

Εργασία: Μηχανή αναζήτησης Wikipedia άρθρων

Οδηγίες παράδοσης εργασίας

Καταληκτική Ημερομηνία Παράδοσης
Παρασκευή 29 Μαΐου 2020, 5μμ (bonus +1.5)
Τετάρτη 3 Ιουνίου 2020, 5μμ

Τρόπος παράδοσης: Η εργασία θα υποβληθεί στο course του μαθήματος. Μια υποβολή ανά ομάδα.

Παραδώστε τον **κώδικα** (link στον κώδικα σε dropbox ή google drive) και μια **αναφορά**.

Το link πρέπει να εμφανίζεται στην αρχή της αναφοράς.

Η αναφορά θα πρέπει να είναι ~2000 λέξεις.

Περιεχόμενο Αναφοράς

Περιγράψτε συνοπτικά το σύστημα αναζήτησης άρθρων της Wikipedia που αναπτύξατε. Η αναφορά σας μπορεί να βασιστεί στην προηγούμενη αναφορά για το σχεδιασμό.

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 της Wikipedia 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 άρθρο της Wikipedia, κλπ, καθώς και αν υποστηρίζετε αναδιάταξη των αποτελεσμάτων και αν ναι με βάση ποιο ή ποια κριτήρια..