

Université de Carthage

# Rapport de Projet de Fin d'Etudes

Parcours : Génie Logiciel Système d'information 3

## *Intitulé*

Développement de l'application UBooK

## *Réalisé par*

Hadj Sassi Mahdi

## *Encadré par*

Mr Ben Mnee Tarek

# U-Book

*Année Universitaire : 2021-2022*

# Rapport Projet Fin d'étude de projet UBooK

HADJ SASSI Hadj Sassi Mahdi  
Section: GLSI3

Encadré par : Mr Ben Mne Tarek

## Résumé

La découverte de la vérité, aussi appelée la recherche de la véracité, à devenu un des besoin dans le mégadonnée (Big Data) c'est le 4ème V. Nous avons développé une application web s'appelle UBooK pour l'échange des documents éducatifs entre le formateur et l'apprenant, cet échange respecte la véracité d'information en vérifiant les métadonnées de ce documents bien précisemement le dicipline, l'auteur et l'année d'établissement de document. Ces Informations sont vérifiés à travers un modèle de machine learning basé sur un ensemble des arbres de décisions. UBooK est aussi conçu pour le gestion des evenement universitaires, et pour garantir la véracité de déroulement des evenements et leurs inscriptions des participants, nous avons raccouri aussi vers un modèle de véracité utilisant le web scrapping.

# DÉDICACE

Tous les mots ne sauraient exprimer la gratitude, le respect, la reconnaissance... Aussi, c'est tout simplement que je dédie ce modeste travail à tous ceux qui me portent dans leurs cœurs.

À ma chère mère **Rahma Ezzedine** et mon père **Samir Hadj Sassi** : Aucune dédicace ne saurait exprimer mon respect, mon amour éternel et ma considération pour les sacrifices que vous avez consenti pour mon instruction et mon bien être. Je vous remercie pour tout le soutien et l'amour que vous me portez depuis mon enfance et j'espère que votre bénédiction m'accompagne toujours.

Que ce modeste travail soit le fruit de vos innombrables sacrifices, bien que je ne vous en acquitte jamais assez. Puisse Dieu, le Très Haut, vous accorder santé, bonheur et longue vie et faire en sorte que jamais je ne vous déçoive.

À mon frère **Mohamed Hadj Sassi** : Aucune dédicace ne saurait exprimer tout l'amour que j'ai pour vous, votre joie et votre gaieté me comblient de bonheur, merci pour vos soutiens et encouragements. vous laissez une très belle trace dans ma vie, vous êtes une fierté pour moi, je vous aime infiniment mon idole.

À mes amis de toujours :

En souvenir de notre sincère et profonde amitié et des moments agréables que nous avons passés ensemble. Vous veuillez trouver dans ce travail l'expression de mon respect le plus profond et mon affection la plus sincère et plus précisément ma chère copine **Houda Ben Souissi**, merci pour ta spontanéité, tes réflexions, tes opinions et ton objectivité. Merci de toujours me donner ton avis, sans filtre et sans retenue. Aussi mon meilleur ami **Firas Bouali**, et mon binôme **Rami Ben Othman**.

*Hadj Sassi Mahdi*

# REMERCIEMENTS

*C'est un grand plaisir que je réserve cette page en signe de ma sincère gratitude envers tous ceux qui m'ont aidé de près ou de loin au bon déroulement de ce travail. Je voudrais exprimer mes sincères remerciements à :*

- *Mon encadreur à la FSB **Mr Tarek Ben Mnée** qui n'ont ménagé aucun effort pour m'assurer un encadrement de qualité tout au long de mon projet, pour ses conseils, ses critiques, son encouragement, sa disponibilité, sa servabilité ainsi que pour m'avoir accueilli et donné les moyens pour accomplir ce stage dans les meilleures conditions.*
- *Mon enseignant d'entrepreneuriat à la FSB **Mr Tarek Lassoued** et mon référent PEEC qui avec un grand dévouement, a consacré beaucoup de temps à suivre de près l'évolution de mon projet et m'a réservé les conditions saines de travail, pour ses directives constructives et l'aide précieuse qu'il m'a apportée.*
- *Le professeur à FSB **Mr Anis Ben Aïcha** pour le temps accordé à la réalisation de ce projet. Son écoute et ses conseils m'ont permis de cibler mes choix vers un bon modèle d'apprentissage automatique et d'orienter mon travail d'une manière appropriée.*
- *Mes référent PEEC **Mme Sameh Akrich** et **Mr Imad Maatouk** pour leurs encouragements, leurs conseils judicieux et tout le temps consacré à me guider dans ce projet au sein de PEEC.*
- *Mon enseignant de développement répartie à la FSB **Mr Khaled Barbaria** pour sa disponibilité, ainsi que pour son aide précieuse dans la création de ce rapport.*

- *Mon cher ami **Fares Mansouri** pour sa présence, sa créativité et l'aide précieuse et talentueuse qu'il m'a apportée pour former le logo UBooK.*

*Je tiens à remercier tous les professeurs, élèves et clubs avec qui j'ai fait de la Découverte Client et pris leurs commentaires, critiques et avis.*

**Mme Manel Zekri**  
(Enseignante à FST)

**Mr Fadi Kacem**  
(Enseignant à FSB)

**Mr Taieb Hadj Sassi**  
(Enseignant à FLS)

**Mme Faiza Cheallougui**  
(Enseignante à FSB)

**Mr Anouer Mahfoudh**  
(Enseignant à ISSAT Ma-  
teur)

**Mr Karim Fathallah**  
(Enseignant à ISGB)

**Mme Nour Beyrem**  
(Enseignante à FSB)

**Mr Chaouki Aouiti**  
(Enseignant à FSB)

**Mme Maryem Meddah**  
(Enseignante à FSB)

**Mme Rim Mahwachi**  
(Enseignante à FSB)

**Mme Hela Mahersia**  
(Enseignante à FSB)

**Mme Cherifa Nakkach**  
(Enseignante à FSB)

**Mme Soumaya Dahi**  
(Enseignante à FSB)

**Clubs**  
(de FSB, FST, INSAT ...)

**Etudiants**  
(de FSS, SUP'COM...)

*Mon dernier mot s'adresse à tous les membres du jury pour l'honneur qu'ils me font de participer à l'examen de notre travail.*

# TABLE DES MATIÈRES

<b>Introduction Générale</b>	<b>7</b>
Cadre de projet . . . . .	7
Problématique . . . . .	8
Solution Proposé . . . . .	8
Objectifs de UBooK . . . . .	9
Organisation du rapport . . . . .	9
<b>Infrastrucutre logicielle</b>	<b>10</b>
<b>I Etude Préable</b>	<b>11</b>
1 Domaine Métier . . . . .	11
1.i Document Educatif . . . . .	11
1.ii Evenement Universitaire . . . . .	16
2 Etude de l'existant . . . . .	18
2.i Analyse de marché . . . . .	18
2.ii Généralisation . . . . .	23
3 Méthodologie de travail . . . . .	24
<b>II Spécification</b>	<b>27</b>
1 Modélisation . . . . .	27
1.i Capture des besoins . . . . .	27
1.ii Diagramme de Use Case . . . . .	27
2 Scénarios des cas d'utilisations . . . . .	27
2.i Scénario d'Edition les métadonnées d'un document . .	29
2.ii Scénario Télécharger un documnet . . . . .	31
2.iii Scénario d'inscription dans un evenement . . . . .	32
2.iv Scénario de gestion de ses evenements . . . . .	34
2.v Scénario Consulter un evenement . . . . .	37
3 Diagramme de classe analyse . . . . .	38

<b>III Conception</b>	<b>39</b>
1    Diagramme de Package . . . . .	40
2    Diagramme de classe conception . . . . .	41
2.i    Relation UBooKer et Document . . . . .	41
2.ii    Relation UBooKer et Evenement . . . . .	42
3    Diagrammes de séquences objets . . . . .	43
3.i    Diagramme de séquence objet Rechercher un document	43
3.ii    Diagramme de séquence objet Téléverser un document	44
3.iii    Diagramme de séquence objet S'inscrire dans un evenement . . . . .	45
<b>IV Realisation et déveleoppement</b>	<b>46</b>
1    Environnement de travail . . . . .	46
1.i    Les languages de programmations et les Framewroks .	46
1.ii    Les systèmes gestions base de données . . . . .	47
1.iii    Outils Pour la modélisation, spécification et conception	47
1.iv    Editeurs de textes . . . . .	47
1.v    Sytème de contrôle . . . . .	48
1.vi    Systéme d'exploitation . . . . .	48
2    Résultat . . . . .	49
<b>Vers un modèle de véracité</b>	<b>52</b>
<b>V Etat de l'art</b>	<b>53</b>
1    Big Data . . . . .	54
1.i    Véracité . . . . .	54
2    Intelligence Artificiel . . . . .	55
2.i    Scrapping . . . . .	56
2.ii    Indexation . . . . .	58
2.iii    ML & Documents . . . . .	59
2.iv    Scrapping & Evenement . . . . .	59
3    Travaux anicennes . . . . .	60
3.i    Modèles de Recherche sur la véracité . . . . .	60
3.ii    Modèles de Recherche sur la véracité pour les documents educatifs . . . . .	63

<b>VI Modéle de UBooK pour la véracité des Documents</b>	<b>65</b>
1 Présentation du modèle . . . . .	65
2 Pseudo-Algorithmes . . . . .	67
2.i Extraction des métadonnées d'un document . . . . .	67
2.ii Recherche de dicipline d'un document . . . . .	68
2.iii Résultat final . . . . .	71
3 Implémentation . . . . .	71
3.i Environnement de travail . . . . .	72
3.ii Collection des Données . . . . .	72
<b>VII Modéle de UBooK pour la véracité des Evenements</b>	<b>73</b>
1 Présentation du modèle . . . . .	73
2 Pseudo-Algorithmes . . . . .	73
3 Implémentation . . . . .	74
<b>VIII Résultat Et Discussion</b>	<b>75</b>

# INTRODUCTION GÉNÉRALE

Ce document a pour objectif de présenter le travail réalisé lors de mon projet de fin d'études universitaires. Concluant le cursus de Licence Fondamentales en Génie Logiciel Système d'information, ce projet a été effectué de Février à Juin 2022 pour la StartUp UBooK.

Tout d'abord, je définirai l'environnement dans lequel s'est déroulé ce projet en présentant la StartUp « UBooK » et certains de ses clients. Puis nous décrivons brièvement la problématique rencontrée dans ce contexte, ensuite nous proposons la solution proposée en la mettant en valeur et nous conclurons par la présentation du plan général du rapport.

Ce projet est basé sur les matières tous au long le parcours académique :

- Algorithme, structure de données et complexité (ASD, ASDC)
- Programmation Java, Java Avancée, J2EE, Programmation Python
- Développement Web, Angular
- Conception des Systèmes d'Information
- Fondements des bases de données, Ingénierie des Bases de Données, Administration des bases de données
- Intélligence artificiel, Machine learning
- Techniques d'indexation et recherche multimedia
- Framework et technologies Big Data
- Architecture SOA et services web
- Tests des logiciels, UX Design
- Système Exploitation, Virtualisation et Cloud
- Management des projets, Gestion d'entreprise, Entrepreneuriat

## Cadre de projet

L'organisme Pôle Étudiant Entrepreneur de Carthage (PEEC) a annoncé les résultat de la 3éme cohorte pour la sélection des candidats au statut Étudiant Entrepreneur le 04 Mars 2021 à l'Institut national des sciences appliquées et de technologie (INSAT), et dans ce cadre là UBooK et déclaré comme étant une StartUp et une société numérique unipersonnelle Tunisienne en cours de développement.

Le Secteur de UBooK est celle de la vie étudiante. La société offre à ses clients des services de partage et échanges des documents éducatifs, aussi le service de poster et de s'inscrire aux événements universitaires.

## Problématique

Des Statistiques (Customer Discovery, Questionnaires, Recherches) lors de la compétition entrepreneuriale Carthage Innov1 2021, indique qu'il y'a

- Manque des espaces de gérance des événements en Tunisie, et bien précisément dans le cadre universitaires.
- Absence des plateformes et des portails qui offrent un service gratuit pour l'accès aux documents éducatifs.
- Manque de crédibilité et de véracité des documents éducatifs sur internet, qui cause la difficulté de trouver le document pertinent qui répond aux besoins.
- Inexistence d'un système tunisien qui traite les examens blancs et qui attribue des badges de connaissances.

## Solution Proposé

Cet problème a conduit vers l'idée de développer une encyclopédie universitaire éducative où on trouve des différentes ressources d'éducation (Cours, Travaux Dirigés, Travaux Pratiques, Examens, Corrigés) avec la possibilité de passer des examens blancs pour gagner des badges des compétences HardSkills ou SoftSkills, plus d'un centre d'événement (Formations, Certifications, Compétitions, Journées) en ligne.

Et ici d'où viens le nom UBooK :

- U correspond à l'Université
- BooK correspond en anglais à livre, document
- UBooK correspond à les documents universitaires (university books)
- BooK est un verbe aussi en anglais pour la réservation
- UBooK pour nous est (you book up in events) vous réservez dans des événements

# Objectifs de UBooK

- Faciliter la recherche des ressources éducatives
- Améliorer le partage des documents éducatives en respectant les ethics
- Augmenter les interactions événementielle
- Expandre le réseau universitaire

## Organisation du rapport

Afin d'illustrer la démarche de notre travail, nous présentons l'organisation générale du rapport, qui s'articule autour des deux grandes parties, la première c'est le développement de l'infrastructure logiciel qui se déroule sur quatre chapitre :

1. Le premier chapitre, intitulé "**Etude préalable**", expose la présentation des domaines métier et l'établissement d'une étude de processus métier, une présentation des solutions similaires, enfin la présentation de notre méthodologie de travail.
2. Le deuxième chapitre, intitulé "**Spécification**", est consacré à l'analyse des besoins, et extraction des besoins fonctionnel pour le reste de projet
3. Le troisième chapitre, intitulé "**Conception**", pour objectif de faciliter le fonctionnement futur du système, afin d'en faciliter la réalisation, de permettre de formaliser les étapes préliminaires du développement d'un système afin de rendre ce développement plus fidèle aux besoins fonctionnel.
4. Le quatrième chapitre, intitulé "**Réalisation et développement**", est pour présenter l'environnement de travail et pour présenter le résultat finale de l'application logiciel

Une deuxième partie pour le développement d'un modèle machine learning qui fonctionne sur l'application développée dans la partie 1, ce modèle est pour but de vérifier la véracité d'information dans l'application, Cette partie est développé sur 3 chapitres :

1. l'état de l'art, où nous présentons les différents techniques utilisées plus les anciens modèles sur la véracité d'information.
2. Modèle de UBooK pour la véracité des documents, consacré pour présenter le logique et l'algorithmique du modèle choisis
3. Résultat et Discussion, analysons les résultats de modèles implementés, finissons par une conclusion de projet

---

# **INFRASTRUCTURE LOGICIELLE**

## **INTRODUCTION**

Dans cette partie, nous nous intéressons sur le développement de l'infrastructure logicielle de l'application UBooK, pour le préparer à un modèle de recherche sur la véracité d'information pour les documents éducatifs et les événements universitaires. Nous allons commencer par l'étude préalable où nous présentons le domaine métier et le marché, puis nous mettons la vision sur l'étude de l'existant où nous positionnerons dans le marché, ensuite nous spécifions les besoins pour faire la conception et terminons par la réalisation de l'infrastructure.

---

# CHAPITRE I

## ETUDE PRÉABLE

Ce Chapitre est déstiné à la présentation du secteur et le marché de l'application, alors nous présentons le domaine métier plus on mettons le points sur les différents volet que l'application traite, enfin nous donnons l'étude de l'existant pour se positionner dans le marché.

### 1 Domaine Métier

les services offert par UBooK sont le partage et l'échange des documents éducatifs entre les différentes membres de l'universités, aussi qu'un centre des evenements en ligne ou les porteur des evenements peuvent poster leurs actions en publique pour y avoir des inscriptions de la part de leurs cibles participants. Ce secteur d'activité présentes des limites qu'on doit respecter et ne pas y dépasser, et dans le cas d'infraction il y'aura des pénalités.

#### 1.i Document Educatif

##### Définitions d'un document

Le nom document vient du verbe latin docere qui signifie " instruire ". Par voie de conséquence, on peut considérer qu'un document est une " chose " qui peut servir à renseigner, à prouver. On utilise le but pour construire la définition. On peut aussi définir la notion de document en s'appuyant sur ses composantes. À ce moment-là, un document est un ensemble d'informations porteur de sens pour un auditoire ciblé[1].

Un document renvoie à un ensemble formé par un support et une information le contenu, celle-ci enregistrée de manière persistante. Il a une valeur explicative, descriptive ou de preuve. Vecteur matériel de la pensée humaine, il joue un rôle essentiel dans la plupart des sociétés contemporaines, tant pour le fonctionnement de leurs administrations que dans l'élaboration de leurs savoirs. Témoin de son époque pour l'historien, pièce à conviction pour le juge, le document pose toujours le problème de sa véracité, mais plus encore de ce qu'il révèle indépendamment de son énoncé ou de son illustration[2].

Le document est un ensemble constitué d'un support matériel sur lequel sont inscrites, par le recours à différents codes de transcription, des informations, structurées et organisées, tant du point de vue intellectuel que de la forme. Il résulte d'une intention de communication. Le document formalise et fixe, de façon plus ou moins stable, des informations dont il permet la transmission, le stockage, la reproduction et le traitement. En tant qu'objet documentaire, il

est le produit d'une création par un ou plusieurs auteur(s). Il est référençable et indexable [3].

D'une manière générale, le document est envisagé comme un ensemble formé par un support et une information qui peut être lu par l'homme ou la machine.

## Définitions d'un document éducatif

Un document représente le support matériel. Il est considéré comme : " un véhicule d'information intelligible, riche dans le fond comme dans la forme. Il est tout à la fois le message, lui-même, sa présentation et sa forme ainsi que son propre véhicule ". Le document qui nous intéresse est le document électronique qui est défini comme " un ensemble cohérent d'objets numériques (texte, graphique, photo, images animées et sons) stockés dans des machines informatiques interconnectées, ou stockés sur des supports informatiques amovibles. Pour le lire, il est nécessaire, soit de l'imprimer sur du papier, soit de le visualiser sur un écran " Un document pédagogique est une instanciation de l'objet pédagogique, ou LO (Learning Object). Il peut être défini comme " toute entité, sur un support numérique ou non, pouvant être utilisée pour l'apprentissage, l'enseignement ou la formation".

Un objet pédagogique peut être réutilisé pour différentes fins. Par exemple, un exercice peut bien servir dans une série de TD (Travaux Dirigé) que dans le cadre d'un examen[4].

Les documents éducatifs s'appellent aussi ressources éducatives libres REL. Les ressources éducatives libres (REL) sont des matériaux d'enseignement, d'apprentissage ou de recherche appartenant au domaine public ou publiés avec une licence de propriété intellectuelle permettant leur utilisation, adaptation et distribution à titre gratuit. Selon l'UNESCO, l'accès universel à une éducation de qualité contribue à la paix, à un développement social et économique durable et au dialogue interculturel. Les REL constituent une opportunité stratégique d'améliorer la qualité de l'éducation et de renforcer le dialogue politique, le partage des connaissances et le renforcement des capacités[5].

## Types de document éducatif

les objets pédagogiques peuvent être, par exemple, des transparents, des notes de cours, des pages Web, des logiciels de simulation, des programmes d'enseignement, des objectifs pédagogiques, etc[6].

Dans ce rapport on va se limiter aux document universitaire qui sont destinés aux étudiants, et ce sont les suivants I.1 :

- Support de Cours : Le support de cours est un élément indispensable de l'enseignement, qui soutient et illustre le discours de l'enseignant pendant le cours magistral. Il n'est compréhensible

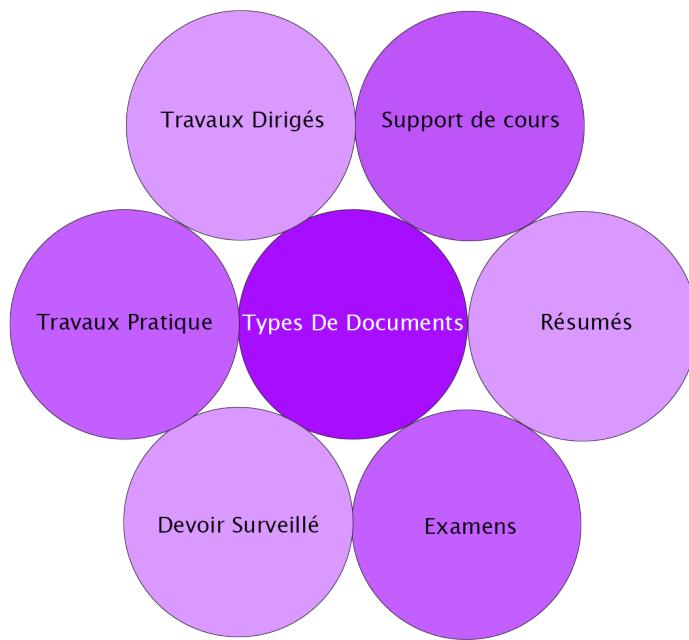


FIGURE I.1 – Liste types des documents éducatives

qu'avec la narration qu'il accompagne[7].

- Travaux Dérigés (TD) : Les TD sont une méthode d'enseignement qui permet aux élèves de mettre en application des connaissances théoriques sous forme d'exercices. Il se déroulent en général en effectifs réduits pour faciliter l'aide du professeur[8].

- Travaux Pratiques (TP) : Les travaux pratiques, souvent abrégés en TP, constituent un type d'enseignement fondé sur l'apprentissage pratique avec en particulier la réalisation d'expériences permettant de vérifier et compléter les connaissances dispensées dans les cours théoriques[9].

- Devoir Survéillés (DS) : Un devoir surveillé, souvent abrégé en DS, est un contrôle de connaissances portant sur un ou plusieurs points abordés au cours de l'année scolaire en cours dans une discipline spécifique. Sa durée varie suivant la difficulté du sujet, de la discipline et du niveau d'étude[10].

- Epreuves et Examens : Un examen est l'action de considérer attentivement avec réflexion un objet. C'est une observation analytique limitée dans l'espace et dans le temps. Un examen est aussi une évaluation orale ou écrite. Son étude fait l'objet de la docimologie. Dans le cadre d'une évaluation certificative, la réussite à un examen peut être consignée par un diplôme. Synonymes : test, contrôle, épreuve, devoir sur table[11].

- Résumés : Un résumé en général est un petit écrit, qui consiste à prendre les points essentiels d'un texte en seulement un ou plusieurs paragraphes[12].

- Corrigé d'un document : La correction d'épreuves est une intervention faite sur un texte dans le but de l'améliorer. Cette étape consiste à vérifier un texte destiné à l'impression,

donc déjà révisé et mis en pages, pour s'assurer qu'il sera exempt de fautes d'impression et de coquilles[13].

## Différent métadonnées des documents éducatifs

Les métadonnées d'un document sont des informations non visuelles contenues dans un document qui fournissent un contexte supplémentaire. Par exemple, l'auteur du document et la date à laquelle il a été créé. Elles peuvent également aider à classer les documents. Par exemple, les utilisateurs peuvent préciser si un document est destiné à un usage interne uniquement ou s'il est accessible au public.

L'ajout de métadonnées à un document aide les organisations à simplifier la recherche et l'extraction de documents. En effet, les outils de recherche peuvent trier les métadonnées des documents beaucoup plus rapidement que le balayage du texte intégral d'un document. En outre, les métadonnées des documents facilitent le tri, l'acheminement, le stockage et le contrôle des documents[14].

Une métadonnée (mot composé du préfixe grec meta, indiquant l'auto-référence ; le mot signifie donc proprement " donnée de/à propos de donnée ") est une donnée servant à définir ou décrire une autre donnée quel que soit son support (papier ou électronique).

Un exemple type est d'associer à une donnée la date à laquelle elle a été produite ou enregistrée, ou à une photo les coordonnées GPS du lieu où elle a été prise.

Les métadonnées sont à la base des techniques du Web sémantique. Elles sont définies dans le cadre du modèle Resource Description Framework (RDF).

Les métadonnées sont, dans le cadre du Web sémantique, des données signifiantes qui permettent de faciliter l'accès au contenu informationnel d'une ressource informatique, une notice de contenu intégrée en quelque sorte (dans l'en-tête des documents HTML côté code source ou en tant que fichier XML autonome par exemple)[15].

Un document numérique est une suite de fichiers : il est décrit par un identifiant unique et un ensemble de métadonnées :

- des métadonnées descriptives pour :
  - donner une description bibliographique approfondie et détaillée dans un format normalisé permettant l'échange de données ;
  - rattacher le document à l'original ou à différentes versions d'un document ;
  - donner accès à la copie numérique.
- des métadonnées de structure pour :
  - rattacher les fichiers d'un même document entre eux ;
  - reconstituer la structure du document : connaître tous les fichiers qui composent un document (fichiers textes, images...) ; connaître la relation physique entre ces fichiers (ordre d'affichage, fichier cible donnant accès à l'ensemble).
- des métadonnées administratives pour :
  - gérer les droits : d'accès (droits d'auteur, confidentialité) et d'usage (droits d'impression, de reproduction, de modification...) ;
  - préserver les informations techniques nécessaires à la lecture des fichiers ;
  - garantir l'intégrité des fichiers et le suivi de leurs éventuelles modifications.

## Les licences Creative Commons

Les licences Creative Commons constituent un ensemble de licences régissant les conditions de réutilisation et de distribution d'œuvres. Élaborées par l'organisation Creative Commons, elles ont été publiées pour la première fois le 16 décembre 2002.

Les licences Creative Commons facilitent l'utilisation d'œuvres et s'adressent aux auteurs qui souhaitent :

Partager et faciliter l'utilisation de leur création par d'autres. Autoriser gratuitement la reproduction et la diffusion (sous conditions). Accorder plus de droits aux utilisateurs en complétant le droit d'auteur qui s'applique par défaut. Faire évoluer une œuvre et enrichir le patrimoine commun. Économiser les coûts de transaction. Légaliser le peer to peer de leurs œuvres (réseau de partage de données poste à poste, chacun jouant tour à tour le rôle de client et de serveur)[16].

Les licences Creative Commons sont un outils pour garantir le partage et l'authentification des œuvres. Avoir la licence est gratuit et simple en quelques clics sur le site[17] Il y'a différent licences et chaque à un sens bien définie. Voir la figure suivante ou il est classés tous les licences avec leurs noms, abréviations, icônes, et les caractéristiques de partage I.2

Nom de la licence	Abréviation	Icône	Attribution requise	Usage Culture remix	Pour usage commercial
"No Rights Reserved"	CC0		Non	Oui	Oui
Attribution	BY		Oui	Oui	Oui
Attribution-ShareAlike	BY-SA		Oui	Oui	Oui
Attribution-NonCommercial	BY-NC		Oui	Oui	✗ Non
Attribution-NonCommercial-ShareAlike	BY-NC-SA		Oui	Oui	✗ Non
Attribution-NoDerivatives	BY-ND		Oui	✗ Non	Oui
Attribution-NonCommercial-NoDerivatives	BY-NC-ND		Oui	✗ Non	✗ Non

FIGURE I.2 – Liste des licences creative commons

## Le secteur de partage des documents

Si une personne procure frauduleusement un faux document à autrui dans le but de lui faire accéder à un droit, une autorisation ou encore pour constater une qualité ou une identité, elle risque 5 ans d'emprisonnement et 75 000 € d'amende (article 441-5 du Code pénal) Si la personne se fait délivrer ce faux document, les sanctions sont ramenées à 2 ans d'emprisonnement et 30 000 € d'amende (article 441-6) [17b]

"L'œuvre audiovisuelle (film, documentaire, émission tv...) est protégée par le droit d'auteur comme un tout, résultant de la combinaison du scénario, des dialogues, de la réalisation, de la musique originale : toute oeuvre audiovisuelle est une oeuvre de collaboration. Ce statut d'oeuvre est posé par la loi et détermine les conditions dans lesquelles les coauteurs

exercent ensemble leur droit patrimonial.” Source : Guide pratique du droit d'auteur : utiliser en toute légalité : textes, photos, films, musiques, Internet + protéger ses créations d'Anne-Laure STERIN, 2ème éd. totalement actualisée. Maxima, 2011. 543 p.

Avant d'analyser un document et d'utiliser les informations qu'il apporte pour répondre à une problématique, il convient de déterminer si le document est fiable (= digne de confiance).

Pour partager un document il faut s'assurer qu'il est conforme et cohérent, aucune altération ou modifacatoin, distorsion du contenu et le plus important c'est de mentionner les sources d'informations, en réspectant les copyrights et les licences creative commons.

## 1.ii Evenement Universitaire

### Définitions d'un evenement universitaire

De manière générale, un événement est un fait important et marquant. Il suscite donc un certain intérêt. Cependant, les événements dont il est question ici sont conçus avec un objectif spécifique pour un public cible bien défini.

Ces événements peuvent avoir différents objectifs. Ils peuvent viser à informer, à permettre l'échange de connaissances, à encourager le réseautage, à recruter du personnel, à motiver ou remercier le personnel existant, ou encore à renforcer l'esprit d'équipe entre les employés[18].

Un événement universitaire est un événement, autre que les cours universitaires prévus dans le cadre du programme, qui se tient dans un bâtiment universitaire ou un espace extérieur sur le campus universitaire[19]

Événement universitaire désigne un événement ou un programme, sur ou hors campus, organisé au nom de l'Université, ou un événement ou un programme qu'une personne raisonnable identifierait comme étant affilié à l'Université[20].

Un evenement alors est un programme organisé au seins de l'environnement universitaire ca peut etre par des clubs, instituts, centre formations dans ou pas l'établissement. L'evenement est pour un objectif bien déterminé selon une thématique, un evenement peut avoir des sponsors pour soutenir l'action.

### Types d'evenement universitaire

Les Types d'événement universitaires sont variés, mais on va s'intéresser sur ces types mentionnés ci dessous I.3

**Journée Universitaire** Tous les établissements, organisations, Clubs organisent des journées portes ouvertes, généralement dès l'hiver, au moment des inscriptions. L'occasion pour vous de voir l'école, ses installations, ses enseignants, ses étudiants, son ambiance[21].

Les journées universitaires sont variés, la plus part ne nécessitent pas d'inscriptions, et la plus part sont gratuites. Une journée universitaire peut etre pour introduire un nouveau



FIGURE I.3 – Les Types des evenement universitaires

club ou déclarer un evenement, ou pour célébrer une fête. Une journée universitaire peut être aussi une conférence, ou un congré selon le domaine d'organisateur.

**Formation** La formation peut se définir d'une manière générale, comme : " l'action d'un formateur s'exerçant sur une ou plusieurs personnes en vue de les adapter technique, physiquement et psychologiquement à leurs futures fonctions. " Il s'agit à la fois d'un apprentissage de connaissances et d'un apprentissage de méthodes de travail et de savoir-faire mais aussi d'une expérimentation de nouvelles attitudes et de nouveaux comportements. Elle permet l'adaptation à l'emploi, le développement du potentiel des individus, le développement intellectuel et rationnel, la croissance des capacités d'adaptation et de régulation de l'individu dans ses rapports avec son environnement professionnel, etc... [21]

Le déroulement des formations c'etait la plus part des temps dans des centre de formations, ou bien dans les instituts universitaires, ou un formateur donne une formation à des etudiants inscrits a cet evenement, la plus part des formations sont gratuites. Mais Dernièrement à cause de la pandémie COVID-19, le déroulement des formations devient en ligne sur les plateforme de meet en ligne, en mentionne comme exemple, Google meet, Microsoft Teams, Zoom, Cisco ...,

**Compétition** Action de chercher à obtenir en même temps que d'autres le même titre, la même charge ou dignité, la même fonction, etc. : La compétition électorale. 2. Action de participer à un championnat, à une coupe, à un tournoi : Faire de la compétition automobile. (Dictionnaire LAROUSSE : Compétition, 1er et 2éme recherche)

Les Compétitions sont des evenements à but de gagner un prix, les compétitions ou en anglais (Contests) se sont des jeux, des concours, des hackathons informatique, marathons mathématiques, robotique ou même dans la littérature et les beaux arts, Il y'a un nombre limite de participants pour se présenter dans l'evenement. Chaque Inscrit a cet compétition

doit payer sont conditature et de respecter les règles de jeux.

**Certification** Selon AFNOR, "la certification est une activité par laquelle un organisme reconnu, indépendant des parties en cause, donne une assurance écrite qu'une organisation, un processus, un service, un produit ou des compétences professionnelles sont conformes à des exigences spécifiées dans un référentiel" [22].

Les Certifications sont des épreuves pour des compétences (HardSkills/SoftSkills), Presque tous les certificatins se déroulent après des formations liés, ou les inscrits dans ces formations peuvent valoriser leur participations par une certificat de connaissance après un passage de l'épreuve. Les certifications peuvent être présentielle sous le garde des inspecteurs ou bien en ligne, comme les certifications informatiques, la sécurité ça sera à travers une session contrôlés qui garantie pas le candidats de sortir et de tricher. La pluspart des certifications sont payés.

### Le secteur de publication des evenements

Le partage des evenements sur un réseau sociale, il faut que le propriétaire de cet evenement est en accord avec l'action de publication. La publication est une étape éventuelle pour les evenements, car c'est à travers les intéressés peuvent avoir l'information sur l'existance de cet action Les limites dans les evenements en ligne c'est la confidentialité des informations des participants et des présents. Car ils vont donner leurs données pour se participer, ces informations vont être utilisé pour l'organisation et la gestion des présences.

Alors les responsables sur les evenements doivent respecter leurs confidentialités et n'utiliser pas ces données dans des buts lucratifs ou justificatifs publique.

## 2 Etude de l'existant

### 2.i Analyse de marché

A ce niveau d'étude , la prospection de marché semble être apporté un plus pour le positionner par rapport aux services existants, et nous aider pour capturer les besoins. Commencant par le secteur d'échange des documents vers le secteur de gestion des documents. Le choix des solutions selon les premiers référencements distincts sur le moteur de recherche Google.

#### Pour le secteur des documents éducatives

**SlideShare** Slideshare est l'un des outils fondamentaux pour faire un bon travail de distribution de contenu en ligne. Il est particulièrement utile pour la mise en ligne, la publication et le partage de PDF, de présentations et de Keynote. La page d'accueil de SlideShare est I.4

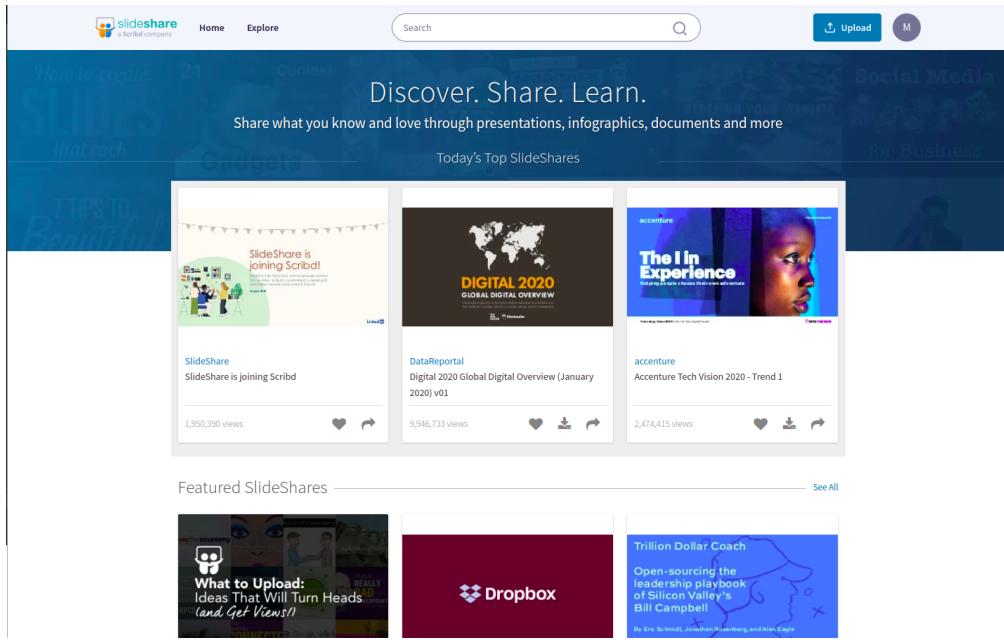


FIGURE I.4 – SlideShare

#### Avantages

- +Télécharger des présentations
- +Intégration facile dans d'autres sites Web
- +Compatible avec les fichiers volumineux
- +Conversion de fichier pour la visualisation

#### Inconvénients

- Aucun moyen intégré de suivre les vues(les interactions utilisateurs)
- Doit utiliser un autre programme pour créer et éditer du matériel(les ressources éducatifs)
- Pas partage de document sur d'autre plate-formes
- Services payant pour 9\$ par mois

**Moodle** Moodle est une plateforme d'apprentissage en ligne (en anglais : Learning Management System ou LMS) libre distribuée sous la Licence publique générale GNU écrite en PHP. Développée à partir de principes pédagogiques, elle permet de créer des communautés s'instruisant autour de contenus et d'activités. Le mot « Moodle » est l'abréviation de Modular Object-Oriented Dynamic Learning Environment : « Environnement orienté objet d'apprentissage dynamique modulaire », La page d'accueil pour Moodle est I.5

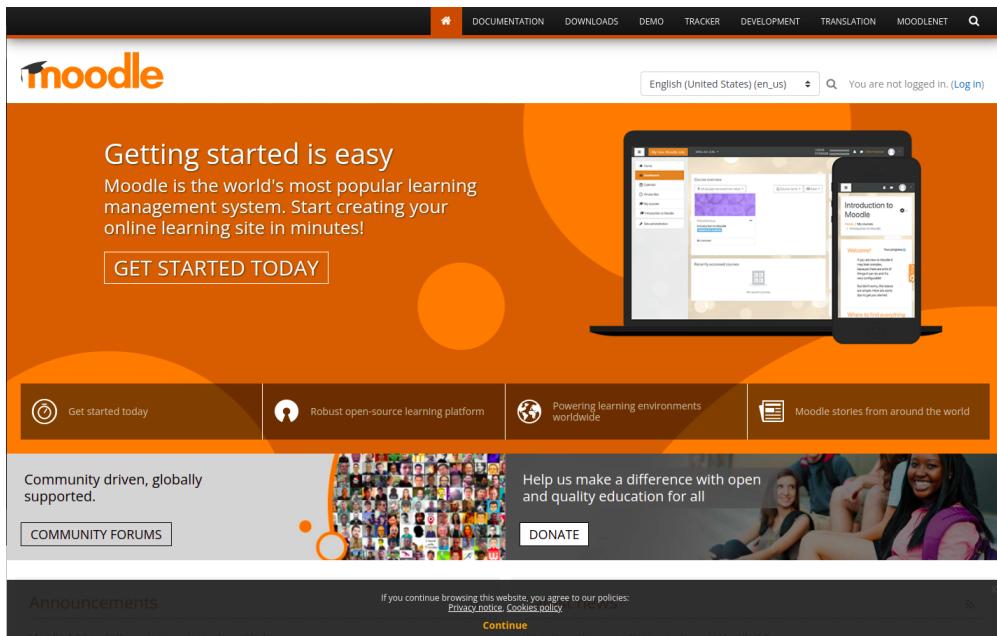


FIGURE I.5 – Moodle

#### Avantages

- +Flexibilité d'apprentissage
- +Élimination des barrières spatio-temporelles de l'éducation en face à face.
- +Maîtrise des TIC par les étudiants
- +Large communication entre étudiants et enseignants
- +Logiciel gratuit.

#### Inconvénients

- Pas simple d'utilisation, pas intuitif.
- difficultés apprêhender toutes les fonctionnalités.
- étapes de configuration fastidieuses,
- multitude de paramètres spécifiques
- Formation préalable nécessaire

**Google Classroom** Google Classroom est une plate-forme d'apprentissage gratuite dédiée aux écoles. Son but est de simplifier la création et la diffusion de cours et d'exercices de façon numérique. La plateforme a été présentée comme une fonctionnalité supplémentaire de Google Apps à la suite d'une publication, le 12 août, 2014. Google Classroom s'axe sur la simplicité d'usage. Elle est accessible à partir de tous les appareils mobiles, la page d'accueil est I.6

#### Avantages

- +Facile à utiliser et accessible depuis tous les appareils.
- +Communication et partage efficaces.
- +Accélère le processus d'affectation
- +Rétroaction efficace.
- +Excellent système de commentaires.

#### Inconvénients

- Gestion de compte difficile.
- Options d'intégration limitées.
- Difficile partage d'apprenants
- Aucune mise à jour automatique.
- Problèmes d'édition

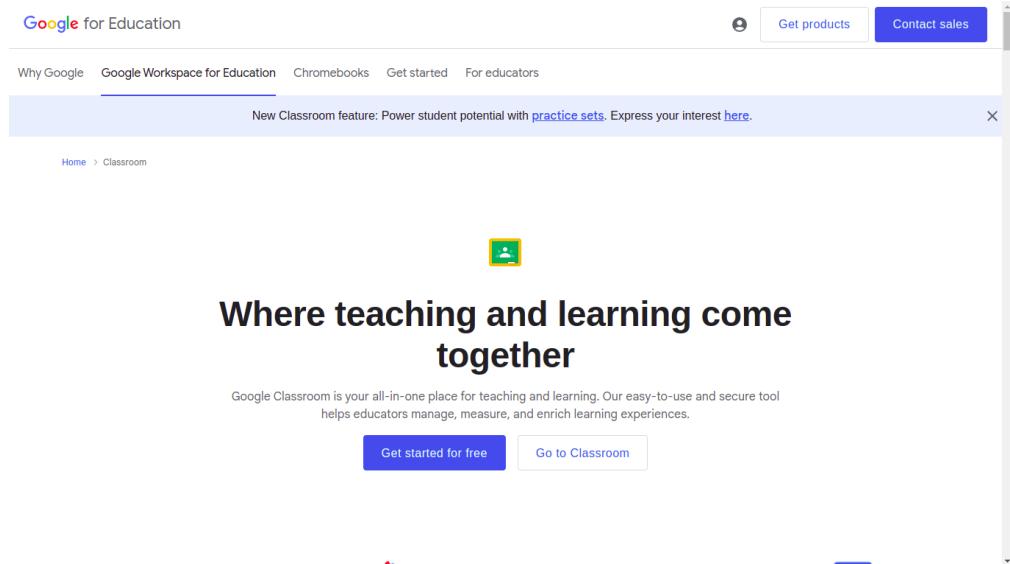


FIGURE I.6 – Google Classroom

**Pages Facebook pour l'échange des Documents** Facebook est un réseau social, qui permet aux utilisateurs de créer des communautés, des groupes pour l'échange et le partage de documents éducatifs, exemple de page d'échange dans Facebook I.7

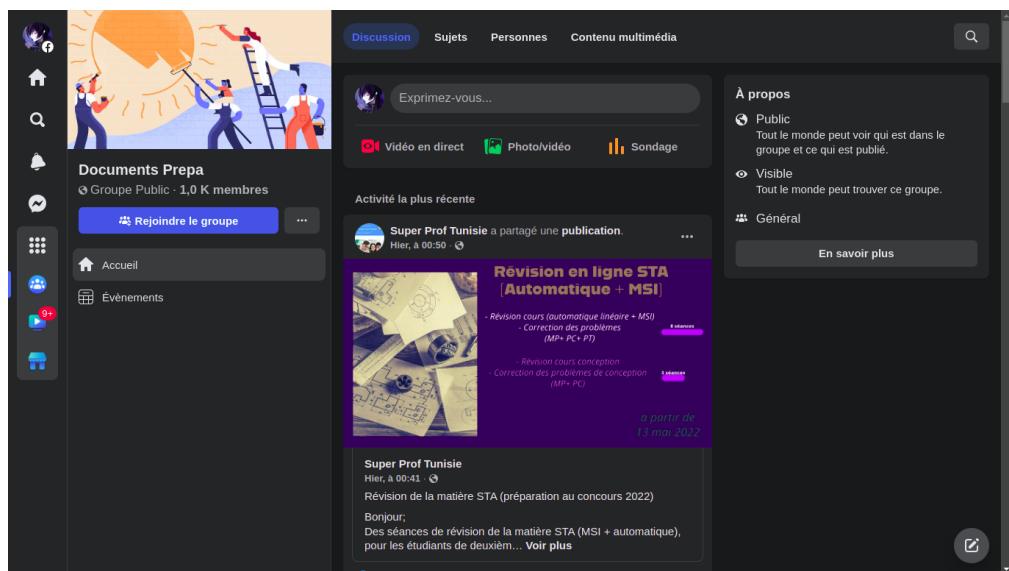


FIGURE I.7 – Les pages facebook pour l'échange des documents

## Pour le secteur des événements universitaires

**4C** Le Centre de Carrières et de Certification des Compétences (4C) est une structure rattachée à la présidence de l'Université ou au doyen /directeur de l'établissement d'enseignement supérieur et de recherche dont la mission est de préparer et d'accompagner ses

<p><b>Avantages</b></p> <ul style="list-style-type: none"> <li>+Familiaire.</li> <li>+Excellent système de commentaires.</li> <li>+Facile à apprendre</li> </ul>	<p><b>Inconvénients</b></p> <ul style="list-style-type: none"> <li>-N'est pas professionnel.</li> <li>-Taille de fichier limité</li> <li>-Publicité gênante</li> <li>-Non organisées</li> <li>redondance d'information</li> </ul>
--	---

usagers, étudiants et diplômés, en vue de faciliter leur insertion sur le marché du travail. Il tend également à jouer le rôle du partenaire privilégié pour toute entreprise désirant recruter un profil professionnel particulier ayant obtenu un diplôme universitaire mais n'ayant pas encore cumulé une expérience confirmée. Le 4C est le maillon entre l'université, l'étudiant et l'entreprise. Le 4C œuvre également à faciliter la certification des compétences afin de renforcer les chances de recrutement de nouveaux diplômés. Il met ses services à la disposition des entreprises afin de renforcer et valoriser les qualifications professionnelles de leurs employés, la page d'accueil pour 4c est I.8

The screenshot shows the homepage of the 4C platform. At the top, there are navigation links for 'Tableau de Bord', 'Actualités 4C' (selected), 'Mes Actualités', 'Offres', 'Création CV', 'Profil', 'Notifications', and 'Chat'. The main content area is titled 'Actualités 4C' and includes tabs for 'FORMATIONS', 'ÉVÉNEMENTS', and 'CERTIFICATIONS'. There are search and filter options for 'Recherche', 'Région', 'Type', and 'Centre 4C'. Below these are two event cards:

- Formation Soft Skills** (4C-ENI CARTHAGE) - Soft Skills, 16/05/2022, by M-Azziz Ben-Hamida, Coach et consultant en Soft Skills. Organisé par Orange Digital Center. Date : du 16 au 18 Mai 2022. Buttons: Nouvelle, Details, Inscription, Partager.
- Atelier de rédaction d'un CV** (4C-ISD) - Techniques de recherche d'emploi, 11/05/2022, by INSTITUT SUPERIEUR DE DOCUMENTATION DE TUNIS. Buttons: Nouvelle, Details, Inscription, Partager.

FIGURE I.8 – 4C

<p><b>Avantages</b></p> <ul style="list-style-type: none"> <li>+Inscription Gratuites.</li> <li>+Utilisation Facile</li> <li>+Réseau bien développé</li> <li>+Fameux dans l'environnement universitaire</li> </ul>	<p><b>Inconvénients</b></p> <ul style="list-style-type: none"> <li>-Les evenements dépassés ne sont pas supprimé</li> <li>-L'ajout des events est limité aux administrateurs</li> <li>-Limité à la region Tunisie</li> <li>-N'est pas assez dynamique</li> </ul>
--	--

**10times** "10times est le plus grand agrégateur d'événements professionnels au monde. Salons professionnels ou conférences, vous nommez un événement et nous l'avons répertorié. Parcourez les événements, errez et devenez fou @ 10 fois. Nous gérons un trafic B2B massif à travers le monde sur cette seule plateforme incroyablement incroyable." [23]

La page d'accueil pour 10times est I.9

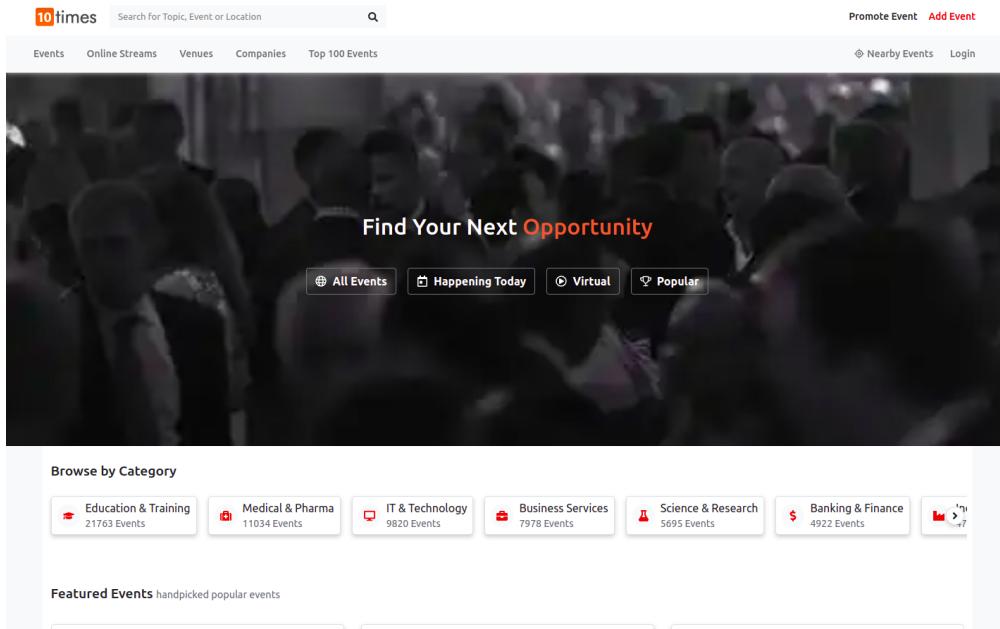


FIGURE I.9 – 10times

#### Avantages

- + Accessible par tous plateformes.
- + Dynamique
- + Réseau mondiale
- + Grande communauté

#### Inconvénients

- Pas d'inscription pour les utilisateurs
- Contenu trop chargé
- Difficile à apprendre
- Ajout d'évenement est de 1500\$ par mois

## 2.ii Généralisation

Nous avons analysées les solutions existants et nous les avons généralisés selon deux tableaux, le premier pour le secteurs des documents I.1 et l'autre pour le secteur des événements I.2. Enfin nous représentons tous les solutions existants sur un repère à 3 axes le premier (le prix du service), le deuxième (l'échange des documents) et le troisième (la gestion des événements). Pour but de nous positionner entre ces solutions existants I.10.

	Forces	Faiblesses
SlideShare [70], Scribd [71]	Populaire, Facile à utiliser	Paiement pour accéder au document, Publicités gênantes
ResearchGate [72]	Facile à utiliser	Manque d'organisation, Paiement pour accéder au document, Publicités gênantes
Rivezli.tn [68]	Simple et Facile à utiliser	Limité au cours d'économie et gestion avec paiement
ZLibrary [69]	Multiple discipline disponibles, contenu riche	Contenu très chargés et n'est pas facile à utiliser
Pages Facebook [66]	Le plus populaire Gratuit	Manque d'organisation Document redondants
UVT [73], Classroom [67]	Gratuit	Difficile à utiliser Accès Restreints
Libraires	Traditionnel et facile à accéder	Ne sont pas accessible tout le temps, Paiement pour achat des documents
Bibliothèques	Gratuit	Ne sont pas accessible tout le temps

TABLE I.1 – Analyse S.W des concurrents dans le secteur partages des documents.

	Forces	Faiblesses
4C[63]	Inscriptions Gratuites	Ajout limité aux administratives, et services non dynamique
10Times [65]	Grande communauté mondiale	Ajout d'un evenement de 1500\$ par moi, contenu très chargés
Google Forms [64]	Simple et facile à utiliser	Toujours gratuits même pour les evenement payants
Pages Facebook [66]	Le plus populaire Gratuit, accessible grand public	Manque d'organisation, aucune véracité pour les evenements ni les inscriptions.
Les médias	Accessible grand public, véracité respecté	Trop cher pour l'annonce.

TABLE I.2 – Analyse S.W des concurrents dans le secteur d'évenement.

### 3 Méthodologie de travail

Dans le cadre de ce projet, nous sommes intéressées à l'adoption le modèle en V. Le cycle en V ou V model en anglais est un modèle utilisé dans différents processus de développement, notamment dans le développement de logiciels. Élaboré dans les années

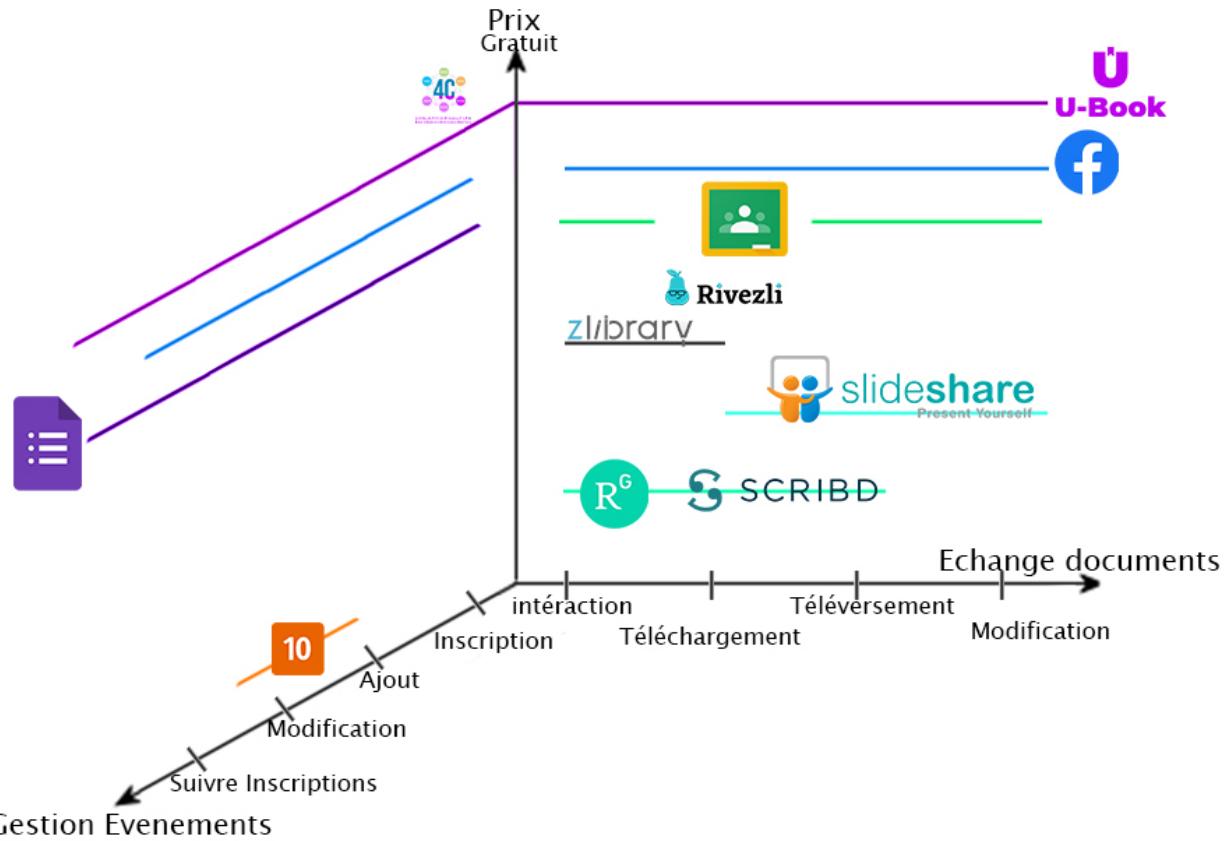


FIGURE I.10 – Positionnement de UBooK par rapport les autres soutions existants

90 sous sa forme originale, il est perfectionné au fil des ans et adapté aux méthodes de développement contemporaines, la représentation de cycle en V est I.11

Dans l'ensemble, ce modèle peut permettre d'éviter les malentendus ainsi que les tâches inutiles. De plus, il permet de s'assurer que toutes les tâches soient exécutées en temps voulu, dans le bon ordre en réduisant les temps morts au maximum.

- Optimisation de la communication entre les parties prenantes grâce à des modalités et des responsabilités clairement définies.
  - Risques maîtrisés et meilleure planification grâce à des fonctions, des structures et des résultats bien définis en amont.
  - Amélioration de la qualité du produit grâce à l'intégration de mesures liées à l'assurance qualité.
  - Réduction des coûts grâce à un processus transparent de l'ensemble du cycle de vie du produit.

Dans un premier temps, le cycle en V définit le déroulement d'un projet en phases distinctes qui sont tour à tour détaillées :

- En début de projet, le modèle prévoit une analyse de l'ensemble des besoins relatifs au système envisagé.
  - Le projet est ensuite enrichi par l'expression des besoins fonctionnels et non fonctionnels liés à l'architecture du système.
  - Puis on passe à la phase de conception du système, lors de laquelle les composants et

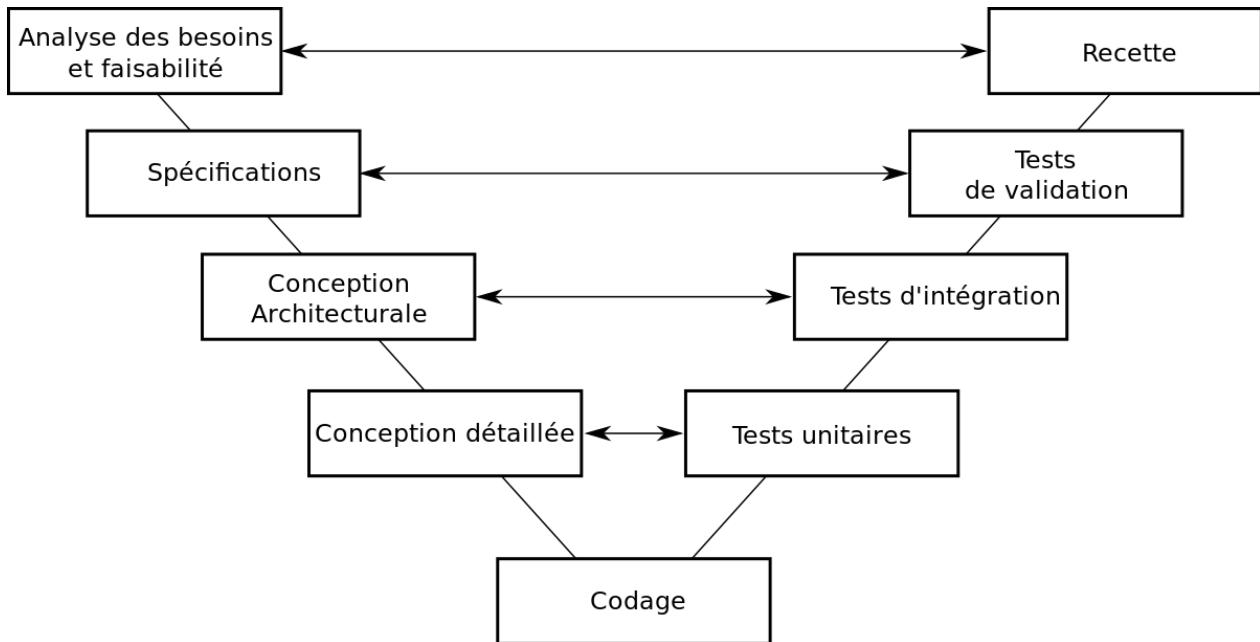


FIGURE I.11 – Cycle de développement : Modèle en V

les interfaces du système sont planifiés.

- Une fois ces étapes franchies, on peut passer à la conception de l'architecture logicielle en détail.

On entre alors dans la phase effective de développement du logiciel en fonction de ce qui a été planifié. Ensuite ce sont les phases d'assurance qualité, qui se réfèrent toujours aux étapes du développement. Le modèle prévoit les tâches suivantes :

- Tests unitaires ;
- Tests d'intégration ;
- Intégration système ;
- La « recette » (ou test d'acceptation).

## Conclusion

Ce chapitre est dédié à la présentation générale de notre solution suivie d'une description des concurrents, dans les deux premières partie, nous avons décrit dans les parties suivantes l'évenement universitaire, les documents educatifs, le domaine métier, l'étude de l'existant et la méthodologie de développement modéle en V .le chapitre suivant sera consacré au spécification , conception et développement de la partie infrastructure logicielle de l'application .

---

# CHAPITRE II

## SPÉCIFICATION

Dans ce chapitre et le suivant nous tenons compte l'aspect technique de l'application, commençant par l'environnement de travail et les technologies utilisés, ensuite la spécification ou on étudiera les diagrammes UML de Use Case avec quelques descriptions textuelles, puis les diagrammes de séquences système pour chaque cas d'utilisation.

### 1 Modélisation

Nous allons présenter notre diagramme de cas d'utilisation globale dans le but de donner une vision globale du comportement fonctionnel et de décrire les interactions du système avec les utilisateurs. Commoncons par l'extraction des Roles et les besoins fonctionnels enfin nous les présentons sur un diagramme de Use Case.

#### 1.i Capture des besoins

Puisque nous sommes en train d'étudier une plateforme sur internet, alors chaque personne à la possibilité d'ouvrir l'application. Donc un premier rôle c'est l'internaute qui peuvent Rechercher des documents, consulter les événements, découvrir l'environnement universitaire en consultant les organismes universitaires (les Clubs, les centres de formations et les instituts universitaires), Et bien évidemment s'inscrire pour rejoindre la communauté des UBooKers. Autre rôle c'est l'utilisateur inscrit qui peut télécharger, téléverser les documents et interagir avec ainsi que s'inscrire et gérer (ajouter, modifier, supprimer) des événements, nous les appelons des UBooKers. Les UBooKers paient leurs inscriptions dans les événements si il n'est pas gratuit alors le système de paiement à un rôle dans l'application.

#### 1.ii Diagramme de Use Case

Dans le diagramme de use case nous avons modélisé les rôles dans l'application et pour chaque un d'eux, nous avons associer les cas d'utilisations appropriés. Voir la figureII.1

### 2 Scénarios des cas d'utilisations

Dans cette section nous analysons quelques cas utilisations dans des scénarios nominales et alternatives.

Et Pour faciliter l'interprétation de comportements de l'application, nous présentons des maquettes pour quelques autres cas utilisations.

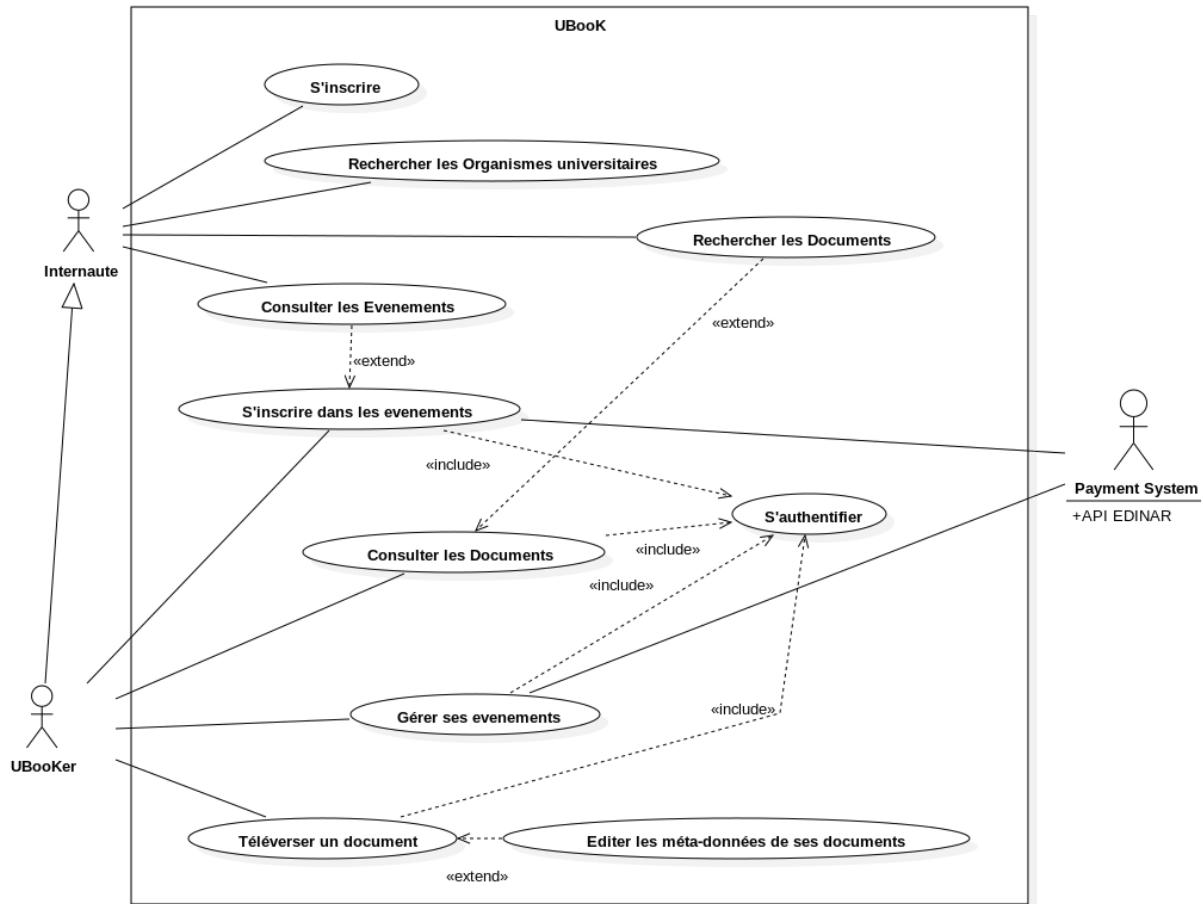


FIGURE II.1 – Le Diagramme de Use Case

## 2.i Scénario d’Edition les métadonnées d’un document

Pour éditer les informations d'un document, il faut que l'utilisateur soit inscrit et il est le propriétaire de ce document, le tableau de scénario d'édition des métadonnées d'un document est II.1

Scénario d'édition d'un document	
Titre	Editer les métadonnées de ses documents
Acteur	UBooKer
Précondition	Le document à éditer ses métadonnées est créé par le UBooKer
Scénario nominale	<ul style="list-style-type: none"> <li>— UBooKer Modifie les métadonnées</li> <li>— UBooKer soumet le formulaire</li> <li>— Le document se met à jour</li> <li>— Une notification par email sur la mise à jour</li> <li>— UBooKer est redirigé vers la page de consultation de document.</li> </ul>
Scénarios alternatives	<ul style="list-style-type: none"> <li>— Si le UBooKer n'a pas rempli les zones nécessaires : Le bouton de soumission de formulaire est bloqué, jusqu' les zones sont bien remplis</li> </ul>
Scénarios d'erreurs	<ul style="list-style-type: none"> <li>— Si le UBooKer a quitté la page de modification des métadonnées sans soumission : Le bouton de soumission de formulaire est bloqué, jusqu' les zones sont bien remplis</li> </ul>
Postconditions	Le UboooKer est dans la page de consultation de document édité.

TABLE II.1 – Scénario Editer les métadonnés de ses documents par un UBooKer

## Téléverser un document

Le téléversement d'un document est pour les inscrits, mais il y'a un processus de vérification de véracité de ce document. s'il est confirmé alors l'ajout de document se fait avec un badge de véracité, sinon un email d'alerte vers le propriétaire de ce document et aucun badge s'ajoutera pour ce moment la. Le diagramme de séquence de téléversement d'un document est II.2

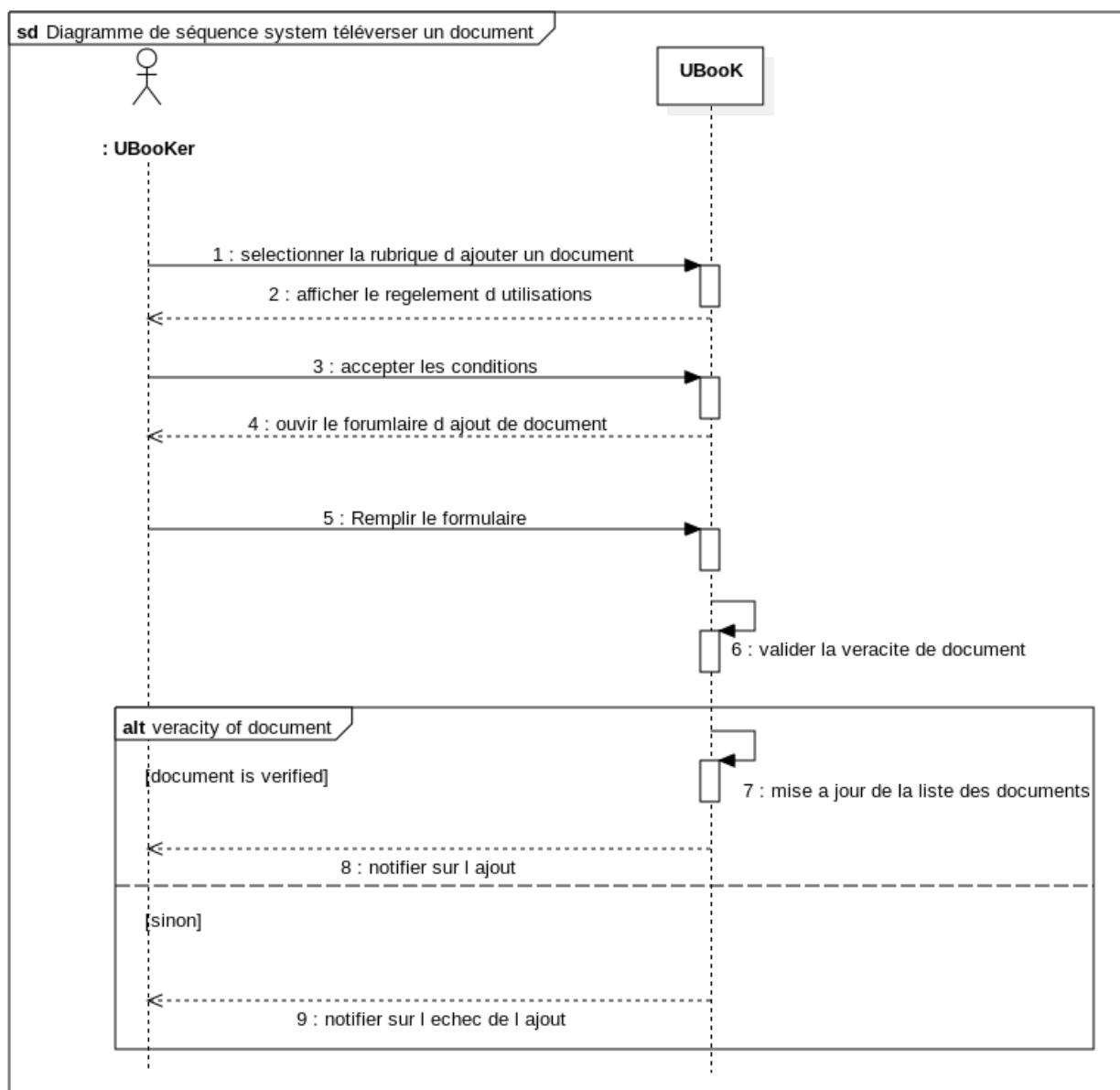


FIGURE II.2 – Diagramme de séquence system upload document par UBooKer

## 2.ii Scénario Télécharger un documnet

Nous avons représenter le scénario de téléchargement d'un document sur un maquête pour faciliter l'interprétation du comportement de système II.3

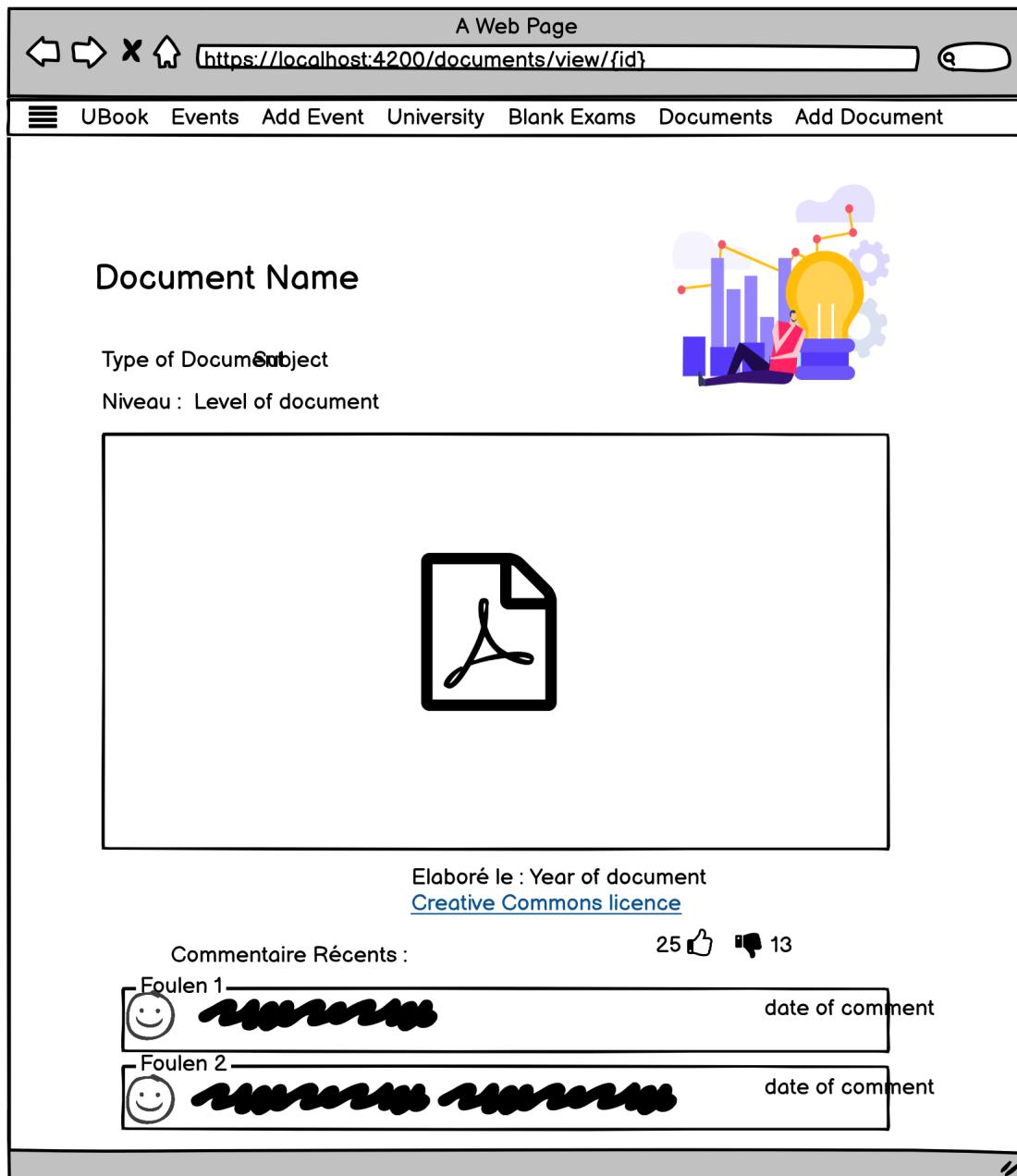


FIGURE II.3 – Maquette download document par UBooKer

### 2.iii Scénario d'inscription dans un evenement

Pour S'inscrire il est nécessaire d'être un UBooKer, et l'inscription réserve une place dans l'évenement. Si l'évenement est payant, la page de paiement s'ouvre pour payer l'inscription et automatiquement après une place sera réservé. Le tableau qui représente les scénarios nominales, alternatives de scénario d'inscription dans un évenement II.2 et le diagramme de séquence système est II.4

Scénario s'inscrire dans des events	
Titre	Inscrire dans des evenements
Acteur	UBooKer
Précondition	<ul style="list-style-type: none"> <li>— Le UBooKer n'est pas inscrit dans l'évenement</li> <li>— L'évenement n'a pas atteint le nombre maximum d'inscription</li> </ul>
Scénario nominale	<ul style="list-style-type: none"> <li>— UBooKer clique sur s'inscrire</li> <li>— UBooKer rempli les informations pour confirmer son identité</li> <li>— UBooKer soumet le formulaire</li> <li>— UBooKer reçoit un alerte de réussite d'inscription</li> </ul>
Scénarios alternatives	<ul style="list-style-type: none"> <li>— Si le lien vers l'inscription de l'évenement n'est pas dans UBooK : le UBooKer est redirigé vers ce lien,</li> <li>— Si les informations soumises ne sont pas correctes, un message d'alerte indique qu'il y'a un erreur dans l'entrée des coordonnées</li> </ul>
Scénarios d'erreurs	<ul style="list-style-type: none"> <li>— Si le UBooKer a quitté la page d'inscription avant soumission : aucune modification pour le utilisateur ni l'évenement</li> </ul>
Postconditions	Une place réservé dans l'évenement.

TABLE II.2 – Scénario S'inscrire dans un évenement par un UBooKer

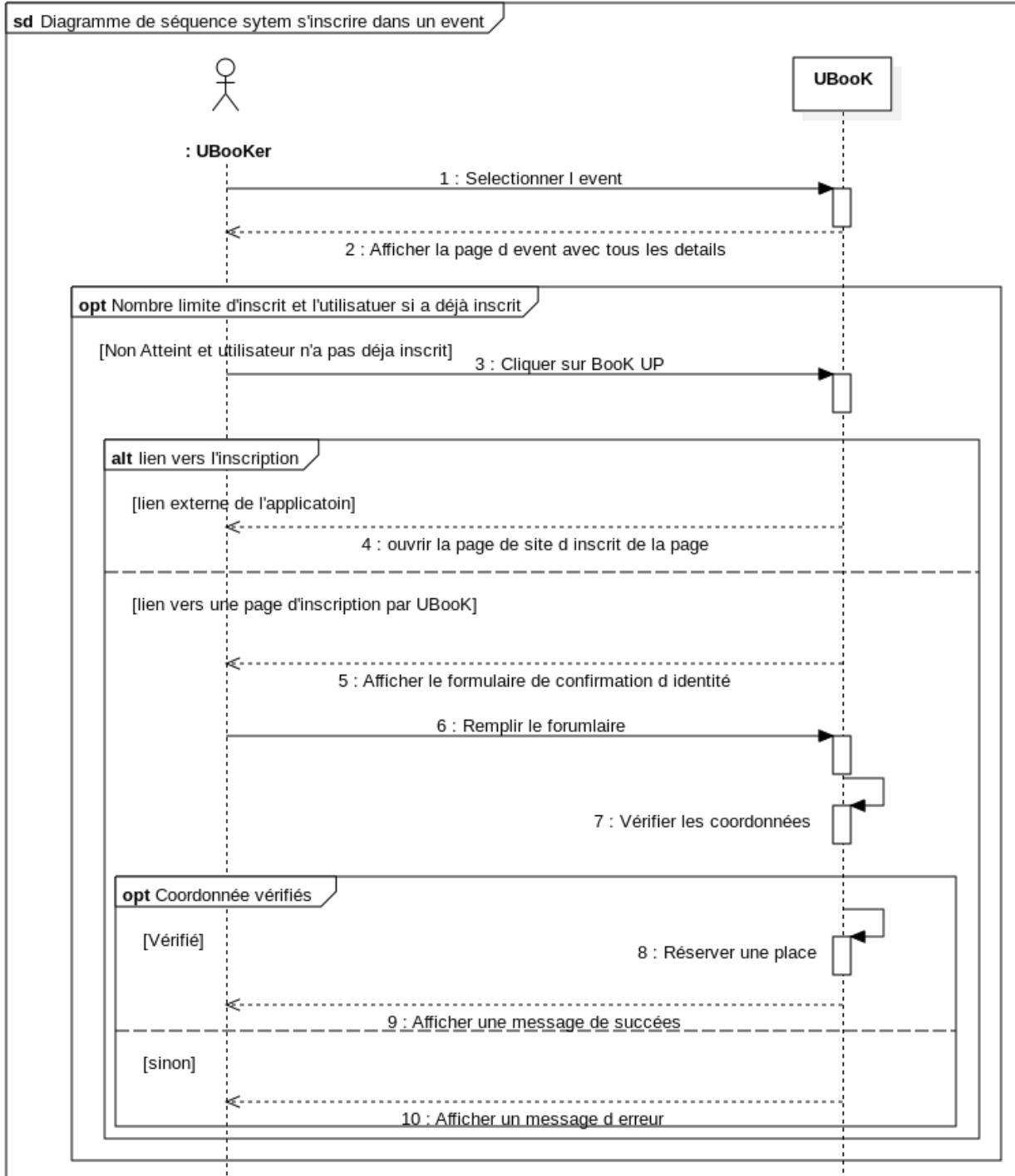


FIGURE II.4 – Diagramme de séquence Inscription dans un event par UBooKer

## 2.iv Scénario de gestion de ses evenements

Pour ajouter un evenement, l'utilisateur doit etre un UBooKer, si l'utilisateur est indépendant de tous les organismes universitaires alors il doit payer pour afficher l'evenement, mais s'il est sous la supervision d'un organisme universitaire alors un mail de vérification sera envoyé vers ce dernier pour valider l'ajout gratuite. Pour les autres gestion : la modification, la suppression et le suivie des inscriptions et limité sur le propriétaire de l'evenement. Voir le tableau II.3 pour les scénarios nominale ,altéranatives pour l'ajout d'un evenement, Pour le diagramme de séquence system voir II.5, et les maquêtes pour l'ajout de l'evenement sont II.6

Scénario Ajouter un evenement	
Titre	Ajouter un evenement
Acteur	UBooKer
Précondition	L'utilisateur est inscrit dans l'applicatoin UBooK
Scénario nominale	<ul style="list-style-type: none"> <li>— UBooKer rempli les informations nécessaire de l'evenement</li> <li>— UBooKer soumet le formulaire</li> <li>— L'evenement s'ajout mais privée pour le UBooKer</li> <li>— L'organisme supérieur de UBooKer recoit un mail de vérification d'ajout d'evenement</li> <li>— L'organisme supérieur de UBooKer confirme l'ajout d'evenement</li> <li>— Un mail d'ajout avec succès envoyé vers le UBooKer</li> <li>— l'evenement se mise à jour en publique</li> </ul>
Scénarios alternatives	<ul style="list-style-type: none"> <li>— Si il y'a un champs nécessaire n'est pas rempli : le bouton de soumission est bloqué</li> <li>— L'organisme supérieur de UBooKer n'a pas confirmé l'ajout : l'evenement reste privé avec possibilité de payement pour le rendre publice</li> <li>— Le UBooKer est indépendants des organismes inscrits dans l'applicatoin : le UBooKer est redirigé vers la page de payement, pour finaliser l'ajout de l'evenement.</li> </ul>
Scénarios d'erreurs	<ul style="list-style-type: none"> <li>— Si le UBooKer a quitté la page d'ajout avant soumission : aucune ajout s'effectue</li> </ul>
PostConditions	Le UBooKer est redirigé vers la page de consultation d'evenement.

TABLE II.3 – Scénario Ajouter un evenement par UBooKer

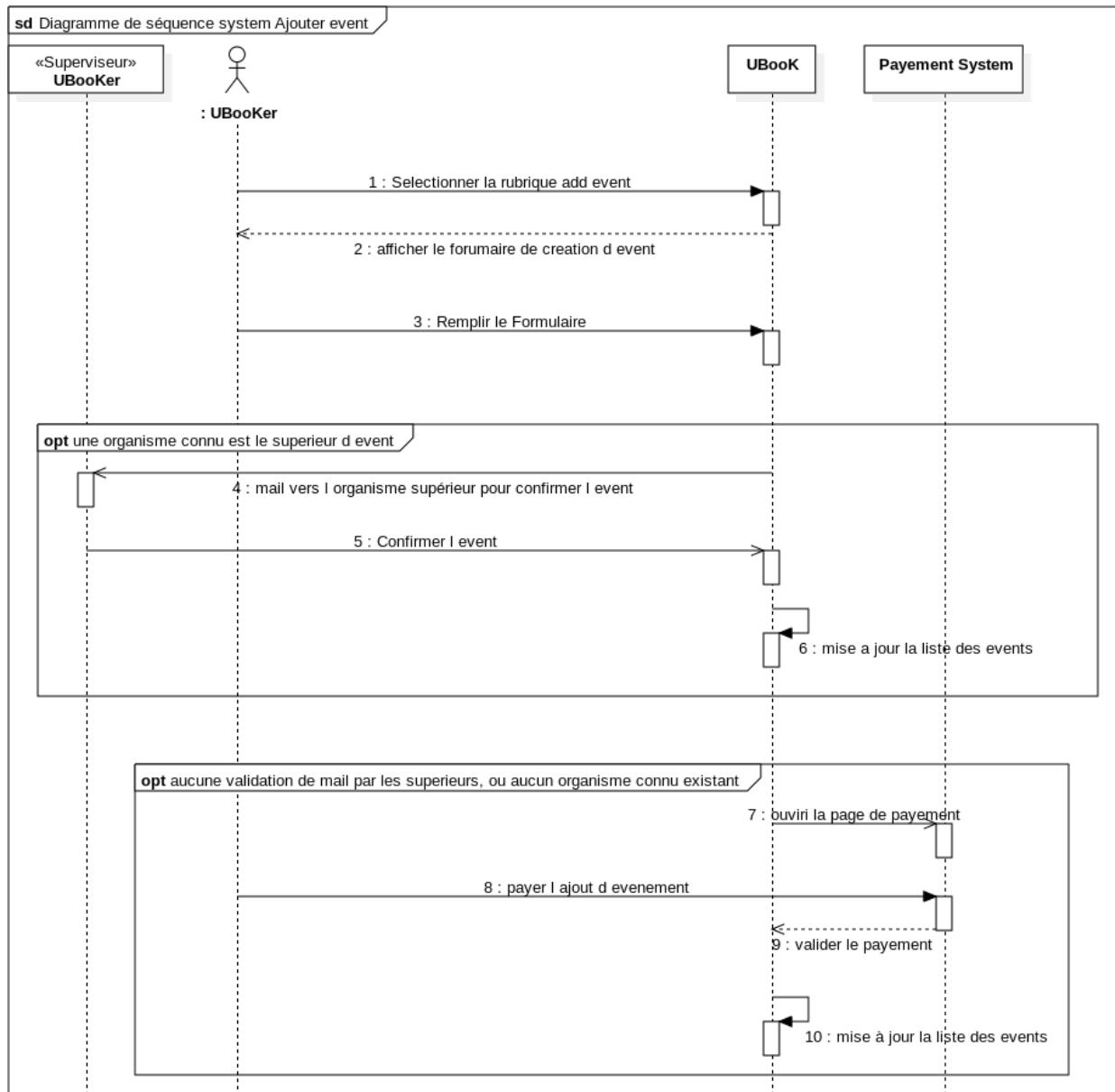


FIGURE II.5 – Diagramme de séquence d'ajouter un event par UBooKer

A Web Page  
<https://localhost:4200/event/new>

UBook Events Add Event University Blank Exams Documents Add Document

## New Event



**Event Type**  Training  Certification  Competition  Journey

**Event Type**

- Arts  Computer Sciences  Entrepreneurship
- Arts  Entrepreneurship  Computer Sciences
- Computer Sciences  Entrepreneurship  ...

**Event Name**

**Select University Organisms**

Instituts : Faculté Des Sciences Bizerte

Clubs : FPT FSB  Art Of Code FSB

When you click on this button it will open the pop up : Select University Organisms la maquette suivante

**Event Capacity 300**

**Event Email**

**Event Tel**

**Registration Link**

  Let UBooK take Booking on Charge

**Event Price 0DT**

Starting Date  Finishing Date  Registration Date

**Event Description**

**Event Display PDF,DOCX,PPTX**

**Finish**

Select the related University Organisms

Instituts :

Faculté Des Sciences Bizerte

Clubs :

Tunivision FSB Enactus FSB FPT FSB

Microsoft FSB Art Of Code FSB

Training Centers :

Void

FIGURE II.6 – Maqués Ajout Evenement Par UBooKer

## 2.v Scénario Consulter un evenement

Pour la consultation d'un evenement, l'internaute ouvre la rubrique event et choisi le type d'evenement qui il en train de chercher, soit Jounrée soit certification soit compétition soit formation. Il trouve les récentes evenements affichés en grille. Il peut chercher dans la barre de recherche ou faire son filtrage, même changer le type d'affichage en liste. Voir la maquête II.7

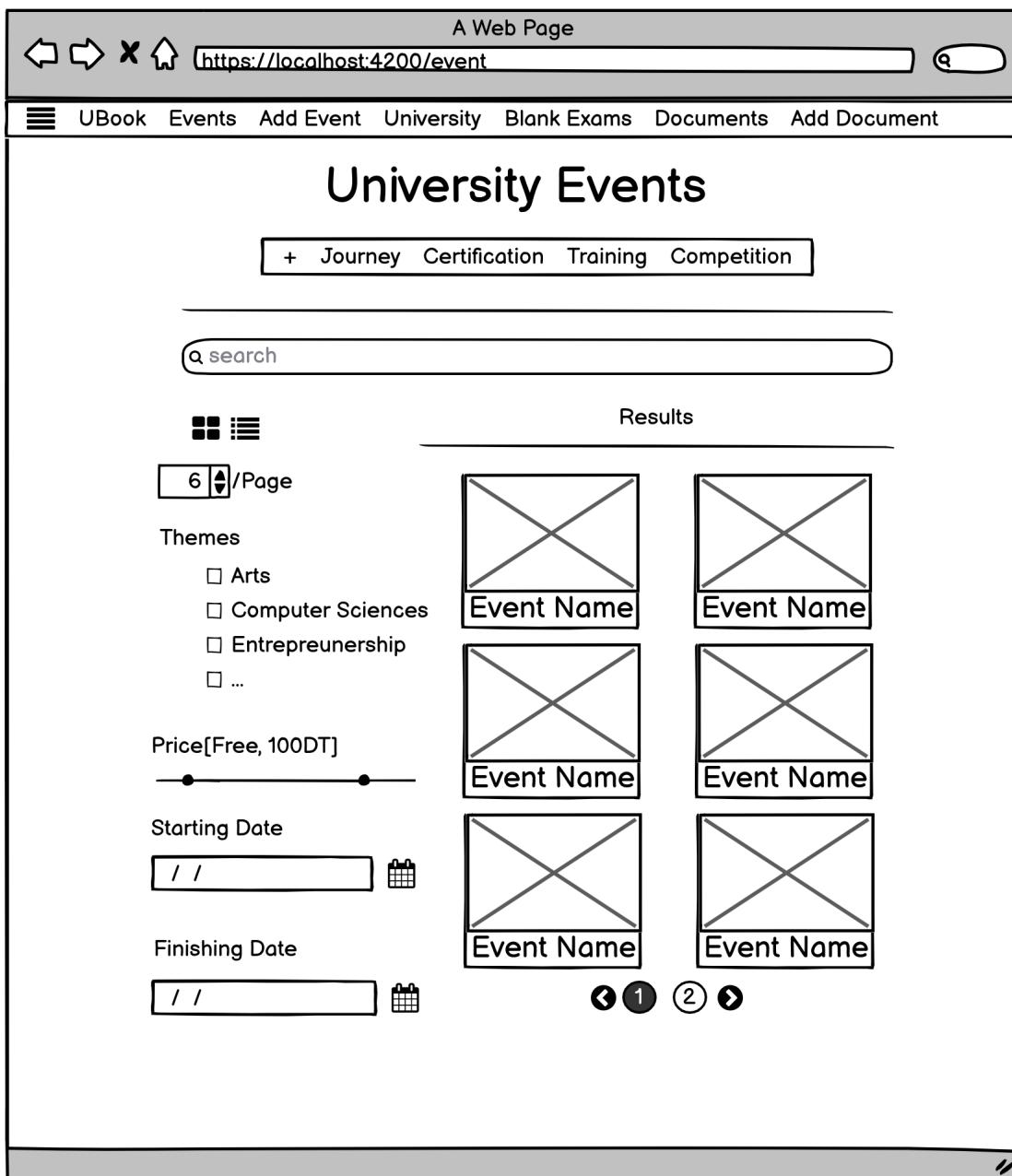


FIGURE II.7 – Diagramme de séquence system consulter evenement par Internaute

### 3 Diagramme de classe analyse

Dans ce diagramme on modélise les entités de l'application, et cette représentation va être le schéma dans la partie ingénierie de base de donnée. Voir II.8

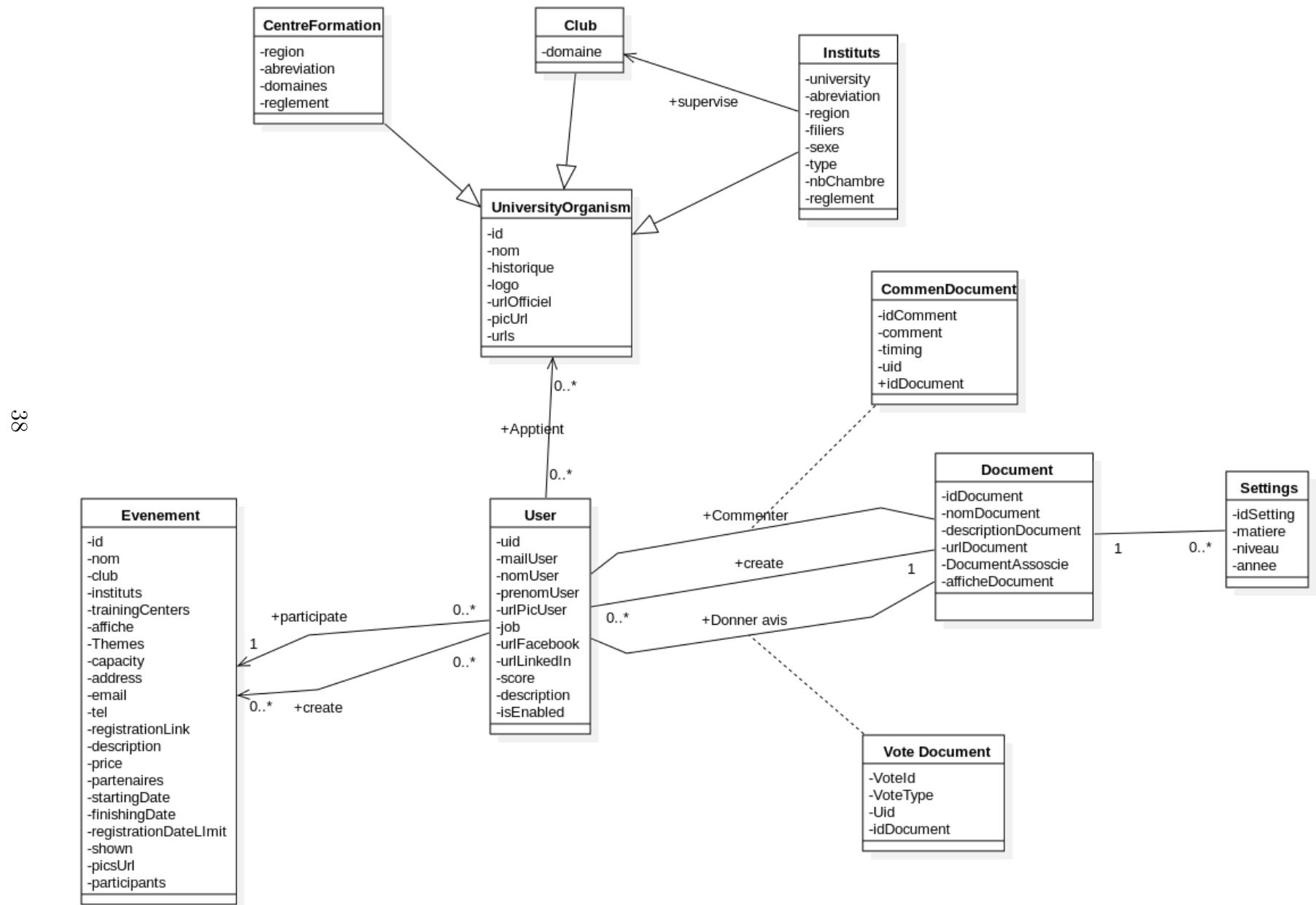


FIGURE II.8 – Diagramme de classe d'Analyse

---

# CHAPITRE III

## CONCEPTION

Dans ce chapitre, nous présentons la partie ingénierie de l'application.

Dans ce projet, nous divisons l'application sur trois grandes parties :

- La partie Front End pour toutes les vues, nous choisissons le couleur rouge pour cette partie
- La partie Back End pour les controlleurs, nous choisissons le couleur bleu pour cette partie
- La partie Model qui représente la couche Entity , nous choisisson le couleur vert pour cette partie.

Commencons par le diagramme de Package III.1 ou on modélise l'interaction entre les différentes parties, puis présentons le diagramme de classe conception sur deux figures la première représente la relation entre l'utilisateur et les documents III.2, la deuxième représente la relation entre l'utilisateur avec les evenements III.3, ensuite nous etudions quelques diagrammes de séquences objets qui sont :

- Rechercher un documentIII.4
- Téléverser un documentIII.5
- S'inscrire dans un evenementIII.6

# 1 Diagramme de Package

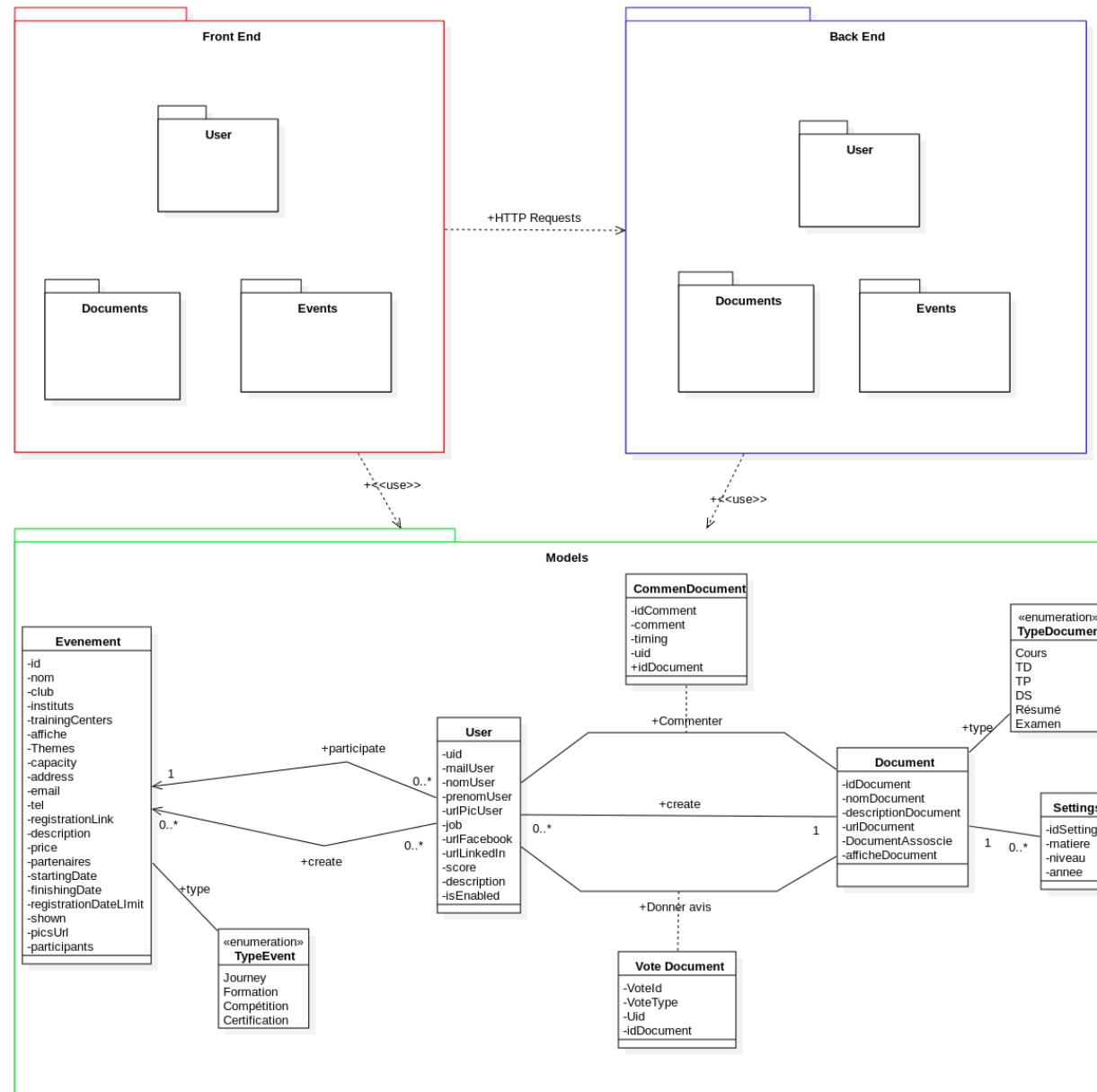


FIGURE III.1 – Diagramme de Package

## 2 Diagramme de classe conception

### 2.i Relation UBooKer et Document

14

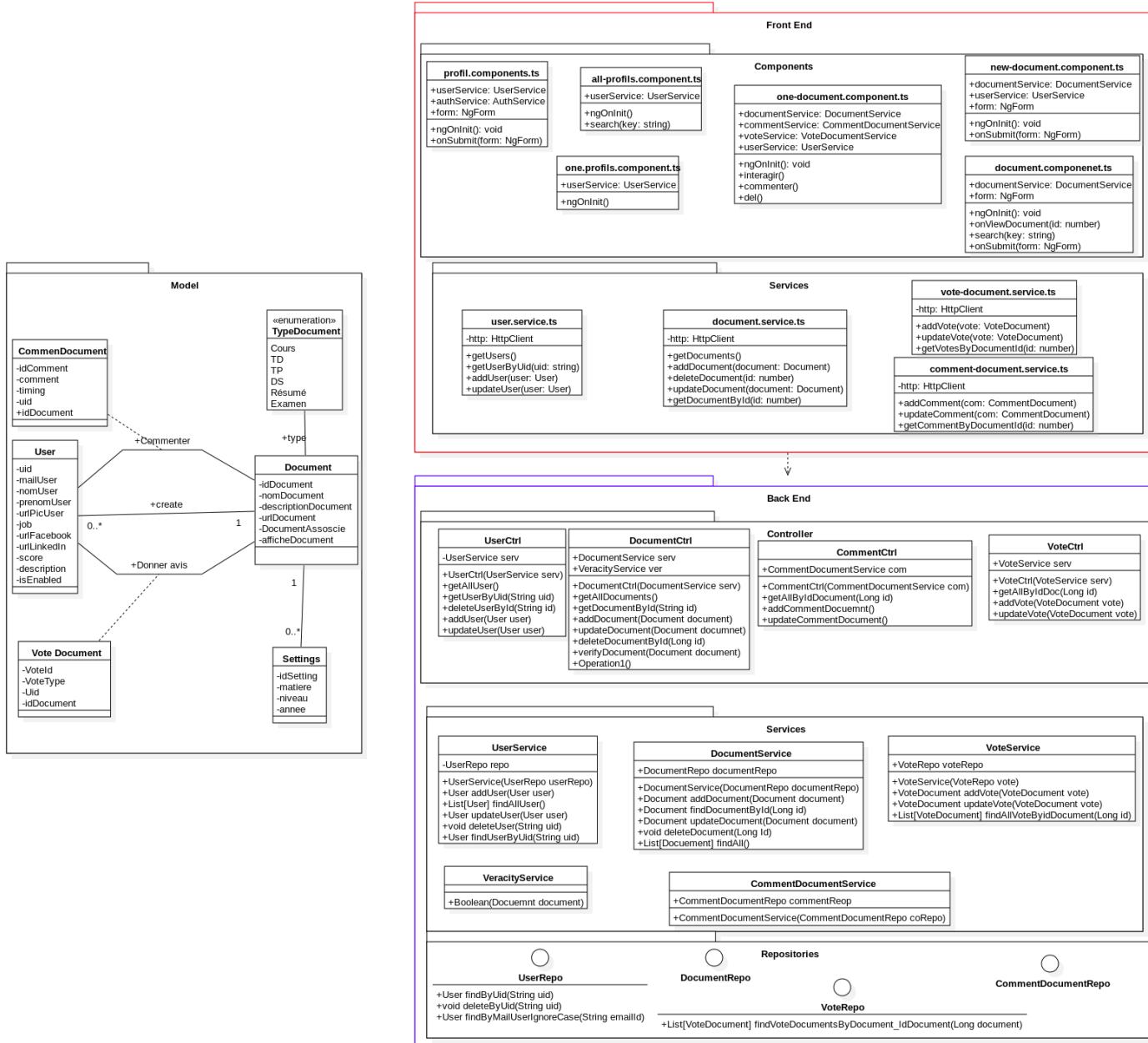


FIGURE III.2 – Diagramme de classe d’Analyse

## 2.ii Relation UBooKer et Evenement

42

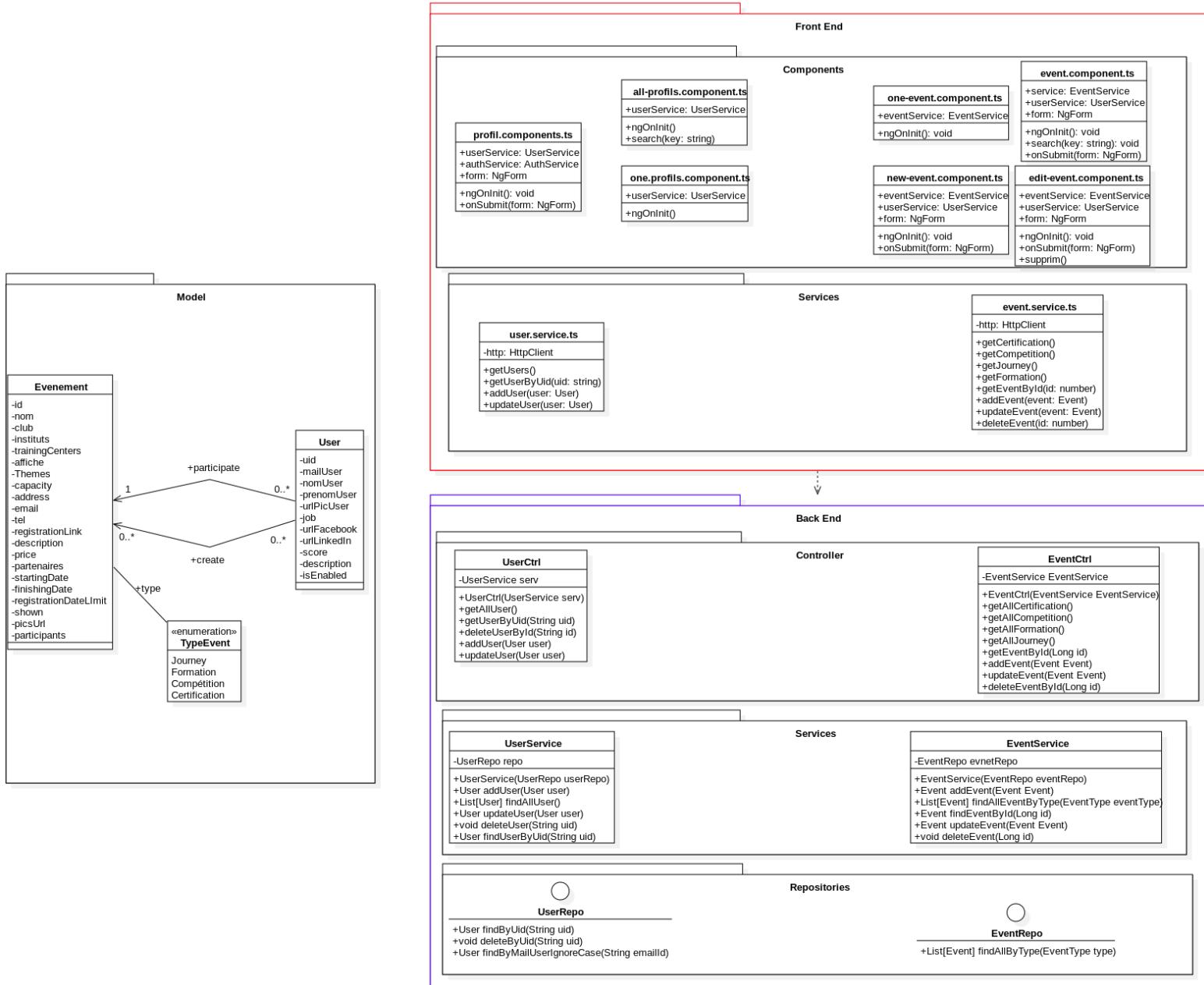
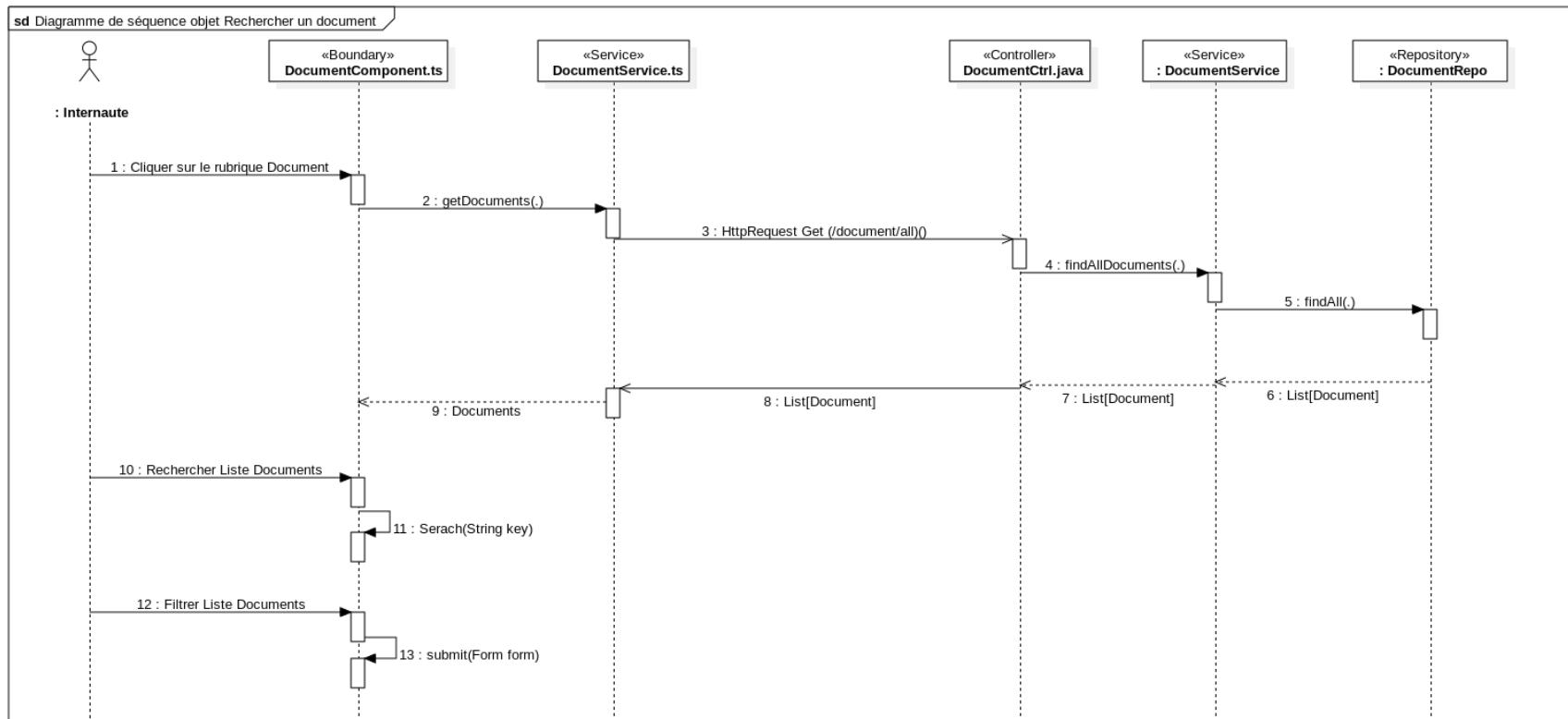


FIGURE III.3 – Diagramme de classe d’Analyse

### 3 Diagrammes de séquences objets

#### 3.i Diagramme de séquence objet Rechercher un document



43

FIGURE III.4 – Diagramme de séquence objet consultation liste documents

### 3.ii Diagramme de séquence objet Téléverser un document

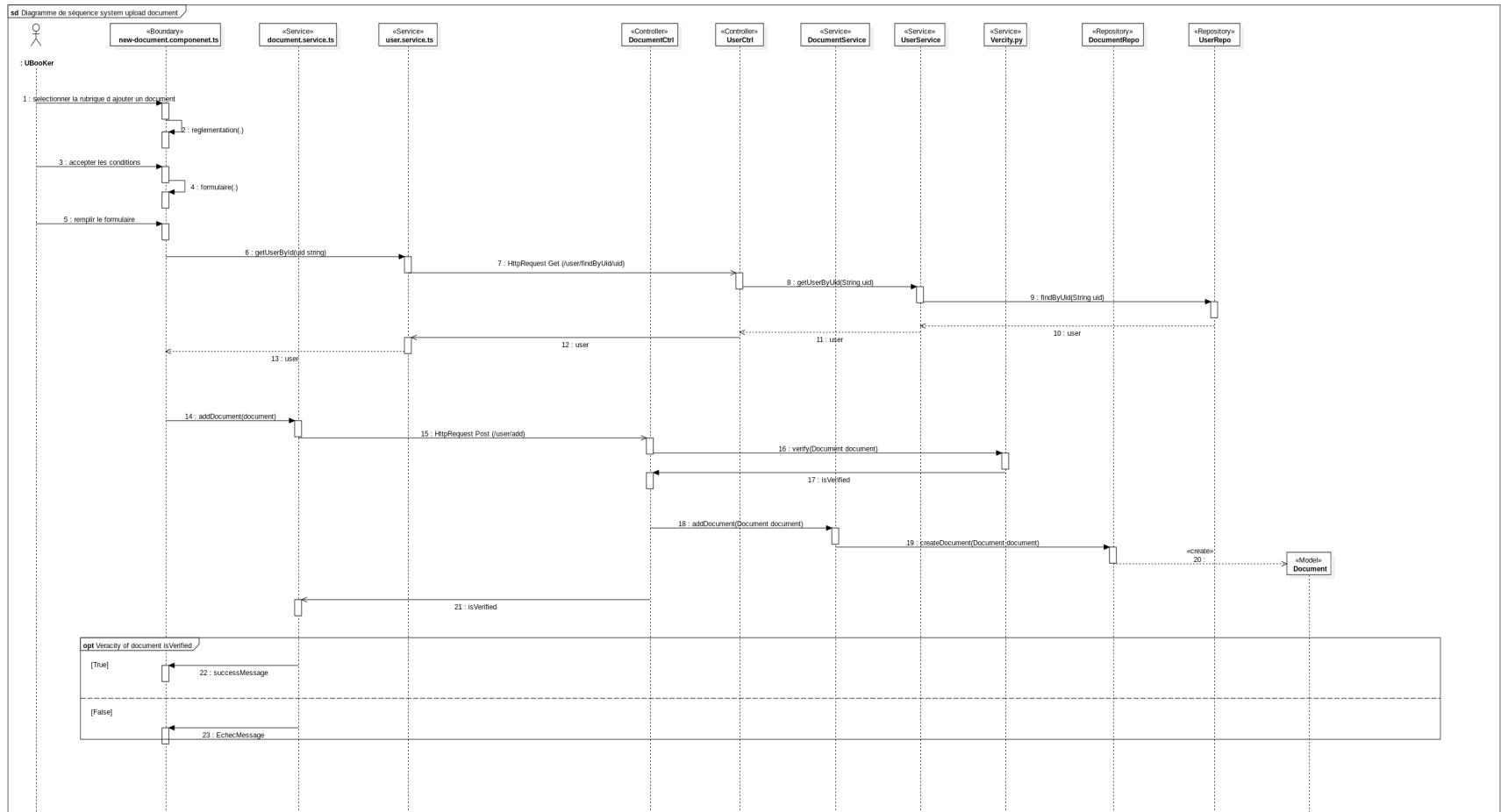


FIGURE III.5 – Diagramme de séquence objet upload document

### 3.iii Diagramme de séquence objet S'inscrire dans un evenement

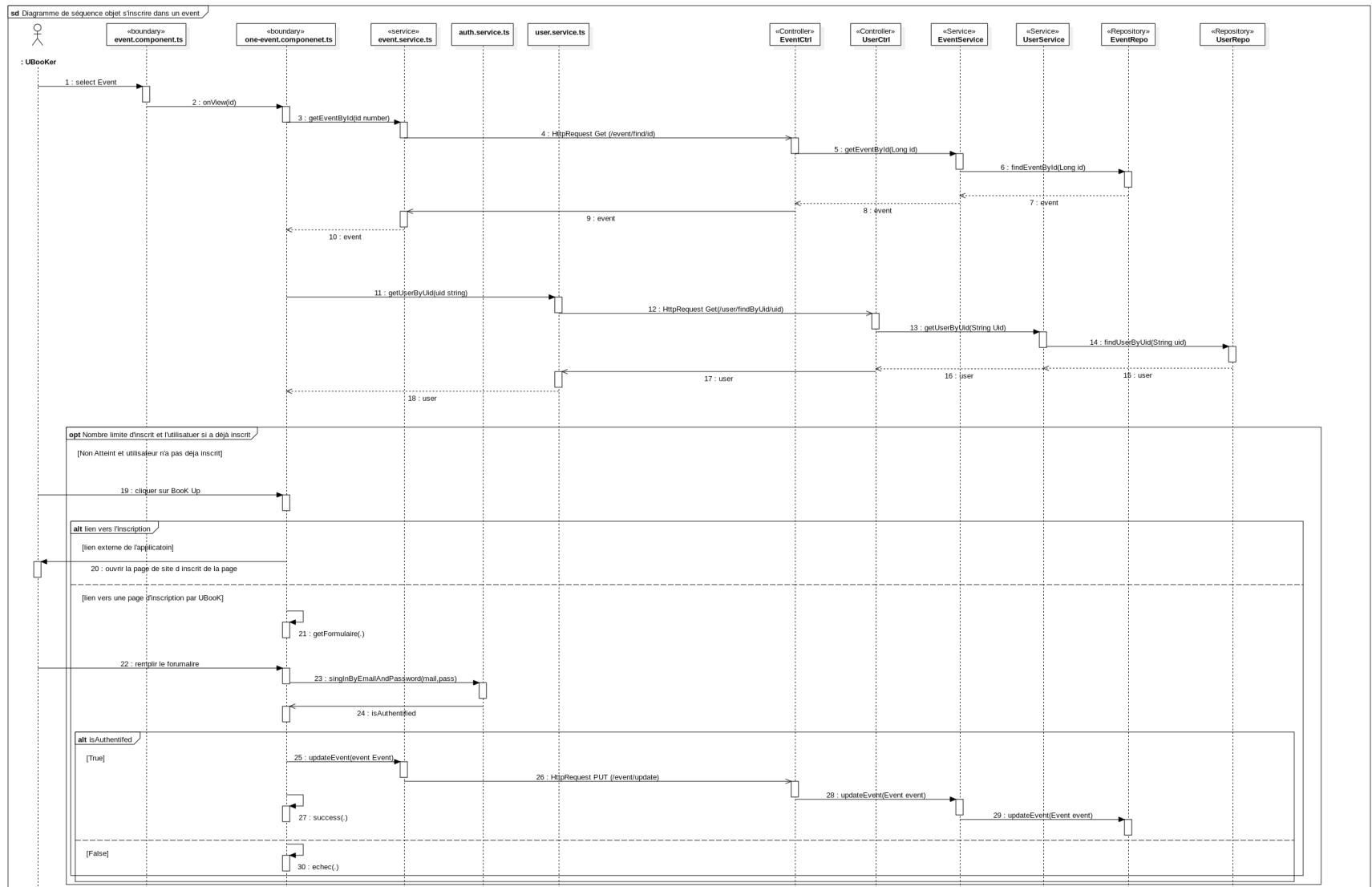


FIGURE III.6 – Diagramme de séquence objet inscription dans un event

---

# CHAPITRE IV

## REALISATION ET DÉVELOPPEMENT

### 1 Environnement de travail

Dans cette section, nous présentons les technologies utilisées pour la réalisation de projet UBooK.

#### 1.i Les langages de programmations et les Frameworks

##### Framework Spring Boot

Spring Boot est un framework de développement JAVA Entreprise Edition (JEE). C'est une déclinaison du framework classique de Spring qui permet essentiellement de réaliser des microservices (ce sont la majeure partie du temps des services web qui sont regroupés en API)[24].

Nous avons choisi Spring Boot(v 2.5.3) avec le jdk1.8 comme notre technologie Back End pour l'infrastructure logiciel associé aux dépendances :

- Spring Boot Starter Web : Un outil pour lancer le serveur web sur un port.
- Lombok : Un outil pour générer les Getters et les setters automatiquement
- Spring Boot Starter Test : Un outil pour lancer tests sur le projet Spring.
- Spring Boot Starter Data JPA : Un outil pour faciliter les requêtes SQL
- Spring Boot Starter Mail : Un outil pour envoyer des emails à travers Spring Boot
- FreeMarker : Un outil pour personnaliser les emails avec le code html et css.

##### Framework Angular

Angular est un framework d'application Web gratuit et open source basé sur TypeScript, dirigé par l'équipe Angular de Google et par une communauté d'individus et d'entreprises. Angular est une réécriture complète de la même équipe qui a construit AngularJS.[25] Nous avons choisi Angular (v 12.2.6) de TypeScript (v 4.0.3) depuis le package manager npm (v 8.4.1) comme Front End pour l'infrastructure logiciel.

##### Python

Le Python est un langage de programmation caractérisé par sa polyvalence : il est utilisé pour le développement web, l'IA, le machine learning, les systèmes d'exploitation, le développement d'applications mobiles, les jeux vidéo et bien d'autres. Nous avons choisi

Python comme un outil de vérification de véracité des documents comme indiqué dans la partie 2 (vers une modéle de véracité).

## 1.ii Les systèmes gestions base de données

### MySql

MySQL est un système de gestion de base de données relationnelle open source. Nous avons utilisé MySql (version 8.0.28) comme notre système de gestion de base de données relationnelle.

### Framework Firebase

Firebase est une plateforme développée par Google pour créer des applications mobiles et web. C'était à l'origine une société indépendante fondée en 2011. En 2014, Google a acquis la plate-forme et c'est maintenant leur offre phare pour le développement d'applications[26]. Nous avons utilisé Firebase comme un outil back end pour bénéficier de ces services d'authentification avec les application Google, Facebook, Twitter, LinkedIn.

## 1.iii Outils Pour la modélisation, spécification et conception

### StarUml

StarUML est un logiciel de modélisation UML (Unified Modeling Language) open source. Étant simple d'utilisation, nécessitant peu de ressources système, supportant UML 2, ce logiciel constitue une excellente option pour une familiarisation à la modélisation[27]. Nous avons utilisé le logiciel StarUml pour modéliser les diagrammes UML de spécification et de conception.

### Blasamiq

Balsamiq est un logiciel de conception de wireframes qui permet aux équipes de créer des maquettes et des prototypes interactifs, mais aussi de réaliser des tests utilisateurs. L'outil est destiné à tous ceux qui ont besoin de créer des wireframes, et son utilisation ne requiert pas d'expérience particulière dans le webdesign[28].  
Nous avons Utilisé Balsamiq comme un outils pour faire la conception graphique, et modéliser les maquétes.

## 1.iv Editeurs de textes

### Framework JetBrains

JetBrains est un éditeur de logiciels pour développeurs. Nous choisissons de travailler sur les editeurs de code de JetBrains qui sont :

- IntelliJ (Editeur de texte pour Java)
- WebStorm (Editeur de texte pour le Web (Angular, Html, Css, Javascript))
- Pycharm (Editeur de texte pour Python)

## **Latex**

LaTeX (dont le logo est L<sup>A</sup>T<sub>E</sub>X) est un langage et un système de composition de documents. Il s'agit d'une collection de macro-commandes destinées à faciliter l'utilisation du « processeur de texte » TeX de Donald Knuth[29].

Nous utilisons l'outil Latex pour écrire les rapports.

## **1.v Système de contrôle**

### **Git et GitHub**

Git est de loin le système de contrôle de version le plus largement utilisé aujourd'hui. Git est un projet open source avancé, qui est activement maintenu[30].

GitHub is a for-profit company that offers a cloud-based Git repository hosting service. Essentially, it makes it a lot easier for individuals and teams to use Git for version control and collaboration[31].

Nous avons utilisé Git et Github comme un système de contrôle de version et de superviser l'avancement de projet.

## **1.vi Système d'exploitation**

### **Linux Ubuntu**

Ubuntu est un système d'exploitation GNU/Linux fondé sur Debian. Il est développé, commercialisé et maintenu pour les ordinateurs individuels (desktop), les serveurs (Server) et les objets connectés (Core) par la société Canonical[32].

Nous avons utilisé Ubuntu version 20.04.4 LTS comme l'environnement de développement de l'application.

## 2 Résultat

Dans cette section nous présentons quelques interfaces développées, ou nous présentons le design choisi, et nous montrons le bon suivi des maquettes.

Nous avons choisis les deux interfaces, Téléverser un document par UBooKer IV.1 et comme résultat de téléchargement nous allons consulter ce document ajouté à travers l'interface de consultation de son document par UBooKer IV.2

The screenshot shows the 'New Document' page of the UBooKer IV.1 application. At the top, there is a navigation bar with links: U-BOOK, EVENTS, ADD EVENT, UNIVERSITY, BLANK EXAMS, DOCUMENTS, ADD DOCUMENT, PROFILE, and LOG OUT. Below the navigation bar, the title 'New Document' is displayed next to a decorative icon of a person sitting at a desk with a lightbulb and gears.

The form fields include:

- Document Name\*: Cours de Programmation en C
- Document Type\*: Cours
- Document Subject\*: Info
- Document Level\*: Licence 1st
- Document Year: 2018
- RESET & SELECT RELATED DOCUMENT button
- Document Description: (empty text area)
- Creative Commons Licenses: A dropdown menu showing 'NonCommercial-ShareAlike 4.0 International License'. There is also an 'AJOUTER' button and a link to 'Regarder la vidéo explicatif'.
- Choose File: ProgC-S1.pdf
- Documents: PDF/DOCX/MSWORD/ File Loaded!
- FINISH button

FIGURE IV.1 – Téléverser un document par UBooKer

# Cours de Programmation en C

## Cours Info

Niveau : Licence 1st

[EDITER LES MÉTADONNÉES OU SUPPRIMER](#)



The image shows a presentation slide with a white background featuring a light blue grid pattern. In the top left corner, there is a small icon of three horizontal lines. The top right corner contains a set of standard presentation control icons: a back arrow, a forward arrow, a double back arrow, a double forward arrow, a search icon, a magnifying glass, and a refresh/circular arrow icon. The title 'Notions Génériques' is centered in a large, dark blue serif font. Below the title, a thin blue horizontal line spans most of the width of the slide. At the bottom center, the text 'Année universitaire 2018/2019 – Semestre 1' is written in a smaller, dark blue sans-serif font. In the bottom right corner, there is a small number '1'. The overall aesthetic is clean and professional, resembling a Microsoft Word document or a LaTeX Beamer presentation.

Test licence by MAHDI is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#).  
Élaboré le : 2018

## Commentaires Récents:

1 0

AJOUTER VÔTRE

## My first Comment

 Mahdi Hadj Sassi  
My first Comment

2022-5-8 - 16:32

FIGURE IV.2 – Consultation de son document par UBooKer

# **CONCLUSION**

Dans cette partie nous avons bien exploré l'infrastructure de l'application UBooK qui traite l'aspect ethic la véracité d'information.

Dans la prochaine partie on étudie la partie recherche vers un modèle de véracité pour les document educatifs et pour les evenements universitaires qui s'execute sur l'infrastructure développé.

---

# VERS UN MODÉLE DE VÉRACITÉ

## INTRODUCTION

UBooK est un réseau sociale qui offre des services de la vie estudiantine. Elle possède un espace de gestion d'événement et un espace de partage des documents éducatifs.

Un des grands challenges c'est que le plateforme respecte les ethics et surtout la véracité d'information publié pour les documents éducatifs et les événements universitaires.

Alors dans ce chapitre on va Passer par les événements, on va les définir, enumérer les différent types et la politique fréquemment utilisé.

Puis on s'intéresse sur les documents éducatives, on va différencier entre les types des documents éducatives et leurs métadonnées, Enfin citons une des politique pour les documents les licences Creative Commons

Après on stop sur la véracité où on comprend mieux le concept et comment elle influence les documents et les événements et pour réaliser cet aspect on va raccourcir vers l'intelligence artificiel et elle même utilise l'indexation comme outil pour finir la tâche, ou on va se concentrer sur les deux derniers en les définir et mentionner les différentes types et modèles.

---

# CHAPITRE V

## ETAT DE L'ART

### Introduction

La Data Science, ou science des données, est un mélange disciplinaire entre la data inférence, le développement d'algorithme et la technologie, dont l'objectif est la résolution de problèmes analytiques complexes. Au cœur de ce grand mélange, on retrouve les données, les quantités massives d'informations brutes stockées dans les data warehouses des entreprises. Concrètement, la science des données permet d'utiliser les données de façon créative pour générer une valeur pour les entreprises.

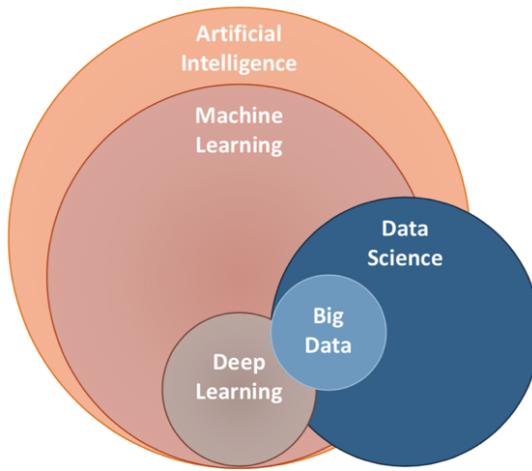


FIGURE V.1 – Data Science : Machine Learning + Big Data

Comme illustré dans la figure ci-dessus V.1, nous pouvons déviser les sciences des données en deux grandes branches : Big Data et L'intelligence artificiel.

Dans ce chapitre nous présentons à la fois le Big Data et le l'artificial intelligence, et comment nous les exploitons pour réaliser la véracité pour les documents et les evenements, à travers les anciens travaux, et terminons vers notre modèle pour répondre à ce besoin.

# 1 Big Data

L'explosion quantitative des données numériques a obligé les chercheurs à trouver de nouvelles manières de voir et d'analyser le monde. Il s'agit de découvrir de nouveaux ordres de grandeur concernant la capture, la recherche, le partage, le stockage, l'analyse et la présentation des données. Ainsi est né le « Big Data ». Il s'agit d'un concept permettant de stocker un nombre indicible d'informations sur une base numérique[33].

Les mégadonnées nous pouvons les présenter sur plusieurs axes, critères, qui sont appelés après les Vs de Big Data.

Le Big data présente maintenant 42V selon[34] parmi eux nous trouvons le Volume, la Vitesse, la Variété, la Véracité, la Variabilité, la Validité, la Vulnérabilité, la Volatilité, la Visualisation et le Valeur.

## 1.i Véracité

Dans ce projet, nous intéressons sur le V Numéro 4 la Véracité, comme une qualité d'information que nous traitons, pour avoir une meilleur résultat.

“C'est l'un des enjeux majeurs de l'exploitation des Big Data. Faux profils sur les réseaux sociaux, fautes d'orthographe, fraudes ... Il est nécessaire de multiplier les précautions (recoupement et enrichissement des données) pour minimiser les biais liés au manque de fiabilité du Big Data[35].

La véracité est le quatrième V du Big Data. Il s'agit de la qualité et de l'exactitude des données. Les données collectées peuvent comporter des éléments manquants, être inexactes ou ne pas fournir d'informations réelles et utiles. La véracité fait globalement référence au niveau de confiance dans les données collectées.

Les données peuvent parfois être confuses et difficiles à utiliser. Si elles sont incomplètes, . Un exemple tiré du domaine médical : si les données sur les médicaments que prend un patient sont incomplètes, la vie du patient peut être en danger.

La valeur et la véracité contribuent toutes deux à définir la qualité et les informations tirées des données. Dans un sens, c'est un facteur d'hygiène. En démontrant la véracité de vos données, vous montrez que vous avez porté un regard critique sur elles.

### Notion de véracité

Qualité morale de celui/celle qui ne trompe pas ou qui n'en a pas l'intention ; en partic., qualité de celui/celle qui se garde de l'erreur et s'emploie à l'éviter dans ses paroles ou dans ses écrits. Synon. bonne foi\*, exactitude, franchise, sincérité, véridicité.

Caractère de ce qui est conforme à la vérité, à la réalité. Synon. authenticité, exactitude. Souci, recherche de l'exactitude, de la fidélité au réel, notamment dans la création artistique et littéraire.

VÉRACITÉ, (Morale.) la véracité ou vérité morale dont les honnêtes gens se piquent, est la conformité de nos discours avec nos pensées ; c'est une vertu opposée au mensonge[36].

La véracité fait référence à la provenance ou à la fiabilité de la source de données, à son contexte et à son importance pour l'analyse qui en découle. Des réponses à ces questions sont nécessaires pour déterminer la véracité de ces informations. La connaissance de la véracité des données nous aide à mieux comprendre les risques associés aux analyses et aux décisions commerciales basées sur cet ensemble de données particulier.

### Véracité Pour les Documents

La véracité dans les documents se fait à travers la conformité des métadonnées avec le contenu du document, en plus un document vide ou bien qui mentionne une information fausse alors il n'est plus vrai. Par exemple un document de grammaire française, ne doit pas être vide ou son contenu est autour de la musique. Si ces conditions sont validés alors ce document est vrai. Aussi que la véracité est lié d'authenticité si l'auteur n'est pas celui le propriétaire de document alors aucune véracité existante.

### Véracité Pour les evenements

La véracité pour les événements signifie que la certitude de déroulement de cet événement, en plus qu'il est bien conforme aux thématiques mentionnées.

Mais la véracité ne se limite pas au événement lui même mais aussi les inscription aux événement doit être vrai car on peut y avoir des faux inscription pour les participants.

## 2 Intéligence Artificiel

### Notion d'IA

L'intelligence artificielle (IA, ou AI en anglais pour Artificial Intelligence) consiste à mettre en œuvre un certain nombre de techniques visant à permettre aux machines d'imiter une forme d'intelligence réelle. L'IA se retrouve implantée dans un nombre grandissant de domaines d'application.

Les domaines d'application de l'intelligence artificielle sont nombreux. Elle est présente dans les appareils photo des smartphones. En mode nocturne, elle permet d'adapter la colorimétrie à l'environnement, et de redonner à une façade éclairée son éclat originel pour le reproduire fidèlement sur votre cliché[37].

En termes simples, l'intelligence artificielle (IA) fait référence à des systèmes ou des machines qui imitent l'intelligence humaine pour effectuer des tâches et qui peuvent s'améliorer en fonction des informations collectées grâce à l'itération[38].

Il y'a qutare écoles en IA :

- Penser comme des humains : Les efforts passionnats pour pousser les ordinateurs à penser d'en fabriquer des machines avec cerveaux au sens le plus latéral.(Haugeland, 1985)
- Penser Rationnellement : L'étude des facultés mentales par l'élaboration des modèles mathématiques.(Charniak et McDermott, 1987)

- Agir comme des humains : L'étude des moyens qui permettent aux ordinateurs de faire des choses qui pour le moment sont mieux faites par des hommes.(Rich et Knight, 1991)
- Agir rationnellement : L'étude et la conception d'agents intelligents.(Poole, 1998)

## **Notion de ML**

L'apprentissage automatique est une branche de l'intelligence artificielle (IA) et de l'informatique qui utilise principalement des données et des algorithmes pour imiter la manière dont les être humains apprennent, en améliorant progressivement sa précision. L'apprentissage automatique est une composante importante du domaine en pleine expansion qu'est la science des données. Grâce à l'utilisation de méthodes statistiques, des algorithmes sont entraînés à effectuer des classifications ou des prévisions, ce qui permet de découvrir des informations essentielles dans le cadre de projets d'exploration des données. Ces informations permettent ensuite de prendre des décisions dans les applications et les entreprises, et ont idéalement un impact sur les principales métriques de croissance. Au fur et à mesure que le big data poursuit son développement et sa croissance, les besoins en spécialistes des données va augmenter, ce qui les obligera à contribuer à l'identification des questions commerciales les plus pertinentes et, par conséquent, des données permettant d'y répondre[39].

En machine learning il y'en a deux branches, l'apprentissage supervisé et non supervisé. Ce dernier consiste à prédire par oui ou non (Vrai ou Faux) à une proposition 'statement' après un entraînement sur deux clusters. Son rôle est la plupart de différencier entre deux groupes. Les fameux modèles sont K-Means, CAH... Et l'apprentissage supervisé lui même a deux sous branches le classification et la régression. Pour le classification son but est de classifier et donner une famille ou un valeur brute pour une donnée, la plupart d'utilisation de classification et de prédire un type de donnée. Les fameux modèles sont : KNN, CNN, Arbre de décision... Mais pour la régression et de donner un intervalle de valeur pour une donnée, la plupart d'utilisation de régression est dans la prédiction des valeur numérique, ou prédiction des pourcentages. Les fameux modèles sont : Linear Régression, Polynomial Regression...

## **2.i Scrapping**

### **Notion de Scrapping**

La capture de données d'écran (screen scraping en anglais) est une technique par laquelle un programme récupère des données normalement destinées à être affichées par un dispositif de sortie vidéo (généralement un moniteur) afin d'en extraire des informations.

Il s'agit souvent de pages web dans lesquelles on souhaite récupérer des informations, mais il peut également s'agir de toute autre forme d'informations qui est formatée avant tout en vue d'être affichée sur un écran. Il peut également s'agir d'informations destinées à un terminal texte, ou encore d'un écran de téléphone cellulaire sur lequel les informations peuvent être analysées après y avoir été affichées par une autre application[40].

Recueillir des données sur le web est parfois compliqué et quand cela est possible, il est difficile de pouvoir les télécharger ou d'effectuer un copier-coller. Le web scraping est une technique permettant l'extraction des données d'un site via un programme, un logiciel automatique ou un autre site. L'objectif est donc d'extraire le contenu d'une page d'un site de façon structurée. Le scraping permet ainsi de pouvoir réutiliser ces données[41].

Le web scrapping est un outil d'extraction des données depuis autres WebSites, à travers un script, et une connaissance sur la structuration de web que nous voulons extraire des informations, nous pouvons réutiliser les données, et les exploiter pour satisfaire notre besoins.

## Modèles de scrapping

Selons le site [42] nous présentons quelques outils de web scrapping :

- Beautiful Soup : c'est vraiment un bel outil pour les scrappers Web en raison de ses fonctionnalités de base. Cela peut aider le programmeur à extraire rapidement les données d'une certaine page Web. Cette bibliothèque nous aidera à extraire les données des fichiers HTML et XML. Mais le problème avec Beautiful Soup est qu'il ne peut pas faire tout le travail tout seul. cette bibliothèque nécessite des modules spécifiques pour fonctionner.
- Selenium : Selenium est conçu pour automatiser les tests des applications Web. Il permet au développeur d'écrire des tests dans un certain nombre de langages de programmation populaires tels que C#, Java, Python, Ruby, etc. Ce framework est développé pour effectuer l'automatisation du navigateur. Jetons un coup d'œil à l'exemple de code qui automatise le navigateur.
- Scrapy : est un framework open source collaboratif qui permet d'extraire les données d'un site web de manière simple et rapide. Développé sous Python, Scrapy dispose d'une grande communauté qui n'hésite pas à créer des modules supplémentaires pour améliorer l'outil.
- Webscraper : est une extension disponible sous Google Chrome qui permet d'extraire les données d'un site internet très rapidement. Web Scraper naviguera sur les sites choisis afin d'en extraire toutes les données. Les données collectées peuvent être exportées sous forme de CSV. L'extension vous permet également de scrapper plusieurs sites à la fois ou même les programmer.
- PhantomBuster : Zéro code et des résultats. Voilà la promesse tenue de PhantomBuster. L'outil offre la possibilité d'extraire les données que vous souhaitez, mais aussi de créer des chaînes d'actions pour générer des pistes d'affaires, des audiences de marketing et une croissance globale. Phantombuster vous donne les outils et le savoir-faire nécessaires pour faire croître votre entreprise plus rapidement. Un outil que nous utilisons chez Wydden et que nous vous présentons dans notre formation Growth Marketing.
- Octoparse :Un des pionniers du scraping web. L'outil Octoparse offre un interface « pointer-cliquer » qui signifie que toute personne qui sait naviguer peut scraper. Aucun code n'est nécessaire. Vous pouvez extraire de données de n'importe quel site web dynamique et récupérer des pages illimitées gratuitement.
- ...

## 2.ii Indexation

### Notion D'indexation

L'indexation automatique de documents est un domaine de l'informatique et des sciences de l'information et des bibliothèques qui utilise des méthodes logicielles pour organiser un ensemble de documents et faciliter ultérieurement la recherche de contenu dans cette collection. La multiplicité des types de documents (textuels, audiovisuels, Web) donne lieu à des approches très différentes, notamment en termes de représentation des données. Elles reposent néanmoins sur un socle de théories communes, telles que l'extraction de caractéristiques, le partitionnement de données (ou clustering), la quantification, et plus généralement la recherche d'information[43].

En général, l'indexation fait référence à l'organisation des données selon un schéma ou un plan spécifique. En informatique, le terme a diverses utilisations similaires, notamment pour rendre les informations plus présentables et accessibles[44].

L'indexation des documents consiste à associer des mots clés et informations à chaque document en fonction du plan de classement, afin de faciliter, ensuite, la recherche, l'accès et le traitement par les utilisateurs.

Alors l'indexation est un ensemble des méthodes et techniques pour l'acquisition, l'organisation, le stockage, la recherche et la sélection d'information pertinente pour un utilisateur

### Modèles D'indexation

Le modèle de recherche d'information RI comprend la fonction de décision fondamentale qui permet d'associer à une requête, l'ensemble des documents pertinents à restituer. Il est étroitement lié au modèle de représentation des documents et requêtes.

L'appariement requête-documents consiste à calculer un score, supposé représenter la pertinence du document vis-à-vis de la requête.

Ce score est souvent calculé à partir d'une fonction ou une probabilité de similarité, en fonction du modèle utilisé, qui tient compte du poids des termes dans les documents. L'assignation d'un score de pertinence à un document permet d'ordonner les documents renvoyés à l'utilisateur. Comme illustré dans V.2 Nous allons présenter quelques modèles, que nous avons besoin après dans l'étude d'ature solutions existants.

- Modèle Booléen : C'est le premier modèle de RI, introduit en 1983 par Salton et McGill, 1 er SRI commercial pour 3 décennies (60-90), Basé sur la théorie des ensembles et l'algèbre de Boole, L'interface d'interrogation de la plupart des moteurs de recherche est basée sur les principes de ce modèle.

Il considère que les termes de l'index sont présents ou absents d'un document

Les poids des termes dans l'index sont binaires :  $w_{i,j} = \{0, 1\}$

Un document est soit pertinent soit non pertinent pour une requête donnée : Pertinence binaire, et jamais partielle.

- Modèle Booléen Pondéré : Extension du modèle booléen en intégrant des pondérations (dénarrant la représentativité d'un terme pour un document)

Fonction de correspondance non binaire (on se passe des implications logiques) basée sur une similarité notée Sim

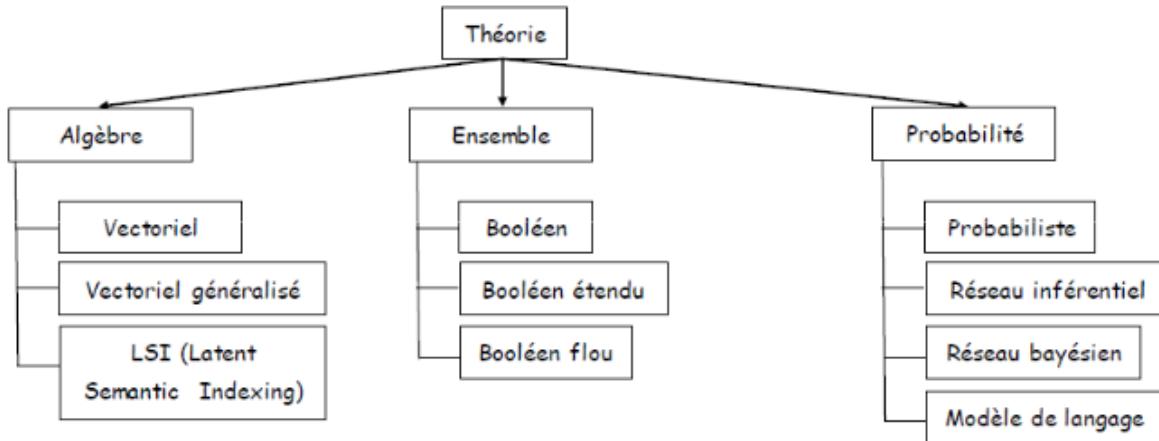


FIGURE V.2 – Classification des modèles selon la théorie

- Modèle vectoriel : Dans des très grandes collections de documents, le résultat de recherche pour une requête donnée dépasse généralement la capacité des utilisateurs à examiner tous les documents de l'ensemble retourné  
Affecter un score (degré de similarité), à chaque document relativement à une requête donnée : le document le plus pertinent devrait avoir le score le plus élevé.
- ...

### 2.iii ML & Documents

Les fameux modèles qui traitent la véracité des documents utilisent le traitement automatique de langage naturel TAN :

Le traitement automatique du langage naturel imite la compréhension humaine des mots et des phrases et permet maintenant aux modèles d'apprentissage automatique de traiter de grandes quantités d'informations avant de fournir des réponses précises aux questions qui leur sont posées. L'utilisation du TAN permet de comprendre le contenu du document et vérifier sa véracité.

Autres modèles qui traitent la véracité sont la classification des documents par l'apprentissage automatique, on mentionne en titre d'exemple KNN, Arbre de décision, Naïve Bayes. Dans la section suivante on suit les travaux anciens et leurs modèles qui traitent la véracité des documents.

### 2.iv Scrapping & Evenement

Le web scraping (parfois appelé harvesting) est une technique d'extraction du contenu de sites Web, via un script ou un programme, dans le but de le transformer pour permettre son utilisation dans un autre contexte, par exemple le référencement [45]

le grattage du web permet de chercher un ensemble d'information sur des sites bien spécifiques, c'est modèle d'IA qui parcours le web sémantique et vérifie les balises HTML. Les meilleurs

bibliothèques dans le web scraping sont : Scrapy (Python), BeautifulSoup (Python), Requests (Python), LXML (Python), Selenium (Python), Jsoup (JAVA) ...

Le grattage du web permet de récupérer les vraies événements en cherchant dans les sites et les pages d'événement confiants

## 3 Travaux anciennes

### Feuille de route

Dans cette partie on va s'intéresser sur les travaux anciens qui est autours la véracité. En premier lieu on étudie les modèles sur la véracité d'information de façon générale En second lieu on étudie les modèles sur la véracité principalement des documents

### 3.i Modèles de Recherche sur la véracité

#### Système d'aide à la décision pour la véracité des données dans le contexte de Big Data

Comme la fiabilité des données est un challenge concrétisé par la dimension de véracité dans le contexte de Big Data, et afin d'extraire des données fiables à partir des systèmes d'information ou les données peuvent incomplète, imprécise, vague,... etc. Ce travail répond à l'objectif de la création d'un système basé sur la théorie de Rough sets\* de véracité de données dans un contexte de Big data.

La théorie des ensembles approximatifs (RST) est une approximation formelle de la théorie des ensembles conventionnelle qui prend en charge les approximations dans la prise de décision. Cette approche peut extraire des connaissances d'un domaine problématique de manière concise et conserver le contenu de l'information tout en réduisant la quantité de données impliquée[46].

#### Modèle probabiliste avec support mutuel des valeurs similaires

Basé sur l'analyse bayésienne, TruthFinder (Yin et al., 2008) calcule la fiabilité de l'information donnée par une source. Il s'appuie sur l'honnêteté de la source d'information et suit l'heuristique selon laquelle si les valeurs fournies sont essentiellement vraies pour de nombreux cas, alors elles seront aussi probablement vraies pour d'autres cas. Mais dans notre cas ce n'est plus fonctionnel car on cherche à un modèle où la source est unique. Et l'information vient d'un seul source, ne peut pas être des sources différents, en respectant la non redondance[47].

#### Découverte de la vérité par corroboration d'informations

C'est un ensemble de trois algorithmes proposés dans (Galland et al., 2010) : COSINE, 2 ESTIMATES et 3-ESTIMATES. COSINE initialise la fiabilité de chaque valeur et de chaque

source. Il calcule de façon itérative la fiabilité d'une source comme une fonction linéaire de sa fiabilité précédenteMais dans notre cas ce n'est plus fonctionnel car on cherche à un modèle ou le source est unique. Et l'information vient d'un seul source, ne peut par être des sources différents, en respectant la non redondance[48].

## Modèle de vérité latente

Ce modèle utilise les réseaux bayésiens pour estimer la fiabilité de l'information. ” Latent Truth Model ” (LTM) est proposé dans (Zhao et al., 2012). Il se base sur deux hypothèses :  
— les données doivent contenir un seul attribut avec une valeur atomique ;  
— la gestion de plusieurs valeurs vraies pour le même attribut. Pour chaque source, LTM considère les probabilités à priori pour qu'elle soit un vrai positif ou un faux positif avec des erreurs négatives. Enfin, les valeurs dont le degré de fiabilité est supérieur à 0.5 sont considérées comme vraies. Il faut noter que le LTM peut ne pas détecter des valeurs vraies pour certains attributs[49].

## Découverte de la vérité par estimation de la vraisemblancemaximale

Basée sur la maximisation de la vraisemblance pour quantifier la fiabilité des sources et la justesse des valeurs, ” Maximum Likelihood Estimation ” (MLE) est proposé dans (Wang et al., 2012). MLE traite seulement les observations booléennes positives et ignorent celles négatives. Dans son algorithme, MLE initialise les paramètres des sources :  
—  $a(s)$  probabilité pour que la source  $s$  signale une valeur vraie et elle est effectivement vraie  
—  $b(s)$  probabilité pour que la source  $s$  signale une valeur vraie alors qu'elle est en réalité fausse. De façon itérative, MLE calcule la probabilité conditionnelle d'une valeur  $v$  d'être vraie sur la base des probabilités ( $a(s)$ ,  $b(s)$ ) de sa source et celle des sources ne fournissant pas la valeur  $v$ . Ensuite la confidence de chaque valeur est calculée itérativement. Après cela, MLE procède à la mise à jour des probabilités ( $a(s)$ ,  $b(s)$ ) de chaque source. Les itérations s'arrêtent à la convergence de  $a(s)$  et  $b(s)$ . Une importante observation faite dans les expériences de l'article (Waguilah and Berti-Equille, 2014) permet de dire que MLE ne peut pas être utilisé avec de grands nombre de sources ( $i \geq 5000$ ). Mais dans notre cas ce n'est plus fonctionnel car on cherche à un modèle ou le source est unique. Et l'information vient d'un seul source, ne peut par être des sources différents, en respectant la non redondance[50].

## Découverte de la vérité avec dépendance de sources par copie

Ici, nous parlerons des algorithmes de recherche de vérité qui tiennent compte de la relation qui existe entre sources. DEPEN est le premier proposé dans (Dong et al., 2009). C'est un modèle bayésien de recherche de vérité qui prend en considération les relations de copie qui peuvent exister entre les sources en pénalisant le vote d'une source si elle est détectée comme étant la copie d'une autre source. Le calcul de la matrice de dépendance entre les sources est très coûteux surtout lorsque les données en entrée sont volumineuses ; c'est un des goulots d'étranglement majeurs de DEPEN et ses extensions[51].

## **Analyse de crédibilité latente**

LCA (" Latent Credibility Analysis ") est un modèle probabiliste proposé dans (Paster-nack and Roth, 2013). Il utilise aussi l'algorithme de maximisation de la vraisemblance pour calculer la probabilité des valeurs en regroupant les mêmes attributs d'un même objet en une même donnée dans un ensemble d'exclusion mutuelle où il n'existe qu'un seul élément vrai. LCA est une approche flexible et puissante pour modéliser le problème de la crédibilité de l'information[52].

## **Découverte de la vérité dans plusieurs domaines de sources contradictoires ayant des vérités multiples**

Ce modèle est un modèle probabiliste et bayésien à la fois, qui intègre le score d'expertise de domaine et le score de confiance pour la détermination de la vérité. DART (Domain AwaRe Truth Discovery) est un modèle proposé dans (Lin and Chen, 2018). L'idée de base est la construction d'un modèle, qui contient deux composantes intégrales : la modélisation de l'expertise du domaine concernant la richesse des données, et la modélisation de l'agrégation de la vérité à partir des réponses de chaque source[53].

## **Découverte de la vérité avec partitionnement des attributs**

Proposée dans (Lamine Ba et al., 2015), c'est une méthode de recherche de vérité qui se base sur le partitionnement des attributs des objets. Son objectif est donc d'estimer une partition optimale de l'ensemble des attributs de telle sorte qu'en appliquant indépendamment n'importe quel algorithme de recherche de vérité de référence sur chaque sous ensemble de la partition, on maximisera la précision de cet algorithme. L'estimation de la partition optimale dans ce cas se base seulement sur la qualité de l'algorithme de recherche[54].

## **Modèle probabiliste pour la découverte de la vérité avec des corrélations d'objets avec des contraintes**

Ce modèle probabiliste se base sur la corrélation entre les objets en tenant compte de certaines contraintes qui lui sont fournies en entrée. " Constrained Truth Discovery " (CTD) est proposé dans (Yang et al., 2019) et formulé comme un problème d'optimisation sous contraintes. Le processus de découverte de la vérité intègre les contraintes de type " Denial Constraints (DCs) " (Chomicki and Marcinkowski, 2005) à l'aide d'un formalisme basé sur la logique du premier ordre universellement quantifié qui peut exprimer un grand nombre de relations effectives et largement existantes entre les objets. Sur cette base les auteurs proposent des algorithmes pour partitionner les objets en groupes disjoints en générant des contraintes arithmétiques pour chaque groupe disjoint séparément. Ensuite, les vraies valeurs des attributs des objets dans chaque groupe disjoint sont dérivées en minimisant une fonction objective sous les contraintes arithmétiques correspondantes[55].

## **Vérification de faits par partitionnement de données**

Dans ce travail, nous intéressons la vérification de faits dans un domaine où les données ont une structure inhérente qui n'est pas connue à l'avance. Pour résoudre ce problème il a été

conçu et proposé un algorithme, appelé TD-AC, de partitionnement de données intelligent basé sur la méthode de clustering des données des k-moyennes du domaine de l'apprentissage automatique. Pour choisir la partition optimale, l'indice de silhouette a été choisis. Ensuite, (OSIAS NOËL) a proposé une validation des performances de l'algorithme. Pour ce faire, il l'a comparé à l'approche proposée dans (Lamine Ba et al., 2015) et aux algorithmes de découverte de la vérité standards sur des jeux de données synthétiques, semi synthétiques et réelles. Enfin, (OSIAS NOËL) a montré que TD-AC a une complexité en temps comparable à celle des algorithmes standards contrairement à l'algorithme brute force dans (Lamine Ba et al., 2015). Cependant, nous avons observé que lorsque le jeu de données est caractérisé par beaucoup de valeurs, cet approche est moins performante : ceci à cause de l'utilisation d'une matrice creuse comme entrée de l'algorithme de clustering entraînant une difficulté à trouver la partition optimale[56].

### **3.ii Modèles de Recherche sur la véracité pour les documents éducatifs**

#### **Incorporation de phrases (SIF)**

Les travaux qui ont combinaient l'incorporation de mots en utilisant des opérations sur des vecteurs et des matrices pour dériver l'incorporation de phrases ou de phrases. Les résultats ont montré que l'exploitation des vecteurs par multiplication par coordonnées permet d'obtenir de très bonnes performances dans les opérations binaires étudiées. Et concentré sur les représentations distribuées des phrases. Ils nécessitaient généralement une analyse syntaxique et il a été démontré que le résultat fonctionnait pour les représentations au niveau de la phrase. Une autre approche a mis en place un algorithme non supervisé pour apprendre les représentations distribuées de phrases ou de documents. Leurs expériences ont indiqué que leur méthode était compétitive avec les méthodes de pointe[57].

#### **Analyse syntaxique**

L'analyse syntaxique est l'une des technologies de base du traitement du langage naturel et la pierre angulaire d'une compréhension approfondie du langage. La tâche de l'analyse syntaxique est d'identifier les composants syntaxiques contenus dans la phrase et la relation entre ces composants, en utilisant généralement des arbres d'analyse pour représenter les résultats de l'analyse syntaxique[58].

#### **Calcul de l'incorporation de phrases en fusionnant l'arbre d'analyse syntaxique et l'incorporation de mots**

L'article propose une méthode d'incorporation de phrases basée sur les résultats de l'arbre d'analyse syntaxique et des vecteurs de mots. L'ordre de fusion des noeuds dans l'arbre d'analyse garantit que la méthode est capable de conserver l'ordre des mots dans les phrases. Et (Yong Wang, Shuixiu Wu) considèrent les balises de l'arbre d'analyse comme des paramètres de poids, qui capturent les informations syntaxiques dans l'incorporation de phrases. Ainsi, la méthode proposée a le potentiel de surmonter la faiblesse des méthodes existantes. De plus, il est rapide à calculer et à apprendre les paramètres, en plus d'obtenir des performances

meilleures ou comparables que SIF traditionnel sur diverses tâches de similarité textuelle. De plus, il ya aussi quelques problèmes dans la méthode qui devraient être améliorées. Par exemple, l'apprentissage des paramètres doit être optimisé, comme la conception d'une fonction cible d'apprentissage non supervisée pour obtenir de meilleures performances[59].

### **Conclusion vers notre modèle**

Nous avons présenté ci-dessus quelques récents travaux sur les méthodes de recherche de vérité. Ces méthodes se basent dans la majeure partie des cas sur le vote pour le calcul de la confidence, et utilisent en retour ces scores de confidence pour estimer la fiabilité des sources. Il faut noter que certaines méthodes y ajoutent la dépendance entre les sources et la similarité entre les valeurs pour booster la précision du processus dans certains types d'applications bien précises.

Mais dans le cas d'un nouveau auteur qui poste un nouveau document, automatiquement la véracité n'est pas réspecté malgré que le source puisse être confiant et fiable, alors nous cherchons d'un modèle qui traite un document selon d'autre critères plus que le source, dans la chapitre suivant nous détaillons notre modèle pour valider la véracité.

---

# CHAPITRE VI

## MODÉLE DE UBOOK POUR LA VÉRACITÉ DES DOCUMENTS

### 1 Présentation du modèle

Pour vérifier la véracité d'un document éducatif, nous analysons, les métadonnées plus précisement : la discipline, l'auteur et l'année de ce document.

L'utilisateur dans UBooK va entrer les champs : matière, année et la licence creative commons qui contient l'information sur l'auteur de ce document, et bien évidemment le document lui même.

Nous allons raccourir à extraire les informations de l'année, et l'auteur à travers le scrapping de document, en utilisons les outils qui lisent les métadonnées de documents d'extension soit PDF soit DOCX ou PPTX.

Et pour la matière de document, nous développons un modèle de machine learning qui vérifie la discipline de ce document.

Et pour réaliser ce modèle, nous sommes besoin d'une Data Set, pour l'entrainer selon un algorithme d'apprentissage.

Enfin nous comparons les champs saisies par le UBooKer et les valeurs trouvées par le grattage de document, si il sont conforme, nous assumons que la véracité est respectée.

#### Data Set

Puisque UBooK est encore dans la phase de départ, nous choisissons comme Data Set les matières globales :

- Matière Santé : Englobe tous ce qui est médecine, pharmacie, infirmerie, anesthésie ...
- Matière SVT : Englobe tous ce qui est sciences de la vie, biologie, la terre, géologie ...
- Matière Chimie : Englobe tous ce qui est chimie organique, agro élémentaire, ...
- Matière Physique : Englobe tous ce qui est mécanique, électrique, astronomie ...
- Matière Mathématique : Englobe tous ce qui est Analyse, Algèbre, probabilité ...
- Matière Informatique : Englobe tous ce qui développement, réseau, télécommunication ...
- Matière Litérature : Englobe tous ce qui est Français, Anglais, Arabe ...
- Matière Gestion : Englobe tous ce qui est finance, économie, fiscalité ...
- Autre : Les reste des matières comme musique, dessin ... que nous allons les ajouter dans d'autre rubrique.

Et pour chaque rubrique nous pouvons le diviser en sous rubrique dans le progression de l'application et ainsi de suite, jusqu'à y avoir une granularité très fine.

## Algortihme d'apprentissage

Nous sommes dans la phase de classification des documents selon les classes mentionnée dans sous section précédente.

Alors les fameux algorithmes de classification sont KNN, Naive Bayes, Arbre de décision, Forêt de décision...

Nous avons éliminer l'algorithme KNN (K Nearest Neighbours) car il consomme trop de ressources informatique et matérielle, Surtout sur une data set détaillé, selon[60] La complexité dans KNN est de l'ordre de  $O(n)$  avec  $(k \parallel n)$  avec  $n$  est le nombre de document dans toute la dataset.

Par conséquence, nous avons 9 classe ( $k = 9$ ) et pour chaque  $k$  supposons qu'il y'a au moins 30 documents alors  $n = k * 30 = 9 * 30 = 270$ , d'où la complexité est exponentielle.

Pour l'algorithme Naive Bayes malgré qui' il est conçu pour le classification mais il implique que chaque fonctionnalité soit indépendante, ce qui n'est pas toujours le cas. Dans notre cas il peut y'avoir des documents dépendant à deux matières, comme le mécanique et l'électronique dans un cours de physique.

Pour l'algorithme Arbre de décision, et totalement pertinent si il réponds à un question de vrai ou faux, mais dans le cas de plusieurs réponses, souvent il donne des résultat médiocre. D'où vient la foret de décision pour ne se limitera pas à la réponse de la question oui et non, l'arbre de décision est facile à interpréter les résultats par l'oeil.

Alors nous choisissons la combinaisons entre l'arbre et la foret des décisions, Un ensemble des arbres de décisions qui forment une foret, comme indiqué dans la figure ci-dessous VI.1.

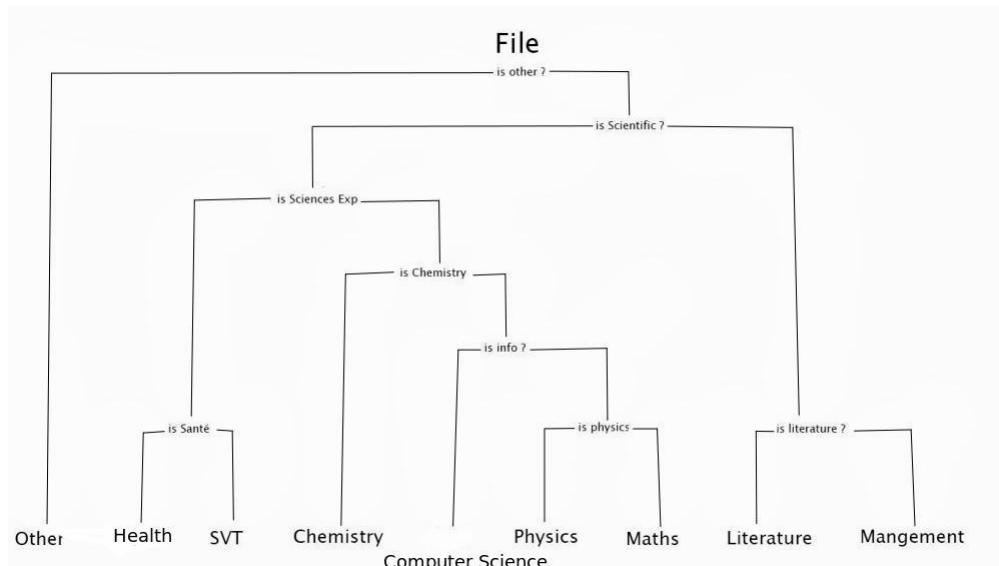


FIGURE VI.1 – Arbres de décision de UBooK

## 2 Pseudo-Algorithmes

### 2.i Extraction des métadonnées d'un document

Pour extraire les informations d'année et d'auteur, nous allons importer et utiliser des bibliothéques spécialisé à extraire les métadonnées des fichiers et des documents, et selon l'extension de fichier nous allons choisir bien quelle outil qu'on va raccourir. Après nous récupérons les champs Année et Auteur.

```
1  \\ Import libraries for PDF, Word, PowerPoint Document
2  import PDF
3  import DOCX
4  import PPTX
5
6  \\ Define variables
7
8  var file = "" \\ this is the link for the file we're treating
9  var metaData = {"Author": "", "Year": ""} \\ this is a dictionary for the
10 metdata
11
12 \\ Method for PDF Files
13
14 Method metaDataFromPdf (file) {
15     pdf = PDF(file)
16     data = pdf.getDocumentInfo()
17     metaData["Author"] = data["Author"]
18     metaData["Year"] = data["Year"]
19 }
20
21 \\ Method for DOCX Files
22
23 Method metaDataFromDocx (file) {
24     doc = DOCX(file)
25     data = doc.getDocumentInfo()
26     metaData["Author"] = data["Author"]
27     metaData["Year"] = data["Year"]
28 }
29
30 \\ Method for PPTX Files
31
32 Method metaDataFromPptx (file) {
33     ppt = PPTX(file)
34     data = ppt.getDocumentInfo()
35     metaData["Author"] = data["Author"]
36     metaData["Year"] = data["Year"]
37 }
38
39 \\ Method for determining the file either PDF, DOCX or PPTX
40
41 Method detectFile (file) {
42     if (file.endsWith('.pdf')) {
        metaDataFromPdf(file)
```

```

43     } if (file.endsWith('.docx')) {
44         metaDataFromDocx(file)
45     } if (file.endsWith('.pptx')) {
46         metaDataFromPptx(file)
47     }
48 }
49
50 // Main Method
51
52 Method init(documentLink) {
53     this.file = documentLink
54 }
```

## 2.ii Recherche de discipline d'un document

Pour la recherche de matière d'un document, nous allons utiliser les mots dans ce document comme featuris pour vérifier son véracité, et pour le réaliser nous choisissons l'indexation comme outil à extraire les features

**Extraction des text** Tout d'abord Nous allons extraire le fichier text des document PDF, DOCX et PPTX.

```

1 // Import libraries for PDF, Word, PowerPoint Document
2 import PDF
3 import DOCX
4 import PPTX
5
6 // Define variables
7
8 var File = "" // this is the link for the file we're treating
9 var txt = "" // this is the result of the txt inside the document
10
11 // Main Method
12
13 Method init(documentLink) {
14     this.file = documentLink
15 }
16
17 // Method for determining the file either PDF, DOCX or PPTX
18
19 Method detectFile (file) {
20     if (file.endsWith('.pdf')) {
21         txtFromPdf(file)
22     } if (file.endsWith('.docx')) {
23         txtFromDocx(file)
24     } if (file.endsWith('.pptx')) {
25         txtFromPptx(file)
26     }
27 }
```

```

28
29     \\ Method to Extract text from PDF
30
31     Method txtFromPdf( file ) {
32         pdf = PDF( file )
33         for i in range ( pdf.numberOfPages ) {
34             page = pdf.getPage(i)
35             txt += page.getText()
36         }
37     }
38
39     \\ Method to Extract text from DOCX
40
41     Method txtFromDocx( file ) {
42         doc = DOCX( file )
43         for i in range ( doc.numberOfParagraphs ) {
44             paragraphe = doc.getParagraph(i)
45             txt += paragraphe.getText()
46         }
47     }
48
49     \\ Method to Extract text from PPTX
50
51     Method txtFromPptx( file ) {
52         ppt = PPTX( file )
53         for i in range ( ppt.numberOfSlides ) {
54             slide = ppt.getSlide(i)
55             txt += slide.getText()
56         }
57     }

```

**Nettoyage de Text** Le nétoyage de text est une étape fondamentale de l'indexation pour avoir une résultat plus pertinent, le nétoyage du texte se fait sur plusieurs étapes.

1. Suppression des espaces extra
2. Segmentation et tokenisation
3. Suppression des mots communs en français et en anglais comme ('with', 'alors', 'why', ...)
4. Suppression des caractères spéciaux, accentués, comme (é -> e, à -> a, ü -> u, ...)
5. Normaliser les mots à travers la racinisation et la lemmatisation

```

1     \\ Import Libraries
2
3     import EnglishFrenchStopWords
4     import removingAccents
5     import stemming
6     import lemmatisation
7
8     \\ Define variables
9
10    var tokens \\ List of tokens

```

```

11
12  \\ Method removing extra spaces
13
14 Method RemovingExtraSpaces(txt) {
15     txt = txt.trim() \\Removing spaces from the beginning and the end of
16     the string
17     txt = txt.replace("\n","") \\Removing back to line from string
18     txt = txt.replace("\t","") \\Removing tabulations from string
19 }
20
21  \\ Method tokenisation
22
23 Method Tokenization(txt) {
24     tokens = txt.split(" ") \\split text into list with the space as the
25     indicator
26 }
27
28  \\ Method Removing stop words
29
30 Method StopWords(tokens) {
31     var tokens2
32     for i in tokens {
33         if (i not exist in EnglishFrenshStopWords)
34             tokens2.add(SpecialCarac(i))
35     }
36     tokens = tokens2
37 }
38
39  \\ Method Removing special caracteres
40
41 Method SpecialCarac (token) {
42     return removingAccent(lemmatisatoin(stemming(token)))
43 }
```

**Vérifier avec le modèle** Nous avons maintenant une liste des mots (tokens) de ce document, nous utilisons le modèle de machine learning qui a entraîné sur un ensemble des token lui aussi, et ce dernier donne le résultat prédit de matière de document .

```

1
2  \\ Import Librairies
3
4 import MachineLearningVeracitySubjectModel
5
6  \\ Declaring variables
7
8 var tokens = "" \\ list of tokens extracted from the last pseudo code
9 var subject = "" \\ result we're searching for
10
11  \\ Main method
12
13 init(tokens) {
```

```

14     this.tokens = tokens
15     model = MachineLearningVeracitySubjectModel.load()
16     subject = model.predict(tokens)
17 }
```

## 2.iii Résultat final

Enfin Maintenant nous avons tous les informations nécessaires pour vérifier la véracité de document.

```

1  \\ Import Librairies
2
3 import MetaDataVeracityService
4 import SubjectVeracityService
5
6 \\ declaring variables
7 var author =""
8 var year =""
9 var subject = "" \\ this is the subject entered by the UBooKer in the field
10 var file =""
11 var TrueSubject = ""\\this is the subject predicted by the model
12 var metada
13 \\ Main Method
14 init(file ,author ,year ,subject) {
15     this.author = author
16     this.year = year
17     this.subject = subject
18     this.metada = MetaDataVeracityService(file).metaData
19     this.TrueSubject = SubjectVeracityService(file).subject
20     if (this.author == metadata['Author'] && this.year == metada['Year'] &&
21         this.TrueSubject == subject) {
22         \\ treatment of veracity is confirmed
23     } else {
24         \\ treatment of veraciy is not confirmed
25     }
}
```

Enfin, nous appliquons ces pseudo algorithmes sur une language de programmation et sur une vrai data Set. Dans la section suivante nous intéresson de plus sur l'implementation de modèle pour la vérification de véracité des documents.

## 3 Implémentation

Dans cette section, nous allons introduire l'environnement de travail sur laquel nous avons implémenter notre modèle de recherche sur la véracité des document educatifs.

### 3.i Environnement de travail

Comme mentionnée dans la partie 1 (l'infrastrucutre logiciel) de ce rapport dans le chapitre Réalisation et développement, dans la section Environnement de travail, que nous allons présenter la langage de programmation Python. Nous avons utilisé Python (v 3.8.10) comme langage pour créer le modèle de recherche, avec les bibliothéques :

- PyPDF2 : La bibliothéque pour traiter les fichier PDF
- python-pptx :La bibliothéque pour traiter les fichier PPTX
- python-docx :La bibliothéque pour traiter les fichier DOCX
- nltk : La bibliothéque pour le traitement automatique de langage naturel
- numpy : La bibliotéhque pour gérer les liste et les tableaux
- pandas : La bibliothéque pour gérer les DataFrame
- joblib : La bibliotéhque pour sauvegarder et charger les modèles atteints.
- sklearn : La bibliothéque de machine learning pour les différent modèles.
- os : La bibliothéque pour les commandes systèmes

Nous avons utilisé PYCharm de JetBrains comme éditeur de texte, avec le framework Jupyter pour travailler sur les notebooks sur le système d'exploitation Ubunutu 20.04 LTS. Et la réalisation de modèle nécessite une data Set pour s'entraîner, et créer les paramètres internes et externes de modèles, dans la sous section suivantes nous présentons notre set de données.

### 3.ii Collection des Données

Les données a colléctés sont : Des différents documents (cours, tp, td, résumé, examen, ds) sur plusieurs dicipline, et les regrouper selon les arbres de décisions montrés dans la section présentation du modèle dans le même chapitre.

Nous avons raccourir à la fameuse application web Kaggle [61] , mais nous avons pas trouvé les documents que nous en train de chercher, d'où nous avons questionnée des étudiants dans des filières différentes pour nous fournir leurs documents éducatifs, en plus de quelques recherches sur internet, nous avons établis notre dataSetVI.2.

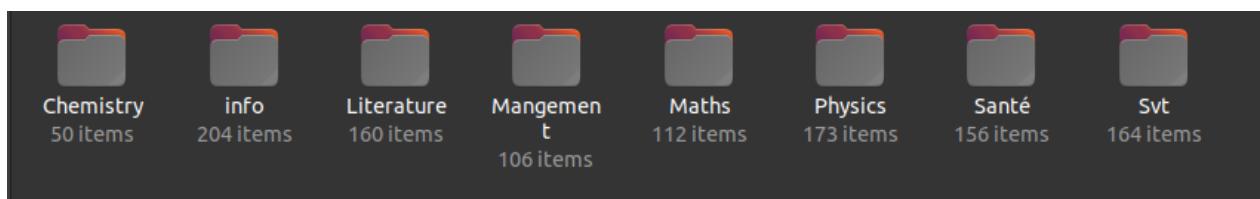


FIGURE VI.2 – DataSet de UBooK

---

# CHAPITRE VII

## MODÉLE DE UBOOK POUR LA VÉRACITÉ DES EVENEMENTS

### 1 Présentation du modèle

La vérification du véracité des evenements se fait à travers la vérification de crédibilité de déroulmenet de l'evenement, et la vérification des inscriptions.

Alors la véracité des evenement est pour vérifier l'inexistance de faux inscripteurs, et des faux evenement.

**Véracité de déroulement d'evenement** Pour garantir la crédibilité des événement présente dans le plateforme UBooK, nous grattons le web et nous affichons les evenements présents sur d'autre sites comme 4C, 10Times, en garantissons leurs véracité, ou en utilisant la coté psychique de l'utilisateur, rarement qu'une personne paye un evenement pour poster un faux evenement.

Alors tous d'abord, si un UBooKer à posté un evenement, et ce UBooKer est supérivsé par un autre UBooKer comme un club ou une institus, alors nous vérifions avec ce supérviseur par mail le déroulement de l'evenement si il a confirmé alors, la véracité est respécté, sinon le supérviseur n'a pas confirmé ou le UBooKer est indépendants, dans ce cas il y'a une option de payer pour poster l'evenement sur le plateforme, si'il paye alors la véracité est respécté, sinon l'évenement restera caché, jusqu'a une des deux solutions accure, soit payement, soit confirmation d'après le supérviseur.

**Véracité de inscriptions** Pour garantir que la pluspart des inscripteurs vont être present dans l'evenement, il est obligatoire d'entrer le login et le mot de passe lors de l'inscription, et comme ca les inscripteurs sont sérieux de rejoindre l'evenement.

### 2 Pseudo-Algorithmes

Nous allons présenter la partie de scrapping the web pour récupérer les evenement sérieux dans les sites et les pages officiels des evenements.

Nous somme besoins d'une connaissance sur la structuration de la page que nous en train de gratter.

```
1    \\\ import librairies
```

```
2
```

```

3 import outilDeGrattage
4 import RequêtesHttp
5
6 \\ declare variables
7 var webpage = "" \\ c'es le lien vers la page
8 var src = "" \\ c'est le contenu HTML de la page
9 var content = ""\\ c'est le resultat de grattage
10 var events = [] \\ la liste des evenement qu'on cherche
11
12 webPage = RequestHttp.get('Le lien vers le site')
13 src = webPage.text
14 content = outilDeGrattage(src)
15 events = content.findall('<div class = "event">')

```

### 3 Implémentation

Pour l'implémentation de web scrapping nous choisissons Python (v 3.8.10) comme langage de programmation avec les outils :

- Beautiful Soup : Outil de grattage, rapide et facile à utiliser
- Requests : Outil pour envoyer les requetes Http, nous l'avons utiliser pour charger la page web qu'on est en train de gratter
- mysql.connector : Pour ajouter dans la base de donnée de l'application UBooK les resultats trouvés.

Implémenter sur l'editeur de texte PYCharm de JetBrains sur le système d'exploitations Linux Ubunutu v20.04 LTS

---

# CHAPITRE VIII

## RÉSULTAT ET DISCUSSION

Dans ce chapitre nous analysons le modèle de recherche de véracité pour les documents éducatifs, nous avons utiliser les métriques de bibliothèques sklearn MAE (Mean Absolute Error), MSE (Mean Squared Error), RMSE (Root Mean Square Error), R2 Score comme résultat finale des métriques.

Nous avons calculé le score de test pour chaque arbre de décision illustré dans le chapitre Modèle de UBooK pour la véracité de document, dans la section Présentation du modèle, sous-section Algorithme d'apprentissage.

Listons quelques exemples de résultat VIII.1 VIII.2 .

```
print("Train score", model.score(X_train,y_train))
print("Test score", model.score(X_test,y_test))

Train score 0.8675078864353313
Test score 0.875
```

FIGURE VIII.1 – Résultat de test de l'arbre de décision IsScientific

```
model.fit(X_train,y_train)
print("Train score", model.score(X_train,y_train))
print("Test score", model.score(X_test,y_test))

Train score 0.9068219633943427
Test score 0.7483443708609272
```

FIGURE VIII.2 – Résultat de test de l'arbre de décision IsInformatique

Les résultats de ce modèle sont assez pertinents à cause de la qualité de dataSet, il n'y a pas une bonne répartition pour les différents matière, comme illustré dans la soussection dataSet de section présentation du modèle, la rubrique chimie contient jusute 50 élément par rapport la section info contient 204 élément.

Comme solution pour résoudre ce problème, il est nécessaire de lancer l'application et commence à collécter les documents de tous les sections, et basant sur la nouvelle base de donnée, nous améliorons notre modèle de recherche de véracité pour les documents éducatifs.

Si l'application UBooK fonctionne parfaitement avec ce modèles, lorsque on donne un document avec des faux métadonnées, une alerte détaillé sera envoyé vers le UBooKer qui a posté le document comme montré dans la figure suivanteVIII.3.

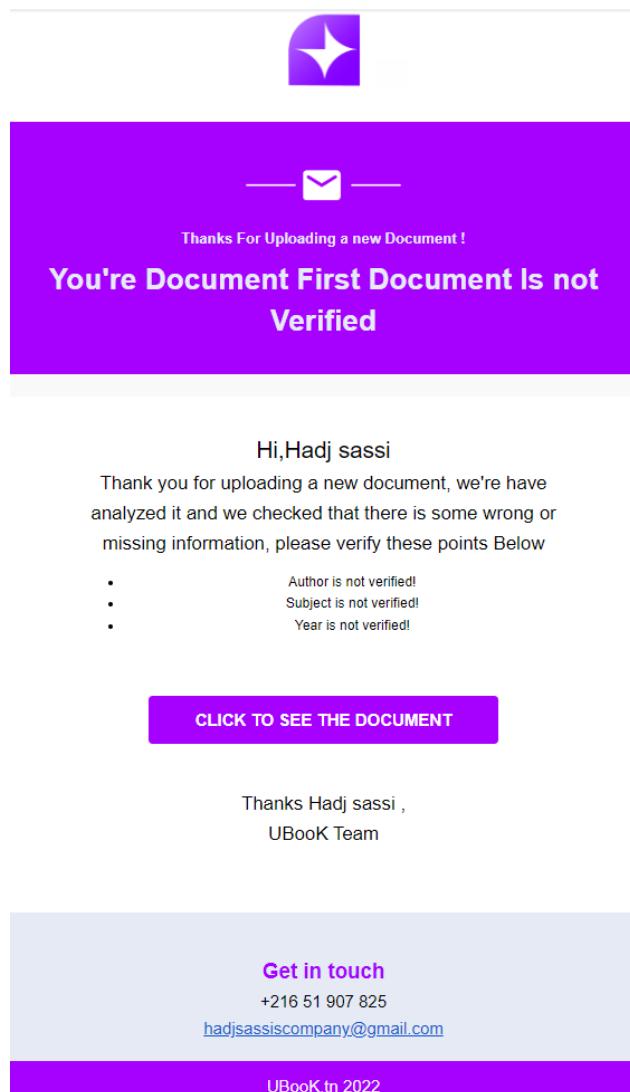


FIGURE VIII.3 – Résultat de mail pour la véracité n'est pas respecté

Et Si les métadonnées sont respectés, alors le document sera labélisé dans l'application Vérifié, et un mail sera envoyé pour remercier le UBooKer comme illustrés dans les figures suivantes VIII.4 VIII.5.

**U-Book**
[EVENTS](#)
[ADD EVENT](#)
[UNIVERSITY](#)
[BLANK EXAMS](#)
[DOCUMENTS](#)
[ADD DOCUMENT](#)
[PROFILE](#)
[LOG OUT](#)

[ ]
[ ]

10 /Page

Type

Subject

Level

Year

Results :

Cours	Cours	TD
First Document MD 7th SVT	Second Document MD 6th Management	fqsdfqsdfsq MD 6th Maths
Examen	TP	TD
3al 7it MD 4th Chemistry	dawa7 ya mbarka MD 6th Management	blablabla MD 6th Info
TD <input checked="" type="checkbox"/>		Cours
First Document 1 MD 2nd Literature		First Document MD 5th SVT

« 1 »

**U-Book**
[EVENTS](#)
[ADD EVENT](#)
[UNIVERSITY](#)
[BLANK EXAMS](#)
[DOCUMENTS](#)
[ADD DOCUMENT](#)
[PROFILE](#)
[LOG OUT](#)

## First Document

1

TD Literature

Niveau : MD 2nd

[EDITER LES MÉTADONNÉES OU SUPPRIMER](#)

FIGURE VIII.4 – Résultat d'un document avec une véracité réspcté

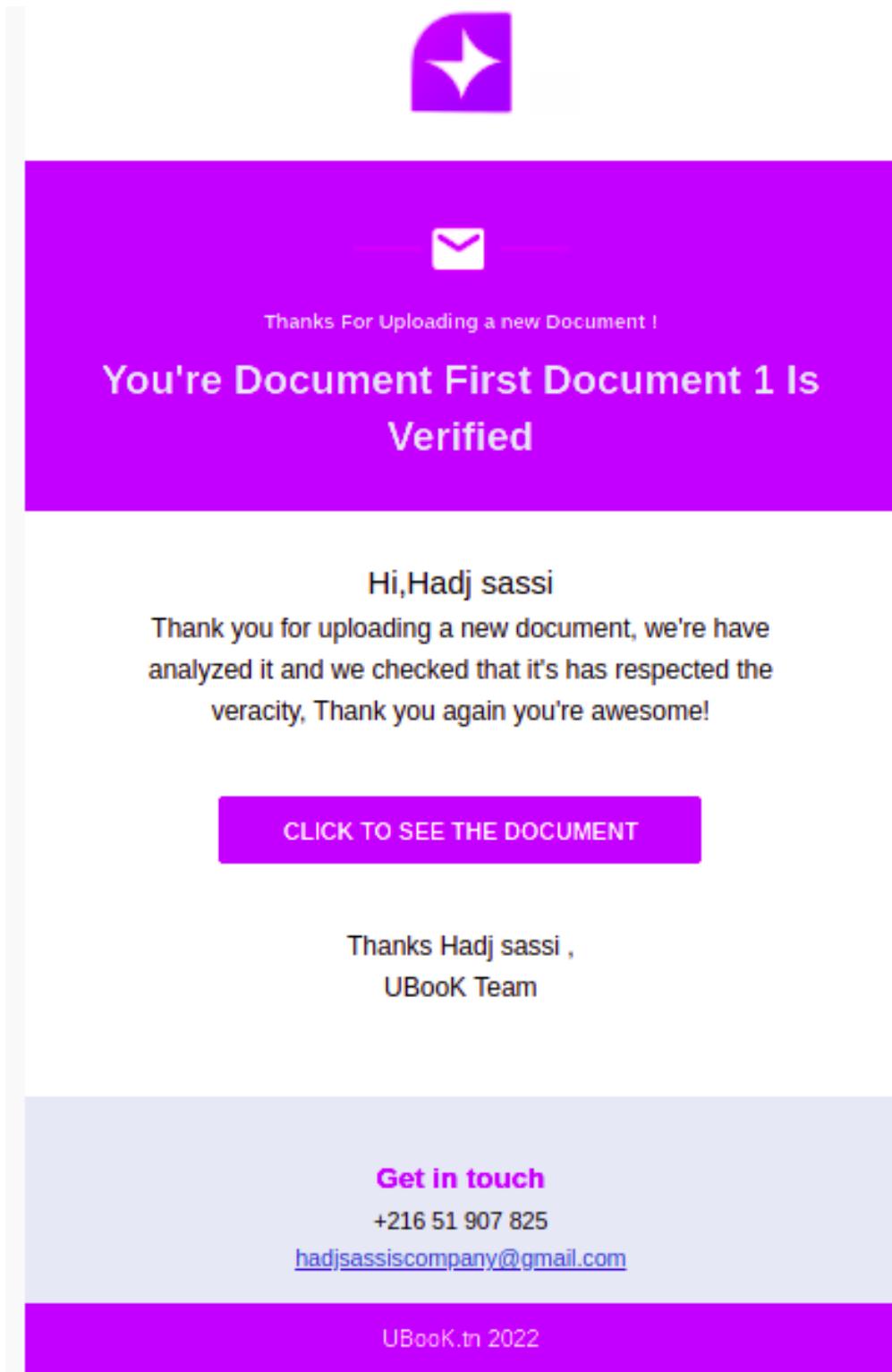


FIGURE VIII.5 – Résultat de mail pour la véracité est résepecté

# CONCLUSION GÉNÉRALE

Dans le cadre du projet de fin d'études, nous avons participé au développement d'une application d'échanges de documents, et de gestion d'évenement en résepéctant les ethics de véracité. Le présent rapport détaillé toutes les étapes par lesquelles nous sommes passées pour arriver au résultat attendu.

Nous avons commencé par comprendre le contexte général du projet et identifier les différentes exigences du futur système. Nous avons préparé, par la suite la conception de l'application, ensuite nous l'avons développée sur Angular comme Front End et Spring comme Back End. Après nous avons ajouté la partie de machine learning pour vérifier la véracité des documents et des evenements, en utilisant principalement Python comme langage de programmation.

Finalement, notre travail ne s'arrête pas à ce niveau. En effet, plusieurs perspectives s'offrent à ce projet on situe le développement de la rubrique examen blancs et la création des document personnels.

# TABLE DES FIGURES

I.1	Liste types des documents educatives . . . . .	13
I.2	Liste des licences creative commons . . . . .	15
I.3	Les Types des evenement universitaires . . . . .	17
I.4	SlideShare . . . . .	19
I.5	Moodle . . . . .	20
I.6	Google Classroom . . . . .	21
I.7	Les pages facebook pour l'échange des documents . . . . .	21
I.8	4C . . . . .	22
I.9	10times . . . . .	23
I.10	Positionnement de UBooK par rapport les autres soultions existants . . . . .	25
I.11	Cycle de développement : Modéle ev V . . . . .	26
II.1	Le Diagramme de Use Case . . . . .	28
II.2	Diagramme de séquence system upload document par UBooKer . . . . .	30
II.3	Maquéte download document par UBooKer . . . . .	31
II.4	Diagramme de séquence Inscription dans un event par UBooKer . . . . .	33
II.5	Diagramme de séquence d'ajouter un event par UBooKer . . . . .	35
II.6	Maquétes Ajout Evenement Par UBooKer . . . . .	36
II.7	Diagramme de séquence system consulter evenement par Internaute . . . . .	37
II.8	Diagramme de classe d'Analyse . . . . .	38
III.1	Diagramme de Package . . . . .	40
III.2	Diagramme de classe d'Analyse . . . . .	41
III.3	Diagramme de classe d'Analyse . . . . .	42
III.4	Diagramme de séquence objet consultation liste documents . . . . .	43
III.5	Diagramme de séquence objet upload document . . . . .	44
III.6	Diagramme de séquence objet inscription dans un event . . . . .	45
IV.1	Téléverser un document par UBooKer . . . . .	49
IV.2	Consultation de son document par UBooKer . . . . .	50
V.1	Data Science : Machine Learning + Big Data . . . . .	53
V.2	Classification des modèles selon la théorie . . . . .	59
VI.1	Arbres de décision de UBooK . . . . .	66
VI.2	DataSet de UBooK . . . . .	72
VIII.	Résultat de test de l'arbre de décision IsScientific . . . . .	75
VIII.	Résultat de test de l'arbre de décision IsInformatique . . . . .	75
VIII.	Résultat de mail pour la véracité n'est pas résepcté . . . . .	76
VIII.	Résultat d'un document avec une véracité respcté . . . . .	77
VIII.	Résultat de mail pour la véracité est résepcté . . . . .	78

# **LISTE DES TABLEAUX**

I.1	Analyse S.W des concurrents dans le secteur partages des documents. . . . .	24
I.2	Analyse S.W des concurrents dans le secteur d'evenement. . . . .	24
II.1	Scénario Editer les métadonnés de ses documents par un UBooKer . . . . .	29
II.2	Scénario S'inscrire dans un evenement par un UBooKer . . . . .	32
II.3	Scénario Ajouter un evenement par UBooKer . . . . .	34

# BIBLIOGRAPHIE

- [1] <https://www.maxicours.com/se/cours/la-notion-de-document/>  
Consulté le (20/03/2022)
- [2] <https://fr.wikipedia.org/wiki/Document>  
Consulté le (20/03/2022)
- [3] <https://wikinotions.apden.org/notions.php?p=consult&nom=editorialisation>  
Consulté le (20/03/2022)
- [4] [https://www.memoireonline.com/07/12/5996/m\\_Annotations-collaboratives-des-documents-pedagogiques3.html](https://www.memoireonline.com/07/12/5996/m_Annotations-collaboratives-des-documents-pedagogiques3.html)  
Consulté le (22/03/2022)
- [5] <https://fr.unesco.org/themes/tic-education/rel>  
Consulté le (25/03/2022)
- [6] [https://www.memoireonline.com/07/12/5996/m\\_Annotations-collaboratives-des-documents-pedagogiques3.html](https://www.memoireonline.com/07/12/5996/m_Annotations-collaboratives-des-documents-pedagogiques3.html)  
Consulté le (25/03/2022)
- [7] [https://www.unil.ch/files/live/sites/ecoledemedecine/files/shared/Enseignants\\_Supports\\_Guidelines\\_UPFBM\\_121120vs2.pdf](https://www.unil.ch/files/live/sites/ecoledemedecine/files/shared/Enseignants_Supports_Guidelines_UPFBM_121120vs2.pdf)  
Consulté le (22/03/2022)
- [8] <https://www.linternaute.fr/dictionnaire/fr/definition/td>  
Consulté le (24/03/2022)
- [9] [https://fr.wikipedia.org/wiki/Travaux\\_pratiques](https://fr.wikipedia.org/wiki/Travaux_pratiques)  
Consulté le (24/03/2022)
- [10] [https://fr.wikipedia.org/wiki/Devoir\\_surveillé](https://fr.wikipedia.org/wiki/Devoir_surveillé)  
Consulté le (24/03/2022)
- [11] <https://fr.wikipedia.org/wiki/Examen>  
Consulté le (24/03/2022)
- [12] <https://fr.wikipedia.org/wiki/Résumé>  
Consulté le (24/03/2022)
- [13] [http://bdl.oqlf.gouv.qc.ca/bdl/gabarit\\_bdl.asp?id=5089](http://bdl.oqlf.gouv.qc.ca/bdl/gabarit_bdl.asp?id=5089)  
Consulté le (23/03/2022)
- [14] <https://www.processmaker.com/fr/blog/document-metadata/>  
Consulté le (26/03/2022)

- [15] <https://fr.wikipedia.org/wiki/Métadonnée>  
Consulté le (26/03/2022)
- [16] [https://fr.wikipedia.org/wiki/Licence\\_Creative\\_Commons](https://fr.wikipedia.org/wiki/Licence_Creative_Commons)  
Consulté le (05/04/2022)
- [17] <https://creativecommons.org/choose>  
Consulté le (06/04/2022)
- [18] <https://www.justifit.fr/b/guides/droit-penal/faux-usage-de-faux-sanctions/>  
Consulté le (06/04/2022)
- [19] <https://www.evenement.com/guides-professionnels/definitions/definition-evenementiel/>  
Consulté le (02/04/2022)
- [20] <https://adminguide.stanford.edu/chapter-8/subchapter-2>  
Consulté le (03/04/2022)
- [21] <https://www.lawinsider.com/dictionary/university-event>  
Consulté le (07/04/2022)
- [22] <https://www.demos.fr/blog/quest-ce-que-la-formation>  
Consulté le (07/04/2022)
- [23] <https://www.dictionnaire-juridique.com/definition/certification.php>  
Consulté le (07/04/2022)
- [24] <https://www.linkedin.com/company/events/>  
Consulté le (07/04/2022)
- [25] <https://www.axopen.com/spring-boot-lyon>  
Consulté le (07/04/2022)
- [26] <https://en.wikipedia.org/wiki/Angular>  
Consulté le (05/05/2022)
- [27] <https://en.wikipedia.org/wiki/Firebase>  
Consulté le (05/05/2022)
- [28] <https://inf1410.teluq.ca/teluqDownload.php>  
Consulté le (05/05/2022)
- [29] <https://www.blogdumoderateur.com/tools/balsamiq/>  
Consulté le (05/05/2022)
- [30] <https://fr.wikipedia.org/wiki/LaTeX>  
Consulté le (05/05/2022)
- [31] <https://www.atlassian.com/fr/git/tutorials/what-is-git>  
Consulté le (05/05/2022)

- [32] <https://kinsta.com/knowledgebase/what-is-github/>  
Consulté le (05/05/2022)
- [33] <https://fr.wikipedia.org/wiki/Ubuntu>  
Consulté le (05/05/2022)
- [34] <https://www.lebigdata.fr/definition-big-data>  
Consulté le (05/05/2022)
- [35] <https://www.kdnuggets.com/2017/04/42-vs-big-data-data-science.html>  
Consulté le (05/05/2022)
- [36] <https://fr.wikipedia.org/wiki/veracite>  
Consulté le (26/03/2022)
- [37] <https://www.lalanguefrancaise.com/dictionnaire/definition/veracite>  
Consulté le (26/03/2022)
- [38] <https://www.futura-sciences.com/tech/definitions/informatique-intelligence-artificielle-555/>  
Consulté le (26/03/2022)
- [39] <https://www.oracle.com/fr/artificial-intelligence/what-is-ai/>  
Consulté le (26/03/2022)
- [40] <https://www.ibm.com/cloud/learn/machine-learning>  
Consulté le (26/03/2022)
- [41] [https://en.wikipedia.org/wiki/Data\\_scraping](https://en.wikipedia.org/wiki/Data_scraping)  
Consulté le (01/05/2022)
- [42] <https://www.rgdesign.fr/blog/web-scraping/>  
Consulté le (01/05/2022)
- [43] <https://wydden.com/10-outils-pour-scaper-des-donnees-sans-coder-ou-presque/>  
Consulté le (01/05/2022)
- [44] [https://fr.wikipedia.org/wiki/Indexation\\_automatique\\_de\\_documents](https://fr.wikipedia.org/wiki/Indexation_automatique_de_documents)  
Consulté le (01/05/2022)
- [45] <https://fr.theastrologypage.com/indexing>  
Consulté le (01/05/2022)
- [46] <https://fr.wikipedia.org/wiki/WebScraping>  
Consulté le (01/05/2022)
- [47] [https://www.academia.edu/42880013/Système daide à la décision pour la véracitédes données dans lecontexte de Big Data](https://www.academia.edu/42880013/Système_daide_à_la_décision_pour_la_véracité_des_données_dans_lecontexte_de_Big_Data)  
Consulté le (02/05/2022)

- [48] Yin, X., Han, J., and Yu, P. Truth discovery with multiple conflicting information providers on the web. *Knowledge and Data Engineering, IEEE Transactions on*, 20 :796 – 808, 07 2008. doi : 10.1109/TKDE.2007.190745  
Consulté le (17/04/2022)
- [49] Galland, A., Abiteboul, S., Marian, A., and Senellart, P. Corroborating information from disagreeing views. In Proceedings of the third ACM international conference on Web search and data mining, pages 131–140, 2010.  
Consulté le (17/04/2022)
- [50] Zhao, B., Rubinstein, B. I., Gemmell, J., and Han, J. A bayesian approach to discovering truth from conflicting sources for data integration. arXiv preprint arXiv :1203.0058, 2012.  
Consulté le (17/04/2022)
- [51] (Wang, D., Kaplan, L., Le, H., and Abdelzaher, T. On truth discovery in social sensing : A maximum likelihood estimation approach. IPSN'12 - Proceedings of the 11th International Conference on Information Processing in Sensor Networks, 04 2012. doi :10.1145/2185677.2185737.)  
Consulté le (17/04/2022)
- [52] Dong, X., Berti-Equille, L., and Srivastava, D. Integrating conflicting data : The role of source dependence. *PVLDB*, 2 :550–561, 08 2009.  
Consulté le (17/04/2022)
- [53] Pasternack, J. and Roth, D. Latent credibility analysis. In Proceedings of the 22nd international conference on World Wide Web, pages 1009–1020, 2013.  
Consulté le (17/04/2022)
- [54] Lin, X. and Chen, L. Domain-aware multi-truth discovery from conflicting sources. *Proceedings of the VLDB Endowment*, 11(5) :635647, 2018.  
Consulté le (17/04/2022)
- [55] Lamine Ba, M., Horincar, R., Senellart, P., and Wu, H. Truth finding with attribute partitioning. In Proceedings of the 18th International Workshop on Web and Databases, WebDB'15,page 27–33, New York, NY, USA, 2015. Association for Computing Machinery. ISBN 9781450336277. doi : 10.1145/2767109.2767118. URL <https://doi.org/10.1145/2767109.2767118>.  
Consulté le (17/04/2022)
- [56] Yang, Y., Bai, Q., and Liu, Q. A probabilistic model for truth discovery with object correlations. *Knowledge-Based Systems*, 165 :360–373, 2019. Chomicki, J. and Marcinkowski, J. Minimal-change integrity maintenance using tuple deletions. *Information and Computation*, 197(1-2) :90–121, 2005.  
Consulté le (17/04/2022)
- [57] OSIAS NOËL NICODÈME FINAGNON TOSSOU 3 Mars 2021Url <https://www.researchgate.net/p>  
par partitionnement de donnees Truth discovery by data partitioning)  
Consulté le (17/04/2022)
- [58] Song, M., Zhao, X., Liu, Y., Zhao, Z. : Text sentiment analysis based on convolutional neural network and bidirectional LSTM model. In : Zhou, Q., Miao, Q., Wang, H., Xie, W., Wang, Y., Lu, Z. (eds.) ICPCSEE 2018. CCIS, vol. 902, pp. 55–68. Springer, Singapore (2018). [https://doi.org/10.1007/978-981-13-2206-8\\_6](https://doi.org/10.1007/978-981-13-2206-8_6)  
Consulté le (17/04/2022)

- [59] Wu, W., Zhou, J., Qu, W. : A survey of syntactic parsing based on statistical learning.J.Chi. Inf. Process. 27(3), 9–19 (2013)  
Consulté le (17/04/2022)
- [60] Yong Wang , Maosheng Zhong, Lan Tao, and Shuixiu Wu 2020 URL <https://ibook.pub/police-an-effective-truth-discovery-method-in-intelligent-crowd-sensing.html>  
Consulté le (17/04/2022)
- [61] <https://www.isnbreizh.fr/nsi/activity/algoReffKnn>  
Consulté le (15/05/2022)
- [62] <https://www.kaggle.com/>  
Consulté le (15/05/2022)
- [63] c : <https://www.4c.tn/>  
Consulté le (16/05/2022)
- [64] oogle Forms : <https://docs.google.com/forms/u/0/>  
Consulté le (16/05/2022)
- [65] 0 Times : <https://10times.com/>  
Consulté le (16/05/2022)
- [66] acebook : <https://www.facebook.com/>  
Consulté le (16/05/2022)
- [67] oogle Classrooms : <https://classroom.google.com/u/0/h>  
Consulté le (16/05/2022)
- [68] rvezli.tn : <https://rivezli.tn/>  
Consulté le (16/05/2022)
- [69] library : <https://z-lib.org/>  
Consulté le (16/05/2022)
- [70] lideshare : <https://www.slideshare.net/>  
Consulté le (16/05/2022)
- [71] cribd : <https://www.scribd.com/home>  
Consulté le (16/05/2022)
- [72] esearch Gate : <https://www.researchgate.net/>  
Consulté le (16/05/2022)
- [73] oodle : <https://moodle.org/>  
Consulté le (16/05/2022)