ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν Πανεπιστήμιον Αθηνών
ΙΔΡΥΘΕΝ ΤΟ 1837

Ανάπτυξη Λογισμικού για Αλγοριθμικά Προβλήματα

Χειμερινο εξαμηνο 2020

1η Προγραμματιστικη Εργασια

Αναζητηση και συσταδοποιηση διανυσματων στη C/C++

Αριθμός Μητρώου(ΑΜ):

1115201700217

1115201700203

Ονοματεπωνυμο:

Ορέστης Στεφανου

Λεωνίδας Εφραιμ

Ακαδημαικη Χρονια 2020-2021

ΠΕΡΙΕΧΟΜΕΝΑ

1

ΕΙΣΑΓΩΓΗ

Στα πλέσια της εργασία είχαμε να υλοποίήσουμε τον αλγόριθμο LSH για διανύσματα στον D-διάστατο χώρο, καθώς και τον αλογόριθμο τυχαίας προβολής στον υπερκύβο βάσης της μετρικής Μανχάταν L1. Στην συνέχεια έπρεπε να εκτελέσουμε κάποια querys στο dataset που μας δώθηκε έτσι ώστε να επαληθεύσουμε την σωστή λειτουργία των αλγορίθμων. Τέλος κληθήκαμε να υλοποιήσουμε τους αλγόριρθμους για την συσταδοποίση διανυσμάτων βάση της μετρικής Μανχάταν όπου η ανάθεση θα έπρεπε να γίνει με τον αλγόριθμο του Lloyd's ή με αντίστροφή ανάθεση μέσω Range Search με LSH. Η υλοιποίση της εργασίας έχει γίνει σε C++

2

ΜΕΤΑΓΛΩΤΤΙΣΗ-ΕΚΤΕΛΕΣΗ

Για τις ανάγκες τις εργασία δημιουργήσαμε 3ις main συναρτήσεις όπου οι δύο είναι υπεύθυνες για του αλγόρθιμους LSH και Hypercube, ενώ η τρίτη είναι υπεύθυνη για το Clustering
Η μεταλγώττιση γίνετε με τις παρακάτω εντολές

- **make lsh**
- **make cube**
- **make cluster**

Ενώ η εκτέλεση των προγραμμάτων γίνετε με τις εντόλες που μας δώθηκαν στην εκφώνηση της εργασίας, δηλαδή:

- **LSH**
  ./lsh –d <input file> –q <query file> –k <int> -L <int> -o <output file> -N<number of nearest> -R <radius>
- **HYPER CUBE**
  ./cube –d <input file> –q <query file> –k <int> -M <int> -probes <int> -o<output file> -N <number of nearest> -R <radius>
- **CLUSTERING**
  ./cluster –i <input file> –c <configuration file> -o <output file> -complete <optional> -m <method: Classic OR LSH or Hypercube>

3

ΥΛΟΠΟΙΗΣΗ

3.1 ΕΙΣΟΔΟΣ ΔΕΔΟΜΕΝΩΝ

Για την είσαγωγή των δεδομένων έχουμε δημιουργήσει μια συνάρτηση με το όνομα **ReadData** η οποία δέχετε σαν όρισμα το path με το αρχείο εικόνων και ένα vector όπου στην συνέχεια το γεμίζει με τις είκονες.
Η συνάρτηση αφού ανοίξει το αρχείο διαβάζει διαδοχικά 4 integers όπου αντιπροσοπεύουν αντοίστοιχα

- Το magic number
- Το ύψος της εικόνας της εικόνας
- Το πλάτος της εικόνας της εικόνας
- Τον αριθμό των εικόνων που υπάρχουν στο αρχείο

Αφού ξέρουμε τις διαστάσεις τον εικόνων τώρα μπορούμε να διαβάζουμε N*N chars και να τους αποθηκεύουμε σε μια γραμμη του vector διαδοχικά.
Για την υλοποίση της συνάρτησης ReadData χρειαστήκαμε να υλοποίσουμε ακόμα μια συνάρτηση με όνομα **NumReverse** η οποία πέρνει ένα interger και του αλλάζει το endian του με μερικά shifts γιατί ο αριθμός που υπάρχει στο αρχείο είναι ανάποδα οποτέ πρέπει να αντιστραφεί.

3.2 ΜΕΤΡΙΚΕΣ

Για τις μετρικές δημιουργίσαμε μια κλάση με το όνομα **Metrics** η οποία έχει μια συνάρτηση με το όνομα **get distance** η οποία δέχετε σαν όρισμα τις 2 είκονες που θέλουμε να βρούμε της απόσταση τους κάθως και ακόμα ενα όριμσα το οποίο είναι το όνομα της μετρική π.χ. L1 για την μετρική Μαχνάταν. Έδω μπορούν να υλοιποιηθούν και άλλές μετρικές αλλα στην εργασία μας ζητήθηκες μόνο η μετρική Μανχάταν.
Για την υλοιπόιση της Μαχάταν μετρικής πήραμε το αθροσμα της απόλυτη τιμή των σημέιων των δύο εικόνων

3.3 LSH

3.4 HYPER CUBE

3.5 CLUSTERING

3.5.1 Lloyds

3.5.2 LSH Range Search