

INSTITUT SUPÉRIEUR DES ÉTUDES TECHNOLOGIQUES DE BEJA

# Projet de fin d'études

## Conception et réalisation d'un module d'extraction intelligente de données

Elaboré par :Yosr Landolsi

Président de jury :Mme. Balkiss Bettoumi

Encadrant Académique : Mme. Wafa Neji

Rapporteur :Mr. Hattab Ayari

Encadrant Professionnel : Mr. Fethi Amara

Année Universitaire 2023-2024

# Plan

**01**

Introduction

**02**

Contexte du projet

**03**

Analyse et spécifications des besoins

**04**

Pilotage du projet avec SCRUM

**05**

Conception de projet

**06**

Réalisation

**07**

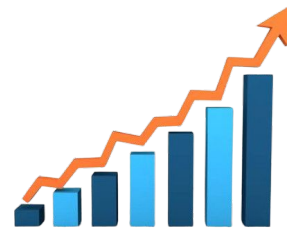
Conclusion et perspectives



# Organisme d'accueil



Fondée depuis  
2017



spécialisée dans la gestion de  
contenu d'entreprise et les  
services de mise en œuvre pour  
la Tunisie et les pays africains.



Installée à Technopark  
Elghazala



# Introduction

- ☐ Classification traditionnelle
- ☐ Utilisation des documents papiers
- ☐ Extraction manuelle des données



# Plan

**01**

Introduction

**02**

Contexte du projet

**03**

Analyse et spécifications des besoins

**04**

Pilotage du projet avec SCRUM

**05**

Conception de projet

**06**

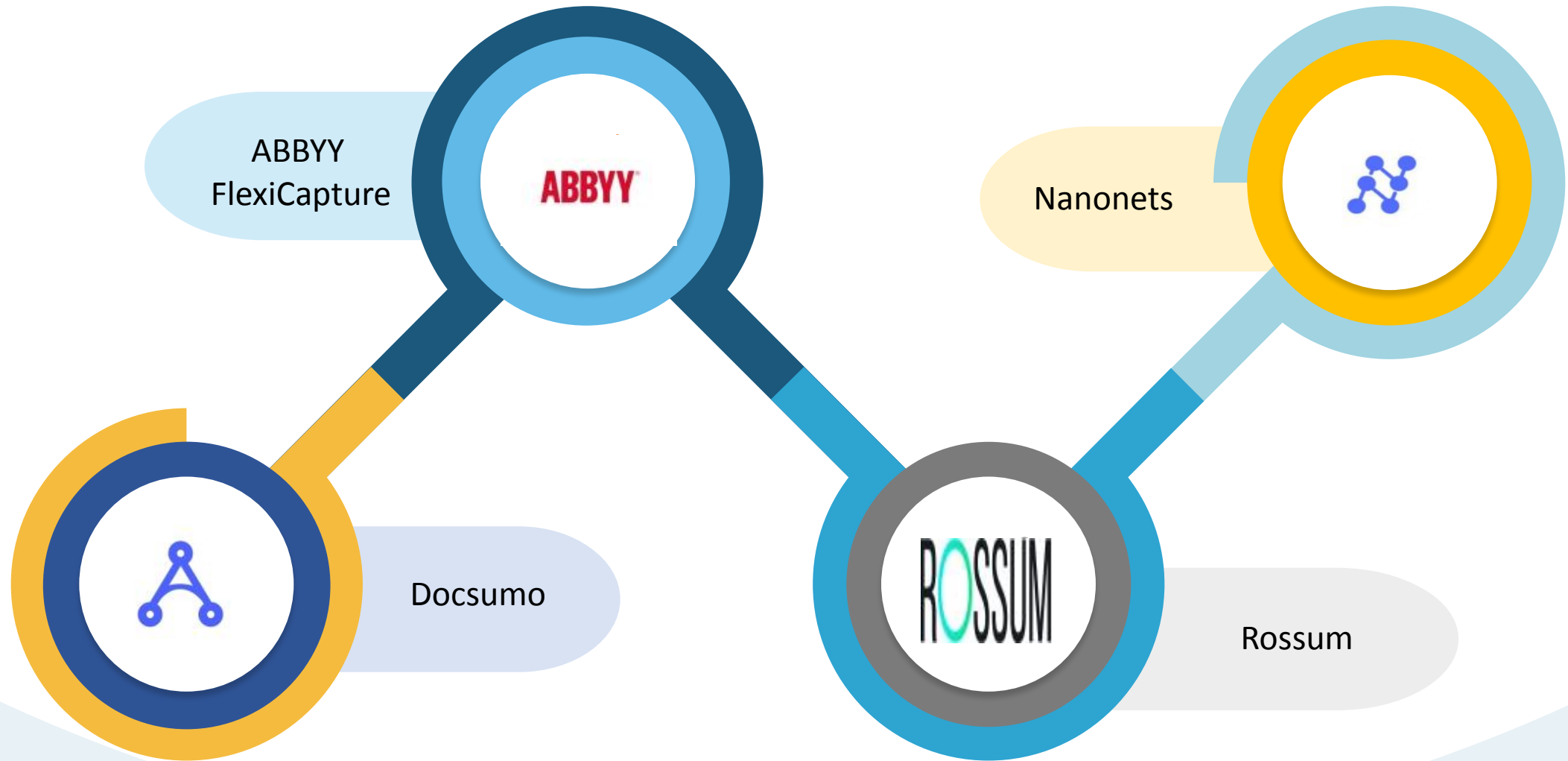
Réalisation

**07**

Conclusion et perspectives



## Contexte de projet / Solutions Existantes



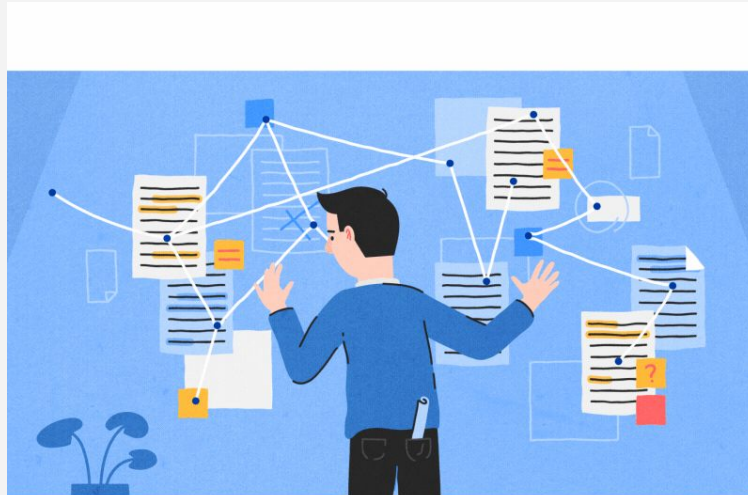
# Contexte de projet / Etude et analyse de l'existant

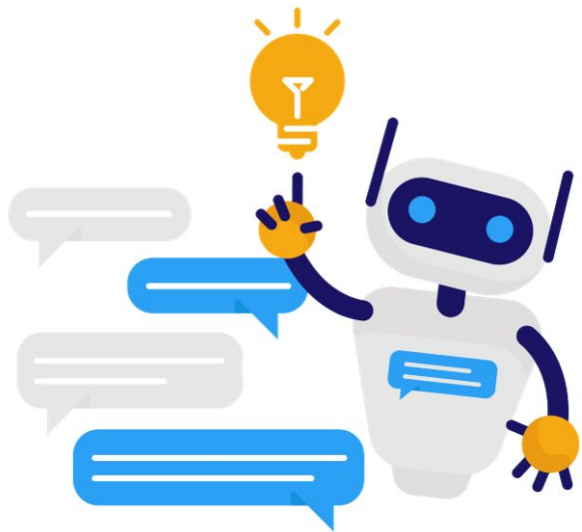


Contraintes de vitesse

Coûts de main-d'œuvre

Performance et erreurs



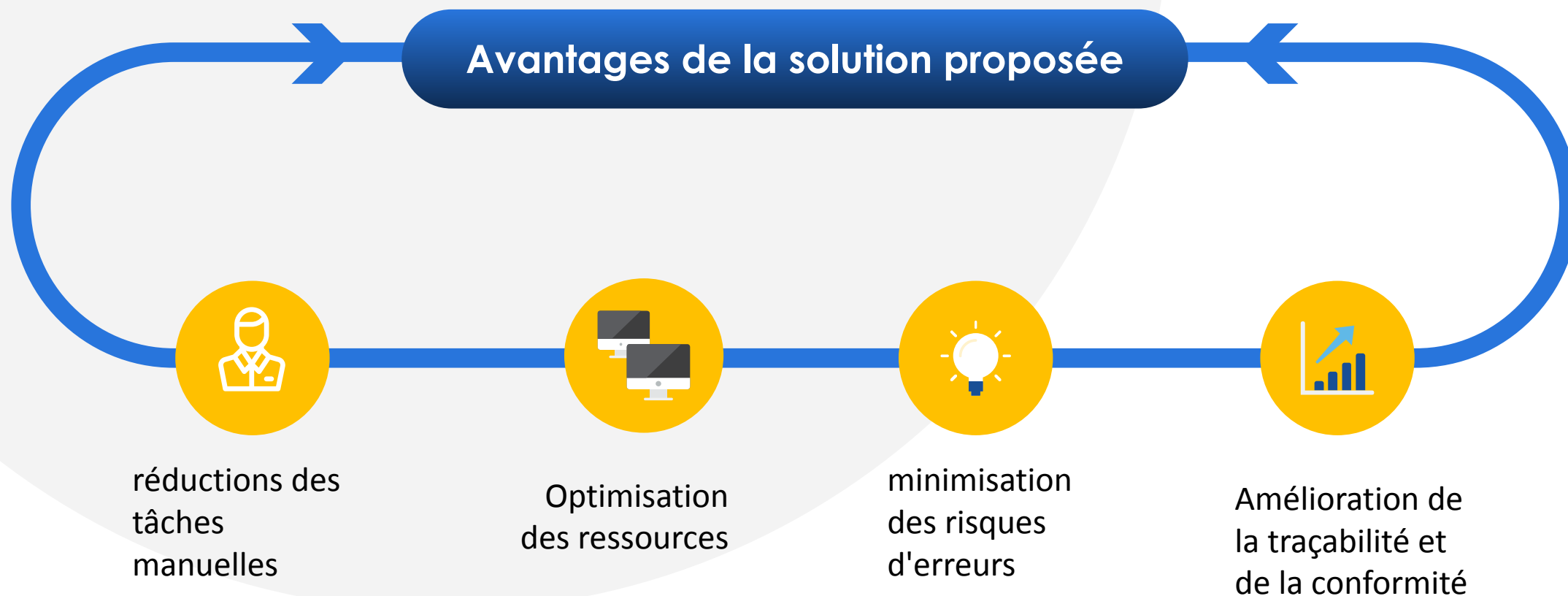


concevoir et réaliser un module d'extraction intelligente de données en utilisant les principes de machine learning.



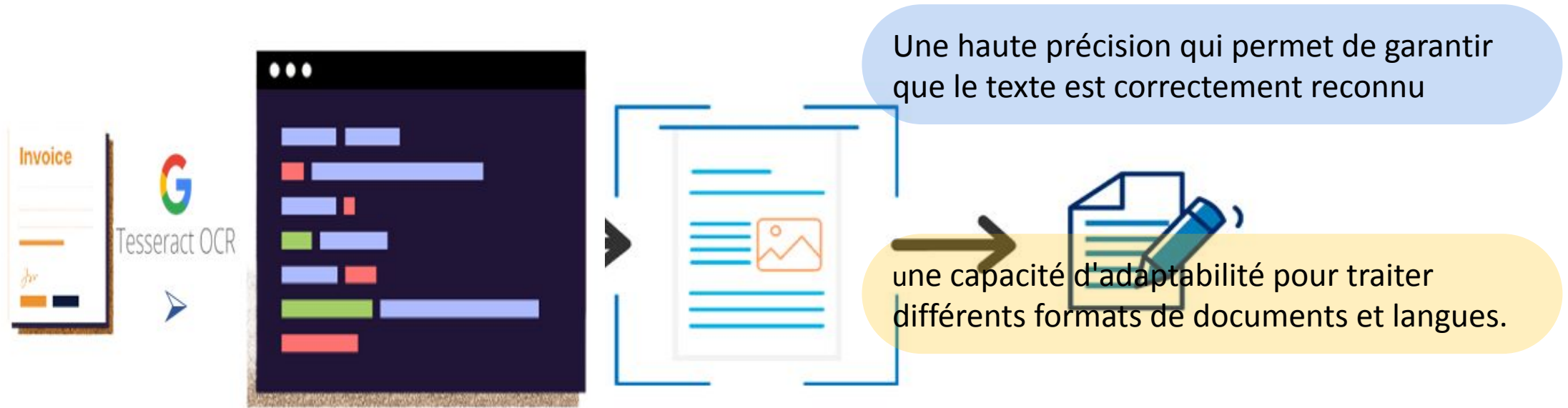


# Avantages de la solution proposée





## OCR





# Classification



Les expression régulières

organiser efficacement les informations

améliorer la précision de l'extraction

optimiser les flux de travail automatisés



# Extraction intelligente de données



L'extraction intelligente des données est un processus qui utilise des techniques d'apprentissage automatique pour identifier et extraire automatiquement des données précieuses à partir des documents



# DocQuery

Modèle entraîné

Multilingue

Document Question-Answer pipeline

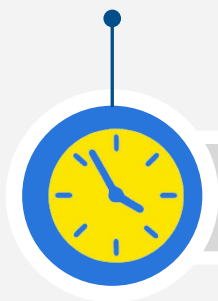
Résultat d'extraction avec un score de certitude



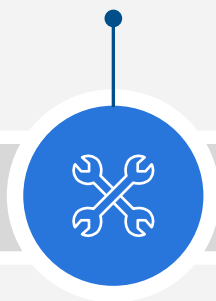


# DocQuery

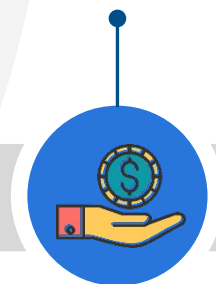
Extraction rapide



Précision et  
flexibilité



Open source



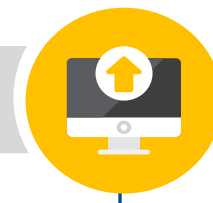
Performance sur  
divers types de  
documents



Compatibilité  
avec divers  
formats de  
documents

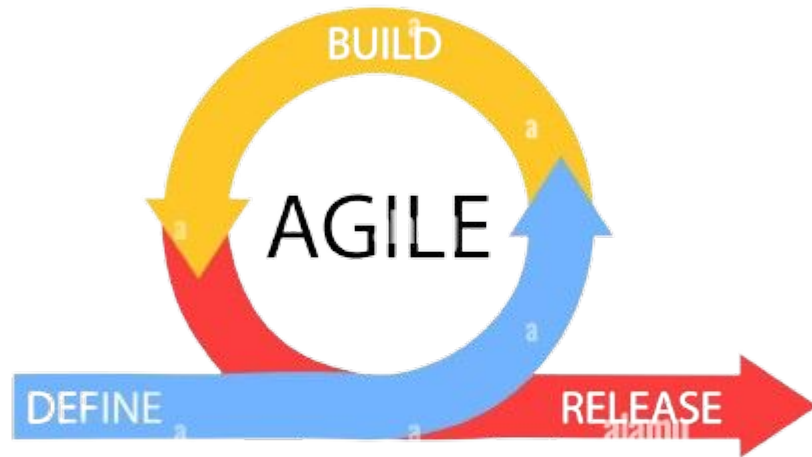


locale





## Contexte du projet /Choix de méthodologie de travail



- ☐ Réduction des risques.
- ☐ Produit flexible et adaptable.
- ☐ Organisation et visibilité

# Plan

**01**

Introduction

**02**

Contexte du projet

**03**

Analyse et spécifications des besoins

**04**

Pilotage du projet avec SCRUM

**05**

Conception de projet

**06**

Réalisation

**07**

Conclusion et perspectives





# Les acteurs



Administrateur

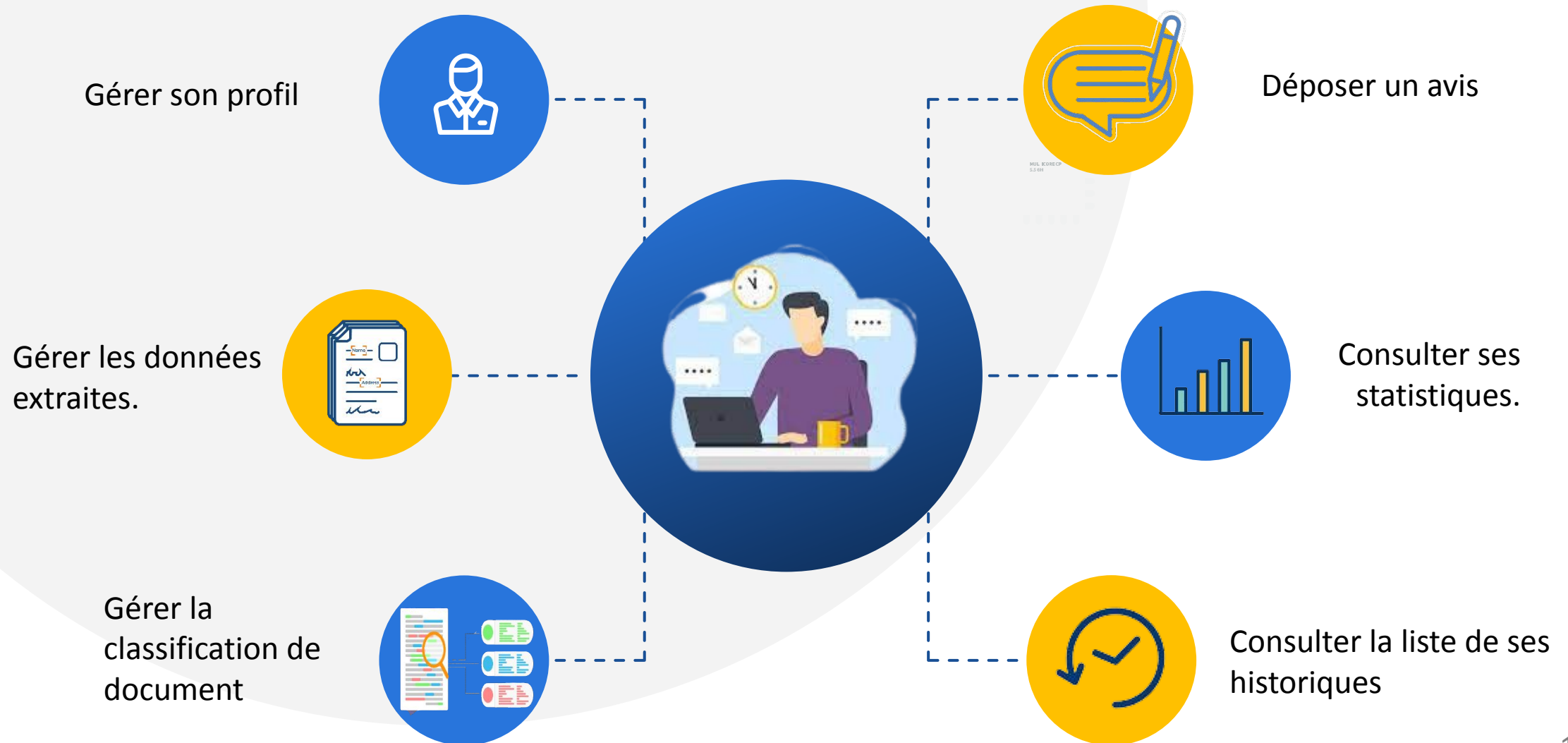


Utilisateur  
CENTRO



Système  
intelligent

# Fonctionnalités d'Utilisateur CENTRO

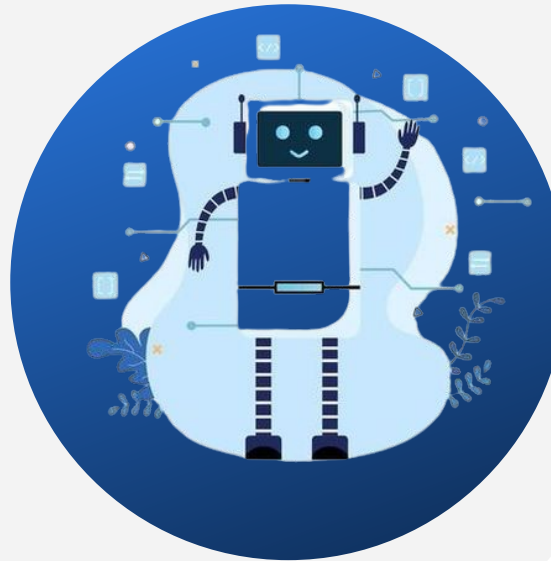


# Fonctionnalités d'administrateur



# Fonctionnalités de système intelligent

Extraire les  
informations  
pertinentes



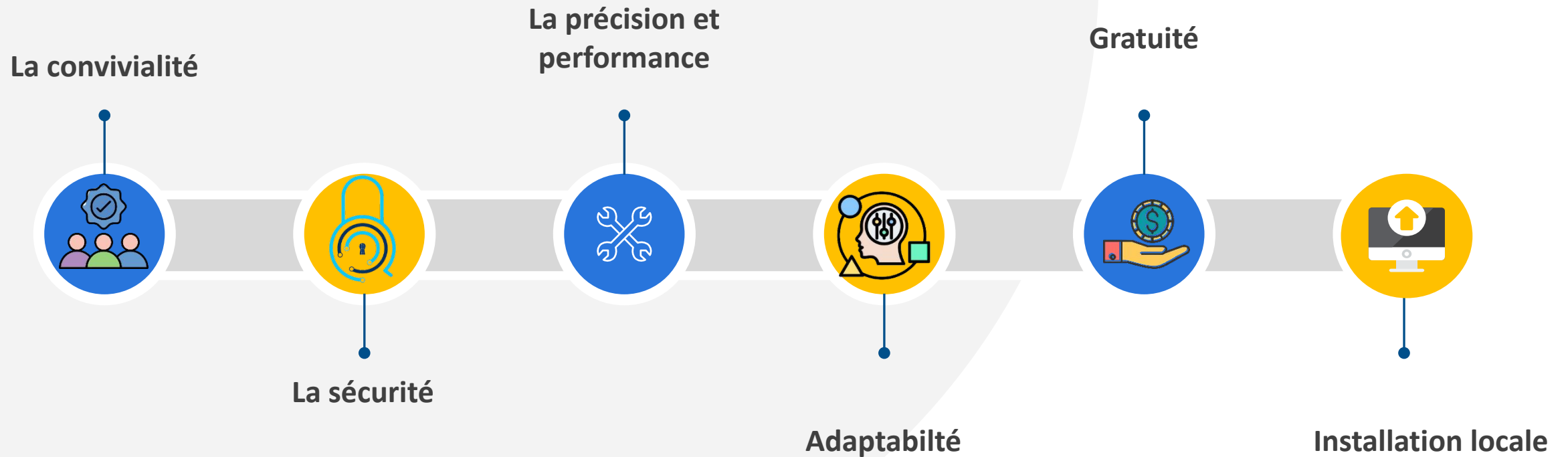
Classifier le document



Effectuer l'OCR

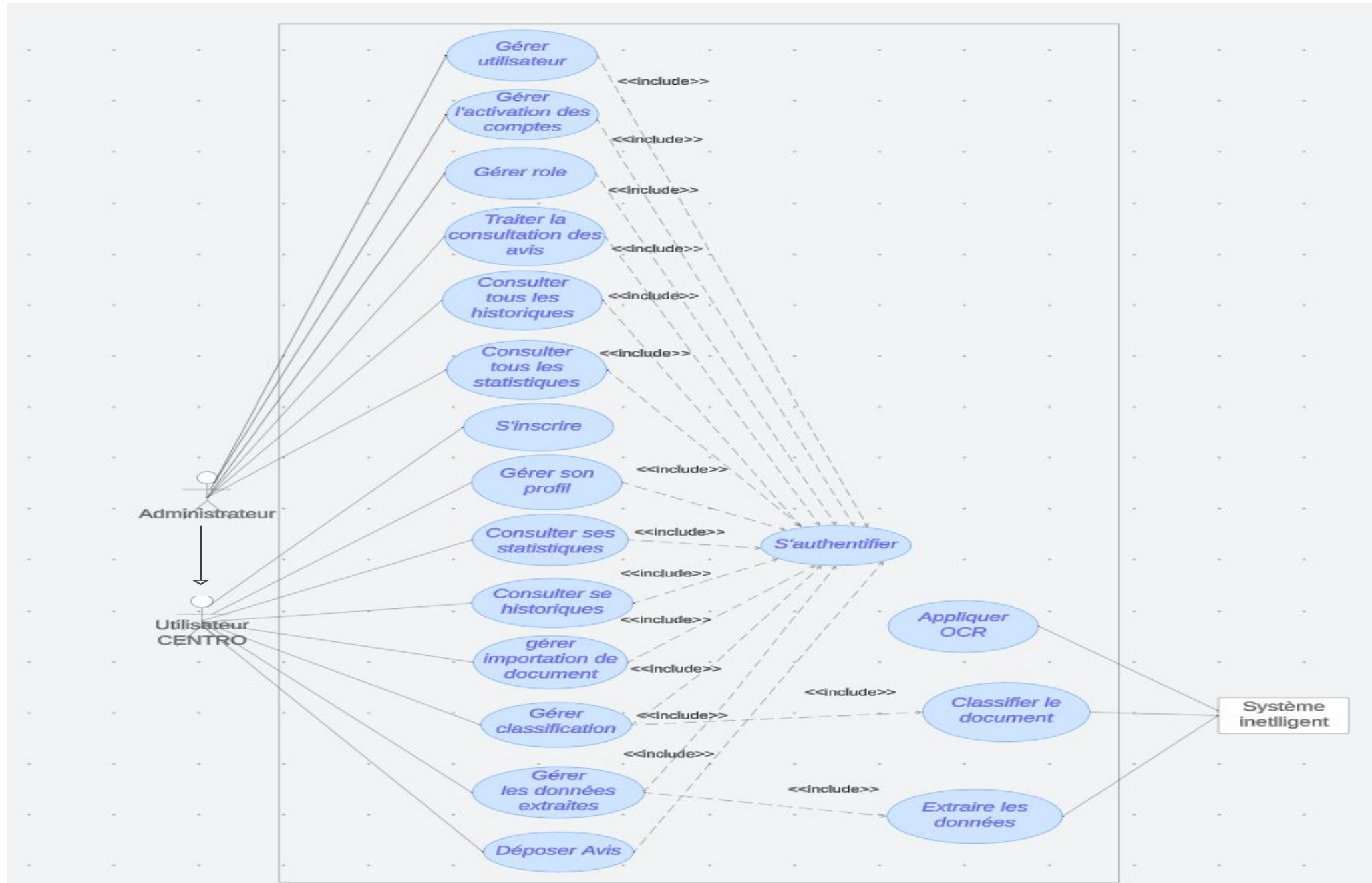


# Analyse et spécifications des besoins/ Les besoins non fonctionnels





# Analyse et spécifications des besoins / Diagramme de cas d'utilisation global



# Plan

**01**

Introduction

**02**

Contexte du projet

**03**

Analyse et spécifications des besoins

**04**

Pilotage du projet avec SCRUM

**05**

Conception de projet

**06**

Réalisation

**07**

Conclusion et perspectives



# Pilotage du projet avec SCRUM/ Backlog Product

				En tant qu'utilisateur CENTRO ou		
<b>Sprint 4</b>		données extraites		des données que j'ai extraites.		
			18	En tant qu'Administrateur CENTRO, je dois pouvoir consulter toutes les statistiques des données extraites.	Basse	2j
	<b>Déposer et consulter avis</b>	Déposer avis	19	En tant qu'Administrateur CENTRO, ou utilisateur je dois pouvoir déposer mon avis à propos les résultats reçu de classification et du modèle d'extraction.	Basse	3j
		Traiter la consultation des avis	20	En tant qu'Administrateur CENTRO, je dois pouvoir consulter toute la liste des avis reçu par les utilisateurs.	Basse	3j
	<b>Statistiques</b>	Consultation les statistiques des	17	CENTRO, je dois pouvoir consulter les statistiques	Basse	5j





# Pilotage du projet avec SCRUM/ Planification des sprints

## **Sprint 1:(2 semaines)**

- Authentification
- Gestion des utilisateurs CENTRO
- Consultation et modification Profil

## **Sprint 2:(4 semaines)**

- Extraction intelligente des documents
- Classification des documents
- Remplissage de formulaire

## **Sprint 3:(3semaines)**

- Implémentation de la sécurité
- Gestion des sessions et des rôles
- Activation des comptes des utilisateurs

## **Sprint 4:(3semaines)**

- Historiques
- Statistiques
- Avis

# Plan

**01**

Introduction

**02**

Contexte du projet

**03**

Analyse et spécifications des besoins

**04**

Pilotage du projet avec SCRUM

**05**

Conception de projet

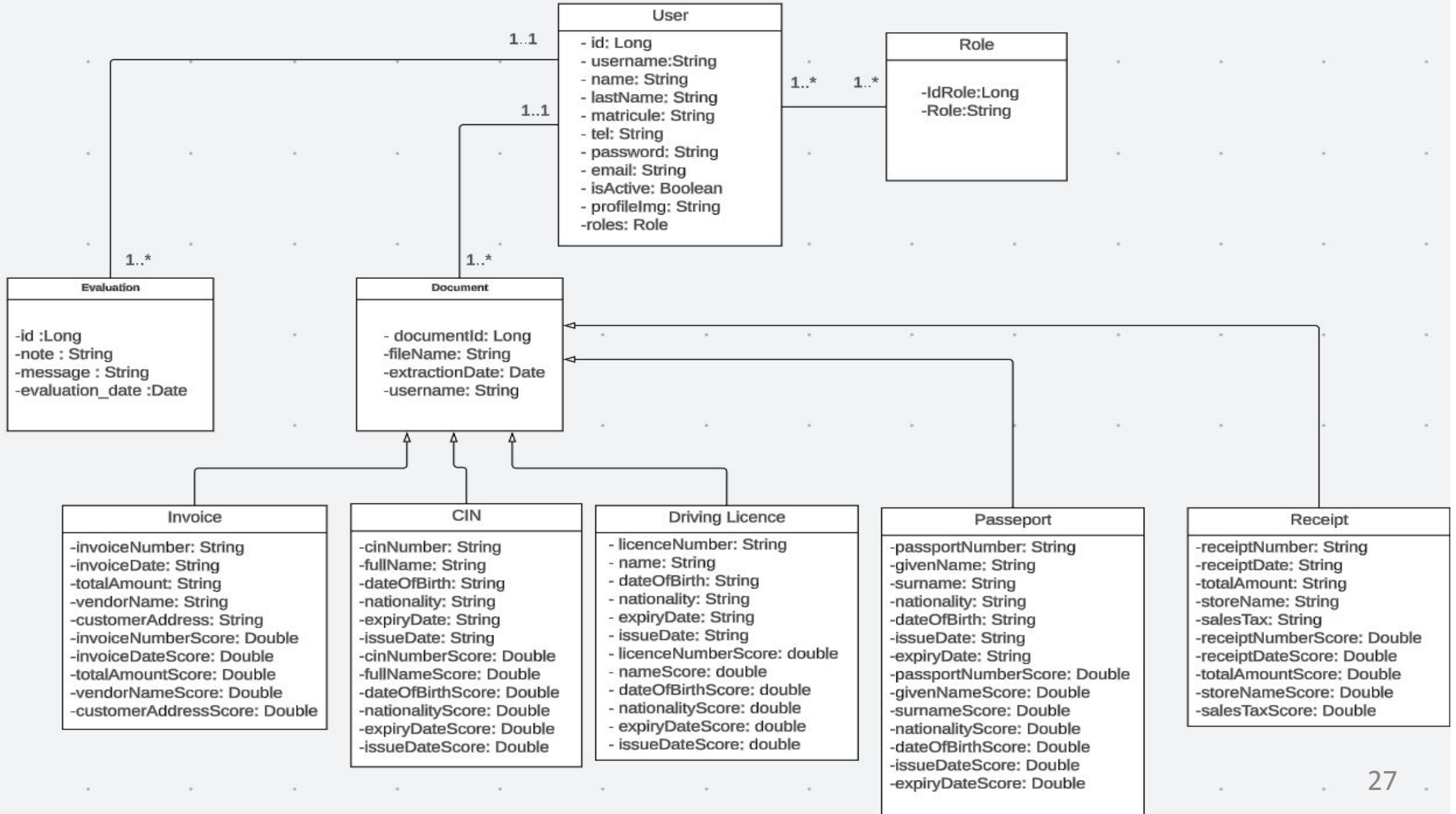
**06**

Réalisation

**07**

Conclusion et perspectives

# Conception de projet/ Diagramme des classes



2 Semaines  
19 Fév -2 Mars

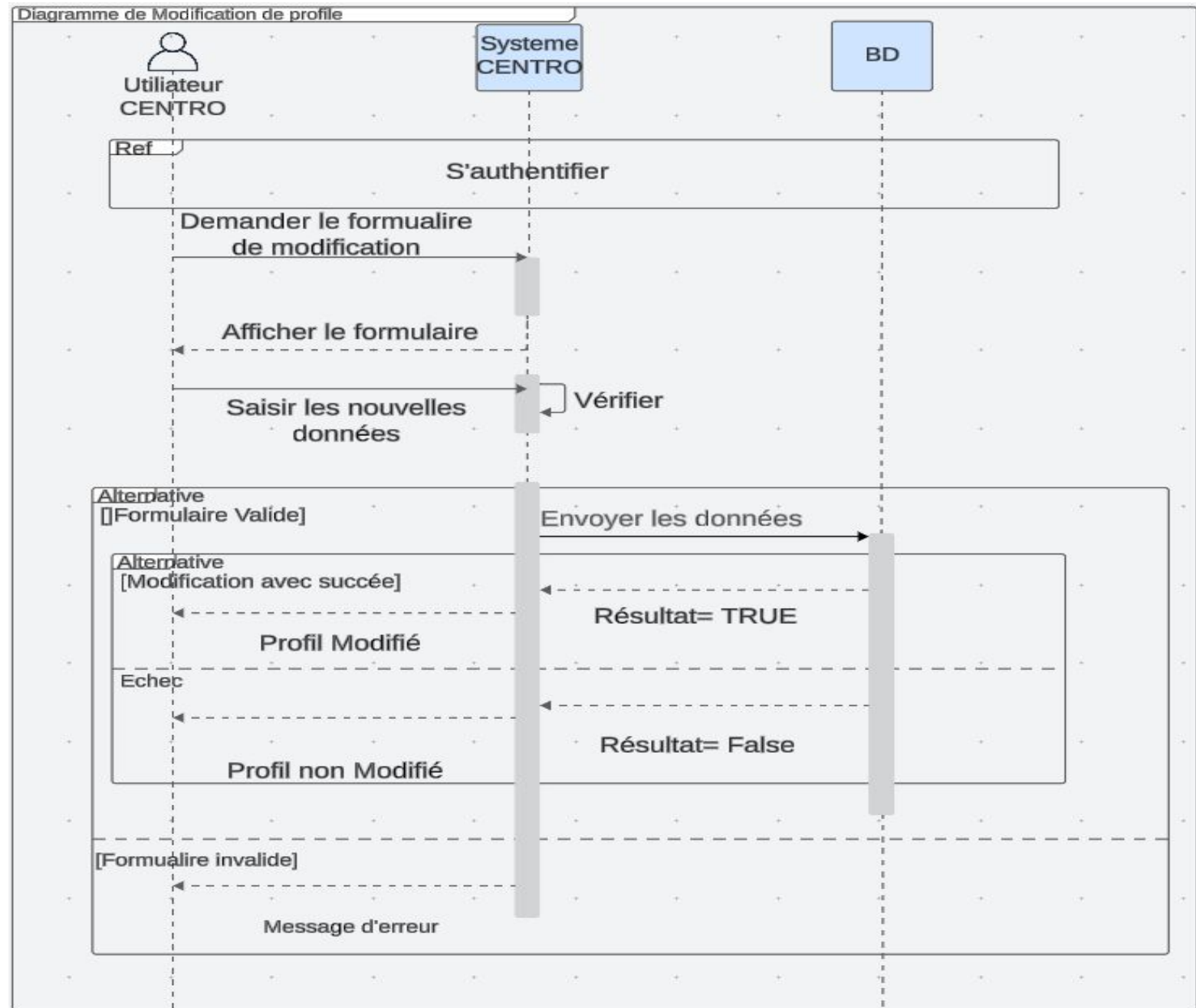
## Sprint 1 : Gestion des utilisateurs CENTRO

- Authentification
- Gestion des utilisateurs CENTRO
- Consultation et modification Profil



# Conception de projet / Diagrammes des séquences

Diagramme de séquences :  
« Modifier son profil »

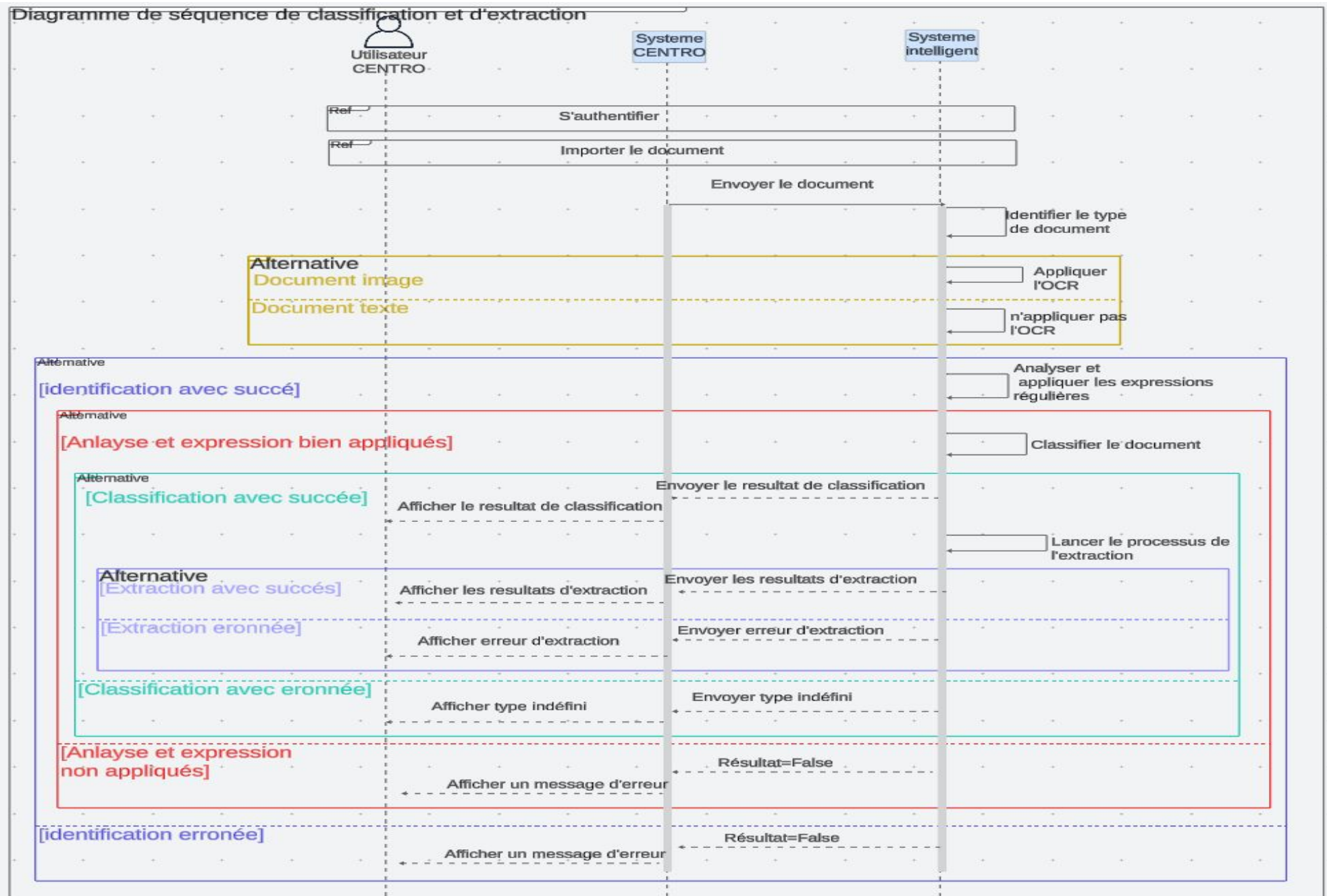






# Conception de projet / Diagrammes des séquences

Diagramme de séquences :  
« classification et extraction »





3 Semaines  
2-21Avril

## Sprint 3 : Implémentation de la sécurité et gestion de l'accès à l'application

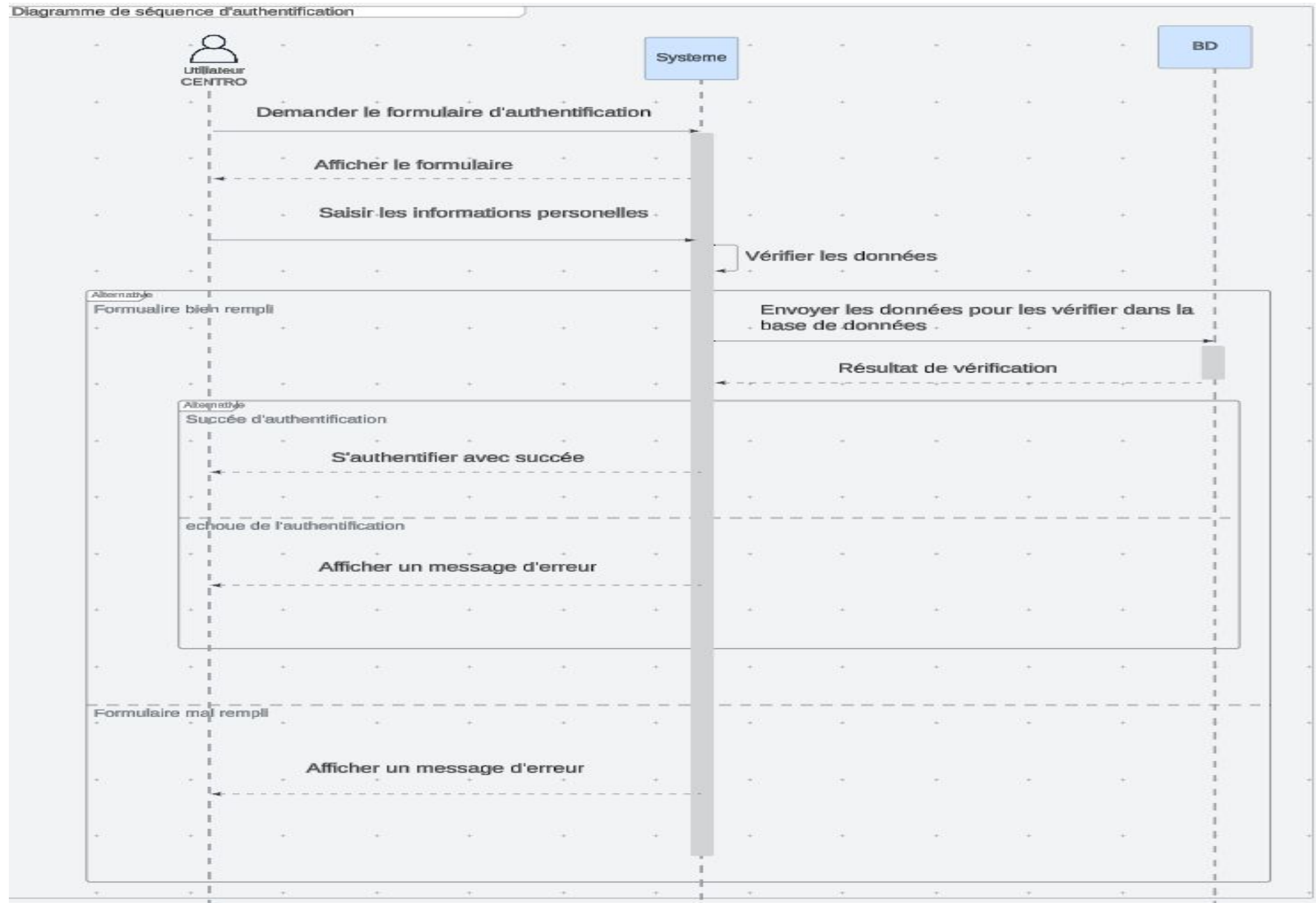
- Implémentation de la sécurité
- Gestion des sessions et des rôles
- Activation de comptes des utilisateurs





# Conception de projet / Diagrammes des séquences

Diagramme de  
séquences :  
« d'authentification »

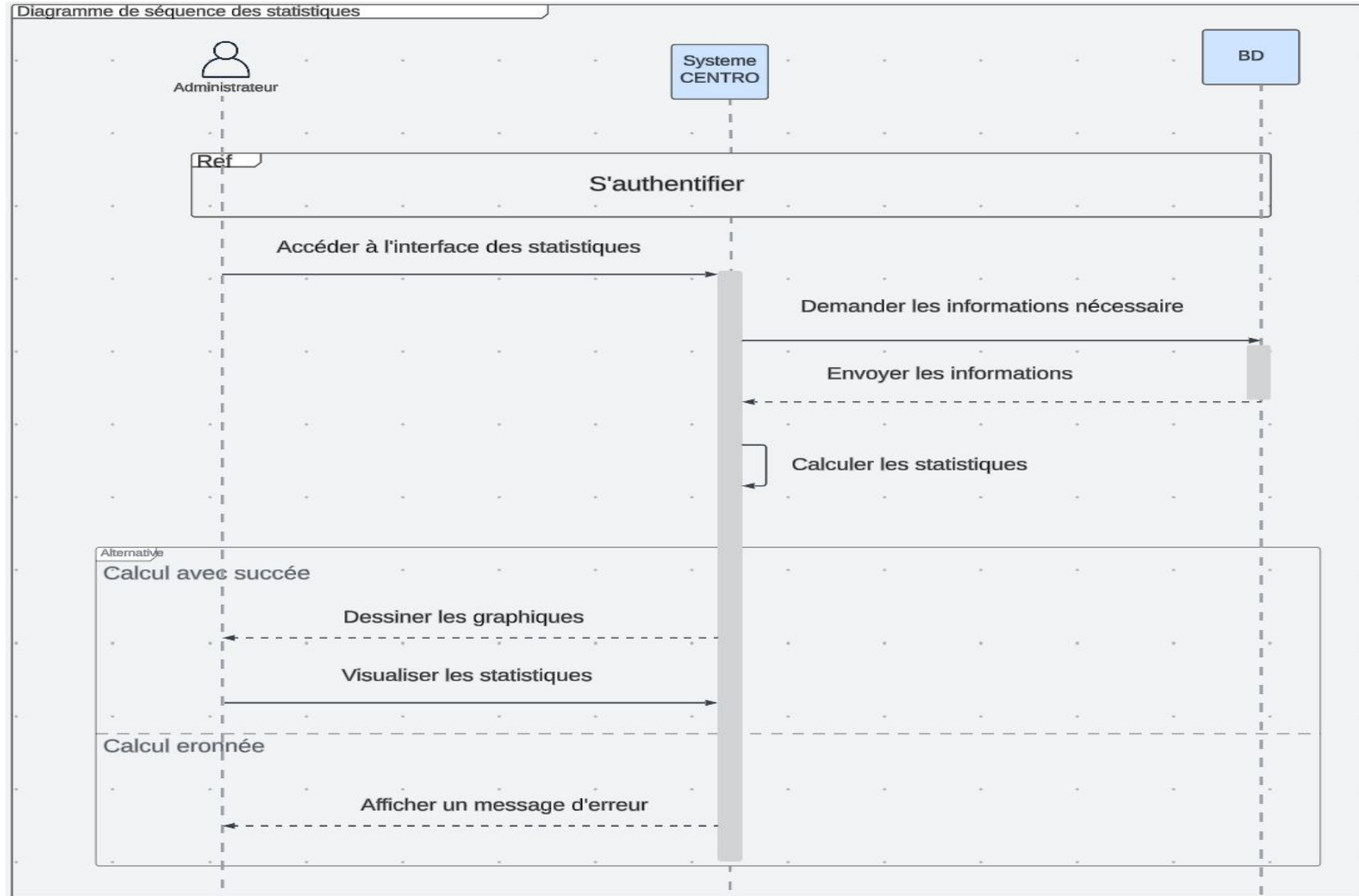






# Conception de projet / Diagrammes des séquences

Diagramme de séquence  
: « Consulter le tableau  
de bord des statistiques »



# Plan

**01**

Introduction

**02**

Contexte du projet

**03**

Analyse et spécifications des besoins

**04**

Pilotage du projet avec SCRUM

**05**

Conception de projet

**06**

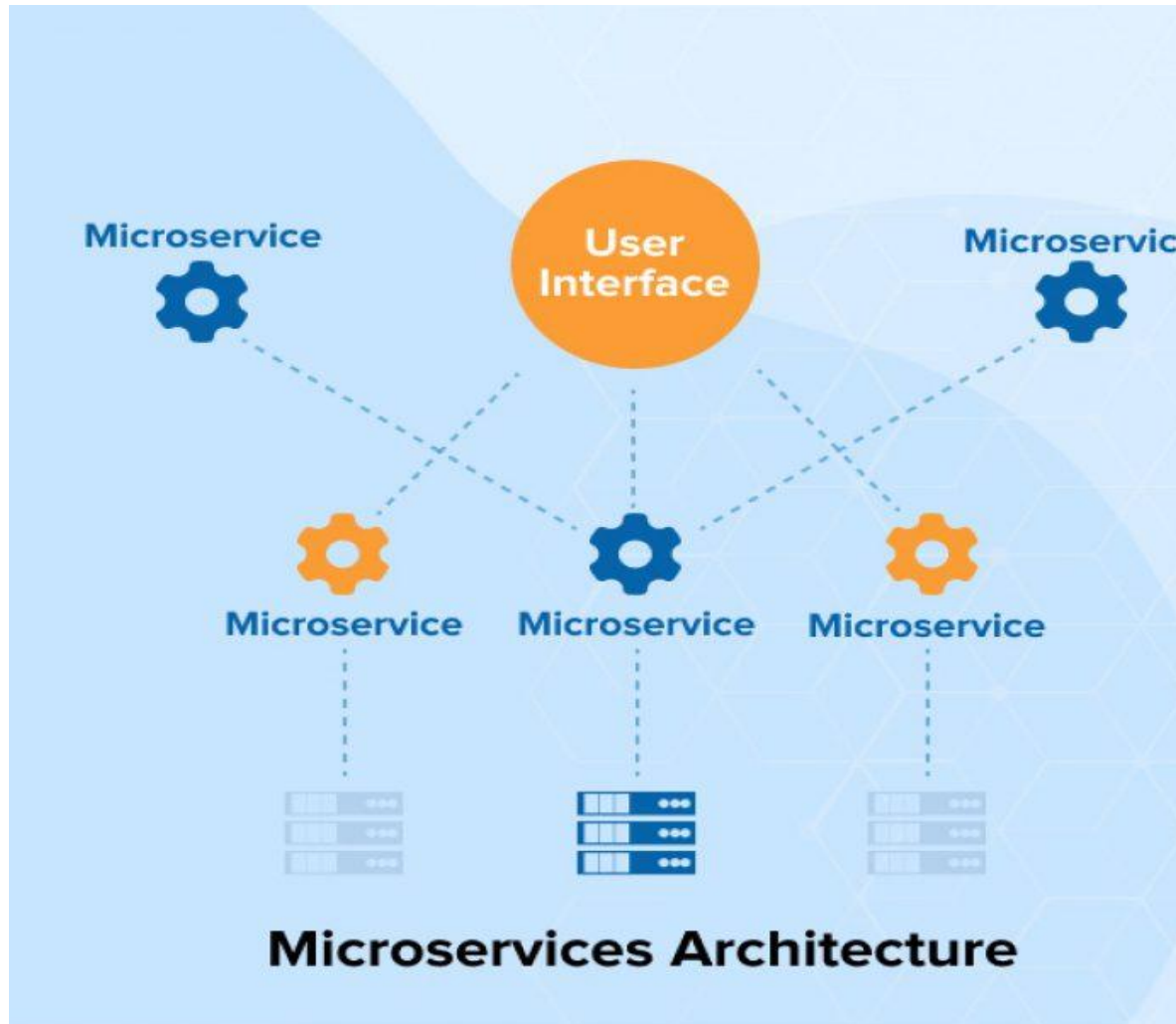
Réalisation

**07**

Conclusion et perspectives



# Architecture Micro-services

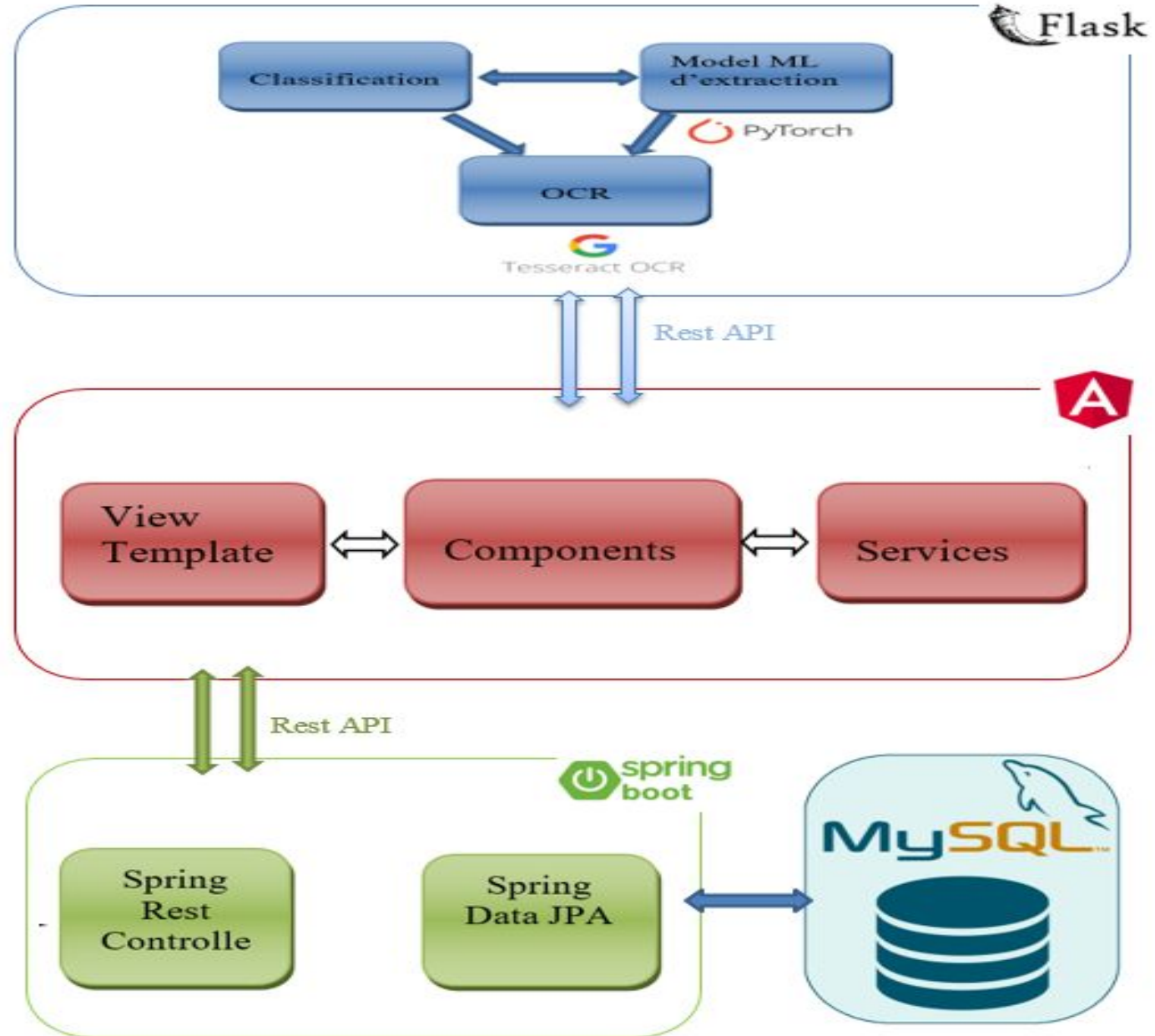


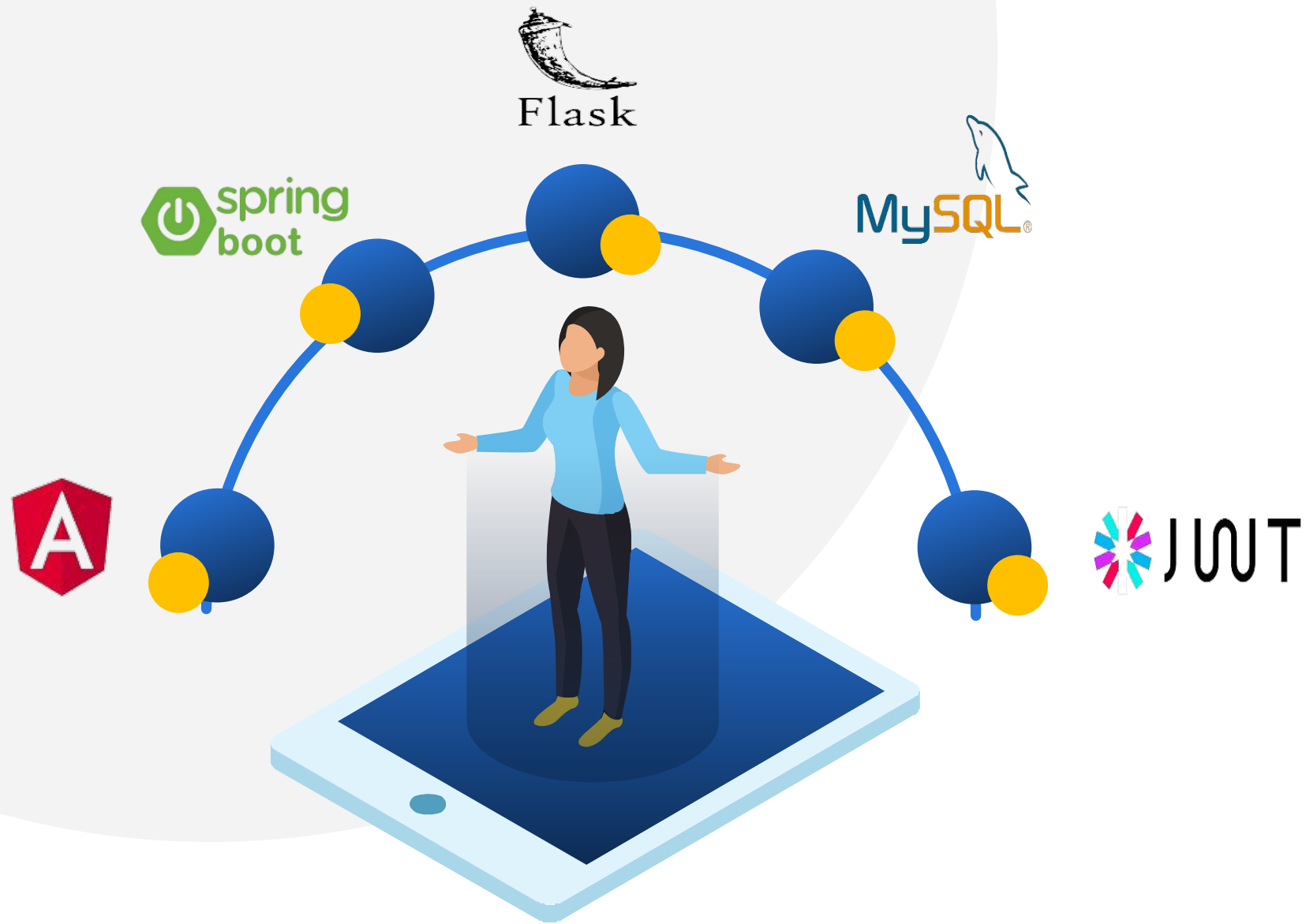
**Flexibilité**

**Réduction des pannes**



# Réalisation / Architecture Logique





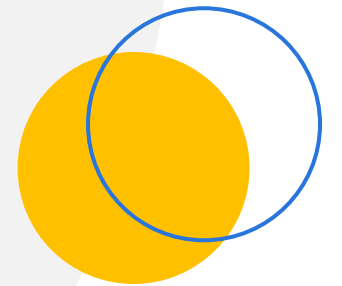


## Réalisation / Environnement logiciels





# Démonstration



# Plan

**01**

Introduction

**02**

Contexte du projet

**03**

Analyse et spécifications des besoins

**04**

Pilotage du projet avec SCRUM

**05**

Conception de projet

**06**

Réalisation

**07**

Conclusion et perspectives



Rendre le modèle d'extraction utilisé autoapprenant

Utiliser un modèle intelligent afin de classifier les documents

**Merci de votre  
attention**

