

Τμήμα Μηχανικών Ηλεκτρονικών Υπολογιστών και Πληροφορικής

Ανάκτηση Πληροφορίας

Μέλη ομάδας:

Γεώργιος Δεληγιώργης

Ελευθέριος-Μάριος Μανίκας

ΑΚΑΔΗΜΑΪΚΟ ΕΤΟΣ 2024

Πέμπτη 23 Μαΐου

Περιεχόμενα

Σύντομη περιγραφή εργασίας	3
Οδηγίες εκτέλεσης προγράμματος.....	4

Σύντομη περιγραφή εργασίας

Η εφαρμογή που δημιουργήσαμε υλοποιεί ένα σύστημα εύρεσης επιστημονικών άρθρων τα οποία πήραμε από την ιστοσελίδα:

<https://www.kaggle.com/datasets/rowhitsuami/nips-papers-1987-2019-updated/data?select=papers.csv>. Επιλέξαμε από ένα σύνολο 9680 άρθρων 488 άρθρα για να τα χρησιμοποιήσουμε ως την βάση δεδομένων μας. Το format του αρχείου είναι .csv(comma separated values).

- Στην πρώτη στήλη έχουμε την χρονιά που δημοσιολογήθηκε το κάθε άρθρο.
- Στην δεύτερη στήλη τον τίτλο του.
- Στην τρίτη στήλη το abstract που είναι μια μικρή περίληψη του άρθρου και μπορεί να παραλείπεται, δεν έχουμε επιλέξει άρθρα που έχουν abstract.
- Στην τέταρτη στήλη είναι όλο το υπόλοιπο κείμενο του άρθρου.

Αρχικά μετά την εισαγωγή του ερωτήματος και την επιλογή του checkbox ο χρήστης πατάει το κουμπί search και έτσι εμφανίζονται στην result area ταξινομημένα κατά φθίνουσα σειρά συνάφειας ανά 10 οι απαντήσεις στην ερώτηση. Δίνεται η δυνατότητα στο χρήστη να πατήσει το κουμπί more με το οποίο θα δει τις παρακάτω σελίδες με τις καταχωρίσεις. Οι λέξεις που ταιριάζουν στην αναζήτηση του χρήστη εμφανίζονται στα αποτελέσματα πρασινισμένες. Αριστερά είναι το πεδίο του search history το οποίο ανανεώνεται αυτόματα μετά από κάθε αναζήτηση. Μέσω των μεθόδων που έχουμε αποθηκεύεται σε ένα αρχείο κάθε αναζήτηση η οποία θα αξιοποιηθεί για να προτείνει σε άλλον χρήστη μελλοντικά ερωτήματα τα οποία εμφανίζονται στην κονσόλα όταν τρέχουμε την εφαρμογή και όχι σε κάποιο jpanel. Τέλος δίνεται η δυνατότητα μέσω στον χρήστη μέσω του κουμπιού Sort by Year να ταξινομήσει τα αποτελέσματα που βλέπει στο παράθυρο με φθίνουσα χρονολογική σειρά.

Οδηγίες εκτέλεσης προγράμματος

Οι βιβλιοθήκες που χρειάζεται να κάνει κάποιος Import για να τρέξει το πρόγραμμα είναι όλα τα jars που βρίσκονται εντός των φακέλων "modules","modules-thirdparty","modules-test-framework" που αφορούν την έκδοση της βιβλιοθήκης Lucene 9.10.0 για Java. Επίσης απαραίτητο βήμα είναι η δημιουργία ενός directory στο οποίο θα αποθηκεύεται το ευρετήριο της Lucene και πρέπει να αντικατασταθεί το path `private static final String INDEX_DIRECTORY` με το αντίστοιχο path και στις 3 κλάσεις όπου χρειάζεται. Το επόμενο βήμα είναι να αντικαταστήσετε τα path που οδηγούν στο csv αρχείο, πιο συγκεκριμένα στο πεδίο `private static final String CSV_FILE_PATH.`

Όταν έχουν ολοκληρωθεί αυτά τα βήματα θα είμαστε σε θέση να τρέξουμε την κλάση LuceneIndex.java μόνο μια φορά έτσι ώστε να δημιουργηθεί το ευρετήριο της Lucene που θα περιέχει όλη την πληροφορία από το csv αρχείο μας.

Τέλος τρέχουμε την κλάση LucenGui.java ως 1 java application και εμφανίζεται το παράθυρο αναζήτησης που είναι υπεύθυνο για τις ερωτήσεις και απαντήσεις που θα θέσει ο χρήστης.

ΣΗΜΕΙΩΣΗ:

Σε περίπτωση που ο χρήστης διαγράψει τον φάκελο του ευρετηρίου(indexluc) θα πρέπει να επαναλάβει την διαδικασία από την αρχή διότι η κλάση LucenGui.java δεν θα μπορεί να τρέξει χωρίς της ύπαρξη αυτού του directory με τα απαραίτητα αρχεία.