

Song_Searcher

To link στο GitHub repository https://github.com/GiannisTsochlas/Song_Searcher

To Readme :

Το Project έχει ως κεντρικό θέμα το σχεδιασμό και την υλοποίηση ενός συστήματος αναζήτησης πληροφορίας σχετική με τραγούδια .Τα δύο κυρία σκέλη του είναι 1ον η δημιουργία μιας συλλογής (corpus) από σχετικά αρχεία csv και η υλοποίηση μιας μηχανής αναζήτησης συγκεκριμένων πληροφοριών από την συλλογή .Συγκεκριμένα υπάρχει διεπαφή του χρήστη με αυτή τη μηχανή στην οποία ο χρήστης θέτει ερωτήματα, το σύστημα με τη σειρά του επιστρέφει τα συναφή με το ερώτημα άρθρα της συλλογής σε διάταξη με βάση τη συνάφεια τους με το ερώτημα .Η μηχανή αναζήτησης βασίζεται πάνω στη βιβλιοθήκη Lucene .

Στόχος του project είναι να αναζητάς τα δεδομένα που επιθυμείς μέσα από μία μεγάλη συλλογή δεδομένων. τέτοιου είδους projects Συναντάμε πολύ συχνά και διευκολύνουν κατά πολύ την καθημερινότητά μας Βλέπε Google ,Bing, YouTube. Τα πεδία στα οποία η μηχανή θα κάνει αναζήτηση είναι το όνομα του καλλιτέχνη ο τίτλος,το άλμπουμ,ημερομηνία,οι στίχοι και η χρονολογία του τραγουδιού.

Η συλλογή των εγγράφων που θα χρησιμοποιήσουμε στο project βρίσκεται στον φάκελο music data .Ο οποίος περιέχει αρχεία csv και json , στη δικιά μας υλοποίηση θα εκμεταλλευτούμε τα αρχεία csv .Η δόμηση των αρχείων είναι η εξής Artist,Title,Album,Date,Lyric,Year (π.χ Ariana Grande,,"thank u, next","thank u, next",2018-11-03,"οι στίχοι του τραγουδιού ",2018).Οπότε ανακεφαλαιώνοντας έχουμε αρχεία csv με το όνομα του κάθε καλλιτέχνη και μέσα σε αυτά υπάρχουν πληροφορίες για τον τίτλο το άλμπουμ την ημερομηνία τους στίχους και χρονολογία του καθε τραγουδιού.Η πηγή των αρχείων βρίσκεται στον παρακάτω σύνδεσμο

<https://www.kaggle.com/datasets/deepshah16/song-lyrics-dataset>.

Η συλλογή ενημερώθηκε τελευταία φορά πριν δύο χρόνια και ο δημιουργός της είναι ο Deep Shah .

Η αναζήτηση γίνεται με λέξεις-κλειδιά όπως σε γνωστές μηχανές αναζήτησης π.χ Google ,ένα ενδεικτικό σενάριο είναι εισάγοντας τη λέξη "Emin" το σύστημα Θα πρέπει να μας επιστρέψει συναφή άρθρα τύπου 1. Eminem rap god 2. Eminem Lose Yourself και τα λοιπά.Επίσης υποστηρίζονται και άλλα είδη ερωτήσεων για παράδειγμα αναζήτηση όρων σε συγκεκριμένα πεδία καθώς και διατήρηση πληροφορίας για την ιστορία των αναζητήσεων.

Το σύστημα θα παρουσιάζει τα αποτελέσματα σε διάταξη με βάση τη συνάφεια τους με το ερώτημα .Επιπρόσθετα θα παρουσιάζει τα αποτελέσματα ανά 10 με δυνατότητα στο χρήστη να προχωρήσει στα επόμενα ,οι λέξεις κλειδιά παρουσιάζονται τονισμένες στο αποτέλεσμα

Ιωάννης Τσόχλας AM 4993

Κωνσταντίνος-Διονυσίος Λαμανιακού AM 5110

και παρέχεται δυνατότητα ομαδοποίησης των αποτελεσμάτων με κάποιο κριτήριο που θα ορίσει ο χρήστης.

κατά την διάρκεια της υλοποίησης του project θα ενημερώσουμε το αρχείο με πιο τεχνικές λεπτομέρειες της μηχανής αναζήτησης και το API της Lucene .