

ΑΝΑΚΤΗΣΗ ΠΛΗΡΟΦΟΡΙΑΣ

‘Μηχανή αναζήτησης κριτικών χρηστών’

1η Φάση: Ορισμός των ομάδων και περιγραφή των δεδομένων

Μέλη Ομάδας

Γκαβαρδίνας Όθωνας, AM 2620

Μπουρλή Στυλιανή, AM 2774

Επιλογή Εστιατορίων

Αρχικά, κατεβάσαμε τη Json μορφή του yelp dataset. Από τη συλλογή αυτή χρησιμοποιήσαμε το αρχείο business.json για την επιλογή των εστιατορίων. Πιο συγκεκριμένα, γράψαμε κώδικα σε Java με τον οποίο διαβάσαμε γραμμή γραμμή το αρχείο. Κάθε γραμμή αφορούσε μία ξεχωριστή επιχείρηση. Στο πεδίο categories της κάθε επιχείρησης ελέγξαμε αν περιλαμβάνεται η λέξη ‘Restaurants’, δηλαδή αν η επιχείρηση είναι εστιατόριο. Από αυτά αποθηκεύσαμε τα 14000, σε ένα νέο αρχείο newfile.json, διατηρώντας όμως μόνο τα πεδία που θεωρήσαμε πιο σημαντικά, δηλαδή τα business_id, name, address, city, state, latitude, longitude, stars, review_count και categories. Το κριτήριο επιλογής μας ήταν οι κριτικές να είναι περισσότερες από 50, ώστε να έχουμε αρκετές κριτικές για τα εστιατόρια.

Στατιστικά Στοιχεία

- | | |
|---|---------|
| ➤ Συνολικός αριθμός επιλεγμένων εστιατορίων: | 14000 |
| ➤ Συνολικός αριθμός επιλεγμένων κριτικών: | 2536053 |
| ➤ Μέσος αριθμός κριτικών για τα εστιατόρια: | 181 |
| ➤ Μικρότερος αριθμός κριτικών για τα εστιατόρια: | 51 |
| ➤ Μεγαλύτερος αριθμός κριτικών για τα εστιατόρια: | 7361 |

