

Projet 5 – Construction d’une typologie des modifications

L’objectif de ce projet est de proposer une solution théorique et éventuellement technique pour construire une typologie des actions de modification que définit un arrêté préfectoral sur les autres arrêtés préfectoraux et textes antérieurs qu’il référence.

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.

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

Exemple d’entrée :

```
<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 préfectoral n° 90/IC/402 du 12 octobre 1999</a> ;</li>
</ul>
```

Pour cet exemple, le but sera alors de détecter que *le présent arrêté remplace partie* des arrêtés listés. À partir de tous les exemples de modifications qui vous sont donnés, il s’agira alors de construire un ensemble de catégories qui vont servir de base pour une classification des opérations (ajout, modification, abrogation) à réaliser automatiquement. Il ne vous est pas demandé de réaliser cette classification automatique.