Ηλιάνα-Χάρις Σιούστη 1115201100107

Κουγιανός Νικόλαος 1115201100060

1/11/2016

Ανάπτυξη Λογισμικού για

Αλγοριθμικά Προβλήματα

1η Προγραμματιστική Εργασία

Υλοποίηση του Locality Sensitive Hashing (LSH) στη γλώσσα C

Αρχεία Κώδικα:

ATD.h, ATD.c: Σε αυτά τα 2 αρχεία περιέχονται η δομή komvos που χρησιμοποιείται στο πρόγραμμα και οι 2 συναρτήσεις καταστροφής και εισαγωγής στοιχείων στην συνδεδεμένη λίστα.

LSHfunctions.h, LSHfunctions.c: Περιέχουν τις συναρτήσεις για ψευδοτυχαία γέννηση παραμέτρων και συντελεστών των συναρτήσεων καθώς και τις συναρτήσεις κατακερματισμού.

LSH.h, LSH.c: Περιέχει την κύρια συνάρτηση του προγράμματος. Η LSH κάνει όλα τα απαραίτητα ανοίγματα και διαβάσματα αρχείων εισόδου, δημιουργεί πίνακες κατακερματισμού και εκτυπώνει τα αποτελέσματα στην οθόνη. Περιέχεται επίσης και η συνάρτηση ταξινόμησης sort.

Έχει υλοποιηθεί ο LSH αλγόριθμος, καθώς και ο DBH όσον αφορά στη μετρική Matrix. Υπολογίζονται οι αποστάσεις Euclidean, Cosine και Matrix, ενώ έχει υλοποιηθεί και η απόσταση Manhattan η οποία διδάχθηκε σε διάλεξη. Γίνεται διάβασμα των αρχείων εισόδου, τοποθετούνται τα αντικείμενα σε buckets βάσει της εκάστοτε συνάρτησης κατακερματισμού, προκύπτουν αποτελέσματα μετά από αναζήτηση από τα Query files και γίνεται εκτύπωση των αποτελεσμάτων σε output file με τον τρόπο που ζητάται στην εκφώνηση. Επιπλέον, υπολογίζεται ο ΝΝ Nearest Neighbour καθώς και όσοι βρίσκονται εντός ακτίνας R. Έχει οριστεί με 4 μέσω #define το W. Υπάρχουν έλεγχοι αποτυχίας δέσμευσης μνήμης και τερματίζεται το πρόγραμμα στην περίπτωση αποτυχίας. Σε αντίθετη περίπτωση, γίνεται αποδέσμευση της μνήμης πριν την έξοδο του προγράμματος.

Το πρόγραμμα είναι πλήρως παραμετρικό.

Οδηγίες μεταγλώττισης:

/\* Makefile \*/

OBJS = main.o ATD.o LSHfunctions.o LSH.o

SOURCE = main.c ATD.c LSHfunctions.c LSH.c

HEADER = ATD.h LSHfunctions.h LSH.h

OUT = lsh

CC = gcc

CFLAGS = -g -c

all: $(OBJS)

$(CC) -g $(OBJS) -o $(OUT) -lm

main.o: main.c

$(CC) $(CFLAGS) main.c

ATD.o: ATD.c

$(CC) $(CFLAGS) ATD.c

LSHfunctions.o: LSHfunctions.c

$(CC) $(CFLAGS) LSHfunctions.c

LSH.o: LSH.c

$(CC) $(CFLAGS) LSH.c

clean:

rm -f $(OBJS) $(OUT)

count:

wc $(SOURCE) $(HEADER)

Το πρόγραμμα μεταγλωττίζεται με την εντολή make –f Makefile.

Ενδεικτικά, το πρόγραμμα θα μπορούσε να εκτελεστεί με κάποιον από τους παρακάτω τρόπους:

–d DataHamming.csv –q QueryHamming.csv –k 4 -L 5 -o outputH.txt

–d DataEuclidean.csv –q QueryEuclidean.csv –k 4 -L 5 -o outputV.txt

–d DistanceMatrix.csv –q QueryDistanceMatrix.csv –k 4 -L 5 -o outputM.txt

–d DataHamming.csv –q QueryHamming.csv -o output.txt (default K=4 και L=5)