

3^η Σειρά Ασκήσεων

Ανάδραση Συνάφειας και Ευρετήρια
Ημερομηνία Παράδοσης: Δευτέρα 4 Μαΐου 2009
Ομάδες των 2 ατόμων

Άσκηση 1

Στο μοντέλο του Rocchio,

- (α) ποιες τιμές των α , β και γ μας δίνουν το αντίστοιχο του “Find pages like this one”; Εξηγήστε γιατί.
(β) σε ποια περίπτωση η ερώτηση q_m μετά την ανάδραση είναι ίδια με την αρχική ερώτηση q_0 ;

Άσκηση 2

Χρησιμοποιείτε το σύστημα ανάκτησης εικόνων με ανάδραση [<http://amazon.ece.utexas.edu/~qasim/cires.htm>] που είδαμε στο μάθημα. Συγκεκριμένα: (1) διατυπώστε μια ερώτηση και (2) επιλέξτε τις εικόνες από τα αποτελέσματα που θεωρείται πιο σχετικές.

Σχολιάστε τα αποτελέσματα για μία ερώτηση και για δύο διαφορετικές επιλογές εικόνων για το βήμα (2).

Άσκηση 3

Θεωρείστε τα παρακάτω κείμενα (έγγραφα):

- d1 : Ο κομήτης του Χάλλεϋ μας επισκέπτεται περίπου κάθε εβδομήντα έξι χρόνια.
d2 : Ο κομήτης του Χάλλεϋ πήρε το όνομά του από τον αστρονόμο Έντμοντ Χάλλεϋ.
d3 : Ένας κομήτης διαγράφει ελλειπτική τροχιά.
d4 : Ο πλανήτης Άρης έχει δύο φυσικούς δορυφόρους, το Δείμο και το Φόβο.
d5 : Ο πλανήτης Δίας έχει 63 γνωστούς φυσικούς δορυφόρους.
d6 : Ένας κομήτης έχει μικρότερη διάμετρο από ότι ένας πλανήτης.
d7 : Ο Άρης είναι ένας πλανήτης του ηλιακού μας συστήματος.

Θεωρείστε ότι επιλέγουμε τους όρους $K = \{\text{κομήτης, Χάλλεϋ, πλανήτης, δορυφόρος, τροχιά και σύστημα}\}$.

- (α) Αναπαραστήστε τα κείμενα χρησιμοποιώντας το διανυσματικό μοντέλο (vector model).
(β) Θεωρείστε, ότι έχετε ένα σύστημα ανάδρασης συνάφειας και ο χρήστης αξιολογεί από τα αποτελέσματα της ερώτησης $q_0 = \{\text{κομήτης}\}$ ως σχετικά τα d3 και d6. Δώστε τη νέα ερώτηση και το αποτέλεσμα της μετά την ανάδραση.
(γ) Επαναλάβετε το ερώτημα (β) χρησιμοποιώντας ψευτο-ανάδραση.
(δ) Κατασκευάστε τον πίνακα αυτό-συσχέτισης (correlation matrix) και τον κανονικοποιημένο πίνακα αυτό-συσχέτισης (normalized correlation matrix) για το σύνολο όρων K και τα έγγραφα της απάντησης της ερώτησης q_0 . Επαναλάβετε το ερώτημα (β) προσθέτοντας έναν όρο στην ερώτηση με βάση τον κανονικοποιημένο πίνακα.
(ε) Επαναλάβετε το ερώτημα (β) χρησιμοποιώντας καθολική ανάλυση.

Άσκηση 4

Θεωρείστε το έγγραφο d2 της Άσκησης 3.

Αγνοώντας τους τόνους, σχεδιάστε

- (α) το trie του λεξιλογίου του
(β) το δέντρο καταλήξεων του θεωρώντας ως σημεία ευρετηρίου (index points) τις αρχές των λέξεων.
(γ) κατασκευάστε το trie του λεξιλογίου του εγγράφου d1 της Άσκησης 3 και στη συνέχεια συγχωνεύστε το με αυτό του ερωτήματος (α).