

M1 Informatique –UE RITAL

Projet de Recherche d'Informations

Noms, prénoms:

Jouve Vincent
Cadiou Antoine

Dans ce projet nous avons fait toute la partie obligatoire qui était demandée, du TME1 au TME3.

Nous avons également réalisé le Bonus du TME2 et TME3, dans lequel nous avons essayé d'optimiser les hyperparamètres b et $k1$ dans le modèle Okapi BM25. Pour cela nous avons utilisé une méthode Grid-Search en faisant varier le b de 0.5 à 1 par pas de 0.05 et en faisant également varier $k1$ de 1.2 à 2.0 par pas de 0.1 (car c'est l'intervalle recommandé sur Wikipédia). Nous avons analysé les paramètres optimaux au regard de plusieurs métrique de scores (MAP, Precision, Recall et MRR)

Nous avons ensuite évalué les scores avec une cross-validation (en divisant l'ensemble en k folds, puis en des ensembles de train et de test).

Nous obtenons des résultats très corrects :

Sur les fichiers « **cisi** », avec un modèle *Okapi*, en cross-validation sur 5 folds

Precision@1 \approx 0.19 (en moyenne, il y a 1 document sur 5 qui est directement pertinent/retourné en tête)

MRR \approx 0.26 (en moyenne, le 1^{er} document pertinent apparaît dans les 4 premiers documents retournés)

Sur les fichiers « **cacm** », avec un modèle *Okapi*($b=0.8$, $k1=1.2$):

Precision@1 \approx 0.32

MRR \approx 0.44

Les résultats sont plutôt bons sur cacm car la base contient 4204 articles et la longueur moyenne des pertinences pour une requête est de 12. Donc il y a en moyenne 12 articles pertinents sur 4204, et le modèle classe en 1^{ère} position un article pertinent dans 32% des cas.

Nous avons également eu l'occasion de tester le modèle de langue et le modèle vectoriel avec les différentes formules de poids.

Nous n'avons cependant pas réalisé le TME 4.