

INDEX:

-INTRODUCCIO (roger)

- CALCUL JACCARD (david)

- . Explicacio teoria
- . Com ho hem implementat nosaltres (explicacio practica)

- MINHASH

- . Explicacio teoria (com hem triat les funcions minhash (ja esta al pdf))
- . Com ho hem implementat nosaltres (explicacio practica)

-LSH

- . Explicacio teoria
- . Com ho hem implementat nosaltres (explicacio practica)

- EXPLICACIO GENERAL DE L'APLICACIO QUE HEM FET (QUE FA CADA CLASSE I FUNCIONS) (TAMBE COM EXECUTAR-LA, QUE FA CADA OPCIO)

- GENERADOR DE PROVES

- ANALISIS/EXPERIMENTACIO:

1. CALCUL JACCARD. (DAVID)

- . Exposar els resultats amb el nostre joc de proves.
- . Comparar amb un altre set de documents i mirar els resultats.
- . Diferencia en funcio de la k que utilitzes (fer una grafica) i grafic de com augmenta el temps en funcio de la k.

. Analisis de com canvia els results (precisio com el temps en diversos documents de mides diferents)

. Analisis de com canvia els results (precisio com el temps en diversos documents de mides diferents)

(Fer servir 3 documents diferents de 50 paraules, 100 i 500). Triem la k que millor funcionava per l'apartat anterior. Mostrem les 3 matrius de resultats i comentem els resultats.

Calculem el temps que triga calcular tota la matriu i fem una grafica comparant-lo i veient si la relacio es lineal o no en funcio de t.

2. MINHASH (formar la matriu per a que et mostri totes les comparacions possibles). (ARANDA)

- . Exposar els resultats
- . Comparativa en funcio de les families de Hashing que utilitzis.
- . Diferencia en funcio de la k que utilitzes (fer una grafica) i grafic de com augmenta el temps en funcio de la k.

. Analisis de com canvia els results (precisio com el temps en diversos documents de mides diferents)

(Fer servir 3 documents diferents de 50 paraules, 100 i 500). Triem la k

que millor funcionava per l'apartat anterior. Mostrem les 3 matrius de resultats i comentem els resultats.

Calculem el temps que triga calcular tota la matriu i fem una grafica comparant-lo i veient si la relacio es lineal o no en funcio de t.

3. LSH (ROGER)

- . Exposar els resultats (del joc de proves inicial (el de 50 paraules))

- . Analitzar els resultats obtinguts veient si els parells que hem obtingut son realment documents semblants o no. Per a veure-ho, s'haura de mirar a la taula.

Opcionalment: Fer una funcio que a partir dels pairs obitnguts, vagi a la matriu obtinguda a de min hash on tenim les similituds entre tots els parells de documents. Aleshores, que comprovi si els parells obtinguts tambe tenen similitud \geq threshold (es a dir si es un fals positiu o no). Aixi al final podem donar un percentatge amb la quantitat de encerts.

- . Un cop hakis obtingut un percentatge que indiqui si encerta mes o menys amb els parells obtinguts, provar amb altres funcions de hash i mirar com canvia aquest percentatge.

- RESULTATS I CONCLUSIONS

-BIBLIOGRAFIA