



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Εθνικόν και Καποδιστριακόν
Πανεπιστήμιον Αθηνών
— ΙΔΡΥΘΕΝ ΤΟ 1837 —

Ανάπτυξη Λογισμικού Για Αλγοριθμικά Προβλήματα

Εργασία 1

Εκπονήσαντες:

- Γιακούμπ Ματσεγιέβски AM 1115201700080
- Αναστάσιος Μελιδώνης AM 1115201700084

Περιεχόμενα Οδηγίες

Περιεχόμενα:

- Γενική περιγραφή προγράμματος σελ. 3
- Λογικές ενότητες προγράμματος 1 σελ. 4
- Λογικές ενότητες προγράμματος 2 σελ. 5

Οδηγίες:

Κατά την υλοποίηση της εργασίας ακολουθήσαμε πιστά τις δοθείσες προδιαγραφές τόσο σχετικά με την μορφή της εισόδου που θα γίνεται απ' τον χρήστη όσο και σε θέματα που σχετίζονται με την αλγοριθμική πτυχή του project.

Ο χρήστης αρκεί να εκτελέσει την εντολή `make`. Έτσι θα παραχθούν τρία εκτελέσιμα αρχεία με όνοματα `lsh`, `cube` και `cluster` για τις διαφορετικές εντολές που θα ζητηθούν. Στην συνέχεια μπορεί να κληθεί οποιοδήποτε απ' τα τρία προγράμματα.

Στην περίπτωση των προγραμμάτων `lsh` και `hypercube` μετά την εκτέλεση των δοθέντων απ' το αρχείο `queries` εντολών, στο `terminal` θα εμφανιστεί ερώτηση σχετικά με το αν ο χρήστης επιθυμεί να εκτελεστούν `queries` πάνω σε νέες φωτογραφίες. Αν πληκτρολογήσει ένα όνομα αρχείου, θα κληθεί να εισάγει το όνομα του αρχείου στο οποίο βρίσκονται οι εικόνες. Στην συνέχεια, το πρόγραμμα εκτελείται απ' την αρχή. Σε διαφορετική περίπτωση, δηλαδή αν εισαχθεί `'exit'`, ακολουθεί αποδέσμευση μνήμης και το πρόγραμμα τερματίζεται.

Ο παραδοτέος φάκελος περιέχει επιμέρους φακέλους, στους όποιους κατανέμονται τα περιεχόμενα του project, συγκεκριμένα υπάρχει διάκριση σε αρχεία επικεφαλίδας, πηγαία και αντικειμενικά.

Παράλληλα, παρέχουμε ένα αρχείο τύπου makeFile, το οποίο χρησιμοποιείται για την μεταγλώττιση του δοθέντα κώδικα. Η εκτέλεση του παράγει τρία εκτελέσιμα προγράμματα ονόματι lsh, cube και cluster, όπως και ζητούνται στην εκφώνηση της εργασίας.

Το lsh αρχικοποιεί μια ακολουθία από hash tables, τα όποια περιέχουν ειδικά διανύσματα διαταραχής, των όποιων οι συντεταγμένες ακολουθούν κανονική, ομοιόμορφη κατανομή βάση των δοθέντων τύπων. Στην συνέχεια, διαβάζονται εικόνες απ' το αρχείο DataSet και τοποθετούνται στα buckets των διαδοχικών hash tables. Στην συνέχεια, διαβάζονται εικόνες απ' το αρχείο QuerySet. Ανάλογα με τις παραμέτρους που δίνονται απ' τον χρήστη, είτε επιστρέφονται οι k κοντινότεροι γείτονες με αναζήτηση στα bucket, όπου κατανέμονται οι νέες εικόνες είτε οι εικόνες των ίδιων buckets, των όποιων η απόσταση απ' τα query images είναι μικρότερη απ' την δοθείσα ακτίνα.

Όμοια στο cube αρχικοποιείται ένα hash table, του όποιου τα buckets αποτελούν τις κορυφές του υπερκύβου. Χρησιμοποιούνται διανύσματα διαταραχής για τον ορισμό hashing συναρτήσεων, των όποιων η έξοδος προβάλλεται σε πέδιο 0 ή 1, παράγοντας έτσι τα διαδοχικά bits, που συνθέτουν τον αριθμό της κορυφής, όπου θα αποθηκευτούν οι εικόνες. Στην συνέχεια, ακολουθεί η προβολή των query images στις κορυφές του υπερκύβου και αναζήτηση των k πλησιέστερων / γειτόνων που βρίσκονται σε ακτίνα r , διατρέχοντας γειτονικές προβολές. Ενδέχεται ο αριθμός των εικόνων ή κορυφών που θα εξεταστούν να περιοριστεί απ' τον χρήστη.

Τέλος στο clustering αρχικοποιείται μια δομή clustering και ανάλογα απ' την επιλογή του χρήστη ενδέχεται να αρχικοποιηθεί και ένα data member του τύπου lsh ή hypercube, που θα χρησιμοποιηθούν για το reverse assignment με την χρήση range search. Στην συνέχεια, λαμβάνονται υπόψη όλες οι εικόνες για τον ορισμό των αρχικών κεντροειδών με την χρήση της τεχνικής k -median++.

Έτσι επιτυγχάνεται επιλογή κεντροειδών, τα όποια βρίσκονται μακριά μεταξύ τους, αυξάνοντας την πιθανότητα καλής αρχικής ομαδοποίησης και μείωσης του πλήθους των επαναλήψεων που απαιτούνται για την σύγκλιση του αλγορίθμου. Τέλος, ανάλογα με την δοθείσα τεχνική γίνεται ανανέωση των clusters είτε με την τεχνική Lloyd είτε με reverse assignment μέσω LSH ή HyperCube.

- **algoAux:** Περιέχει βοηθητικές συναρτήσεις για αλγοριθμικούς υπολογισμούς, που χρησιμοποιούνται σε πολλά διαφορετικά σημεία του project, όπως η manhattan μετρική απόστασης που χρησιμοποιείται στο υπολογισμό αποστάσεων μεταξύ των εικόνων
- **bucket:** Περιέχει βασικές συναρτήσεις εξαγωγής εικόνων απ' τα buckets των hash tables
- **cluster:** Ορίζει συναρτήσεις που βοηθούν στην διεπαφή του set που περιέχει τις εικόνες με την γενική κλάση clustering. Ορίζουν στην εισαγωγή εικόνων στα cluster και την εκένωση τους σε κάθε επανάληψη.
- **clusterEntry:** Περιέχει συναρτήσεις της βοηθητικής κλάσης μέλους συστάδας, η οποία περιέχει πληροφορία σχετικά με την τρέχουσα κατάσταση κάθε εικόνας στην προκείμενη φάση της συσταδοποίησης με την χρήση του reverse assignment. Έτσι γνωρίζουμε ποιες εικόνες βρέθηκαν στην τρέχουσα επανάληψη και το κοντινότερο κεντροειδές τους.
- **clustering:** Το πιο γενικό αρχείο για την λειτουργικότητα της συσταδοποίησης. Ορίζει κλάση, η οποία περιέχει αναφορές στα τρέχοντα κεντροειδή και συστάδες. Εκτελεί την βασική λειτουργία της διαδοχικής συσταδοποίησης, ενημερωμένη για την απαιτούμενη μέθοδο. Περιέχει συναρτήσεις υπολογισμού σιλουέτας.
- **hashTable:** Ορίζει την κλάση του πίνακα κατακερματισμού, ο οποίος περιέχει πίνακα με buckets, τα διανύσματα διαταραχής που ορίζουν τις τμηματικές συναρτήσεις κατακερματισμού h , καθώς και τις τιμές των υπολοίπων διαίρεσης των εκθετών του παράγοντα m , που συμμετέχουν στον υπολογισμό των πολυώνυμων των συναρτήσεων κατακερματισμού (για πλήρη κατανόηση ανατρέξτε στον εκτενή σχολιασμό των **mExponential συναρτήσεων**). Ορίζονται συναρτήσεις για την εισαγωγή εικόνων στην δομή LSH και HyperCube αντίστοιχα.
- **hyperCube:** Περιέχει συναρτήσεις που αρχικοποιούν τον πίνακα κατακερματισμού του υπερκύβου, καθώς και τις συναρτήσεις για την αναζήτηση πλησιέστερων k γειτόνων ή αυτών που βρίσκονται εντός της δοθείσας ακτίνας.
- **image:** Περιέχει τον ορισμό της κλάσης εικόνας, η οποία περιέχει την μαθηματική αναπαράσταση εικόνας υπό την μορφή διανύσματος pixels και το προσδιοριστικό της εικόνας στο data set.

- **imageDuplet:** Ορίζει την βοηθητική κλάση της δυάδας, η οποία περιέχει μια φωτογραφία και την απόσταση της από το query image. Έτσι επιτρέπεται η πιο ευανάγνωστη ανανέωση του σωρού, που χρησιμοποιείται στις συναρτήσεις εύρεσης k πλησιέστερων γειτόνων.
- **LSH:** Ορίζει την κλάση του LSH, αρχικοποιώντας την δοθείσα ποσότητα από πίνακες κατακερματισμού ακολουθώντας τις οδηγίες του χρήστη ως προς τις τιμές διαφόρων παραμέτρων. Περιέχονται και οι συναρτήσεις εύρεσης k πλησιέστερων γειτόνων / αυτών που βρίσκονται εντός της δοθείσας ακτίνας.
- **mainAux:** Περιέχει βοηθητικές συναρτήσεις που είναι κοινόχρηστες για τις διάφορες εκδοχές της κύριας συνάρτησης που εκτελεί τον κώδικα της εργασίας. Οι συναρτήσεις σχετίζονται κυρίως με την λήψη δεδομένων απ' τον χρήστη, την εξαγωγή πληροφορίας απο αρχεία και εκτύπωση τελικών αποτελεσμάτων.
- **mainHyperCube/mainLSH/mainCluster:** Ο κώδικας των τριών διαφορετικών εκδόσεων της κύριας συνάρτησης main για το LSH, hypercube και συσταδοποίησης αντίστοιχα. Παρουσιάζουν ελάχιστες διαφοροποιήσεις, υπάρχουν κυρίως για την καλύτερη λογική οργάνωση του συνολικού project.