

Προαιρετική Εργασία

Επέκταση της μηχανή αναζήτησης για τραγούδια για υποστήριξη σημασιολογικής αναζήτησης

Καταληκτικές Ημερομηνίες

Παρασκευή 19 Μαΐου	Παράδοση εργασίας
Εβδομάδα 22 Μαΐου	Προφορική εξέταση (οι ακριβείς ημέρες και ώρες θα ανακοινωθούν αργότερα)

Η προαιρετική εργασία μπορεί να γίνει σε ομάδες έως 2 ατόμων.

Η εργασία μετράει σε ποσοστό 20% στο βαθμό σας στο μάθημα, συγκεκριμένα σε αυτήν την περίπτωση ο τελικός σας βαθμός, θα είναι 50% εργασία, 20% προαιρετική εργασία, και 30% τελική εξέταση.

Επίσης, δίνει +0.5 στο τελικό βαθμό.

Στόχος της εργασίας είναι η επέκταση της μηχανικής αναζήτησης σας ώστε να υποστηρίζει σημασιολογική αναζήτηση.

Συγκεκριμένα, χρησιμοποιείστε κάποιο μοντέλο παραγωγής διανυσματικών αναπαραστάσεων (embedding) όπως word2vec, BERT, κλπ ή υπάρχουσες διανυσματικές αναπαραστάσεις από pretrained σύνολα. Οι διανυσματικές αυτές αναπαραστάσεις πρέπει να αποθηκευτούν σε κάποια δομή, ιδανικά σε κάποιο ευρετήριο της Lucene. Στη συνέχεια, η ανάκτηση εγγράφων θα γίνει εξετάζοντας την ομοιότητα μεταξύ των διανυσματικών αυτών αναπαραστάσεων.

Η εργασία δίνει ελευθερία σχετικά με το πώς θα αποθηκεύσετε και πως θα χρησιμοποιείτε τις διανυσματικές αναπαραστάσεις στην αναζήτηση.