|  |  |  |
| --- | --- | --- |
| Etapes | Objectif |  |
| 1ère Etape | Générer et formater le fichier datset2 avec les titres des articles et keywords |  |
| 2ème Etape | Faire la matrice Termes/Document et réaliser le clustering |  |
| 3ème Etape | Extraire un article , sortir les n-grammes à partir de ses mots-clés |  |
| 4ème Etape | Interroger l’API Babelnet et extraire les domaines/catégorie dont l’occurrence est la plus grande parmi les mots-clés |  |

Tableau Comparatif

Prendre un article dans chaque cluster, obtenir sa catégorie et ainsi cette catégorie sera la sémantique du groupe.