

1^η Σειρά Ασκήσεων (Αξιολόγηση Αποτελεσματικότητας Ανάκτησης)

Ανάθεση: 1 Μαρτίου
Παράδοση: 14 Μαρτίου

Άσκηση 1 (4 βαθμοί)

Θεωρείστε μια συλλογή αξιολόγησης που αποτελείται από 40 έγγραφα βάσει της οποίας θέλουμε να αξιολογήσουμε την αποτελεσματικότητα τριών συστημάτων S1, S2 και S3. Για το λόγο αυτό υποβάλλουμε σε κάθε σύστημα μια επερώτηση q και λαμβάνουμε τις εξής απαντήσεις:

$$\text{Ans}(S1,q) = \langle R N N N R R N N N N R N \rangle$$

$$\text{Ans}(S2,q) = \langle N N N N R N N N N R R R \rangle$$

$$\text{Ans}(S3,q) = \langle R R R N N N \rangle$$

Το αριστερότερο στοιχείο της κάθε απάντησης παριστάνει το υψηλότερα διαβαθμισμένο έγγραφο, αυτό που το σύστημα υπολόγισε ως το πιο συναφές με την επερώτηση q. Συμβουλευόμενοι την συλλογή αξιολόγησης διακρίνουμε τα στοιχεία των απαντήσεων σε συναφή (R) και μη συναφή (N). Έστω ότι ξέρουμε ότι το σύνολο των εγγράφων της συλλογής που είναι συναφή με την επερώτηση q είναι 5.

Συγκρίνεται τα τρία αυτά συστήματα ως προς τα εξής μέτρα:

- (α) Ακρίβεια (Precision)
- (β) Ανάκληση (Recall)
- (γ) F-Measure
- (δ) R-Ακρίβεια (R-Precision)
- (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 έχουμε μια συλλογή N εγγράφων και K συστήματα ανάκτησης πληροφοριών. Θέλουμε να αξιολογήσουμε την αποτελεσματικότητα των συστημάτων αυτών, ώστε να επιλέξουμε το καλύτερο, αλλά δυστυχώς δεν υπάρχει καμία συλλογή αξιολόγησης. Επίσης δεν μπορούμε να κάνουμε οι ίδιοι μια άτυπη αξιολόγηση (ήτοι να υποβάλλουμε σε κάθε σύστημα ένα σύνολο επερωτήσεων και να κρίνουμε τις αποκρίσεις τους ως προς την ακρίβεια τους) διότι είτε δεν έχουμε τον απαιτούμενο χρόνο για κάτι τέτοιο (π.χ. φανταστείτε την περίπτωση που K=1000), ή διότι δεν μπορούμε να το κάνουμε (π.χ. τα έγγραφα είναι γραμμένα στην κινεζική γλώσσα).

(α) Προτείνετε τρόπους αντιμετώπισης αυτού του προβλήματος και δικαιολογείστε τις απαντήσεις σας (συγκεκριμένα τις υποθέσεις υπό τις οποίες αυτό που προτείνετε θα είχε νόημα).

(β) Έστω ότι έχετε X Ευρώ στη διάθεση σας και ότι υπάρχει ένας ... κινέζος ο οποίος με 1 Ευρώ μπορεί να σας απαντήσει αν ένα έγγραφο d είναι συναφές ή όχι με μια επερώτηση q. Πως θα τον χρησιμοποιούσατε για την αξιολόγηση των συστημάτων;

Καλή εργασία