Projet 2 – Détection de segments modificatifs

L'objectif de ce projet est de proposer une solution technique pour la détection 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. Vous pouvez pour cela développer différentes règles et algorithmes, par exemple en vous basant sur les termes du paragraphe qui englobe la référence ou sur les mots proches.

Votre résultat devra se présenter sous la forme d'une liste d'opérations avec une origine et une ou plusieurs cibles avec à chaque fois la partie du segment modificatif qui permettra de classifier l'action à mener. La partie du segment modificatif que vous arrivez à extraire pourra être plus grande que le strict nécessaire, à partir du moment où elle contient ce qui permet de classifier.

Exemple de sortie :

```
modifications = [
   ''arrêté préfectoral n° 23/IC/41 du 23 février 2023'' : {
      ''arrêté préfectoral n° 87/IC/123 du 1er avril 1987'' : ''Les prescriptions du présent arrêté se substituent aux dispositions imposées par les actes administratifs suivants'',
   },
]
```