



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ

Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών

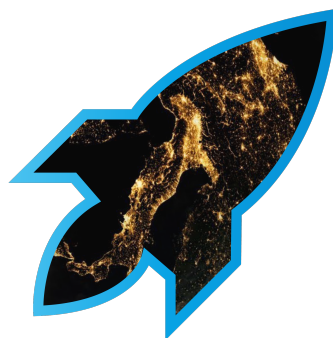
— ΙΔΡΥΘΕΝ ΤΟ 1837 —

ΑΝΑΠΤΥΞΗ ΛΟΓΙΣΜΙΚΟΥ ΓΙΑ ΑΛΓΟΡΙΘΜΙΚΑ ΠΡΟΒΛΗΜΑΤΑ

ΧΕΙΜΕΡΙΝΟ ΕΞΑΜΗΝΟ 2020

1η ΠΡΟΓΡΑΜΜΑΤΙΣΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΝΑΖΗΤΗΣΗ ΚΑΙ ΣΥΣΤΑΔΟΠΟΙΗΣΗ ΔΙΑΝΥΣΜΑΤΩΝ ΣΤΗ C/C++



Αριθμός Μητρώου(ΑΜ):

1115201700217

1115201700203

Ονοματεπώνυμο:

Ορέστης ΣΤΕΦΑΝΟΥ

Λεωνίδας ΕΦΡΑΙΜ

ΑΚΑΔΗΜΑΪΚΗ ΧΡΟΝΙΑ 2020-2021

ΠΕΡΙΕΧΟΜΕΝΑ

1	ΕΙΣΑΓΩΓΗ	3
2	ΜΕΤΑΓΛΩΤΤΙΣΗ-ΕΚΤΕΛΕΣΗ	4
3	ΥΛΟΠΟΙΗΣΗ	5
3.1	ΕΙΣΟΔΟΣ ΔΕΔΟΜΕΝΩΝ	5
3.2	LSH	5
3.3	HYPER CUBE	5
3.4	CLUSTERING	5
3.4.1	Lloyds	5
3.4.2	LSH Range Search	5

ΕΙΣΑΓΩΓΗ

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

ΜΕΤΑΓΛΩΤΤΙΣΗ-ΕΚΤΕΛΕΣΗ

Για τις ανάγκες της εργασίας δημιουργήσαμε 3ις main συναρτήσεις όπου οι δύο είναι υπεύθυνες για του αλγόριθμους LSH και Hypercube, ενώ η τρίτη είναι υπεύθυνη για το Clustering

Η μεταγλώττιση γίνεται με τις παρακάτω εντολές

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

Ενώ η εκτέλεση των προγραμμάτων γίνεται με τις εντολές 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>
```

ΥΛΟΠΟΙΗΣΗ

3.1 ΕΙΣΟΔΟΣ ΔΕΔΟΜΕΝΩΝ

3.2 LSH

3.3 HYPER CUBE

3.4 CLUSTERING

3.4.1 LLOYDS

3.4.2 LSH RANGE SEARCH