

1^ο Σύνολο Ασκήσεων

Καταληκτική Ημερομηνία Παράδοσης: Τετάρτη 3 Απριλίου 2013, στην αρχή του μαθήματος

Υλη: Κεφάλαια 1-4 από το βιβλίο

Μπορεί να γίνει προφορική εξέταση

Άσκηση 1

Θεωρείστε τα παρακάτω συστήματα Ανάκτησης Πληροφοριών:

<http://www.rhymezone.com/shakespeare/>

<https://www.google.com/>

α. Υποστηρίζουν Boolean ή σταθμισμένη (ranked) ανάκτηση;

β. Υποστηρίζουν ερωτήματα φράσεων και θέσεων/εγγύτητας; Αν ναι, δώστε από ένα παράδειγμα.

γ. Υποστηρίζουν Boolean ερωτήματα με χρήση and και or; Ειδικά για το Google απαντήστε και την Άσκηση 1.13 (σελίδα 39) του βιβλίου.

δ. Υποστηρίζουν ερωτήματα με * και αυτόματη διόρθωση μεμονωμένης λέξης; Αν ναι, δώστε από ένα παράδειγμα.

ε. Υποστηρίζουν διόρθωση με βάση τα συμφραζόμενα; Αν ναι, δώστε από ένα παράδειγμα.

Άσκηση 2

Θεωρείστε την παρακάτω συλλογή εγγράφων

Έγγραφο 1: good times to work on information retrieval or on infomatics

Έγγραφο 2: vacation time during Easter time is informative

Έγγραφο 3: information on summer vacations

Έγγραφο 4: retrieval of new information

(α) Δώστε τα tokens και τους όρους (terms) μετά από απλή περιστολή (stemming) που αγνοεί μόνο τον ενικό/πληθυντικό αριθμό.

(β) Δώστε τη μήτρα σύμπτωσης (incidence matrix) χρησιμοποιώντας τους όρους του ερωτήματος (α).

(γ) Δώστε το αντεστραμμένο ευρετήριο (inverted index) χρησιμοποιώντας τους όρους του ερωτήματος (α) και χρησιμοποιήστε για να υπολογίσετε το αποτέλεσμα των (Boolean) ερωτημάτων: (i) information (ii) information and retrieval.

(δ) Δώστε το ευρετήριο θέσεων και δείξτε πως χρησιμοποιείται για την απάντηση της φράσης “information retrieval”.

(ε) Τώρα δώστε τους όρους που προκύπτουν αφού εφαρμόσετε το Porter stemmer (π.χ, <http://tartarus.org/martin/PorterStemmer/>) και το σχετικό αντεστραμμένο ευρετήριο.

(στ) Η εφαρμογή του stemmer επηρεάζει και αν ναι πως την ανάκληση και την ακρίβεια (precision, recall); Π.χ., θεωρείστε το αποτέλεσμα στο ερώτημα informatics.

Άσκηση 3

(α) Τροποποιήστε τον αλγόριθμο υπολογισμού της απόστασης διόρθωσης μεταξύ δύο λέξεων ώστε να χρησιμοποιεί βάρη.

(β) [προαιρετικό] Η απόσταση διόρθωσης είναι αντιμεταθετική; Εξηγήστε. Ισχύει το ίδιο και για τη σταθμισμένη απόσταση;

Άσκηση 4

Άσκηση 4.7 (σελίδα 109) του βιβλίου αλλά για ένα ανανεωμένο Πίνακα 4.1 με βάση τα σημερινά δεδομένα. Δώστε και τον ανανεωμένο πίνακα.