

analyser resultat

djeser kordon et Wei Liu

November 2022

1 resultat

Resultat de base	hachage	casier
0.017271973756973612	0.017449254718420917	0.009591115598182737
0.43648444614279863	0.3794054762397407	0.29157248950597353
0.017147498852489958	0.017449254718420917	0.00943824761520214
0.017154128790881934	0.017552785550750447	0.009591115598182737
0.019341030939412775	0.019212149008816185	0.010866818296689411
0.9377423900044074	0.954461057363432	0.9357336430507163
0.9796380555014308	0.9954105557218398	0.9976028765481423
0.019109115924521775	0.02108541379486394	0.011173466808230953
0.019152864086271124	0.021346134204882034	0.011122345803842264
0.9570809622489082	0.957138663274293	0.937046004842615

On voit sur l'image les résultats de la similarité de jaccard renvoyé par les 3 méthodes. Nous avons pris uniquement celle là car la similarité est trop biaisée car nos deux ensembles ont exactement la même taille. Les valeurs sont donc les mêmes pour les deux ensembles.

On voit que le hachage a des résultats beaucoup plus proches que l'algorithme de hachage avec le casier. Cependant, celui-ci est plus rapide que le précédent et peut être un bon trade off entre l'efficacité et le temps de calcul.

Cette expérience montre que dans ce cas, un sous-ensemble (ici 10000) peut permettre d'approcher les scores réels tout en économisant la quantité de données à enregistrer.