

Αλγόριθμοι και Πολυπλοκότητα

2η Σειρά Γραπτών Ασκήσεων

Γιαννέλος Γιάννης
ΑΜ:03108088

19 Δεκεμβρίου 2011

Άσκηση 4: Χωρισμός κειμένου σε γραμμές

Όπως ορίζεται και στην εκφώνηση, το κείμενο που θέλουμε να χωρίσουμε αποτελείται από n λέξεις μήκους l_1, l_2, \dots, l_n χαρακτήρες η κάθε μια. Οι γραμμές θα πρέπει να έχουν μέγεθος το πολύ C χαρακτήρες και κάθε γραμμή στοιχίζεται στα αριστερά. Ακόμα κάθε λέξη χωρίζεται από την επόμενη με τον κενό χαρακτήρα χωρίς να επιτρέπεται να χωριστεί σε 2 γραμμές και δεν επιτρέπεται η αναδιάταξη τους. Άρα συνολικά, αν στη γραμμή k εμφανίζονται οι λέξεις i, \dots, j τότε η γραμμή έχει αριθμό κενών χαρακτήρων στα δεξιά που δίνεται από τον τύπο $s_k = C + 1 - \sum_{p=i}^j (l_p + 1)$. Το ζητούμενο της άσκησης είναι να βρούμε αλγόριθμο που να χωρίζει κατάλληλα το κείμενο έτσι ώστε να ελαχιστοποιείται το άθροισμα των τετραγώνων του πλήθους των κενών χαρακτήρων που έχουν οι γραμμές στα δεξιά, δηλαδή το $\sum_{k=1}^m s_k^2$. Για την επίλυση θα εργαστούμε με την μέθοδο σχεδίασης bottom-up χρησιμοποιώντας δυναμικό προγραμματισμό:

- Η κύρια ιδέα στηρίζεται στο ότι αν μπορούμε να υπολογίσουμε το ελάχιστο $\sum_{k=1}^{\lambda} s_k^2$ για τις πρώτες λ σειρές, τότε αν προσθέσουμε τις λέξεις για την $\lambda + 1$ σειρά, δεν αλλάζει κάτι στην στοίχιση των προηγούμενων σειρών παρα μόνο της τελευταίας.
- Υπολογίζουμε τον πίνακα S όπου για κάθε i, j το στοιχείο $S[i, j]$ δείχνει το $s_k = C + 1 - \sum_{p=i}^j (l_p + 1)$.
- Με την παραπάνω ιδέα είναι εφικτό να φτιάξουμε μια αναδρομική σχέση για το "κόστος" της διάταξης των πρώτων i λέξεων του κειμένου. Η σχέση αυτή θα είναι:

$$C[i] = \begin{cases} 0, & i = 0 \\ \min_{1 \leq k \leq i} C[k-1] + S[k, i], & i > 0 \end{cases}$$