**Σχολιασμός αποτελεσμάτων των προηγούμενων ερωτημάτων:**

Όσον αφορά τα πρώτα δυο ερωτήματα η διαφορά τους είναι στο κομμάτι της αναζήτησης. Στο πρώτο οι ερωτήσεις γίνονται μόνο στο κομμάτι του τίτλου σε κάθε ένα από τα 1400 έγγραφα ενώ στο δεύτερο η αναζήτηση για το αν η ερώτηση αφορά κάποιο από τα αρχεία κειμένου γίνεται και στα υπόλοιπα κομμάτια (συγγραφέας, περίληψη ). Έτσι και τα αποτελέσματα που πήραμε ήταν διαφορετικά. Ο αριθμός των σχετικών αρχείων για κάθε ερώτηση είναι συγκεκριμένος και δεν αλλάζει από ερώτημα σε ερώτημα. Το διαφορετικό είναι ο αριθμός των αρχείων που επιστράφηκαν από το πρόγραμμα μας διότι είναι και διαφορετικός ο χώρος αναζήτησης. Ψάχνοντας σε όλα τα κομμάτια του κάθε κειμένου αρχικά το recall θα είναι μεγαλύτερο από το precision διότι είναι πολύ λιγότερα τα σχετικά έγγραφα για κάθε ερώτηση από αυτά που ανακτήθηκαν γι’ αυτό σε αυτά που ανακτήθηκαν θα είναι και τα περισσότερα σχετικά. Κάνοντας αναζήτηση μόνο με βάση τον τίτλο του κάθε κειμένου αυτό που πετυχαίνουμε είναι να μειώσουμε τα έγγραφα που ανακτώνται κατά μεγάλο βαθμό γι’ αυτό βλέπουμε σε όλες τις περιπτώσεις στα αποτελέσματα το recall να είναι μικρότερο σε σχέση με το recall της άλλης αναζήτησης (μείωση των ανακτημένων άρα και μείωση του αριθμητή του κλάσματος) αλλά το precision μεγαλύτερο (μείωση των ανακτημένων, μείωση του αριθμητή αλλά και ταυτόχρονη μείωση κατά πολύ και του παρανομαστή). Βέβαια αυτά ισχύουν για την πλειοψηφία των ερωτήσεων αλλά όχι για όλες.

Τα τελευταία δυο ερωτήματα μεταξύ τους έχουν πολύ μικρές αποκλίσεις διότι ο standard analyzer που χρησιμοποιήθηκε στο ένα και το analyzer που φτιάχτηκε από μένα ναι μεν έχουν διαφορές αλλά τα stopwords που παίζουν ρόλο στα αποτελέσματα υπάρχουν και στους δυο analyzer.

Όσον αφορά τα αποτελέσματα της αναζήτησης με μεγαλύτερο βάρος στον τίτλο παρά στο κείμενο τα αποτελέσματα θα είναι προφανώς ίδια με αυτά με τα ίδια βάρη αφού το σύστημα επιστρέφει το ίδιο πλήθος αρχείων. Διαφορά θα υπήρχε αν τα αποτελέσματα μας έβγαιναν άμα το σύστημα επέστρεφε μόνο ένα συγκεκριμένο αριθμό αρχείων (top10) όπου ανάλογα με τα βάρη θα ήταν και διαφορετικά.

Δεν ήξερα ποιος τρόπος ήταν σωστός και τι ζητούσατε γι’ αυτό και όλα τα αποτελέσματα σε κάθε ερώτημα υπολογίστηκαν λαμβάνοντας υπ όψη όλα τα αρχεία που μας επιστρέφονται από το σύστημα και όχι μόνο κάποια πρώτα. (top10). Για την εμφάνιση αποτελεσμάτων σε κάθε ερώτημα υπολογίζοντας μόνο κάποια top αρχεία (για να βγουν και διαφορετικά αποτελέσματα στα ερωτήματα 3 και 4 ) το μόνο που χρειάζεται να αλλάξει σε κάθε ερώτημα είναι το δεύτερο όρισμα του TopDocs hits στην κλάση Searcher, και από 1400 να γίνει π.χ 10.