# Εργασία: Μηχανή αναζήτησης tweets
## (αρχική περιγραφή)

**Καταληκτικές Ημερομηνίες Παράδοσης**
**Τετάρτη 29 Μαρτίου 2017,** Ορισμός Ομάδων και Περιγραφή Δεδομένων
**Πέμπτη 6 Απριλίου 2017**, Περιγραφή Αρχικού Σχεδιασμού
**Τετάρτη 24 Μαΐου 2017** Παράδοση Εργασίας
**Παρασκευή 26 Μαΐου 2017**, 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 ανάκτησης πληροφορίας για μια συλλογή εγγράφων. Τα έγγραφα θα είναι αποθηκευμένα στο δίσκο.

Ως μέρος της εργασίας θα συλλέξετε τα δεδομένα της συλλογής σας. Συγκεκριμένα τα δεδομένα σας θα είναι tweets.

Ελάχιστες απαιτήσεις:
- 10.000 tweets.
- Τα tweets να έχουν κάποια σχέση μεταξύ τους - πχ να είναι τα tweets ενός ή περισσότερων συγκεκριμένων χρηστών (πχ, ειδησεογραφικών πρακτορείων, ποδοσφαιρικών ομάδων, διασημοτήτων, κλπ) ή να αφορούν ένα ή περισσότερα συγκεκριμένα θέματα (;όπως αυτά προσδιορίζονται από κάποιο hashtag).

Το σύστημα σας θα πρέπει να υποστηρίζει αναζήτηση tweets με βάση
- λέξεις κλειδιά
- hashtags
- όνομα του συγγραφέα του tweet
- τοποθεσία του συγγραφέα του tweet
- συνδυασμό των παραπάνω

Ως απάντηση ο χρήστης θα πρέπει να βλέπει τα αντίστοιχα tweets με τονισμένους τους όρους της ερώτησης.

Ο βασικός τρόπος διάταξης θα πρέπει να είναι με βάση το κείμενο.

Πέρα από το βασικό τρόπο διάταξης, το σύστημα σας θα πρέπει να δίνει τη δυνατότητα αναδιάταξης των αποτελεσμάτων με βάση το χρόνο (τα πιο πρόσφατα tweets πρώτα).

Επίσης, το σύστημα σας θα πρέπει να διατηρεί πληροφορία για την ιστορία των αναζητήσεων (π.χ., click-through-rate, δημοφιλείς ερωτήσεις, κλπ). Χρησιμοποιείστε αυτήν την πληροφορία για να:
- αναδιατάξετε τα αποτελέσματα της αναζήτησης, και
- προτείνετε εναλλακτικά ερωτήματα

Επιπρόσθετη λειτουργικότητα όπως διόρθωση ορθογραφικών λαθών, stemming, υποστήριξη συνωνύμων, φράσεων κλπ, καθώς και λειτουργική διεπαφή χρήστη (ευχρηστία, διαδραστικότητα) θα μετρήσουν θετικά.

Για την υλοποίηση, θα χρησιμοποιήστε το σύστημα Lucene[1] μια βιβλιοθήκη ανοικτού κώδικα για την κατασκευή μηχανών αναζήτησης κειμένου.

[1] https://lucene.apache.org/