|  |  |  |
| --- | --- | --- |
| NOTE DE CADRAGE | | |
|  |  |  |
| Named entity extraction from law texts: AI approach | | |

# Plan

# Définition

Les modèles d'extraction d'entités reconnaissent des données spécifiques dans le texte qu’on cible en fonction des besoins. Le modèle identifie les éléments clés du texte, puis les classe dans des catégories prédéfinies. Cela peut nous aider à transformer des données non structurées en données structurées lisibles par machine. Nous pouvons ensuite appliquer le traitement pour récupérer des informations, extraire des faits et répondre à des questions.

# Contexte du projet

De nos jours, en raison de la quantité massive de jurisprudence publiée, il est physiquement impossible pour les chercheurs juridiques de lire, d'analyser et de systématiser tous les documents internationaux et décisions des juridictions nationales. À l'ère du big data légal, de plus en plus les chercheurs commencent à s'apercevoir que la combinaison des méthodes juridiques doctrinales traditionnelles et des méthodes quantitatives empiriques est une voie prometteuse qui nous aidera à donner un sens à toute la jurisprudence disponible.

Cependant, depuis 2015, le nombre d'articles portant sur la prédiction décisions utilisant l'apprentissage automatique a augmenté rapidement, dans le monde entier. Dans le cadre de notre projet d’expertise nous allons donc nous baser sur les différentes approches existantes basées sur l’IA afin d’extraire des données spécifiques.

# Objectif du projet

L'objectif de ce projet d’expertise serait donc d'implémenter des algorithmes d’IA pour extraire des entités légales dans des textes juridiques.

# Périmetre du projet

* Extraction des entités afin de savoir le contexte général d’un texte, souvent c’est le process de répondre à la question : Qui a fait quoi à qui et quand ?
* Il s’agit premièrement d’identifier et puis catégoriser en (entités, temps et quantités)

# Pré-requis

* Rnn (Recurrent Neural Network)
* NLP (Natural Language Processing)
* NLTK (tokenizing,stemming,topis segmentation..)
* Spacy

# Ressources