Exercice du BloomFilter

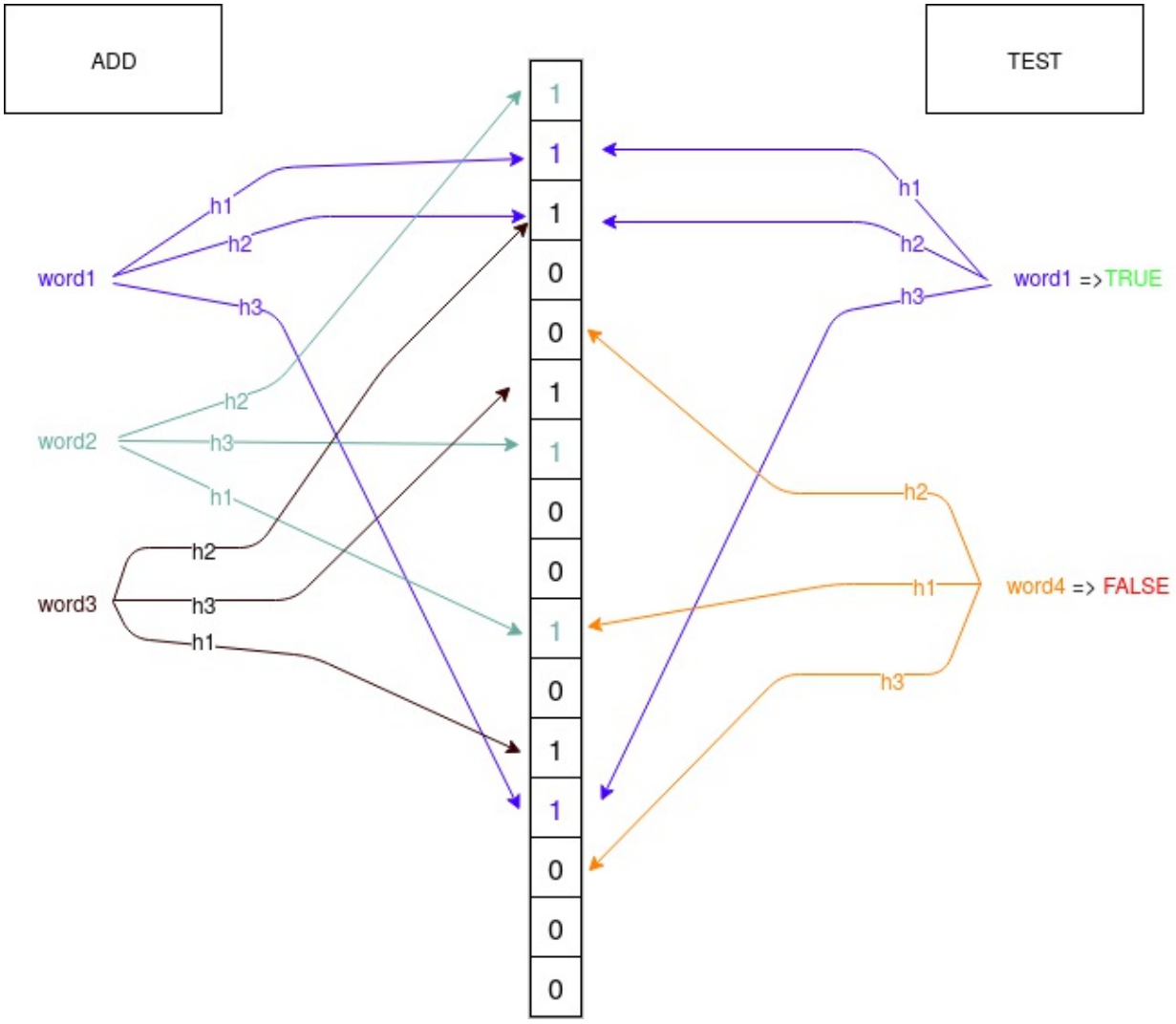
Le "BloomFilter" propose de faire de la recherche d'élément ().
Pour savoir si un élément est présent dans un tableau de taille N (dans notre exercice N = 128), cette méthode doit implémenter 2 fonctions :

La méthode d'ajout : add(String entry)

Pour ajouter un élément au tableau, l'entrée est hashée avec K algorithmes différents. (dans notre exercice, K = 3)
De chacun de ces hashs, on trouve un moyen intelligent de marquer une case du tableau, la seule condition étant que la répartition sur les cases doit être uniforme.

La méthode de test : test(String testWord)

Cette méthode teste si l'argument qui lui a été passé en paramètre à déjà été inséré dans le tableau. Pour cela, on hash testWord avec les mêmes K algorithmes que ceux de la fonction add, et on vérifie si toutes les cases sont marquées. Elle renvoie ou affiche simplement un booléen.



Exercice

Pour cet exercice, commencera par utiliser un tableau de 128 cases.
On utilisera aussi les trois algortihmes de hashage suivants: SHA-1, SHA-256 et SHA-512.

scénario de test:

```
test("word1"); -> false
add("word1");
test("word1") -> true
test("word2") -> false
```

Évolution :

On aimerait créer une fonction `logPresent(String searched)` qui permet de savoir quelles autres entrées ont déjà marqué les cases de chacun des hashs.
Exemple : je reçois le mot `word1`, je trouve les indices correspondant aux trois hashs: 2, 6, 4.
Je veux afficher quelles sont les entrées qui on eu un de leurs hashs qui a donné 2, celles qui ont eu un hash qui a donné 6, et celles qui ont eu un hash qui a donné 4.

Évolution n°2:

Écrire un programme qui permet de répondre aux questions suivantes :
1 - Je veux insérer 20 éléments dans un tableau de 128 cases. Combien d'algorithmes de hashage K dois-je utiliser pour minimiser les chances de collisions ? (avec 0 < K <= 3);
2 - Même question mais avec l'insertion de 80 éléments

On considère qu'il y a collision, si, avant d'ajouter un nouveau mot au tableau, toutes les cases de chacun de ses hashs sont déjà marquées.