

Εργασία: Μηχανή αναζήτησης άρθρων σχετικών με τον COVID-19

Οδηγίες παράδοσης εργασίας

Καταληκτική Ημερομηνία Παράδοσης
Παρασκευή 28 Μαΐου 2020, 5μμ

Τι θα πρέπει να παραδώσετε:

- (1) αναφορά,
- (2) τον source κώδικα
- (3) (προαιρετικά) σύντομο βίντεο με demo της εργασίας (από 2 έως 5 λεπτά). + 0.5 στο βαθμό της εργασίας.

Τρόπος παράδοσης: Τα παραπάνω θα είναι στη σελίδα του github. Υποβάλετε στο ecourse του μαθήματος μόνο ένα αρχείο με τα ονόματα και τους ΑΜ της ομάδας και το link στο github.

Περιεχόμενο αναφοράς: Η αναφορά θα πρέπει να είναι ~2000 λέξεις.

Περιγράψτε συνοπτικά το σύστημα αναζήτησης άρθρων σχετικών με το covid-19 που αναπτύξατε.

Συμπεριλάβετε πληροφορίες για τα παρακάτω:

Το σχεδιασμό και την υλοποίηση των τμημάτων της μηχανής αναζήτησης, δηλαδή (α) της συλλογής των άρθρων, (β) της προ-επεξεργασίας τους, (β) των ευρετηρίων, (γ) της επεξεργασίας ερωτημάτων.

Αναφερθείτε ανάμεσα σε άλλα και στα παρακάτω:

1. Στη συλλογή των άρθρων: ποια και πόσα άρθρα συλλέξατε, με ποιο κριτήριο, με ποιόν τρόπο, το είδος της προ-επεξεργασίας (πχ ποια πεδία κρατήσατε κλπ).
2. Στην ευρετηριοποίηση: ποια είναι η μονάδα εγγράφου και τα αντίστοιχα πεδία (fields), τι ευρετήρια δημιουργήσατε (σε ποια πεδία, και τι είδους, κλπ)
3. Στην αναζήτηση άρθρων: περιγράψτε το είδος των ερωτήσεων, καθώς και ποια είδη ερωτήσεων επιπρόσθετα της αναζήτησης με λέξεις κλειδιά υποστηρίζει το σύστημα σας, τι είδους πληροφορία για την ιστορία αναζητήσεων διατηρείτε και πως την χρησιμοποιείτε για να προτείνετε εναλλακτικά ερωτήματα, πως χρησιμοποιείτε τα embeddings στην αναζήτηση.
3. Στην παρουσίαση των αποτελεσμάτων : πχ ανά 10, τονισμένοι οι όροι της αναζήτησης, σύνδεσμος στο αρχικό άρθρο, κλπ, καθώς και την ομαδοποίηση των αποτελεσμάτων και με ποια κριτήρια γίνεται.