

K23γ: Ανάπτυξη Λογισμικού για Αλγοριθμικά Προβλήματα
Χειμερινό εξάμηνο 2018-19
3 η Προγραμματιστική Εργασία στη γλώσσα C/C++

Απόστολος Φλωράκης-115201400217

Οργάνωση Προγράμματος: Έχω διαχωρίσει τη δημιουργία των dataset και την υλοποίηση των αλγορίθμων recommendation ως δυο ξεχωριστές διαδικασίες. Το data_preproc.cc είναι υπεύθυνο για τη δημιουργία των dataset. Μεταγλωττίζεται με τη βοήθεια του Makefile.data_preproc και εκτελείται ως ./data_preproc <config_file>. Το config file αρχείο που δίνεται ως όρισμα στη γραμμή εντολών περιέχει όλες τις απαραίτητες παραμέτρους (κυρίως αρχεία) που απαιτούνται για την δημιουργία των datasets. Παράδειγμα τέτοιου αρχείου έχει παραδοθεί μαζί με την εργασία. Μετά την εκτέλεση του προγράμματος δημιουργούνται δυο αρχεία που περιέχουν τα διανύσματα των χρηστών που δημιουργήθηκαν καθώς και τα νομίσματα για τα οποία κάθε χρήστης έχει η όχι άποψη. Επίσης η επιτυχημένη δημιουργία των αρχείων εξαρτάται από τις τιμές που έχουν κάποιες defined σταθερές του προγράμματος μας για τις οποίες πρέπει να γίνει σωστή παραμετροποίηση. Συγκεκριμένα, για τη δημιουργία εικονικών χρηστών θα πρέπει στο αρχείο tweet_preproc.cc η σταθερά ICONIC_MODE να έχει τιμή 1, και η CLUSTERING_FILE_SIZE ίση με το πλήθος γραμμών του αρχείου στο οποίο οποίο θα γίνει το clustering. Επίσης στο utils.h το DATA_VECTOR_SIZE πρέπει να είναι ίσο με 203 και τα DATA_NUMBER και TOTAL_DATA_NUMBER επίσης ίσα με το CLUSTERING_FILE_SIZE. Για τη δημιουργία κανονικών χρηστών αρκεί ICONIC_MODE = 0. Τα αρχεία που παράγονται δίνονται ως είσοδος στο recommendation.cc όπου είναι υπεύθυνο για την πραγματοποίηση των υποδείξεων. Η μεταγλώττιση του γίνεται μέσω του Makefile.data_recomm και η εκτέλεση του με την εντολή ./recommendation -d <input file> -f <flags file> -o <output file> -c <config file> -validate. Το flagsfile υποδεικνύει σε ποιά νομίσματα έχει αναφερθεί κάθε user, το configfile αφορά την παραμετροποίηση του προγράμματος και το -validate είναι προαιρετικό ανάλογα με το άμα θέλουμε να γίνει validation. Αναλογα με το πλήθος των δεδομένων πρέπει να παραμετροποιηθούν σωστά η σταθερές DATA_NUMBER και TOTAL_DATA_NUMBER. Τα αρχεία δίνονται ως μια ιεραρχία φακέλων που αφορούν τόσο το συγκεκριμένο κομμάτι της εφαρμογής όσο και τα τμήματα των προηγούμενων εργασιών που αυτή χρησιμοποιεί.

Αρχεία:

- data_preproc.cc: Υπεύθυνο για τη δημιουργία των datasets και την προεπεξεργασία των δεδομένων.
- recommendation.cc: Αφορά το διάβασμα των αρχείων και την εκτέλεση των αλγορίθμων.
- recomm_methods.h: Δηλώσεις που αφορούν τους αλγορίθμους.
- recomm_methods.cc: Υλοποιήσεις αλγορίθμων.

- [tweet_preproc.cc](#): Προεπεξεργασία δεδομένων.
- [tweet_preproc.h](#): Δηλώσεις για την προεπεξεργασία δεδομένων.