

Εργασία 2 – Αξιολόγηση Συστημάτων Ανάκτησης Πληροφορίας

1. Θεωρήστε δύο συστήματα **S1** και **S2** και τρεις ερωτήσεις, **q1**, **q2** και **q3**. Τα σχετικά έγγραφα για το ερώτημα **q1** είναι **50**, για το **q2** **20** και για το **q3** **200**. Το σύστημα **S1** επιστρέφει πάντα **50** αποτελέσματα, ενώ το **S2** επιστρέφει πάντα **200**. Αναλυτικά για τα ερωτήματα: για το **q1** το **S1** εντόπισε **30** σχετικά έγγραφα, ενώ το **S2** εντόπισε **50**. Για το **q2** το **S1** εντόπισε **3** σχετικά έγγραφα, ενώ το **S2** εντόπισε **20** σχετικά έγγραφα. Τέλος, για το **q3** το σύστημα **S1** εντόπισε **50** σχετικά έγγραφα, ενώ το **S2** **150** σχετικά έγγραφα.

Συγκρίνετε τα δύο συστήματα και σχολιάστε χρησιμοποιώντας:

α) ακρίβεια (precision), ανάκληση (recall), και F-measure με macro-averaging, και

α) ακρίβεια (precision), ανάκληση (recall), και F-measure με micro-averaging.

2. Θεωρείστε μια συλλογή εγγράφων που χρησιμοποιείται για την αξιολόγηση συστημάτων ανάκτησης πληροφορίας που αποτελείται από **40** έγγραφα $\{d1, ..., d40\}$. Η συλλογή αξιολόγησης περιλαμβάνει ένα ερώτημα **q** για το οποίο γνωρίζουμε ότι τα έγγραφα της συλλογής που είναι συναφή είναι **5**, συγκεκριμένα τα $\{d1, d11, d18, d21, d33\}$. Θέλουμε να αξιολογήσουμε την αποτελεσματικότητα τριών συστημάτων **S1**, **S2** και **S3**. Για το λόγο αυτό υποβάλλουμε σε κάθε σύστημα το ερώτημα **q** και λαμβάνουμε τις εξής απαντήσεις με διάταξη από τα πιο συναφή στο λιγότερο από τα αριστερά προς τα δεξιά:

Result(S1): **d11, d4, d18, d30, d21, d33, d9, d7, d8, d6, d1, d5**

Result(S2): **d9, d7, d5, d6, d11, d4, d8, d30, d1, d33, d18, d21**

Result(S3): **d18, d33, d11, d1, d12, d2, d30**

Συγκρίνετε τα τρία αυτά συστήματα ως προς τα εξής μέτρα:

(α) F-Measure

(β) Precision@5, Recall@5,

(β) R-Precision

(γ) MAP

(δ) Καμπύλες ακρίβειας/ανάκλησης (P/R curves). Για κάθε σύστημα δώστε 2 γραφήματα: ένα που να απεικονίζει τα P/R σημεία όπως προκύπτουν από τις απαντήσεις, και ένα χρησιμοποιώντας κανονικοποιημένα επίπεδα ανάκλησης (standard recall levels).

Σχολιάστε την καταλληλότητα των 3 συστημάτων, ποιο θα επιλέγατε και γιατί.

3. Θεωρείστε τα δύο παρακάτω συστήματα στα οποία τα αποτελέσματα που επιστρέφονται έχουν βαθμολογηθεί ως προς τη συνάφεια (rel) με ένα ερώτημα (**q**) σε μια κλίμακα 0 έως 3.

| S1 | | S2 | |
|----|-----|----|-----|
| | rel | | rel |
| 1 | 1 | 1 | 2 |
| 2 | 3 | 2 | 1 |
| 3 | 0 | 3 | 0 |
| 4 | 2 | 4 | 0 |
| 5 | 0 | 5 | 3 |
| 6 | 3 | 6 | 3 |
| 7 | 0 | 7 | 1 |
| 8 | 1 | 8 | 0 |

Συγκρίνετε τα δύο συστήματα ως προς: (α) discounted cumulative gain και (β) normalized cumulative gain και σχολιάστε τα αποτελέσματά σας.