

Εργασία Συστήματα Ανάκτησης Πληροφορίας

Σπυρόπουλος Οδυσσέας – 3200183

Βλάχος Γεώργιος – 3200021

2^η Φάση :

Αρχικά στο αρχείο της εργασίας μας, έχουμε έναν φάκελο που ονομάζεται **IR2024** και περιέχει τα αρχεία **documents.txt** και **queries.txt**, τον οποίον δεν συμπεριλάβαμε στο τελικό παραδοτέο για διευκόλυνση στο μέγεθος του zip αρχείου μας.

Για την υλοποίηση της δεύτερης φάσης της εργασίας μας, έχουμε προσθέσει πάνω στο υλικό του προηγούμενου παραδοτέου έναν φάκελο με το όνομα **data_analysis** που περιέχει το αρχείο **data_analysis.ipynb**, στο οποίο εκτελούνται αυτοματοποιημένα τα πειράματα με το εργαλείο **trec_eval** και τα αποτελέσματα των οποίων οπτικοποιούνται με πίνακες και γραφήματα. Οι αλλαγές στους υπάρχοντες φακέλους και αρχεία αναγράφονται παρακάτω :

- **Phase2** : Το αρχείο Phase1 μετονομάστηκε σε Phase2, στην `main` μέθοδος του οποίου αρχικά ορίζουμε τις τιμές των παραμέτρων του πιθανοτικού και γλωσσικού μοντέλου που θα χρησιμοποιήσουμε σε πίνακες. Στην συνέχεια καλούμε για κάθε παράμετρο του γλωσσικού μοντέλου `LMJelinekMercerSimilarity` και για κάθε συνδυασμό των παραμέτρων του πιθανοτικού μοντέλου `BM25Similarity` την νέα συνάρτηση `runExperiment`, η λειτουργία και οι πληροφορίες της οποίας γράφονται παρακάτω.
- **runExperiment (Similarity similarity, String experimentName, String resultsDir) throws IOException, ParseException** : Η συνάρτηση αυτή δέχεται ως παραμέτρους την ομοιότητα του μοντέλου που θέλουμε να τρέξουμε, το όνομα του πειράματος που θέλουμε, το οποίο

έχει δημιουργηθεί στην επαναληπτική **for** που καλεί την συγκεκριμένη συνάρτηση και είναι της μορφής: `"BM25_k1_" + k1 + "_b_" + b`, όπου τα `k1`, `b` διαφέρουν για κάθε πείραμα. Με ανάλογο τρόπο και για το γλωσσικό μοντέλο. Η παράμετρος **results_dir**, έχει να κάνει με την τοποθεσία του φακέλου `txt` που θα γράφουμε τα αποτελέσματα. Η λειτουργία της μεθόδου αυτής είναι σαν της `main` του `Phase1` αρχείου, απλά κάθε φορά με χρήση της ανάλογης ομοιότητας και των αντίστοιχων παραμέτρων.

Αφού καταγραφούν τα αποτελέσματα στον φάκελο **experiments** προχωράμε στο αρχείο **data_analysis.ipynb**. Σε αυτό το αρχείο θα βρείτε την ανάλυση των αποτελεσμάτων και τις επεξηγήσεις του κώδικά μας.