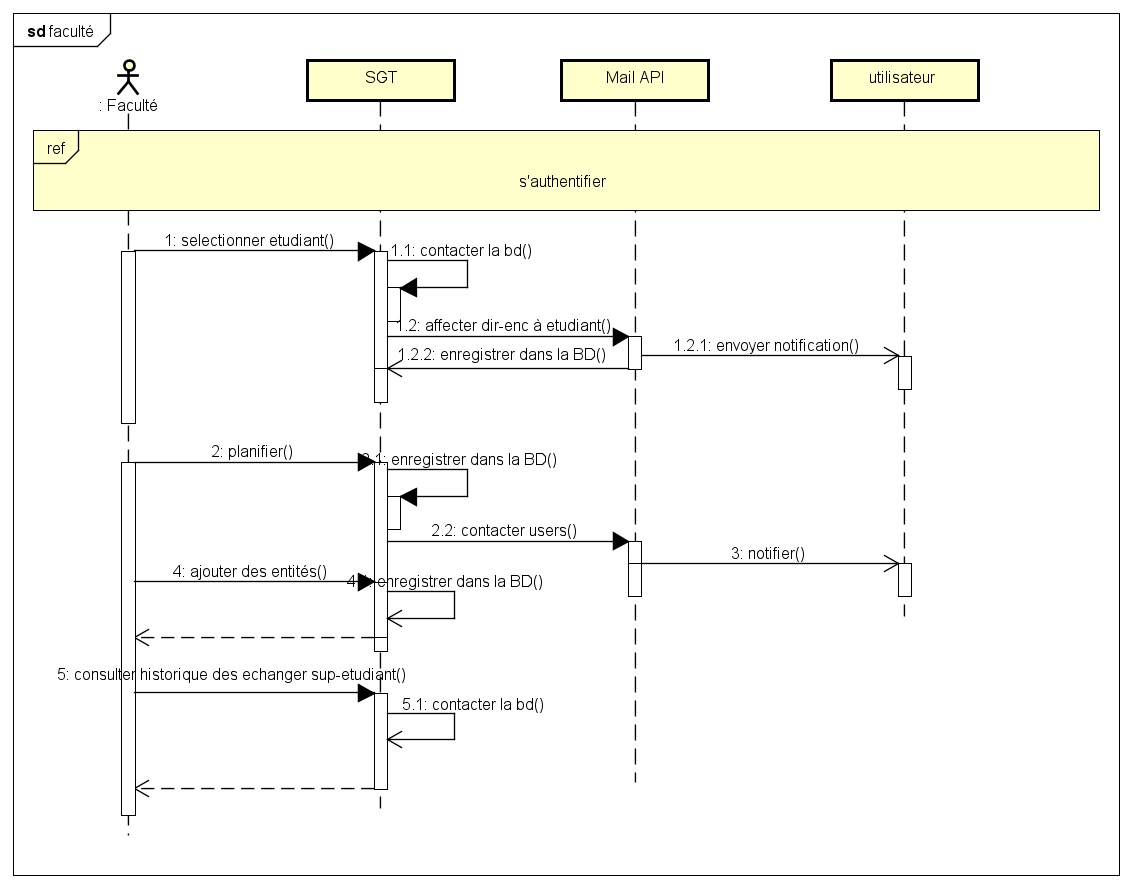


Figure 2.8 Diagramme de séquence faculté

### Diagramme de séquences administrateur

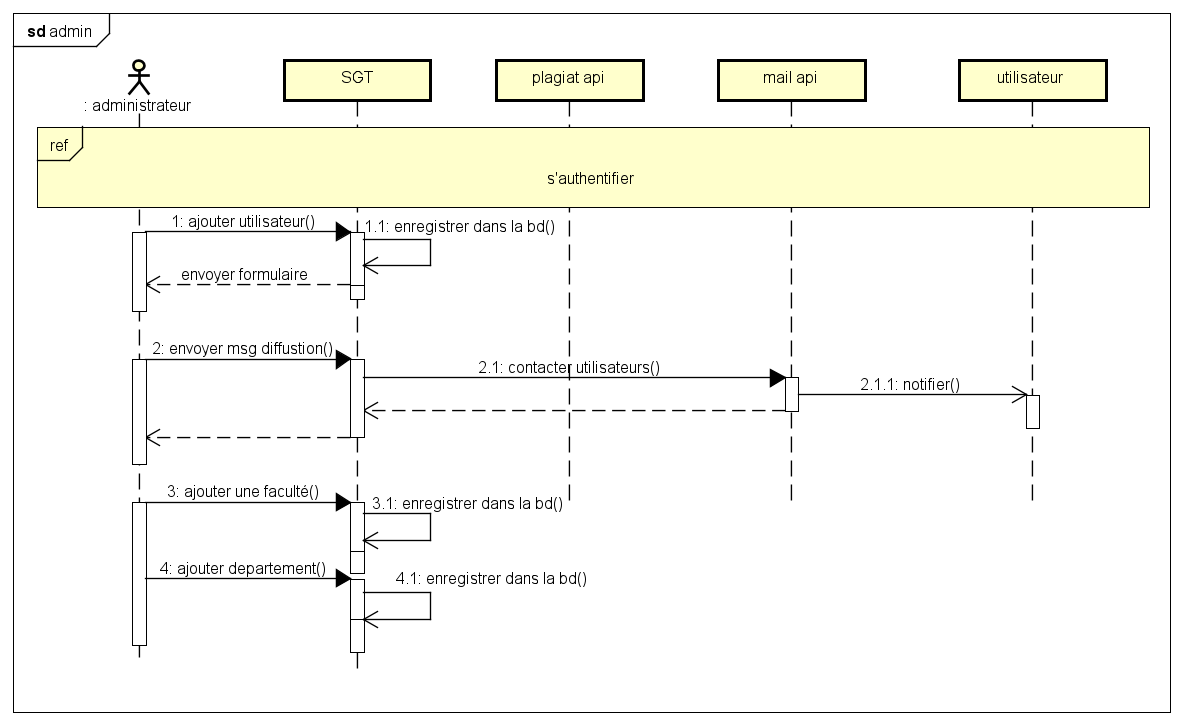


Figure 2.9 Diagramme de séquences administrateur

## 2.1.3. Diagrammes des classes

Le diagramme de classes est souvent considéré comme l'élément central de la modélisation orientée objet car il représente la structure statique du système. Il montre les classes du système, leurs attributs, leurs méthodes ainsi que les relations entre elles, ce qui permet de comprendre les interactions internes du logiciel. Cette vue globale est essentielle pour concevoir et organiser la logique du code, d'où son importance primordiale dans toute modélisation orientée objet. En tant que diagramme obligatoire, il sert de base pour d'autres diagrammes UML comme ceux de séquence ou de cas d'utilisation, qui détaillent davantage les interactions dynamiques (Al-Fedaghi, 2021).

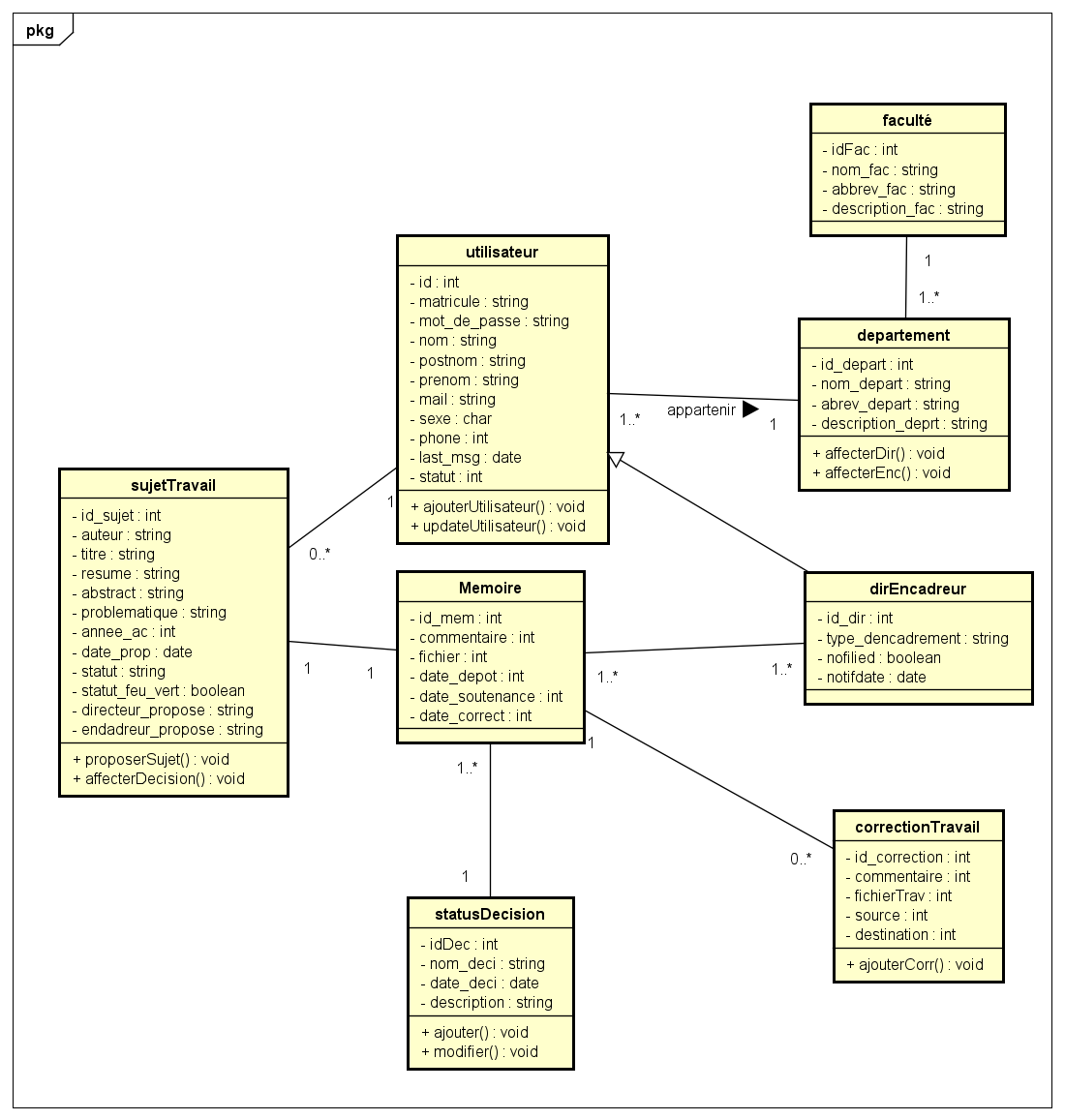


Figure 2.10 Diagramme des classes du système

## 2.1.4. Diagrammes de déploiement

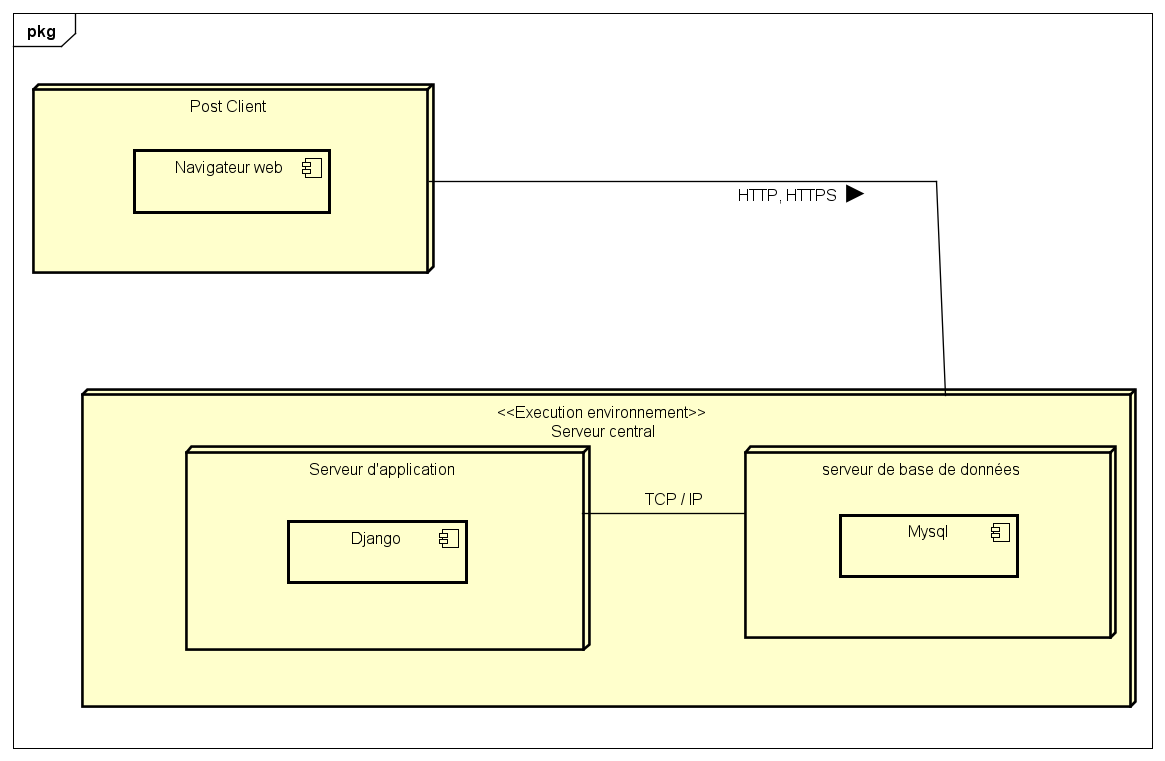


Figure 2.11 Digramme de déploiement

# 2.2. Matériels et méthodes utilisés

Dans cette section, nous présentons les matériels et méthodes employés pour concevoir et implémenter notre système de gestion des travaux de mémoire et de recommandation basé sur le contenue. Pour atteindre nos objectifs, nous avons utilisé une combinaison d'outils de développement, d'algorithmes d’apprentissage automatique et de services externes. La plateforme a été réalisée à l’aide de **Django**, un framework web Python, reconnu pour sa robustesse, sa sécurité, et son évolutivité. Django facilite le développement rapide d’applications web complexes tout en offrant des fonctionnalités intégrées comme l’authentification, la gestion des sessions, et l’administration des bases de données. De plus, Django suit l'architecture MTV (Model-Template-View), ce qui permet de séparer clairement la logique de l'application, l'interface utilisateur, et la gestion des données(Chen et al., 2020).

Pour la partie front-end, nous avons utilisé **Bootstrap**, un framework CSS très populaire qui permet de créer facilement des interfaces utilisateurs modernes et responsives. Bootstrap propose une grande variété de composants préconstruits (comme les barres de navigation, les boutons, et les formulaires), ce qui réduit le temps de développement tout en assurant une mise en page cohérente et adaptable à différents types d'écrans. Nous avons également utilisé **HTML5**, le langage de base pour structurer les pages web, permettant de définir la sémantique des éléments comme les en-têtes, les paragraphes, les images et les liens hypertextes. **CSS3** a été employé en complément de Bootstrap pour affiner le style et la présentation des éléments, et ainsi personnaliser l’interface en fonction des besoins spécifiques du projet.

En ce qui concerne la gestion des données, nous avons choisi **MySQL**, un système de gestion de bases de données relationnelles reconnu pour sa rapidité, sa fiabilité, et sa compatibilité avec Django via l’ORM (Object-Relational Mapping). L'intégration de MySQL dans Django est facilitée par le puissant ORM de ce dernier, qui permet d’interagir avec la base de données à travers des objets Python plutôt que par des requêtes SQL classiques. Ainsi, la manipulation des données, comme les opérations CRUD (Create, Read, Update, Delete), devient plus intuitive et sécurisée, tout en réduisant le risque d'injections SQL.

L’architecture globale de la plateforme repose sur cette combinaison d’outils robustes : Django pour la gestion des fonctionnalités back-end et la connexion avec la base de données MySQL, Bootstrap et HTML5/CSS3 pour garantir une interface utilisateur moderne et responsive. Cette infrastructure nous a permis de développer un système évolutif, sécurisé et optimisé pour répondre aux besoins de gestion des travaux de mémoire à l'UCBC.

## 2.2.1. Description des mémoires analysés

Les mémoires analysés proviennent principalement de la Faculté de Technologie et Sciences de l'Ingénieur (FTSI) de l’Université Chrétienne Bilingue du Congo (UCBC). Nous avons concentré notre étude sur les résumés des mémoires, qui fournissent une vue synthétique et pertinente des travaux réalisés. Les données ont été récoltées à travers plusieurs méthodes : par OCR pour les documents physiques, via Google Forms pour les soumissions électroniques, et à partir de CD contenant les fichiers des mémoires déposés par les étudiants, conformément à la procédure de dépôt du document électronique.

## 2.2.2. Méthode appliquée pour détecter le plagiat

Afin de détecter les plagiats dans le travaux soumis, nous avons envisagé l'intégration d'un système de détection de plagiat en utilisant l’API **Plagscan by Turnitin**, une solution réputée pour la vérification des similarités dans les textes académiques. Toutefois, au stade actuel de développement, nous n’avons pas encore intégré cette API. Nous avons principalement concentré nos efforts sur l'analyse des résumés des mémoires à l'aide de l'algorithme **cosine-similarity** et des **sentence-transformers** pour mesurer la similarité entre les différents résumés. Les sentence-transformers sont des modèles d'apprentissage automatique basés sur TensorFlow, tels que ceux disponibles dans la bibliothèque **tensorflow-hub**, qui permettent de convertir des phrases ou des textes en vecteurs numériques denses. Ces vecteurs capturent les relations sémantiques entre les phrases, facilitant ainsi la mesure de similarités entre différents textes.

Il est à noter que l’Université Chrétienne Bilingue du Congo (UCBC) prévoit d'intégrer **Turnitin** dans un avenir proche. Si cette intégration est réalisée, notre système pourrait alors utiliser l’API **Turnitin** pour effectuer une vérification approfondie de l’intégralité des documents soumis. En outre, bien que notre approche actuelle se concentre sur les résumés, il est possible d'étendre l'analyse aux titres, hypothèses, et problématiques des mémoires pour une évaluation plus complète des similarités.

## 2.2.3. Le prototypage

Pour le développement de notre plateforme nous avons utilisé Bootstrap pour créer une interface utilisateur claire, intuitive et responsive, assurant une expérience homogène sur différents types d’écrans. Cette phase de prototypage nous a permis de tester et d'affiner les fonctionnalités du système, telles que le fonctionnalité de rechercher les sujets similaires par résumé, la soumission de nouveaux sujets de mémoire, la recherche parmi les mémoires existants, et la consultation des recommandations générées par l’algorithme etc. Django a été le socle central de notre développement, assurant la gestion des données et l’intégration des diverses API et algorithmes, tout en facilitant la mise en œuvre des fonctionnalités back-end.

# 2.3. Processus d’analyse et de traitement

Le processus d’analyse et de traitement des données textuelles constitue une étape essentielle de notre approche. Nous avons utilisé des outils avancés de traitement automatique du langage naturel (NLP) pour convertir les résumés des mémoires en vecteurs numériques. Cette transformation a été réalisée à l'aide des sentence-transformers, avec les librairies numpy, pandas etc. facilitant le traitement et la gestion des données. Pour mesurer les similarités entre ces vecteurs, nous avons appliqué l'algorithme de cosine-similarity, ce qui nous a permis d'identifier les thèmes redondants efficacement.

Parallèlement, afin de recommander de nouveaux sujets innovants, nous avons développé une méthode

# 2.4. Conclusion partielle

Ce chapitre a détaillé les méthodologies et la conception de notre système. Nous avons d'abord présenté les divers diagrammes conçus avec le langage de modélisation UML pour illustrer de manière abstraite notre système. Parmi ces diagrammes, nous avons inclus le diagramme de cas d'utilisation, les diagrammes de séquences, le diagramme de classes, le diagramme déploiement. Ces représentations permettent de visualiser les différentes composantes et interactions au sein de notre système, offrant une vue d'ensemble de sa structure et de son fonctionnement.

Ensuite, nous avons présenté les outils et techniques employés, en mettant l'accent sur l'utilisation de Django pour le développement back-end, Bootstrap pour la conception de l'interface utilisateur, et MySQL pour la gestion des données. Nous avons également abordé le prototypage, qui a permis d'itérer rapidement sur les fonctionnalités du système.

Et enfin nous avons décrit les différentes méthodes de collecte des données, en nous concentrant principalement sur les mémoires de la Faculté de Technologie et Sciences de l'Ingénieur (FTSI), récoltées via OCR, Google Forms, et CD. Nous avons détaillé l'approche pour l'analyse des résumés de mémoire, en utilisant l'algorithme cosine-similarity et les sentence-transformers pour mesurer la similarité des textes.

TROISIEME CHAPITRE *:* IMPLÉMENTATION ET PRÉSENTATION DU SYSTÈME

# 3.0. Introduction

Après la conception théorique de notre système dans le chapitre précédent, dans ce chapitre, nous abordons principalement la partie effective du système. Nous présentons ici l'architecture du système, certaines difficultés rencontrées et, enfin, présentons et discutons les résultats obtenus par rapport aux hypothèses énoncées en introduction.

# Références

*Accueil · Dante · Dante*. (s. d.). Consulté 14 août 2024, à l’adresse https://dante.univ-tlse2.fr/s/fr/page/accueil

Aggarwal, C. C. (2016). *Recommender Systems*. Springer International Publishing. https://doi.org/10.1007/978-3-319-29659-3

Al-Fedaghi, S. (2021). *Classes in Object-Oriented Modeling (UML) : Further Understanding and Abstraction*. https://doi.org/10.48550/ARXIV.2106.00267

Alpasan, B. G. B., Mary Jane V. Esimos, Sharra Mae B. Fernandez, Audrey Karyl P., & Sheila Mae S. (2021). E-Document Archiving System with SMS Support. *International Journal of Multidisciplinary Research and Analysis*, *04*(12). https://doi.org/10.47191/ijmra/v4-i12-24

Alvi, F., Stevenson, M., & Clough, P. (2021). Paraphrase type identification for plagiarism detection using contexts and word embeddings. *International Journal of Educational Technology in Higher Education*, *18*(1), 42. https://doi.org/10.1186/s41239-021-00277-8

Béchet, N. (2012). État de l’art sur les Systèmes de Recommandation. *Projet AxIS de l’INRIA, dans le cadre du projet Addictrip*. http://people.irisa.fr/Nicolas.Bechet/Publications/EtatArt.pdf

Bunduki Phd Honoré, CT BORA UZIMA Innocent, & CT KATIMIKA SEREKA VAVASI Felix. (2017, décembre 14). *Comment réaliser un travail scientifique de qualité à l’UCBC : De la rédaction à la soutenance*.

Cheddak, A., Ait Baha, T., Es-Saady, Y., El Hajji, M., & Baslam, M. (2024). BERTopic for Enhanced Idea Management and Topic Generation in Brainstorming Sessions. *Information*, *15*(6), 365. https://doi.org/10.3390/info15060365

Chen, S., Ahmmed, S., Lal, K., & Deming, C. (2020). Django Web Development Framework : Powering the Modern Web. *American Journal of Trade and Policy*, *7*(3), 99‑106. https://doi.org/10.18034/ajtp.v7i3.675

Dávila Morán, R. C. (2022). The Perception of Academic Plagiarism in Industrial Engineering Students at a Public University in Lima. *Publications*, *10*(4), 41. https://doi.org/10.3390/publications10040041

Fitriyani, A., Handayani, D., Noeman, A., Mahbub, A., Salkiawati, R., & Fathurrozi, A. (2022). *E-Archive Document Clustering Information System Using K-Means Algorithm*. 1‑5. https://doi.org/10.1109/ICIC56845.2022.10006935

Gonzalez, A., Ortega, F., Perez-Lopez, D., & Alonso, S. (2022). Bias and Unfairness of Collaborative Filtering Based Recommender Systems in MovieLens Dataset. *IEEE Access*, *10*, 68429‑68439. https://doi.org/10.1109/ACCESS.2022.3186719

Han, M., Zhang, X., Yuan, X., Jiang, J., Yun, W., & Gao, C. (2021). A survey on the techniques, applications, and performance of short text semantic similarity. *Concurrency and Computation: Practice and Experience*, *33*(5), e5971. https://doi.org/10.1002/cpe.5971

Hutson, J. (2024). Rethinking Plagiarism in the Era of Generative AI. *Journal of Intelligent Communication*, *4*(1). https://doi.org/10.54963/jic.v4i1.220

Isinkaye, F. O., Folajimi, Y. O., & Ojokoh, B. A. (2015). Recommendation systems : Principles, methods and evaluation. *Egyptian Informatics Journal*, *16*(3), 261‑273. https://doi.org/10.1016/j.eij.2015.06.005

Jena, K. K., Bhoi, S. K., Malik, T. K., Sahoo, K. S., Jhanjhi, N. Z., Bhatia, S., & Amsaad, F. (2022). E-Learning Course Recommender System Using Collaborative Filtering Models. *Electronics*, *12*(1), 157. https://doi.org/10.3390/electronics12010157

Kalota, F. (2024). A Primer on Generative Artificial Intelligence. *Education Sciences*, *14*(2), 172. https://doi.org/10.3390/educsci14020172

Koç, H., Erdoğan, A. M., Barjakly, Y., & Peker, S. (2021). UML Diagrams in Software Engineering Research : A Systematic Literature Review. *The 7th International Management Information Systems Conference*, 13. https://doi.org/10.3390/proceedings2021074013

Liang, K., Zhang, G., Guo, J., & Li, W. (2023). An Actor-Critic Hierarchical Reinforcement Learning Model for Course Recommendation. *Electronics*, *12*(24), 4939. https://doi.org/10.3390/electronics12244939

McIntire, A., Calvert, I., & Ashcraft, J. (2024). Pressure to Plagiarize and the Choice to Cheat : Toward a Pragmatic Reframing of the Ethics of Academic Integrity. *Education Sciences*, *14*(3), 244. https://doi.org/10.3390/educsci14030244

McManus, R., Logan, A., Wilders, D., & Pennycook, C. (2024). “A World of Possibilities” : The Future of Technology in Higher Education, Insights from the COVID-19 Experience. *Education Sciences*, *14*(1), 63. https://doi.org/10.3390/educsci14010063

Mpia, H. N., Mburu, L. W., & Mwendia, S. N. (2023). CoBERT : A Contextual BERT model for recommending employability profiles of information technology students in unstable developing countries. *Engineering Applications of Artificial Intelligence*, *125*, 106728. https://doi.org/10.1016/j.engappai.2023.106728

Pirnau, M., Botezatu, M. A., Priescu, I., Hosszu, A., Tabusca, A., Coculescu, C., & Oncioiu, I. (2024). Content Analysis Using Specific Natural Language Processing Methods for Big Data. *Electronics*, *13*(3), 584. https://doi.org/10.3390/electronics13030584

Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, *40*(3), 56‑58. https://doi.org/10.1145/245108.245121

Rumanovská, Ľ., Lazíková, J., Takáč, I., & Stoličná, Z. (2024). Plagiarism in the Academic Environment. *Societies*, *14*(7), 128. https://doi.org/10.3390/soc14070128

*SGT ULPGL*. (s. d.). SGT ULPGL. Consulté 14 août 2024, à l’adresse http://www.sgt.ulpgl.net/searchSubect.php

Suleman, R. M., & Korkontzelos, I. (2021). Extending latent semantic analysis to manage its syntactic blindness. *Expert Systems with Applications*, *165*, 114130. https://doi.org/10.1016/j.eswa.2020.114130

Yuan, H., & Hernandez, A. A. (2023). User Cold Start Problem in Recommendation Systems : A Systematic Review. *IEEE Access*, *11*, 136958‑136977. https://doi.org/10.1109/ACCESS.2023.3338705