Εργασία: Μηχανή αναζήτησης *κριτικών* χρηστών
(αρχική περιγραφή)

Καταληκτικές Ημερομηνίες
Πέμπτη 29 Μαρτίου 2018, Ορισμός ομάδων και περιγραφή δεδομένων
Πέμπτη 26 Απριλίου 2018, Περιγραφή σχεδιασμού
Τετάρτη 23 Μαΐου 2018, Παράδοση εργασίας
Πέμπτη 24 ή/και Παρασκευή 25 Μαΐου 2018, Εξέταση εργασίας

Η εργασία μπορεί να γίνει σε ομάδες έως 2 ατόμων.
Η εργασία μετράει σε ποσοστό 50% στο βαθμό σας στο μάθημα.

Η εργασία αφορά στο σχεδιασμό και υλοποίηση ενός συστήματος αναζήτησης πληροφορίας για εστιατόρια και κριτικές εστιατορίων.

Συγκεκριμένα τα δεδομένα σας θα είναι πληροφορίες για εστιατόρια και κριτικές τους από το σύστημα Yelp (https://www.yelp.com/).

Ως πρώτο βήμα δημιουργείστε τη συλλογή σας κατεβάζοντας δεδομένα από το Yelp Open Dataset (https://www.yelp.com/dataset).

Διαλέξτε ένα υποσύνολο των διαθέσιμων δεδομένων τα οποία να αφορούν εστιατόρια και κριτικές αυτών των εστιατορίων.

Ελάχιστες απαιτήσεις:
- 10000 εστιατόρια.
- 1000000 κριτικές αυτών των εστιατορίων

Λειτουργικότητα 1: Αναζήτηση εστιατορίων
Το σύστημα σας θα πρέπει να υποστηρίζει αναζήτηση εστιατορίων (δηλαδή, το αποτέλεσμα θα είναι ένα ή περισσότερα εστιατόρια) *τουλάχιστον* με βάση:
- Τη γεωγραφική θέση του εστιατορίου,
- Το *πλήρες κείμενο* των κριτικών του εστιατορίου (για παράδειγμα εστιατόρια των οποίων οι κριτικές περιλαμβάνουν τη λέξη «sesame»),
- Συνδυασμό των παραπάνω.

Ο βασικός τρόπος διάταξης θα πρέπει να είναι με βάση το κείμενο. Επιπρόσθετα, θα παρέχετε η δυνατότητα αναδιάταξης με βάση τουλάχιστον (1) τον αριθμό των κριτικών που έχει λάβει κάθε εστιατόριο και (2) τον αριθμό αστεριών του.

Λειτουργικότητα 2: Αναζήτηση κριτικής
Το σύστημα σας θα πρέπει να υποστηρίζει αναζήτηση κριτικών εστιατορίων (δηλαδή, το αποτέλεσμα θα είναι μια ή περισσότερες κριτικές εστιατορίων) *τουλάχιστον* με βάση
- Το όνομα του εστιατορίου,
- Λέξεις κλειδιά,
- Συνδυασμό των παραπάνω.

Ο βασικός τρόπος διάταξης θα πρέπει να είναι με βάση το κείμενο. Επιπρόσθετα, θα παρέχετε η δυνατότητα αναδιάταξης με βάση *τουλάχιστον* (1) το πόσο σημαντική είναι η κριτική (πχ, useful count), και (2) το χρόνο: η πιο πρόσφατη κριτική να εμφανίζεται πρώτη.

Λειτουργικότητα 3: Παρουσίαση αντιπροσωπευτικών αποτελεσμάτων

Επεκτείνετε την αναζήτηση εστιατορίων και κριτικών εστιατορίων ώστε να παρουσιάζεται ως απάντηση στο χρήστη ένα αντιπροσωπευτικό υποσύνολο των αποτελεσμάτων.

Για παράδειγμα στην περίπτωση των εστιατορίων, μερικές ιδέες είναι να παρουσιάζονται εστιατόρια σε: (1) σε διαφορετικές τοποθεσίες, (2) με διαφορετικές κουζίνες, (3) με διαφορετικό αριθμό αστεριών (4) οποιοδήποτε συνδυασμός των παραπάνω.

Για παράδειγμα στην περίπτωση των κριτικών εστιατορίων, μερικές ιδέες είναι (1) ένα μείγμα θετικών και αρνητικών κριτικών, (2) κριτικών που να αναφέρονται σε διαφορετικά θέματα, (3) πρόσφατων και λιγότερων πρόσφατων κριτικών, (4) οποιοδήποτε συνδυασμός των παραπάνω.

Μπορείτε επίσης να χρησιμοποιείστε και τα δεδομένα για τους χρήστες ώστε να παρέχετε κριτικές από «διαφορετικούς» χρήστες.

Για την υλοποίηση, θα χρησιμοποιήστε το σύστημα Lucene (https://lucene.apache.org/) μια βιβλιοθήκη ανοικτού κώδικα για την κατασκευή μηχανών αναζήτησης κειμένου.