Πανεπιστήμιο Ιωαννίνων
Τμήμα Μηχανικών Η/Υ & Πληροφορικής

Εργασία: Μηχανή αναζήτησης ταινιών

Καταληκτικές Ημερομηνίες

Παρασκευή 15 Απριλίου 2022	Σύντομη περιγραφή του σχεδιασμού και της συλλογής δεδομένων
Παρασκευή 27 Μαΐου 2022	Παράδοση εργασίας
Τελευταία εβδομάδα του Μαΐου	Προφορική εξέταση (οι ακριβείς ημέρες και ώρες θα ανακοινωθούν αργότερα)

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 ταινιών ή/και κριτικών για ταινίες. Για την υλοποίηση, θα χρησιμοποιήστε τη βιβλιοθήκη **Lucene** https://lucene.apache.org/, μια βιβλιοθήκη ανοικτού κώδικα για την κατασκευή μηχανών αναζήτησης κειμένου.

Συλλογή εγγράφων (corpus). Αρχικά, πρέπει να συλλέξετε τα έγγραφα που θα αποτελούν τη συλλογή σας. Το έγγραφα σας θα είναι έγγραφα σχετικά με ταινίες ή σειρές.

Μπορείτε να κατασκευάσετε τη συλλογή από τα άρθρα με όποιο τρόπο θέλετε, όπως να χρησιμοποιείστε έτοιμες συλλογές εγγράφων, ή να κατεβάσετε ιστοσελίδες (π.χ., με scrapping), ή να συλλέξετε δημοσιεύσεις από κοινωνικά δίκτυα. Τα έγγραφα θα πρέπει απαραίτητα να περιέχουν κείμενο.

Η συλλογή πρέπει να περιλαμβάνει *τουλάχιστον* 500 έγγραφα στην περίπτωση άρθρων και 2.000 έγγραφα στην περίπτωση σύντομων κειμένων από κοινωνικά δίκτυα.

Ανάλυση κειμένου και κατασκευή ευρετηρίου. Η Lucene παρέχει τη δυνατότητα για stemming, απαλοιφή stop words, επέκταση συνωνύμων, κλπ.

Επίσης, κάποιες λειτουργίες, όπως η διόρθωση τυπογραφικών λαθών, ή η επέκταση ακρωνύμων, μπορούν να γίνουν εναλλακτικά κατά τη διάρκεια της αναζήτησης (τροποποιώντας το ερώτημα).

Επιλέξτε το είδος της ανάλυσης που θεωρείτε κατάλληλο και εξηγείστε την επιλογή σας.

Αναζήτηση. Το σύστημα σας θα πρέπει να υποστηρίζει αναζήτηση εγγράφων με λέξεις κλειδιά.

Επιπρόσθετα, θα πρέπει
(1) Να υποστηρίζει και άλλα είδη ερωτήσεων, για παράδειγμα αναζήτηση πεδίου, δηλαδή, την εμφάνιση όρων σε

συγκεκριμένα πεδία (πχ. στον τίτλο).
(2) Να διατηρεί πληροφορία για την ιστορία των αναζητήσεων. Χρησιμοποιείστε αυτήν την πληροφορία για να προτείνετε εναλλακτικά ερωτήματα.

Παρουσίαση Αποτελεσμάτων. Το σύστημα σας θα πρέπει να παρουσιάζει τα αποτελέσματα σε διάταξη με βάση τη συνάφεια τους με το ερώτημα.

Επιπρόσθετα, θα πρέπει
(1) Να παρουσιάζει τα αποτελέσματα ανά 10, με δυνατότητα στο χρήστη να προχωρήσει στα επόμενα.
(2) Οι λέξεις κλειδιά να παρουσιάζονται τονισμένες στο αποτέλεσμα.
(3) Να παρέχει δυνατότητα αναδιάταξης των αποτελεσμάτων με κάποιο άλλο κριτήριο όπως η ημερομηνία, η βαθμολογία, κλπ.