

Projet 4 – Reconnaissance de références juridiques

L'objectif de ce projet est de proposer une solution technique pour la détection de références dans les textes juridiques.

Le jeu de données qui vous est fourni est une base d'arrêtés au format HTML. Cette base a été constituée à partir d'arrêtés préfectoraux scannés au format PDF, OCRisés, puis convertis en documents structurés (HTML) facilement exploitables par une machine.

Exemple d'entrée :

```
<p>Les prescriptions du présent arrêté se substituent aux dispositions imposées par les actes administratifs suivants :</p>
<ul>
<li>arrêté préfectoral n° 87/IC/123 du 1er avril 1987 ;</li>
<li>annexe 2 de l'arrêté ministériel du 12 octobre 1999 ;</li>
</ul>
```

Les références à d'autres textes devront être insérées dans le document HTML de l'arrêté sous forme de balises `<a>` qui pourront donc être facilement retrouvées par un programme.

Exemple de sortie :

```
<p>Les prescriptions du présent arrêté se substituent aux dispositions imposées par les actes administratifs suivants :</p>
<ul>
<li><a>arrêté préfectoral n° 87/IC/123 du 1er avril 1987</a> ;</li>
<li>annexe 2 de l'<a>arrêté ministériel du 12 octobre 1999</a> ;</li>
</ul>
```

Votre code doit couvrir un certain nombre de textes juridiques différents :

- Code de l'environnement
- Directive (européenne)
- Loi
- Arrêté ministériel
- Décret
- Circulaire
- Arrêté préfectoral
- Arrêté (sans précision)