

En premier lieu, nous avons effectué des recherches approfondies sur les différentes mesures de similarités entre chaînes de caractères, pour pouvoir les implémenter par la suite en langage de programmation « Python », car c'est celui qui nous semblaient le plus simple à utiliser.

Ensuite, nous nous sommes concentrés avec notre encadreur à étudier la structure des bibliothèques de la bnf et rameau, afin d'en arriver au constat que l'élément le plus pertinent à aligner de la bnf se trouvait à la fin de la ligne c'est-à-dire après le dernier « \$ » et dans ce cas le pays ou l'appartenance. Contrairement au fichier de rameau là où l'information qu'on l'on voudrait aligner se trouvait au début de la ligne c'est-à-dire après le premier « \$ ». De ce fait, nous nous sommes mis à tester les mesures en question notamment celle de « jaro » et de « levenstein », et nous en sommes arrivé à un point où il nous faudrait désigner un seuil de similarité requis, pour pouvoir considérer une correspondance possible entre deux chaînes de caractères, outre bien évidemment la correspondance parfaite.