

# Rapport Projet Fin d'étude de projet UBook

Hadj Sassi Mahdi

April 2022

## Résumé

.

# Table des matières

<b>I</b>	<b>Infrastrucutre logicielle</b>	<b>6</b>
<b>1</b>	<b>Etude Préable</b>	<b>7</b>
1.1	Evenement Universitaire . . . . .	7
1.1.1	Définitions d'un evenement universitaire . . . . .	7
1.1.2	Types d'evenement universitaire . . . . .	8
1.1.3	Politique des evenements universitaires . . . . .	9
1.2	Document Educatif . . . . .	10
1.2.1	Définitions d'un document . . . . .	10
1.2.2	Définitions d'un document educatif . . . . .	11
1.2.3	Types de document educatif . . . . .	12
1.2.4	Différent métadonnées des documents educatifs . . . . .	13
1.2.5	Les licences Creative Commons . . . . .	15
1.3	Domaine Métier . . . . .	16
1.3.1	Le secteur de partage des documents . . . . .	16
1.3.2	Le secteur de publication des evenements . . . . .	17
1.4	Etude de l'existant . . . . .	17
1.4.1	Analyse de marché . . . . .	17
1.4.2	Analyse de la concurrence . . . . .	18
1.5	Méthodologie de travail . . . . .	20
<b>2</b>	<b>Spécification</b>	<b>23</b>
2.0.1	Environnement de travail . . . . .	23
2.1	Diagramme Use Case . . . . .	25
2.2	Les Diagrammes de séquences systèmes . . . . .	29
<b>3</b>	<b>Conception et développement</b>	<b>38</b>
3.1	Diagramme de classe analyse . . . . .	38
3.2	Diagramme de classe conception . . . . .	39

3.3	Diagrammes de séquences objets . . . . .	40
-----	--	----

## II Vers un modèle de véracité 48

4	Etat de l'art	49
4.1	Véracité . . . . .	50
4.1.1	Notion de véracité . . . . .	50
4.1.2	La véracité en Big Data . . . . .	50
4.1.3	La véracité pour les documents . . . . .	51
4.1.4	La véracité pour les evenements . . . . .	51
4.2	Indexation . . . . .	51
4.2.1	Notion d'indexation . . . . .	51
4.2.2	Modèles de recherche d'information . . . . .	52
4.3	IA et ML . . . . .	53
4.3.1	Notion d'Intelligence Artificiel . . . . .	53
4.3.2	Notion d'Apprentissage Automatique . . . . .	54
4.3.3	IA Et Evenements . . . . .	55
4.3.4	Les modèles ML qui traitent la véracité des documents	56
4.4	Travaux anicennes . . . . .	56
4.4.1	Feuille de route . . . . .	56
4.4.2	Système d'aide à la décision pour la véracité des données dans le contexte de Big Data . . . . .	56
4.4.3	Modèle probabiliste avec support mutuel des valeurs similaires . . . . .	57
4.4.4	Découverte de la vérité par corroboration d'informations	57
4.4.5	Modèle de vérité latente . . . . .	58
4.4.6	Découverte de la vérité par estimation de la vraisem- blancemaximale . . . . .	58
4.4.7	Découverte de la vérité avec dépendance de sources par copie . . . . .	59
4.4.8	Analyse de crédibilité latente . . . . .	59
4.4.9	Découverte de la vérité dans plusieurs domaines de sources contradictoires ayant des vérités multiples . . .	60
4.4.10	Découverte de la vérité avec partitionnement des attri- buts . . . . .	60
4.4.11	Modèle probabiliste pour la découverte de la vérité avec des corrélations d'objets avec des contraintes . . .	61

4.4.12	Vérification de faits par partitionnement de données . .	61
4.4.13	Incorporation de phrases (SIF) . . . . .	62
4.4.14	Analyse syntaxique . . . . .	62
4.4.15	Calcul de l'incorporation de phrases en fusionnant l'arbre d'analyse syntaxique et l'incorporation de mots . . . .	63
4.4.16	Conclusion vers notre modèle . . . . .	63
<b>5</b>	<b>Modèle de Machine Learning</b>	<b>64</b>
<b>6</b>	<b>Résultat</b>	<b>65</b>
<b>7</b>	<b>Discussion</b>	<b>66</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 concluons par la présentation du plan général du rapport.

## Cadre de projet

L'organisme Pôle Étudiant Entrepreneur de Carthage (PEEC) a annoncé les résultats de la 3<sup>ème</sup> 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 celui 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 gestions 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 problématique a conduit vers l'idée de développer une encyclopédie universitaire éducative ou on trouve des différentes ressources d'éducatifs (Cours, Travaux Dirigés, Travaux Pratiques, Examens, Corrigés) avec la possibilité de passer des examens blanc pour gagner des badges des compétences HardSkills ou SoftSkills, plus d'un centre d'événement (Formations, Certifications, Compétitions, Journées) en ligne.

## **Objectifs de UBook**

- Faciliter la recherche des ressources éducatives
- Améliorer le partage des documents éducatifs en respectant les éthiques
- Augmenter les interactions événementielles
- Expandre le réseau universitaire

## **Organisation du rapport**

c'est pour le moment it's in the final step!

# Première partie

## Infrastrucutre logicielle



# Chapitre 1

## Etude Préable

### Introduction

UBook présente des solution dans l'échange des documents et le partage des evenements alors dans ce chapitre, nous allons présenter en premier partie les evenement universitaire puis les document educatives, en deuxième lieu, le domaine du métier et le secteur de UBook, suivi d'une étude de l'existant qui concerne l'analyse de marché. Ensuite, nous analysons les besoins fonctionnels et non fonctionnels. Enfin, nous terminons avec la présentation de la méthode de travail suivi durant le projet et une petite conclusion.

### 1.1 Evenement Universitaire

#### 1.1.1 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.

(<https://www.evenement.com/guides-professionnels/definitions/definition-evenementiel/>)

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  
(<https://adminguide.stanford.edu/chapter-8/subchapter-2/policy-8-2-1#:text=1.-,Definition,space%20on%20the%20University%20campus.>)

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.  
(<https://www.lawinsider.com/dictionary/university-event>)

Un événement alors est un programme organisé au sein de l'environnement universitaire ça peut être par des clubs, instituts, centres de formations dans ou pas l'établissement. L'événement est pour un objectif bien déterminé selon une thématique, un événement peut avoir des sponsors pour soutenir l'action.

### 1.1.2 Types d'événement universitaire

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

#### 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 techniquement, 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. . .  
(<https://www.demos.fr/blog/quest-ce-que-la-formation>)

## **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 2ème recherche)

## **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".  
(<https://www.dictionnaire-juridique.com/definition/certification.php#:~:text=Selon%20AFNOR%2C%20%22la%20certification%20est,exigences%20sp%C3%A9cifi%C3%A9es%20dans%20un%20r%C3%A9f%C3%A9rentiel%22.>)

## **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.  
(<https://www.studyrama.com/formations/orientation-reorientation/journees-portes-ouvertes-le-meilleur-moyen-de-22722>)

### **1.1.3 Politique des evenements universitaires**

La création d'une politique événementielle reste donc la première étape pour implémenter une stratégie événementielle.

## **Inscriptions**

La plupart des événement nécessitent une inscription pour réserver la liste des participants ce qui peut garantir le soutien des sponsors dans l'action. Les inscriptions peuvent être payantes et elles ne le sont pas, et ils sont fréquemment via les plateforme des inscriptions et des formulaires comme

google Forms, mais aussi les inscriptions se font manuellement sur une papier. Il y'a des événements sont ouverts aux groupes bien fermés, autres ouvert aux publics et ne nécessitent pas d'inscriptions.

## **Protocoles**

Le protocole, dans l'organisation événementielle, est l'ensemble de normes qui fixe un certain cadre. Généralement, on les applique dans des événements de type diplomatique ou politique, mais pas seulement. Le protocole est de plus en plus fréquent, dans l'organisation de rendez-vous et de divers congrès, car il facilite les aspects suivants :

L'organisation de l'arrivée et de la sortie des participants, un point particulièrement important dans le cadre d'événements où les invités sont nombreux, surtout au temps de la COVID-19.

(<https://blog.visitacostadelsol.com/fr/protocole-evenement>)

## **1.2 Document Educatif**

### **1.2.1 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.

(<https://www.maxicours.com/se/cours/la-notion-de-document/>)

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.

(<https://fr.wikipedia.org/wiki/Document>)

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.

(<https://wikinotions.apden.org/notions.php?p=consult&nom=document>)

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

### 1.2.2 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 Dirigés) que dans le cadre d'un examen.

([https://www.memoireonline.com/07/12/5996/m\\_Annotations-collaboratives-des-documents-pedagogiques3.html#:~:text=Un%20document%20p%C3%A9dagogique%20est%20une,%20enseignement%20ou%20la%20formation%C2%BB.](https://www.memoireonline.com/07/12/5996/m_Annotations-collaboratives-des-documents-pedagogiques3.html#:~:text=Un%20document%20p%C3%A9dagogique%20est%20une,%20enseignement%20ou%20la%20formation%C2%BB.))

Les documents éducatifs s'appellent aussi ressources éducatifs 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.

(<https://fr.unesco.org/themes/tic-education/rel>)

### 1.2.3 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.

([https://www.memoireonline.com/07/12/5996/m\\_Annotations-collaboratives-des-documents-pedagogiques3.html#:~:text=Un%20document%20p%C3%A9dagogique%20est%20une,%20enseignement%20ou%20la%20formation%C2%BB.](https://www.memoireonline.com/07/12/5996/m_Annotations-collaboratives-des-documents-pedagogiques3.html#:~:text=Un%20document%20p%C3%A9dagogique%20est%20une,%20enseignement%20ou%20la%20formation%C2%BB.))

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

- 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 qu'avec la narration qu'il accompagne.

([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))

- Travaux Dirigés (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.

(<https://www.linternaute.fr/dictionnaire/fr/definition/td/#:~:text=TD%20signifie%20Travaux%20Dirig%C3%A9s.,faciliter%20l'aide%20du%20professeur.>)

- 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. ([https://fr.wikipedia.org/wiki/Travaux\\_pratiques](https://fr.wikipedia.org/wiki/Travaux_pratiques))

- Devoir Surveillés (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. ([https://fr.wikipedia.org/wiki/Devoir\\_surveill](https://fr.wikipedia.org/wiki/Devoir_surveill))

- 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. (<https://fr.wikipedia.org/wiki/Examen>)

- 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. (<https://fr.wikipedia.org/wiki/R%C3%A9sum%C3%A9>)

- 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. ([http://bdl.oqlf.gouv.qc.ca/bdl/gabarit\\_bdl.asp?id=5089#:text=La%20correction%20d'%C3%A9preuves%20est,d'impression%20et%20de%20coquilles.](http://bdl.oqlf.gouv.qc.ca/bdl/gabarit_bdl.asp?id=5089#:text=La%20correction%20d'%C3%A9preuves%20est,d'impression%20et%20de%20coquilles.))

#### **1.2.4 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.

(<https://www.processmaker.com/fr/blog/document-metadata/>)

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 significatives 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).

(<https://fr.wikipedia.org/wiki/M%C3%A9tadonn%C3%A9e>)

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.

### 1.2.5 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).

([https://fr.wikipedia.org/wiki/Licence\\_Creative\\_Commons](https://fr.wikipedia.org/wiki/Licence_Creative_Commons))

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 <https://creativecommons.org/choose/> Il y'a différent licences et chaque à un sens bien définie.








Nom de la licence ↕	Abréviation ↕	Icône ↕	Attribution requis ↕	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 1.1 – Liste des licences creative commons

## 1.3 Domaine Métier

les services offert par UBook sont le partage et l'échange des documents éducatives entre les différentes membres de l'universités, aussi q'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.

### 1.3.1 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)

[<https://www.justifit.fr/b/guides/droit-penal/faux-usage-de-faux-sanctions/>]

"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).

<https://www.etudier.com/dissertations/La-Fiabilit>

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

### **1.3.2 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 etape eventuelle pour les evenements, car c'est à travers les intéressés peuvent avoir l'information sur l'existence 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.

## **1.4 Etude de l'existant**

### **1.4.1 Analyse de marché**

Dans cette sous sections nous listons les cibles clients pour UBook et les concurents .

#### **Les Clients**

- les apprenant : Les etudiants, les enseignants , les internautes et tous les personnes qui cherchent à un document educatives (support de cours, TD, TP...) pour réviser, avoir une idée, préparer un document similaire.

- les formateurs : Les enseignants, les étudiants, les centres de formations, les instituts et toute personne qui a un document éducatif et veut de le publier en public, pour but de partager l'information, avoir une réputation et notoriété, avoir des gains créditeurs ou avantageux de son travail.
- Les organismes universitaires : Les Clubs, les instituts, les centres de formations et toutes personnes veulent poster un événement en public pour y'avoir des inscriptions depuis son cible.

### **Les Concurrents**

Les concurrents dans le secteur de partage des documents éducatifs sont pléthores, nous vous mentionnerons juste les plus célèbres.

- SlideShare
- Scribd
- ResearchGate
- Academia
- Moodle UVT
- Pages Facebook
- Google Classrooms
- les bibliothèques et les librairies
- ...

Et pour les concurrents dans le secteur événement universitaires, nous listons :

- 4C
- 10times
- Pages Facebook
- Les médias télévisés et les radios

### **1.4.2 Analyse de la concurrence**

Dans cette sous-section nous comparons les services offerts par les clients en intéressant sur la qualité des services. Commençons par l'analyse S.W (Strength, Weakness) (Forces, Faiblesses), puis nous faisons le tableau de critères pour tous les concurrents.

	Forces	Faiblesses
SlideShare, Scribd	Populaire, Facile à utiliser	Païement pour accéder au document, Publicités gênantes
ResearchGate	Facile à utiliser	Manque d'organisation, Païement pour accéder au document, Publicités gênantes
Pages Facebook	Le plus populaire Gratuit	Manque d'organisation Document redondants
UVT, Classroom	Gratuit	Difficile à utiliser Accès Restreints
Libraires	Traditionnel et facile à accéder	Ne sont pas accessible tout le temps Païement pour achat des documents
Bibliothèques	Gratuit	Ne sont pas accessible tout le temps

TABLE 1.1 – Analyse S.W des concurents.

	Gratuit	Accessible par tout	Facile	Organisé	Publicités	Redondance des documents
SLideShare, Scribd		●	●	●	●	
ResearchGate		●	●		●	
Pages Facebook	●	●	●			●
UVT, Google Classroom	●					●
U-Book.tn	●	●	●	●		

FIGURE 1.2 – Comparaison Entre les concurrents.

## 1.5 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 90 sous sa forme originale, il est perfectionné au fil des ans et adapté aux méthodes de développement contemporaines.

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

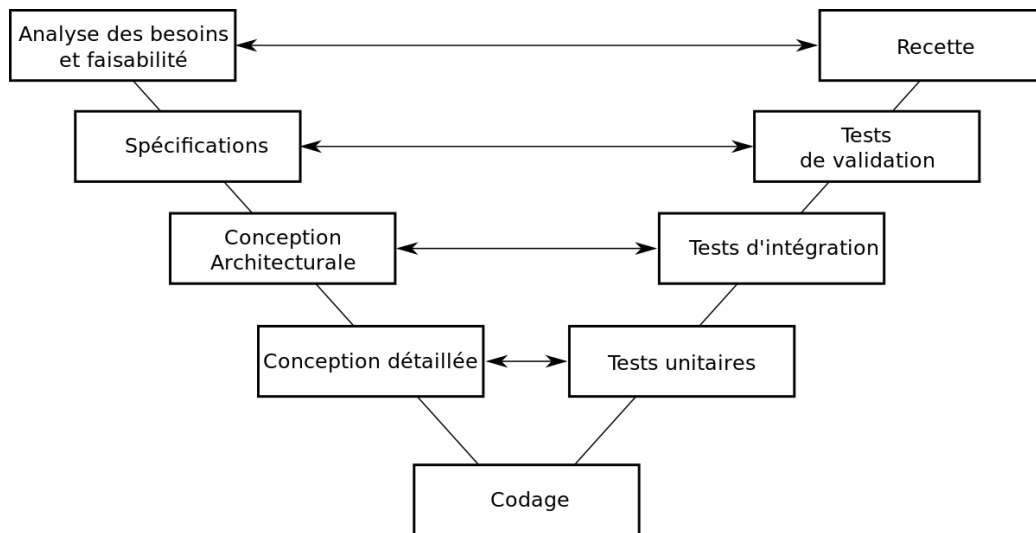


FIGURE 1.3 – Cycle de développement : Modèle ev V

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 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 parties, nous

avons décrit dans les parties suivantes l'événement universitaire, les documents éducatifs, le domaine métier, l'étude de l'existant et la méthodologie de développement modèle en V. le chapitre suivant sera consacré à la spécification, conception et développement de la partie infrastructure logicielle de l'application.



# Chapitre 2

## Spécification

### Introduction

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 etudera les diagrammes UML de Use Case avec quelques descriptions textuelles, puis les diagrammes de séquences systeme pour chaque cas d'utilisation.

#### 2.0.1 Environnement de travail

Dans ce projet j'ai développer l'application sur deux partie BackEnd en utilisant le Framework de JavaEE (Java Entreprise Edition) Spring Boot version 2.5.3 et pour le FrontEnd en utilisant le Framwork de JavaScript Angular version 12.2.6.

#### Spring Boot

En informatique, Spring est un framework open source pour construire et définir l'infrastructure d'une application Java3, dont il facilite le développement et les tests.

En 2004, Rod Johnson a écrit le livre Expert One-on-One J2EE Design and Development<sup>4</sup> qui explique les raisons de la création de Spring.



FIGURE 2.1 – Le Framework Spring Boot

## Angular

Angular (communément appelé "Angular 2+" ou "Angular v2 et plus")<sup>2,3</sup> est un framework côté client, open source, basé sur TypeScript, et co-dirigé par l'équipe du projet « Angular » à Google et par une communauté de particuliers et de sociétés. Angular est une réécriture complète d'AngularJS, cadriciel construit par la même équipe. Il permet la création d'applications Web et plus particulièrement de ce qu'on appelle des « Single Page Applications » : des applications web accessibles via une page web unique qui permet de fluidifier l'expérience utilisateur et d'éviter les chargements de pages à chaque nouvelle action. Le Framework est basé sur une architecture du type MVC et permet donc de séparer les données, le visuel et les actions pour une meilleure gestion des responsabilités. Un type d'architecture qui a largement fait ses preuves et qui permet une forte maintenabilité et une amélioration du travail collaboratif.



FIGURE 2.2 – Le Framework Angular

## 2.1 Diagramme Use Case

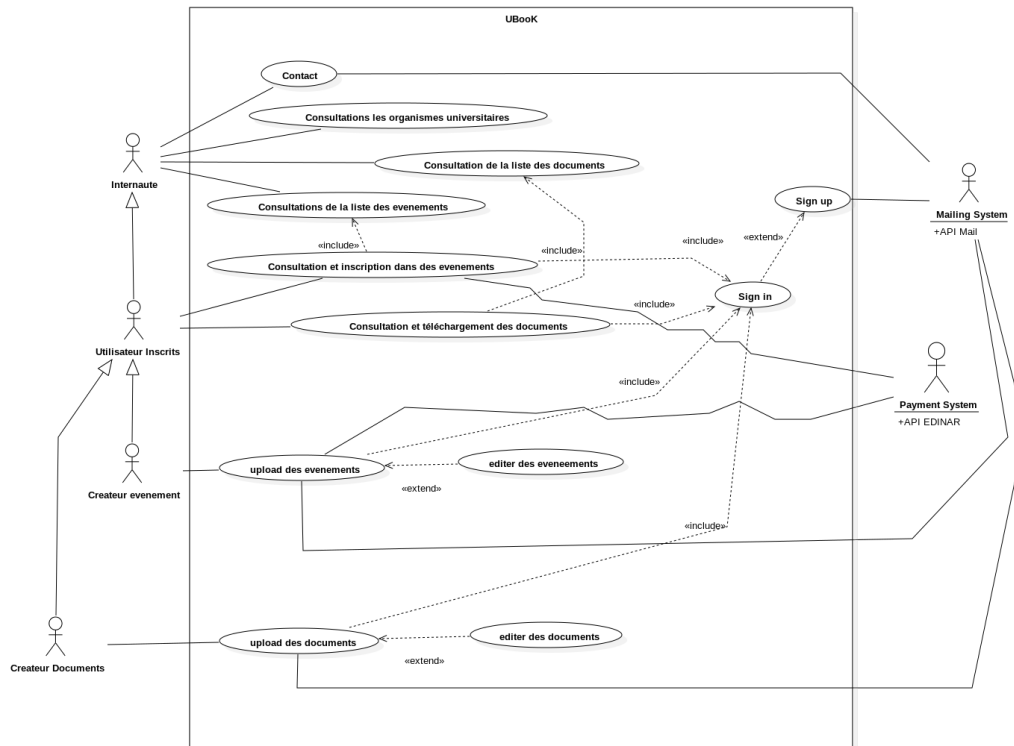


FIGURE 2.3 – Le Framework Angular

Sommaire d'édition des documents	
Titre	Editer ou supprimer un document
Acteur	Créateur de document
Pré-Condition	<ol style="list-style-type: none"> <li>1. L'utilisateur est inscrit dans l'application</li> <li>2. L'utilisateur est dans la consultation d'un document qui l'a déjà créé</li> <li>3. L'utilisateur clique sur modifier ou supprimer</li> <li>4. Connexion internet satisfaisante</li> </ol>
Scénario Nominale	<ol style="list-style-type: none"> <li>1) L'utilisateur faire les modification nécessaire</li> <li>2) L'utilisateur soumettre le formulaire</li> <li>3) L'application envoie un mail de vérification à l'utilisateur</li> <li>4) L'utilisateur vérifie et confirme le mail</li> <li>5) Le document se mise à jour</li> <li>6) En dérière un algorithme de vérification vérifie la véracité de ce document</li> <li>7) Si il est vérifié un mail de satisfaction et de merci envoyé vers l'utilisateur</li> </ol>
Scénario Alternative	<ol style="list-style-type: none"> <li>1) Le formalaire ne contient pas tous les informations nécessaires bloque le soumission</li> <li>2) L'utilisateur ne vérifie pas le mail, le document ne change pas</li> <li>3) Si le document n'est pas vrai et il n'y'a pas de véracité: le document se cache et un mail d'alerte vers l'utilsateur</li> </ol>
Scénario d'Erreur	<ol style="list-style-type: none"> <li>1) Connection internet se coupe : tous changement s'annule</li> </ol>

FIGURE 2.4 – Le cas d'utilisation Editer des documents

Sommaire de Consultation et inscription dans un evenement	
Titre	Consultation et Inscription dans un evenement
Acteur	Utilisateur Inscrits
Pré-Condition	<ol style="list-style-type: none"> <li>1) L'utilisateur est dans la page des evenement ou se trouve la liste des evenement</li> <li>2) L'utilsateur à cliqué sur un evenement</li> <li>3) Connexion internet satisfaisante</li> </ol>
Scénario Nominale	<ol style="list-style-type: none"> <li>1) L'utilisateur clique sur le bouton BooK UP pour s'inscrire dans l'evenement</li> <li>2) Si le lien vers l'inscription d'evenement n'est pas fait par UBook, l'application redirige l'utilisateur vers la page d'inscription</li> <li>3) Sinon l'application redirige le user vers le formulaire d'inscription</li> <li>4) L'utilisateur confirme son nom</li> <li>5) L'utilisateur confirme son mail</li> <li>6) L'utilisateur clique sur soumettre</li> <li>7) L'application vérifie les informations ecrites et envoie un mail de vérification vers l'utilsateur pour confirmer son présence</li> <li>8) L'utilisateur confirme son présence sur le mail</li> <li>9) L'application réserve une place pour cet utilisateur</li> </ol>
Scénario Alternative	<ol style="list-style-type: none"> <li>1. Les information de l'utilisateur ne sont pas confirmé, l'application bloque la soumission de fomulaire</li> <li>2. Les information de l'utilisateur ne sont pas conforme, l'application bloque la soumission de formulaire</li> <li>3. L'utilisateur n'a pas vérifié son présence par mail, l'application ne resérve pas son présence et l'enregistre dans la liste d'attente</li> </ol>
Scénario d'Erreur	<ol style="list-style-type: none"> <li>1. Le lien vers l'inscription est erronée : l'applicatoin envoie un mail pour le createur d'evenement pour vérifier le lien et demander l'excuse de l'utilisateur</li> </ol>

FIGURE 2.5 – Le cas d'utilisation Consultation et inscription dans un evenement

Sommaire de upload des event	
Titre	Téléverser un evenement
Acteur	Créateur d'évenement
Pré-Condition	<ol style="list-style-type: none"> <li>1) L'utilisateur est inscrit dans l'application</li> <li>2) L'utilisateur a cliqué sur un ajout evenement</li> <li>3) Connexion internet satisfaisante</li> </ol>
Scénario Nominale	<ol style="list-style-type: none"> <li>1) L'utilisateur saisie tous les informations sur l'évenement</li> <li>2) L'utilisateur clique sur finir</li> <li>3) Si l'utilisateur est sous une organisme inscrit dans l'application :</li> <li>4) L'application envoie un mail de confirmation vers cet organisme</li> <li>5) L'organisme a confirmé l'évenement</li> <li>6) L'évenement s'ajout dans la liste des evenements</li> <li>7) 3) Sinon, l'utilisateur clique sur payer pour afficher cet evenement</li> <li>8) Il paye la facture</li> <li>9) L'évenement s'affiche dans la liste des evenements</li> </ol>
Scénario Alternative	<ol style="list-style-type: none"> <li>1) L'utilisateur ne saisie pas tous les informations nécessaire l'application bloque la soumission de formulaire</li> <li>2) Si l'organisme n'a pas confirmé l'évenement , l'application n'affiche pas l'évenement</li> <li>3) L'application affiche un bouton payer pour passer la vérification</li> <li>4) L'utilisateur paye la facture l'application ajout cet evenement dans la liste des evenements</li> <li>5) S'il ne paye pas l'évenement restera une semaine dans l'attente pour la validation ou le payement et se supprimera</li> </ol>
Scénario d'Erreur	<ol style="list-style-type: none"> <li>1) Le payement n'est pas finis, l'application n'affiche pas l'évenement et envoie un mail vers l'utilisateur pour l'informer</li> <li>2) Une connexion coupé, aucune transaction se fait et l'évenement se supprimera</li> </ol>

FIGURE 2.6 – Le cas d'utilisation Créer un evenement et le publier

## 2.2 Les Diagrammes de séquences systèmes

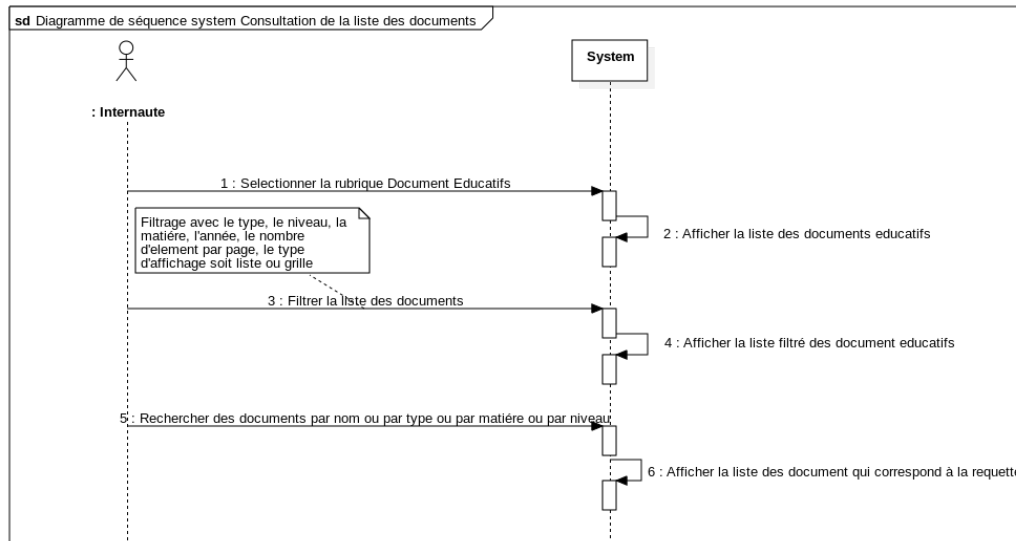


FIGURE 2.7 – Diagramme de séquence system Consultation de la liste des documents

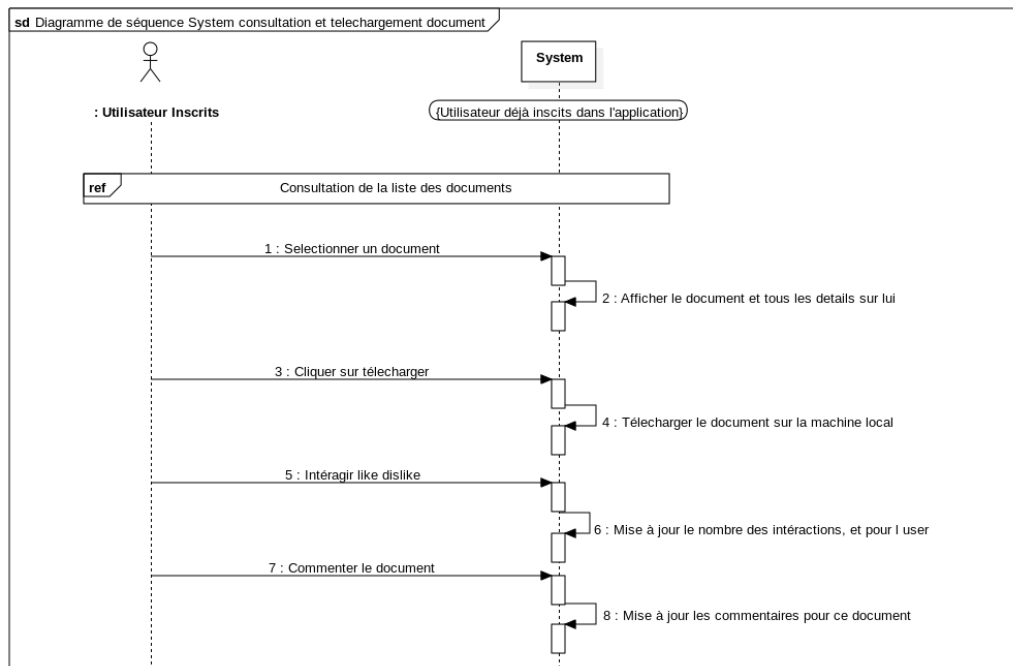


FIGURE 2.8 – Diagramme de séquence System consultation et telechargement document



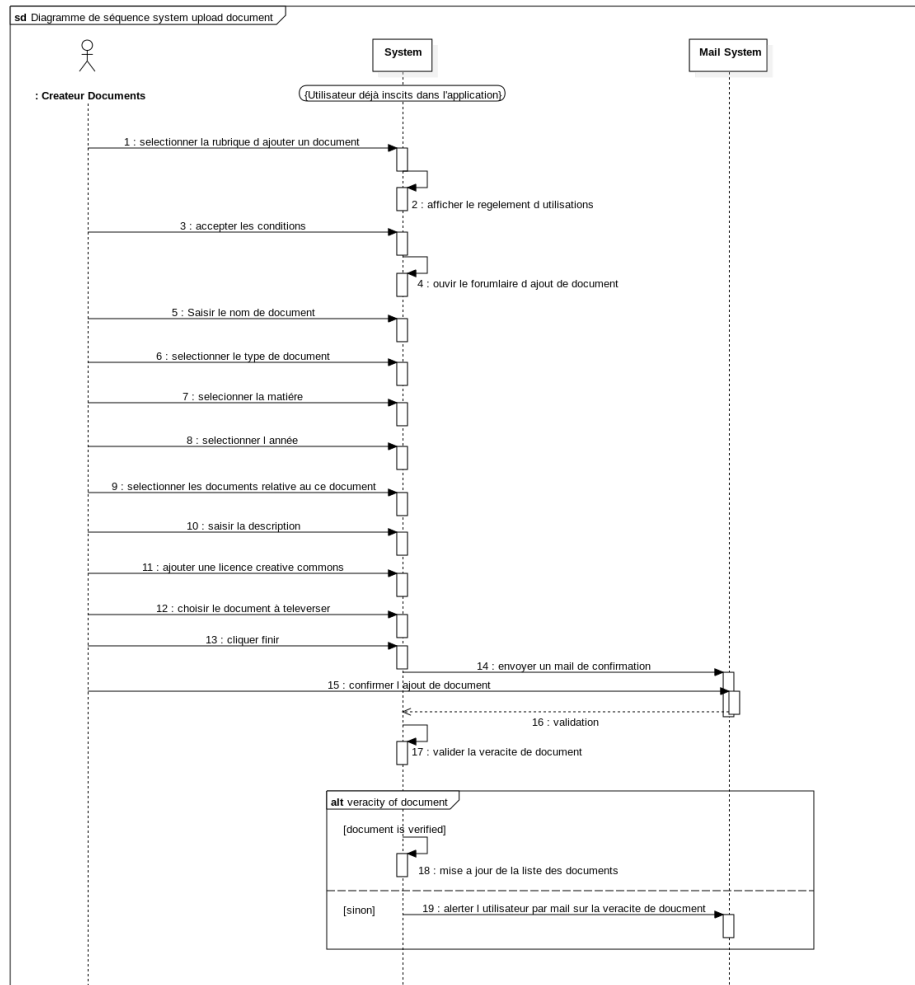


FIGURE 2.9 – Diagramme de séquence system upload document

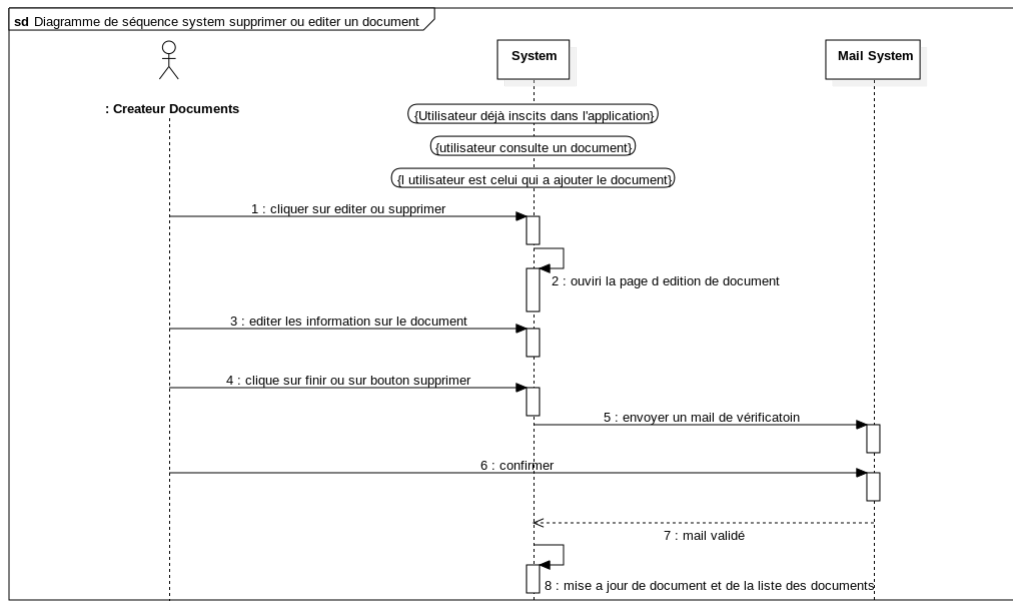


FIGURE 2.10 – Diagramme de séquence system supprimer ou editer un document

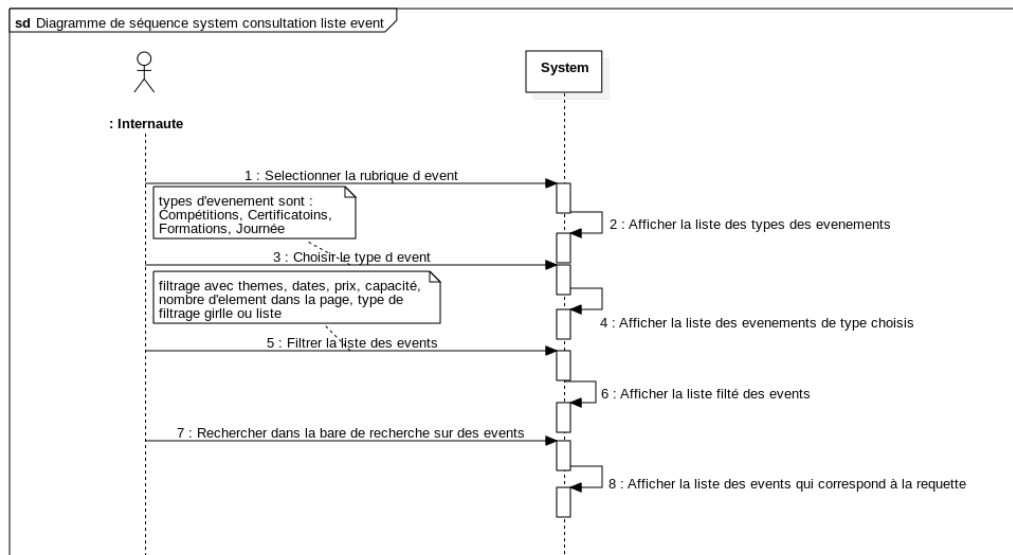


FIGURE 2.11 – Diagramme de séquence system consultation liste event

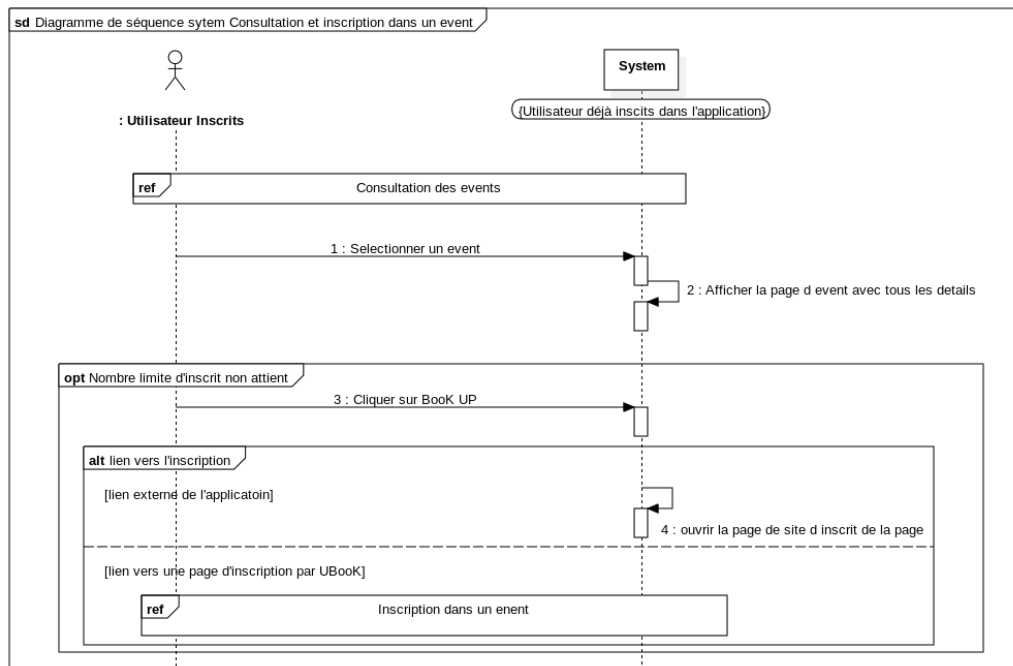


FIGURE 2.12 – Diagramme de séquence sytem Consultation et inscription dans un event

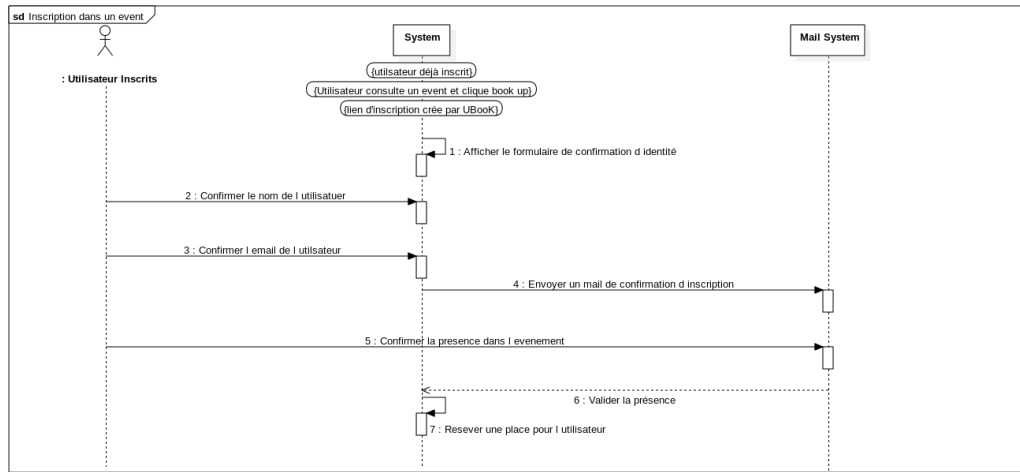


FIGURE 2.13 – Diagramme de séquence Inscription dans un event

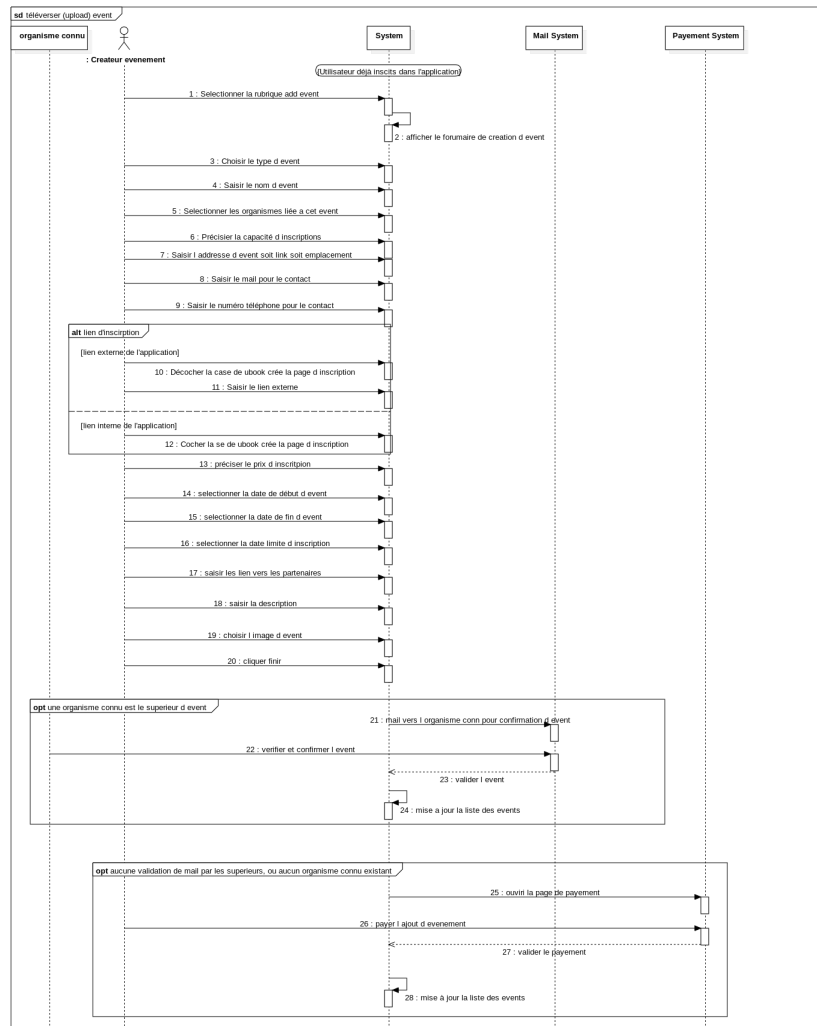


FIGURE 2.14 – Diagramme de séquence Téléverser un event

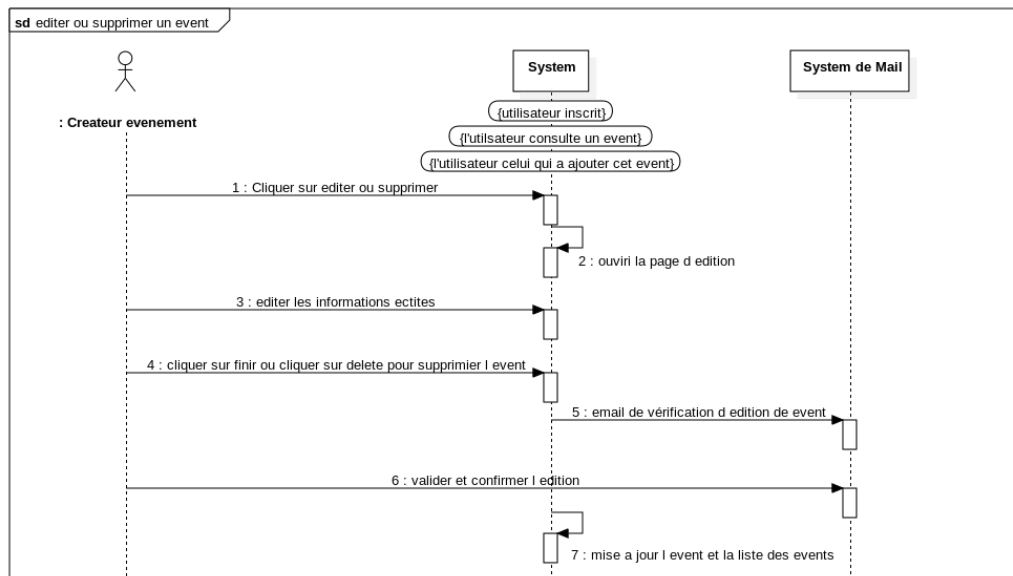


FIGURE 2.15 – Diagramme de séquence Editer un event

# Chapitre 3

## Conception et développement

### 3.1 Diagramme de classe analyse

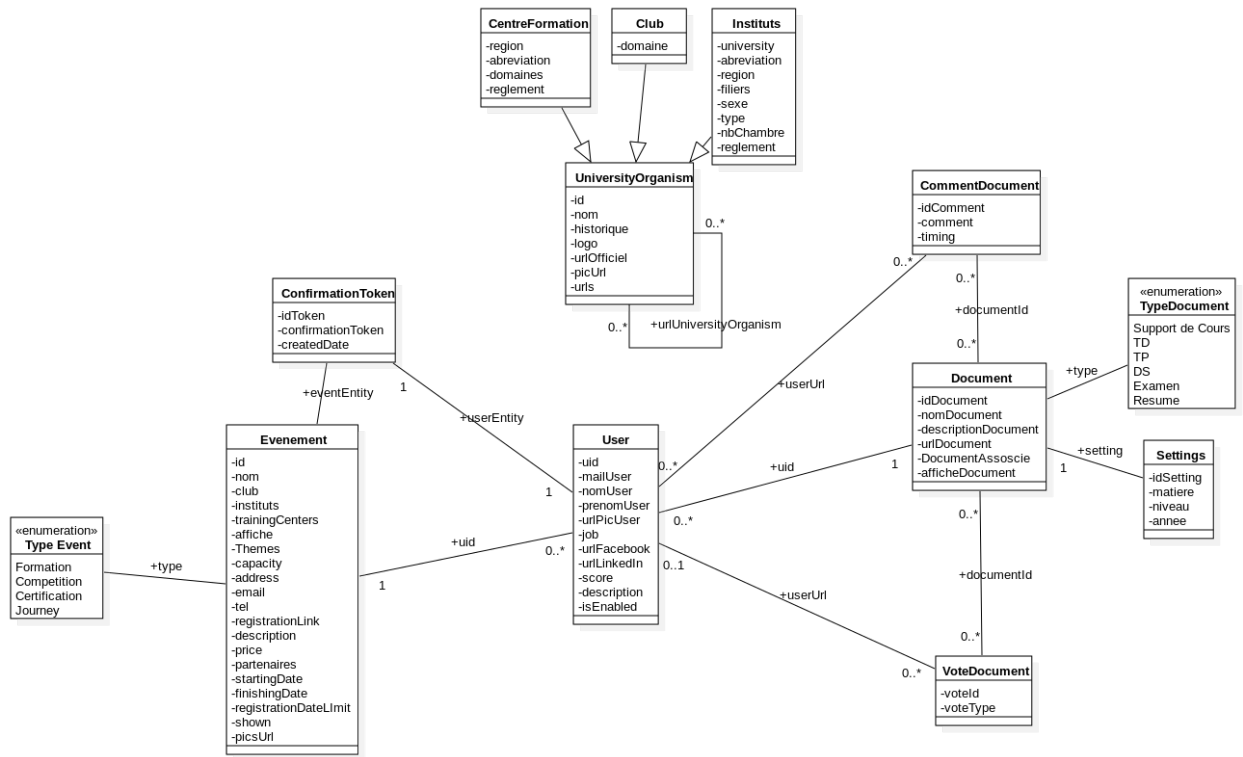


FIGURE 3.1 – Diagramme de classe d'Analyse



### 3.2 Diagramme de classe conception

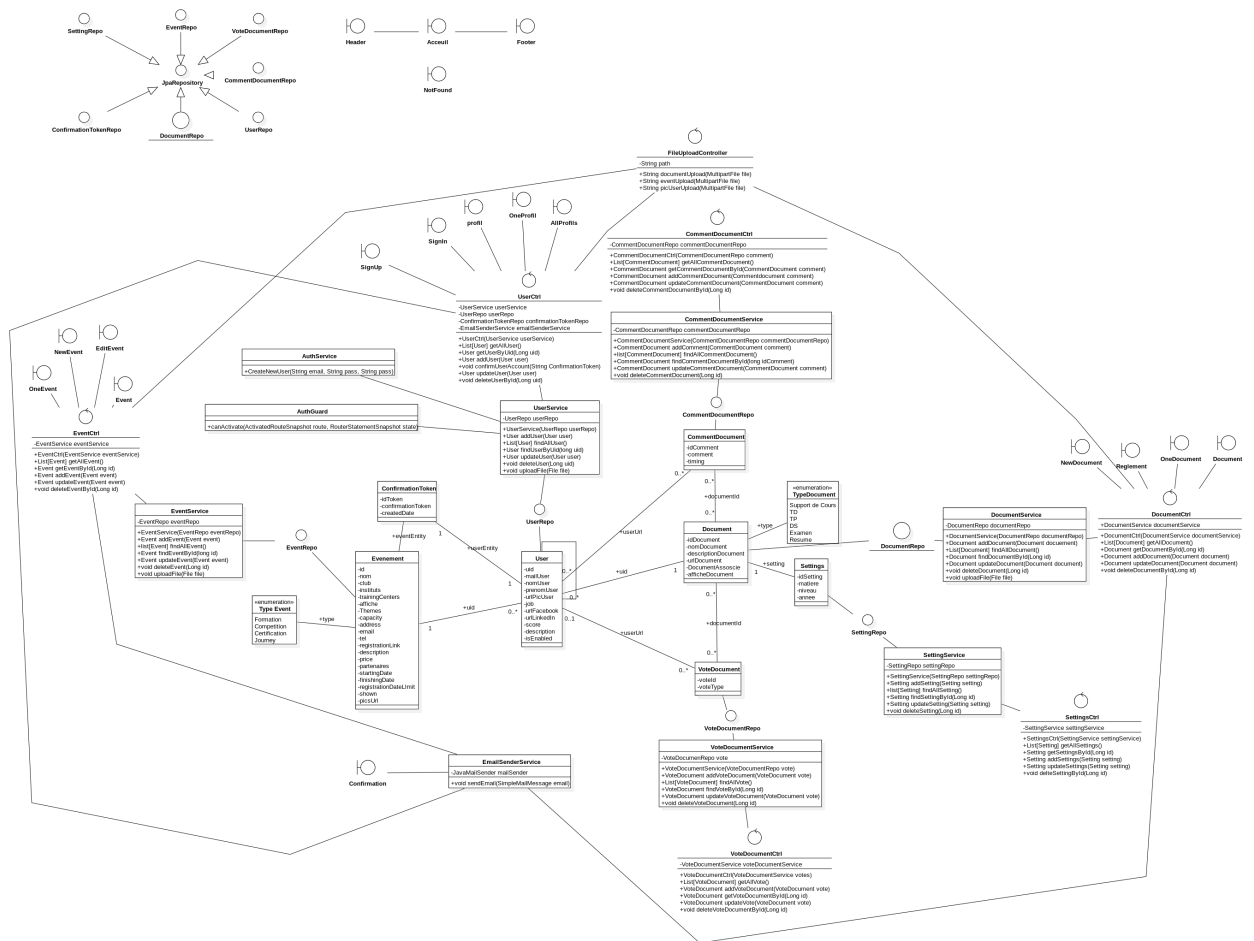


FIGURE 3.2 – Diagramme de classe d'Analyse

### 3.3 Diagrammes de séquences objets

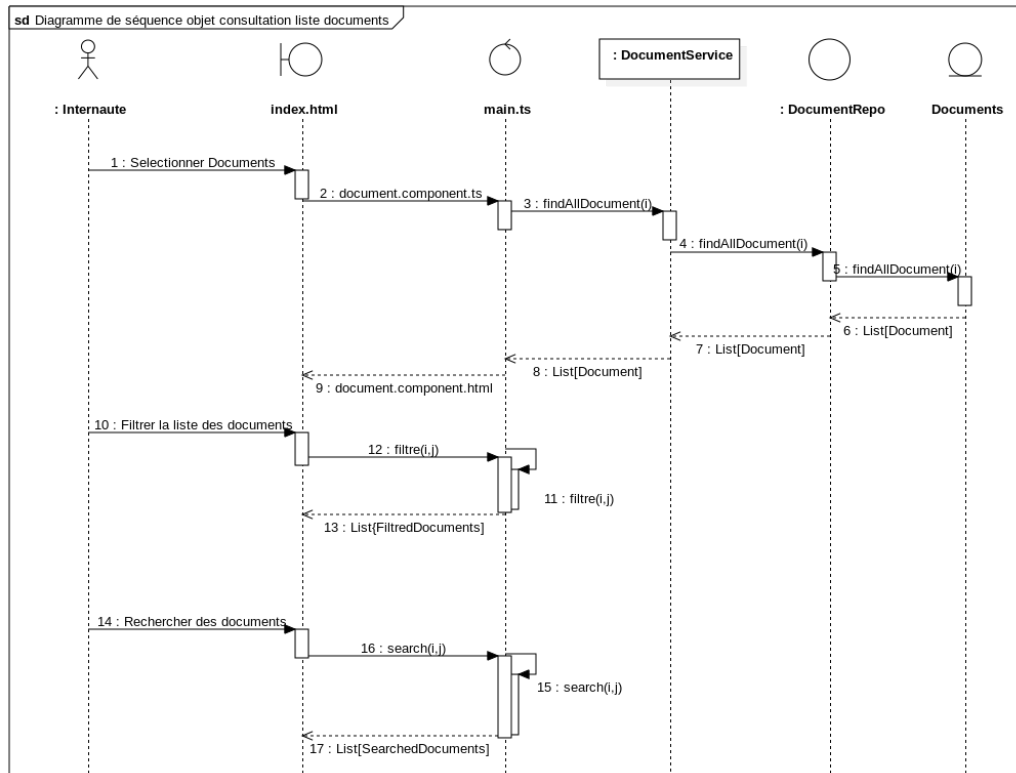


FIGURE 3.3 – Diagramme de séquence objet consultation liste documents

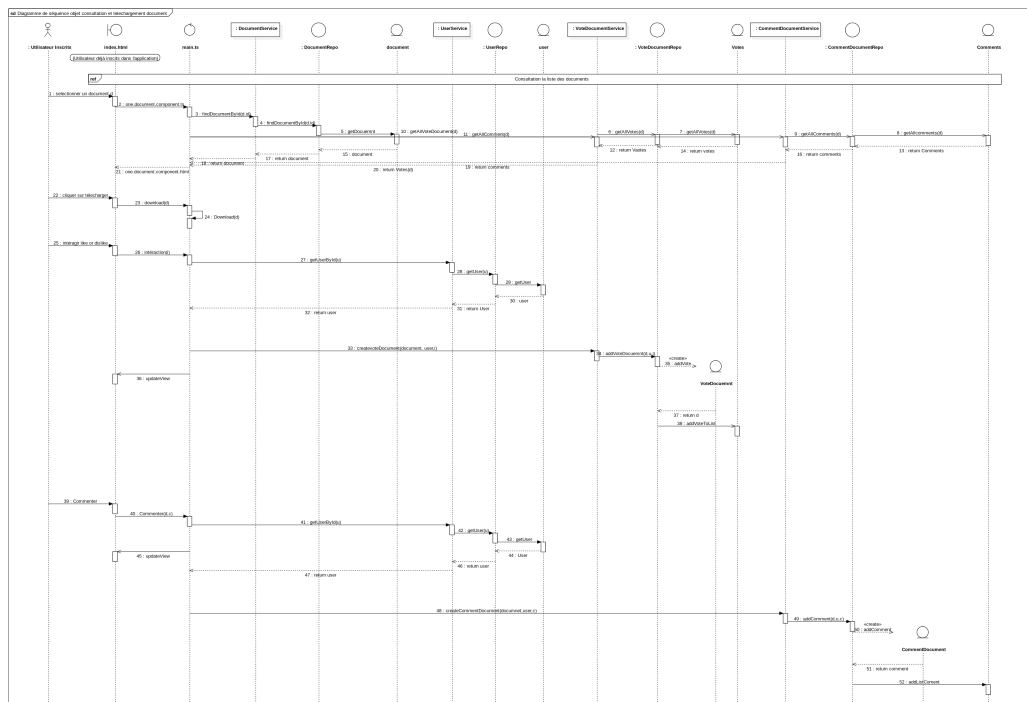


FIGURE 3.4 – Diagramme de séquence objet consultation et telechargement document

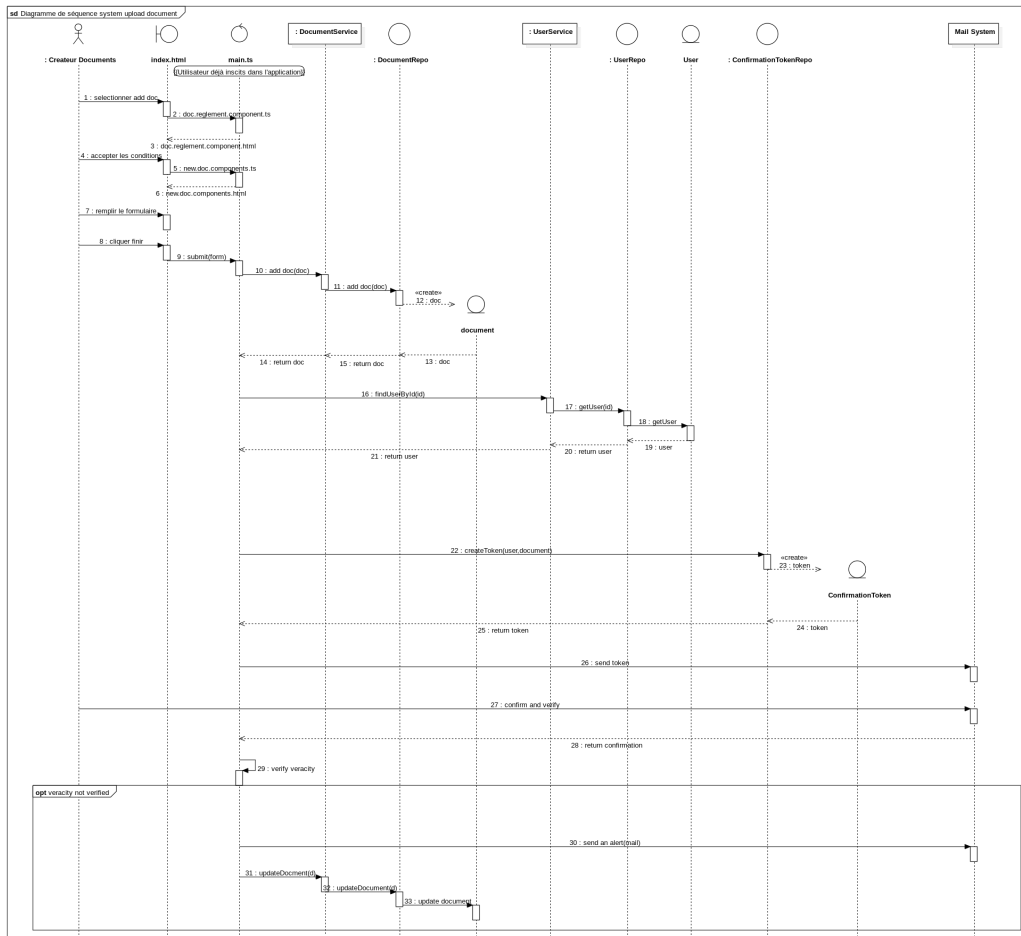


FIGURE 3.5 – Diagramme de séquence objet upload document

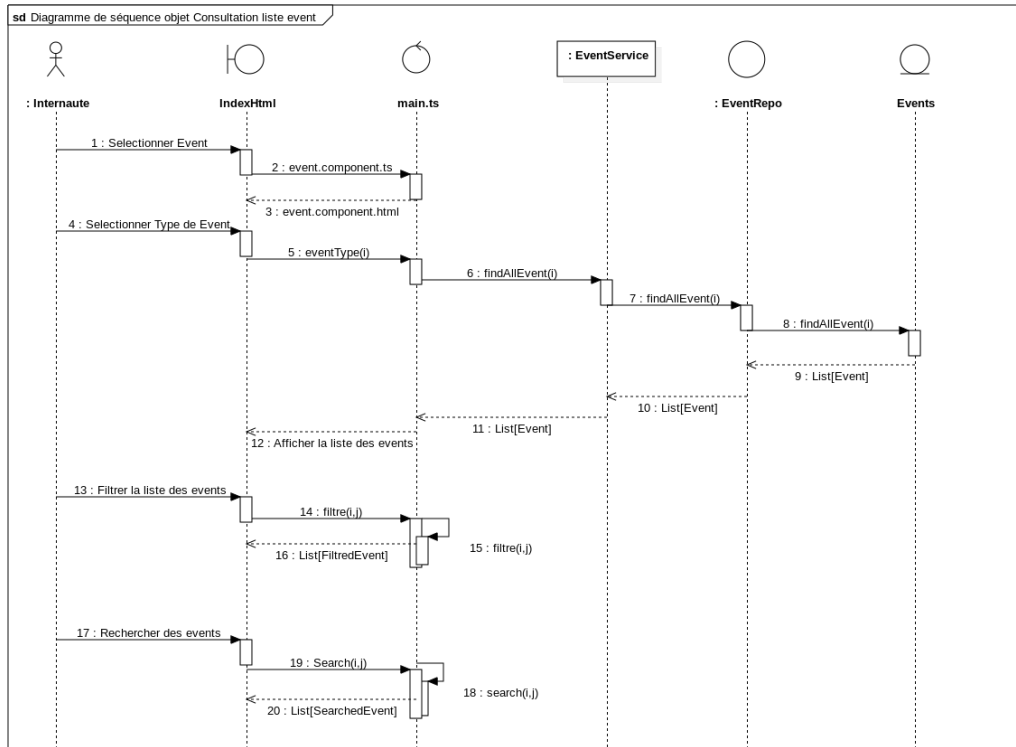


FIGURE 3.6 – Diagramme de séquence objet Consultation liste event

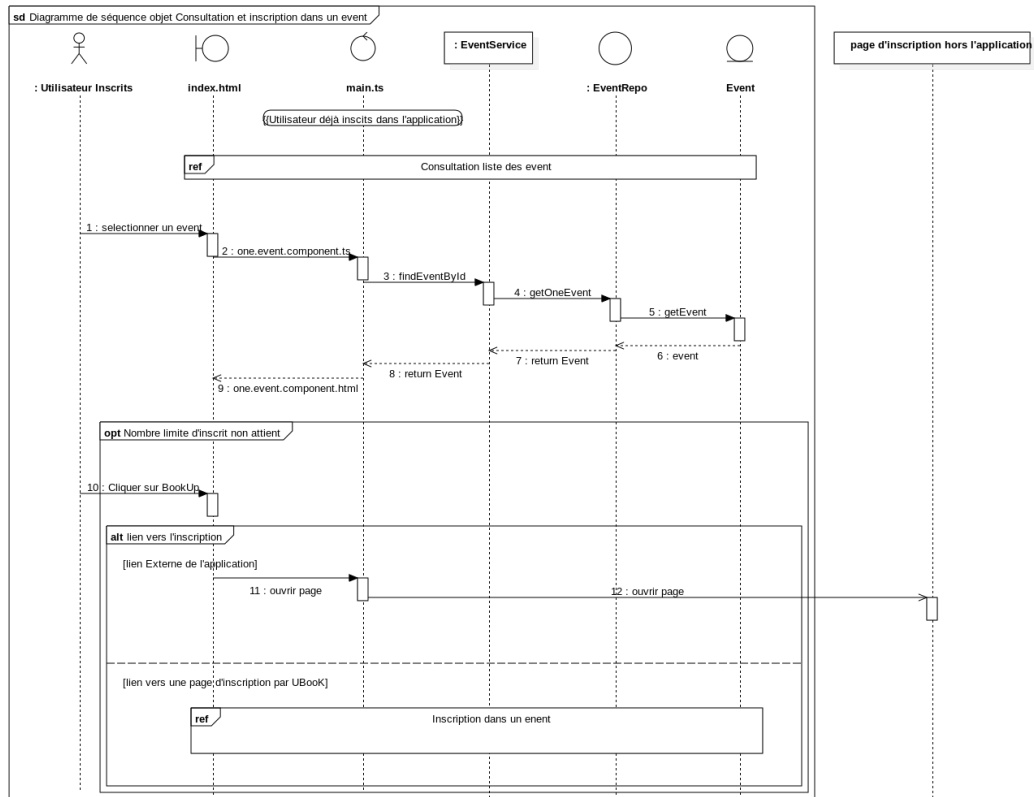


FIGURE 3.7 – Diagramme de séquence objet Consultation et inscription dans un event

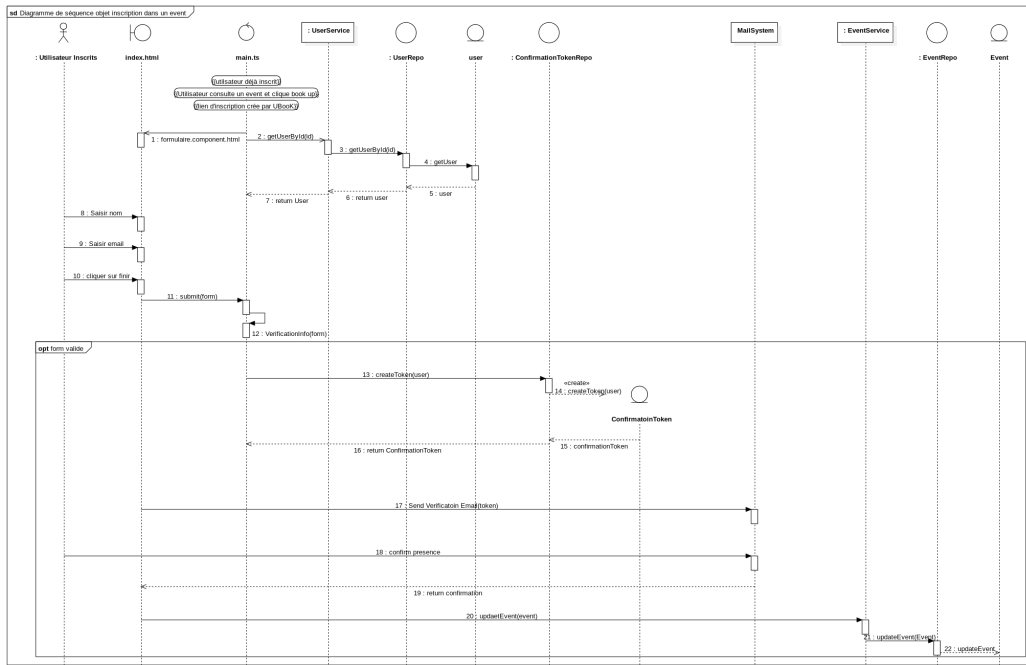


FIGURE 3.8 – Diagramme de séquence objet inscription dans un event

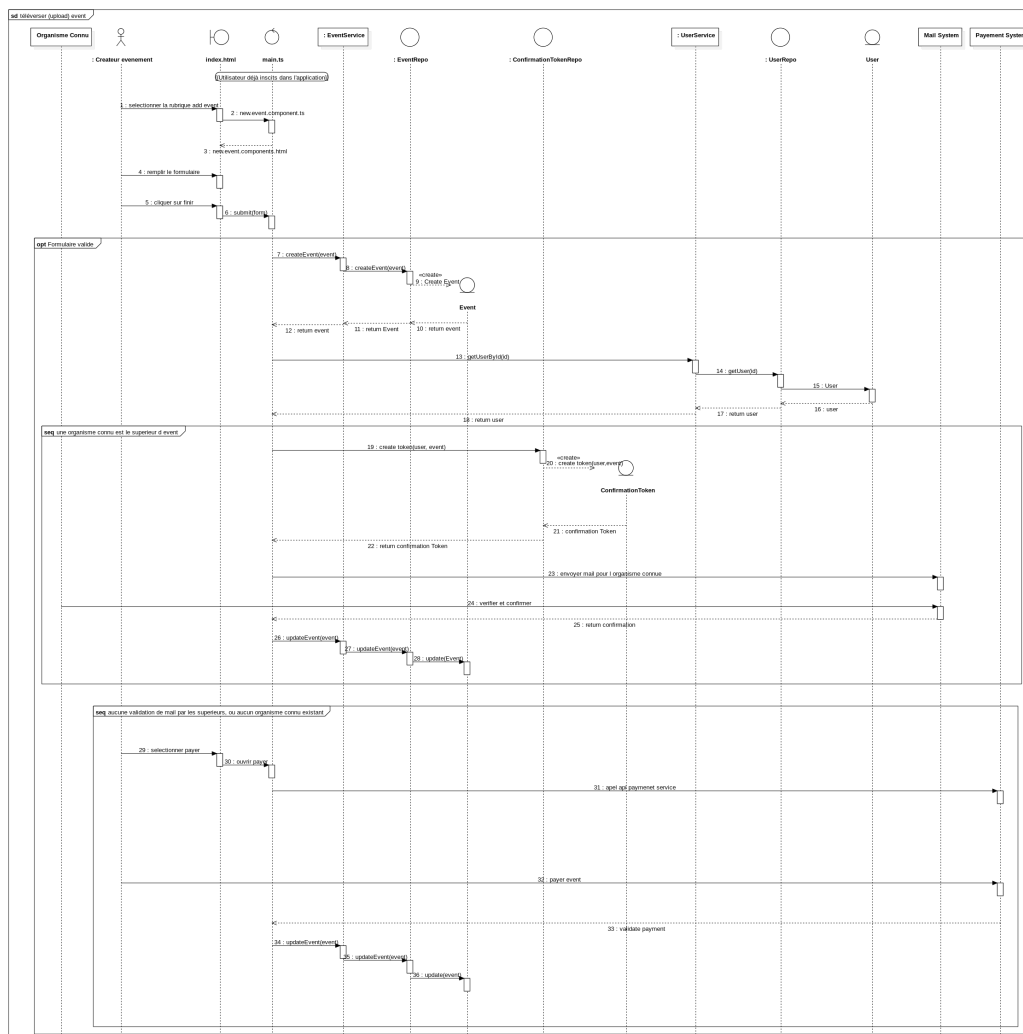


FIGURE 3.9 – Diagramme de séquence objet upload event



# Conclusion

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

Dans la prochaine partie on étudie la partie recherche vers un modèle de véracité pour les documents éducatifs et pour les événements universitaires.

# Deuxième partie

## Vers un modèle de véracité

# Chapitre 4

## Etat de l'art

### 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, énumérer 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.

## 4.1 Véracité

### 4.1.1 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.

(<https://www.lalanguefrancaise.com/dictionnaire/definition/veracite>)

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.

### 4.1.2 La véracité en Big Data

“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.”

(<https://fr.wikipedia.org/wiki/veracite>)

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.

### **4.1.3 La 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.

### **4.1.4 La 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.

## **4.2 Indexation**

### **4.2.1 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.  
([https://fr.wikipedia.org/wiki/Indexation\\_automatique\\_de\\_documents](https://fr.wikipedia.org/wiki/Indexation_automatique_de_documents))

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.  
(<https://fr.theastrologypage.com/indexing>)

L'indexation des documents consiste à associer des mots clés et informations à chaque documents 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

## 4.2.2 Modèles de recherche d'information

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

- Modèle Booléen : C'est le premier modèle de RI, Introduit en 1983 par Salton et McGill, 1<sup>er</sup> 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 :  $w_{i,j} = \{0, 1\}$

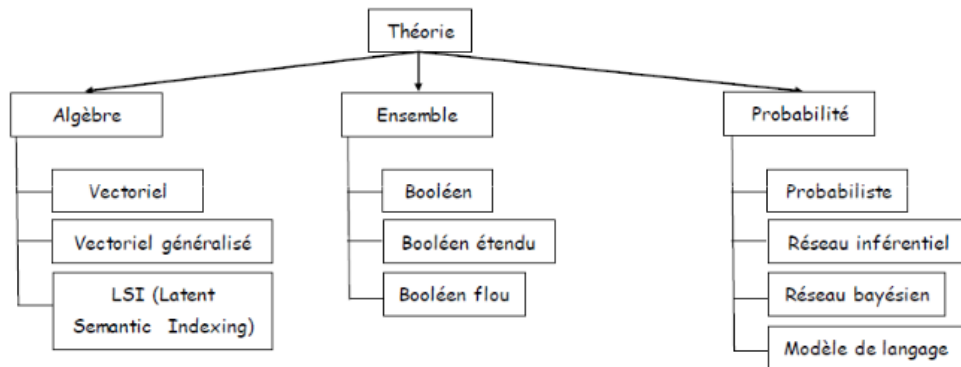


FIGURE 4.1 – Classification des modèles selon la théorie

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 (dénnotant 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

- 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.

— ...

## 4.3 IA et ML

### 4.3.1 Notion d'Intelligence Artificiel

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

implémentée 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. (<https://www.futura-sciences.com/tech/definitions/informatique-intelligence-artificielle-555/>)

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.

(<https://www.oracle.com/fr/artificial-intelligence/what-is-ai/>)

Il y'a quatre écoles en IA :

- Penser comme des humains : Les efforts passionnés 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)

### 4.3.2 Notion d'Apprentissage Automatique

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.

(<https://www.ibm.com/cloud/learn/machine-learning>)

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 la classification et la régression. Pour la classification son but est de classer et donner une famille ou une 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 valeurs numériques, ou prédiction des pourcentages. Les fameux modèles sont : Linear Régression, Polynomial Regression...

### 4.3.3 IA Et Evenements

#### Web Scrapping

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

(<https://fr.wikipedia.org/wiki/WebScrapping>)

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

#### **4.3.4 Les modèles ML qui traitent la véracité des 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.

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

### **4.4 Travaux anciens**

#### **4.4.1 Feuille de route**

Dans cette partie on va s'intéresser sur les travaux anciens qui est autour de 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

#### **Véracité des informations de façon générale**

#### **4.4.2 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 où les données peuvent être incomplètes, imprécises, vagues, 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.

Url : [https://www.academia.edu/42880013/Syst%C3%A8me\\_daide\\_%C3%A0\\_la\\_d%C3%A9cision\\_pour\\_la\\_v%C3%A9racit%C3%A9\\_des\\_donn%C3%A9es\\_dans\\_le\\_contexte\\_de\\_Big\\_Data](https://www.academia.edu/42880013/Syst%C3%A8me_daide_%C3%A0_la_d%C3%A9cision_pour_la_v%C3%A9racit%C3%A9_des_donn%C3%A9es_dans_le_contexte_de_Big_Data))

### 4.4.3 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'une seule source, ne peut pas être des sources différents, en respectant la non redondance.

(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.)

### 4.4.4 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 précédente. 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'une seule source, ne peut pas être des sources différents, en respectant la non redondance.

(Galland, A., Abiteboul, S., Marian, A., and Senellart, P. Corroborating information from disagreeing views. In Proceedings of the third ACM inter-

national conference on Web search and data mining, pages 131–140, 2010.)

#### 4.4.5 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.

(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.)

#### 4.4.6 Découverte de la vérité par estimation de la vraisemblance maximale

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 confiance 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 (Waguih and Berti-Equille, 2014) permet de dire que MLE ne peut pas être utilisé avec de grands nombre de sources ( $>5000$ ). 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'une seule source, ne peut pas être des sources différents, en respectant la non redondance.

(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.)

#### **4.4.7 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. DEPEND 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 DEPEND et ses extensions.

(Dong, X., Berti-Equille, L., and Srivastava, D. Integrating conflicting data : The role of source dependence. PVLDB, 2 :550–561, 08 2009.)

#### **4.4.8 Analyse de crédibilité latente**

LCA (" Latent Credibility Analysis ") est un modèle probabiliste proposé dans (Pasternack 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.

(Pasternack, J. and Roth, D. Latent credibility analysis. In Proceedings of the 22nd international conference on World Wide Web, pages 1009–1020, 2013.)

#### **4.4.9 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.

(Lin, X. and Chen, L. Domain-aware multi-truth discovery from conflicting sources. *Proceedings of the VLDB Endowment*, 11(5) :635647, 2018.)

#### **4.4.10 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.

(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> .)

#### 4.4.11 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.

(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.)

#### 4.4.12 Vérification de faits par partitionnement de données

Dans ce travail, nous intéresser 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 choisi. 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 (La-

mine 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.

(URL : OSIAS NOËL NICODÈME FINAGNON TOSSOU 3 Mars 2021Url  
[https://www.researchgate.net/publication/349733401\\_Verification\\_de\\_faits\\_par\\_partitionnement\\_de\\_donnees\\_Truth\\_discovery\\_by\\_data\\_partitioning](https://www.researchgate.net/publication/349733401_Verification_de_faits_par_partitionnement_de_donnees_Truth_discovery_by_data_partitioning))

## **Véracité des documents**

### **4.4.13 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.

(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) )

### **4.4.14 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.

(Wu, W., Zhou, J., Qu, W. : A survey of syntactic parsing based on statistical learning. J.Chin. Inf. Process. 27(3), 9–19 (2013) )

#### **4.4.15 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 nœuds 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.

(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>)

#### **4.4.16 Conclusion vers notre modèle**

# Chapitre 5

## Modèle de Machine Learning

# Chapitre 6

## Résultat

# Chapitre 7

## Discussion