TP1 Structure interne de données

Emilie Caillerie et Véronique Demianenko

15 mars 2024

1 Introduction

Dans ce TP, nous utilisons le document texte_Shakespeare. Nous allons essayer de trouver une fonction de hachage avec le plus petit nombre de collisions possible et une bonne distribution dans la table de hachage.

Avant de commencer à élaborer nos fonctions, nous devons déterminer M. Pour trouver M on commence par compter le nombre de mot dans le texte. Ensuite, sachant qu'on veut un taux d'occupation de 30%, on prend M valant le triple du nombre de mot dans le texte. Au cas où, on prend le nombre premier suivant pour éviter tout problème dans le cas du hachage par division.

Ainsi, ici on prend $M = 68729$

2 Fonction de hachage par division

Pour cette méthode, on convertit les lettres de l'alphabet en leur code utf8, et pour le mot demandé, on lui associe un nombre obtenu en ajoutant à chaque itération l'indice du caractère multiplié par 26 à la puissance de la position du caractère dans le mot, comme vu dans le cours. On applique le modulo à ce nombre, ce qui donne l'emplacement du mot dans la table de hachage. La fonction utilisée est donc h(x)=x modM.

3 Fonction de hachage par multiplication

Pour cette méthode, on réutilise le même principe, mais au lieu d'utiliser les codes utf8, on prend simplement les indices des lettres dans l'alphabet. La fonction utilisée ici est floor(M * (xAmod1)), avec $A = \frac{\sqrt{5}-1}{2}$.

4 Nombre de collisions

Dans ce TP nous nous occupons pas de créer la table de hachage à proprement parler. On veut obtenir le nombre de collision pour les deux fonctions.

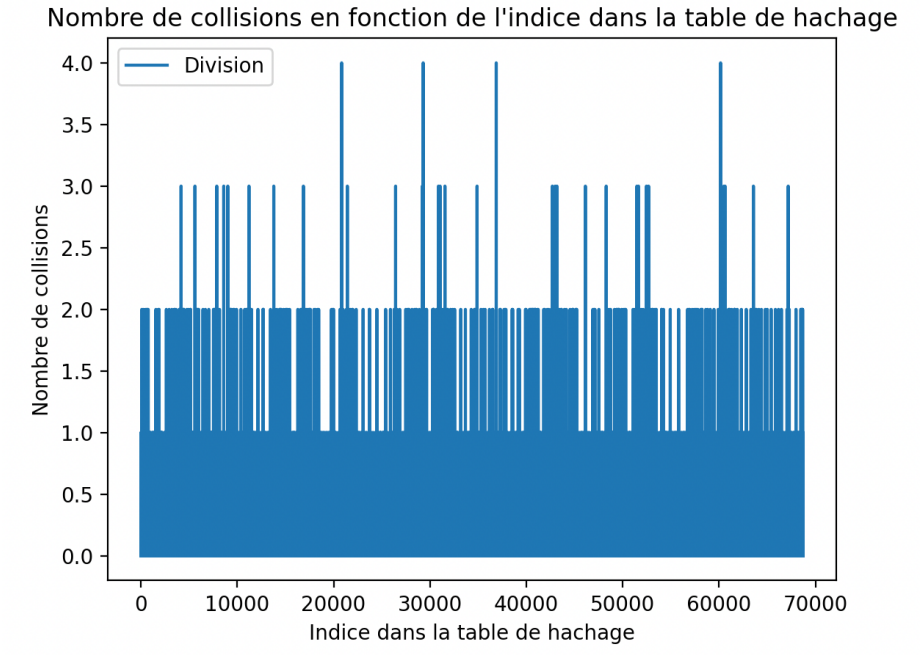
Pour cela on crée une table de taille M remplie de 0 et on applique la fonction de hachage pour chaque mot du texte. A chaque mot, on regarde si l'élément d'indice h(x) dans la table est de valeur nulle. Dans ce cas, on remplace le 0 par le mot. Le cas échant on ajoute 1 à notre compteur de collision.

Pour notre fonction de hachage par division on trouve un nombre de collisions de 3421 (soit 15%) tandis que pour notre fonction de hachage par multiplication on trouve un nombre de collisions de 5950 (soit 26%).

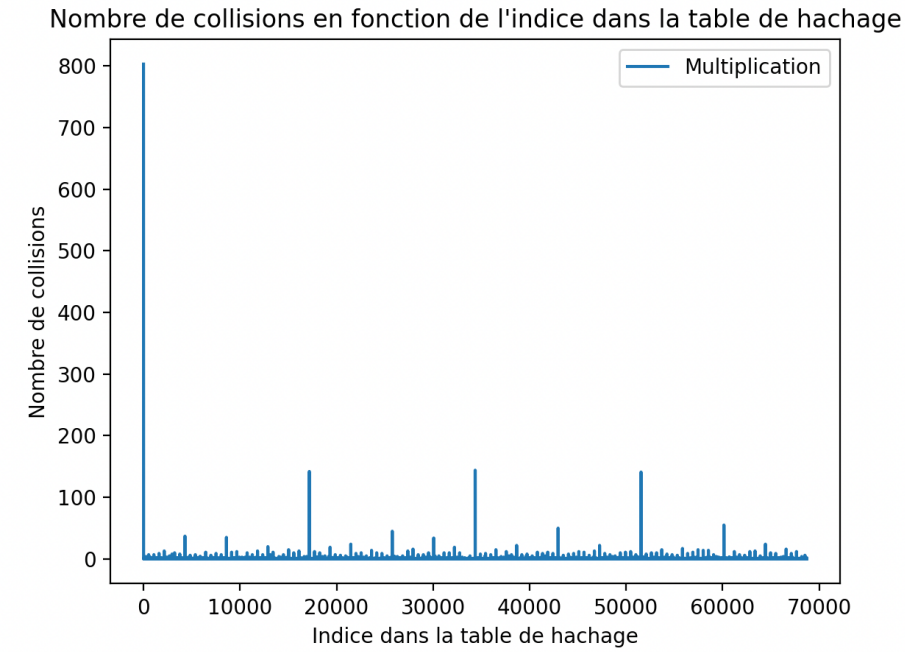
5 Uniformité des fonctions

On cherche aussi à savoir laquelle des deux fonctions offre une meilleure distribution des mots, pour cela on observe l'uniformité des deux en affichant un graphe de l'histogramme des valeurs pour les deux fonctions.

Pour la division, on obtient :



Pour la multiplication, on obtient :



On observe une valeur aberrante au niveau de 0 pour le hachage par multiplication, mais à part cela, les deux fonctions sont uniformes. La fonction par division présente cependant une meilleure uniformité.