

1^η Σειρά Ασκήσεων (Αξιολόγηση της Αποτελεσματικότητας της Ανάκτησης)

Ανάθεση: 21 Μαρτίου
Παράδοση: 16 Απριλίου

Άσκηση 1 (4 βαθμοί)

Θεωρείστε μια συλλογή αξιολόγησης που αποτελείται από 40 έγγραφα $\{d_1, \dots, d_{40}\}$. Η συλλογή αξιολόγησης περιλαμβάνει μια επερώτηση q για την οποία γνωρίζουμε ότι τα έγγραφα της συλλογής που είναι συναφή με αυτήν είναι 5, συγκεκριμένα τα $\{d_1, d_{11}, d_{18}, d_{21}, d_{33}\}$. Θέλουμε να αξιολογήσουμε την αποτελεσματικότητα τριών συστημάτων $S1, S2$ και $S3$. Για το λόγο αυτό υποβάλλουμε σε κάθε σύστημα την επερώτηση q και λαμβάνουμε τις εξής απαντήσεις:

$$\text{Ans}(S1, q) = \langle d_{11}, d_4, d_{18}, d_2, d_{21}, d_{33}, d_9, d_7, d_8, d_6, d_1, d_5 \rangle$$

$$\text{Ans}(S2, q) = \langle d_9, d_7, d_5, d_6, d_{11}, d_4, d_8, d_2, d_1, d_{33}, d_{18}, d_{21} \rangle$$

$$\text{Ans}(S3, q) = \langle d_{18}, d_{33}, d_{11}, d_1, d_5, d_2 \rangle$$

Το αριστερότερο στοιχείο της κάθε απάντησης παριστάνει το υψηλότερα διαβαθμισμένο έγγραφο, αυτό που το σύστημα υπολόγισε ως το πιο συναφές με την επερώτηση q . Συγκρίνετε τα τρία αυτά συστήματα ως προς τα εξής μέτρα: (α) Ακρίβεια (Precision), (β) Ανάκληση (Recall), (γ) F-Measure, (δ) R-Ακρίβεια (R-Precision) και (ε) Fallout. Για κάθε μέτρο σχολιάστε το αποτέλεσμα της σύγκρισης.

Άσκηση 2 (4 βαθμοί)

Σχεδιάστε τις καμπύλες ακρίβειας/ανάκλησης (P/R curves) των συστημάτων της προηγούμενης άσκησης. 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). Αν βλέπατε μόνο αυτά τα γραφήματα (και όχι τις απαντήσεις) θα μπορούσατε να επιλέξετε το καλύτερο σύστημα;

Άσκηση 3 (2 βαθμοί)

Έστω ότι η συλλογή αξιολόγησης αποτελείται από 200 έγγραφα $\{d_1, \dots, d_{200}\}$ και γνωρίζουμε ότι υπάρχουν 3 έγγραφα της συλλογής, συγκεκριμένα τα $\{d_1, d_2, d_3\}$, που είναι συναφή με την επερώτηση q . Θέλουμε να αξιολογήσουμε την αποτελεσματικότητα τριών συστημάτων $S1, S2$ και $S3$ τα οποία επιστρέφουν ως απάντηση έγγραφα συνοδευμένα από ένα βαθμό συνάφειας. Υποβάλλουμε σε κάθε σύστημα την επερώτηση q και λαμβάνουμε τις εξής απαντήσεις:

$$\text{Ans}(S1, q) = \langle d_1, \{d_2, d_{100} - d_{200}\}, d_3 \rangle$$

$$\text{Ans}(S2, q) = \langle d_1, d_2, d_3 \rangle$$

$$\text{Ans}(S3, q) = \langle \{d_1, d_8\}, d_2, d_3 \rangle$$

Η απάντηση $\langle \{d_1, d_8\}, d_2, d_3 \rangle$ σημαίνει ότι τα d_1, d_8 ισοβαθούν στην πρώτη θέση (άρα έλαβαν τον μεγαλύτερο βαθμό συνάφειας). Η απάντηση $\langle d_1, \{d_2, d_{100} - d_{200}\}, d_3 \rangle$ σημαίνει ότι το d_1 έλαβε το μεγαλύτερο βαθμό, ενώ μετά ακολουθεί μια ομάδα από 102 έγγραφα τα οποία ισοβαθούν, και στο τέλος της κατάταξης βρίσκεται το d_3 . Για κάθε ένα από τα 3 συστήματα απαντήστε τα ακόλουθα ερωτήματα:

(α) Ποια είναι η R-ακρίβεια (R-precision);

(β) Ποιο είναι το αναμενόμενο μήκος αναζήτησης για να βρούμε 2 συναφή;

(γ) Ποιο είναι το μέσο αναμενόμενο μήκος αναζήτησης;

.-