

Εργασία: Μηχανή αναζήτησης ειδησεογραφικών άρθρων

Φάση 1: Περιγραφή δεδομένων και αρχικός σχεδιασμός

Καταληκτική Ημερομηνία Παράδοσης: Δευτέρα 14 Απριλίου 2025, 11μμ

Στόχοι αυτής της φάσης είναι: (1) η εξοικείωση με τη συλλογή δεδομένων, (2) η κατανόηση των βασικών βημάτων της εργασίας και (3) ένας αρχικός σχεδιασμός της.

Γενικά για την εργασία θα παραδώστε ένα **GitHub repository**.

Δημιουργείτε το repository ως PRIVATE και δώστε μου προσπέλαση.

Στα Settings του repository επιλέξτε Manage Access και κάντε με collaborator.

GitHub username: pitoura

Μπορείτε να έχετε το repository public μετά το τέλος τους μαθήματος.

Περιεχόμενο του GitHub repository για την Φάση 1:

Readme με σύντομη περιγραφή (500-1000 λέξεις) με πληροφορίες για τα παρακάτω

(α) Ποια θα είναι η συλλογή των εγγράφων που θα χρησιμοποιείτε στην εργασία: Περιγράψτε τα πεδία του εγγράφου που θα χρησιμοποιείτε.

(β) Σύντομη περιγραφή του σχεδιασμού του συστήματος σας (ο σχεδιασμός μπορεί να αλλάξει κατά την υλοποίηση). Ενδεικτική δομή για αυτή την περιγραφή:

Εισαγωγή: Ποιος είναι ο στόχος και η λειτουργικότητας του συστήματος.

Ανάλυση κειμένου και κατασκευή ευρετηρίου: Ενδεικτικά, ποια θα είναι η προεπεξεργασία των άρθρων για τη δημιουργία του εγγράφου, ποια είναι η μονάδα εγγράφου και τα αντίστοιχα πεδία (fields), τα ευρετήρια που σκοπεύετε να δημιουργήσετε (σε ποια πεδία, και τι είδους), ώστε να υποστηρίζονται οι διάφοροι τρόποι αναζήτησης, κλπ

Αναζήτηση: Πως θα γίνετε η αναζήτηση, τα είδη των ερωτημάτων

Παρουσίαση Αποτελεσμάτων: Πως θα παρουσιάζονται τα αποτελέσματα.

Στις απαντήσεις στα παραπάνω όπου είναι δυνατόν αναφερθείτε και στα αντίστοιχα API της Lucene

Τρόπος παράδοσης: Υποβολή ενός αρχείο pdf μέσω του ecourse του μαθήματος. Το αρχείο θα περιέχει μόνο

1. Τα ονόματα και ΑΜ των μελών της ομάδας.
2. Το link στο GitHub repository της εργασίας σας.
3. Το Readme που θα υπάρχει και στο GitHub repository