

Document de cadrage  
Projet de Fin d'Étude

Baptiste Hudyma - Martin Olivier

7 octobre 2019

## Table des matières

<b>1</b>	<b>Compréhension du problème</b>	<b>3</b>
1.1	Contexte . . . . .	3
1.2	Analyse du problème . . . . .	3
1.3	Critères de satisfaction . . . . .	3
<b>2</b>	<b>Objectifs</b>	<b>4</b>
<b>3</b>	<b>Interlocuteurs</b>	<b>5</b>
<b>4</b>	<b>Règles de fonctionnement</b>	<b>6</b>
<b>5</b>	<b>Moyens</b>	<b>7</b>
<b>6</b>	<b>Méthode envisagée</b>	<b>8</b>
<b>7</b>	<b>Organisation de l'équipe</b>	<b>9</b>
<b>8</b>	<b>Planning</b>	<b>10</b>

# **1 Compréhension du problème**

## **1.1 Contexte**

Ce projet est proposé par la Prefecture de la Mayenne dans un contexte de dématérialisation des documents administratifs. En effet dans un effort de numérisation de leur base documentaire, il devient complexe de gérer la grande quantité/diversité de documents informatisés.

La Prefecture de Mayenne envisage donc un projet de numérisation et classification automatisée avec l'aide des nouvelles technologies.

La mise en oeuvre de ce projet permettra de réduire la masse documentaire de l'État.

## **1.2 Analyse du problème**

La Prefecture dispose d'un grand nombre de documents manuscrits qui nécessitent d'être numérisés et classés. Ce travail laborieux est actuellement réalisé par des humains, ce qui rend cette tâche couteuse en temps et en ressources humaines.

La Prefecture envisage donc la mise en place d'un système automatisant toute la chaine de travail, de la numérisation à la classification. Afin de faciliter l'accès aux documents numérisés, la Prefecture désire aussi mettre en place un moteur de recherche afin de réduire le temps de recherche documentaire.

## **1.3 Critères de satisfaction**

TODO

## 2 Objectifs

TODO : en rajouter

- Classer automatiquement des documents numérisés
- Mettre en place un moteur de recherche

### 3 Interlocuteurs

## 4 Règles de fonctionnement

## 5 Moyens

## 6 Méthode envisagée

Le projet est découpé en plusieurs parties principales :

- **Découpage du document en blocs caractéristiques** : Selon l'organisation du document, repérer les zones de texte caractéristiques
- **Reconnaissance des caractères** : Les blocs repérés lors de l'étape précédente doivent être analysés afin d'en retirer le texte exploitable.
- **Analyse du contenu des blocs** : Le contenu textuel des blocs de texte doit être analysé sémantiquement afin d'en retirer le sens principal du document et ses informations utiles.
- **Classification du document en fonction de son contenu** : Le contenu global du document doit être analysé afin de le classer.
- **Création d'une interface de recherche** : Une interface de recherche sera ensuite réalisée afin de faciliter l'accès aux documents classés.



## 7 Organisation de l'équipe

Notre équipe est composée de deux membres :

- Baptiste Hudyma : SPOC, représentant de l'équipe
- Martin Olivier : Spécialiste machine learning (auteur publié dans le domaine)

8    Planning