

ΘΕΜΑΤΑ ΔΙΠΛΩΜΑΤΙΚΩΝ ΕΡΓΑΣΙΩΝ
Εργ. Συστημάτων Βάσεων Γνώσεων & Δεδομένων

ΥΛΟΠΟΙΗΣΗ ΕΥΡΕΤΗΡΙΩΝ ΓΙΑ ΒΙΟΛΟΓΙΚΑ ΔΕΔΟΜΕΝΑ

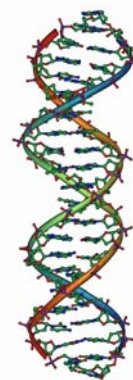
ΠΛΗΡΟΦΟΡΙΕΣ: Θανάσης Βεργούλης, 210 772 1436, bergoulis@dblab.ece.ntua.gr
Παναγιώτης Μπούρος, 210 772 1436, pbour@dblab.ece.ntua.gr

ΠΕΡΙΛΗΨΗ: Στόχος της διπλωματικής εργασίας είναι η υλοποίηση δομών ευρετηρίου για ερωτήματα ταιριάσματος προτύπων σε βιολογικά δεδομένα.

ΑΤΟΜΑ: 1-2

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Ανοιχτή

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Η Μοριακή Βιολογία είναι η επιστήμη της μελέτης των βιολογικών δομών σε μοριακό επίπεδο. Αντικείμενα μελέτης της είναι τα διάφορα βιομόρια όπως για παράδειγμα τα μόρια DNA (εικόνα δεξιά), τα μόρια RNA και οι πρωτεΐνες. Τα συγκεκριμένα μόρια μπορούν να αναπαρασταθούν από ακολουθίες συμβόλων και επομένως υπό αυτή τη μορφή να αποθηκευτούν και να επεξεργαστούν από υπολογιστές. Ο όγκος αυτών των δεδομένων είναι μεγάλος και δημιουργεί την ανάγκη για αποδοτική αποθήκευση και διαχείρισή τους.



Μια χρήσιμη λειτουργία πάνω σε τέτοιου τύπου ακολουθιακά δεδομένα είναι η εύρεση θέσεων στις οποίες εμφανίζονται συγκεκριμένες υπακολουθίες που παρουσιάζουν ενδιαφέρον. Πιο συγκεκριμένα, αν έχουμε μια βάση δεδομένων με N αποθηκευμένες ακολουθίες, έστω S_1, S_2, \dots, S_N , και μια ακολουθία ενδιαφέροντος S_Q , τότε θέλουμε να βρούμε τις θέσεις μέσα στις S_1, S_2, \dots, S_N στις οποίες ταιριάζει η S_Q . Το ταιρίασμα γίνεται με βάση κάποιους προκαθορισμένους κανόνες.

Προκειμένου να επιταχυνθεί η συγκεκριμένη λειτουργία, μπορεί να κατασκευαστεί πάνω στις ακολουθίες δεδομένων κάποια δομή ευρετηρίου (index). Τύποι ευρετηρίων που χρησιμοποιούνται συχνά σε τέτοιες περιπτώσεις είναι τα Δένδρα Επιθεμάτων (Suffix Trees) και οι Πίνακες Επιθεμάτων (Suffix Arrays).

Στα πλαίσια της διπλωματικής εργασίας οι υποψήφιοι θα μελετήσουν τους διάφορους τύπους ευρετηρίων για ακολουθιακά, βιολογικά δεδομένα, θα προχωρήσουν στην υλοποίησή τους και θα πραγματοποιήσουν την πειραματική αξιολόγησή τους.

ΠΕΡΙΣΣΟΤΕΡΕΣ ΠΛΗΡΟΦΟΡΙΕΣ:

Εισαγωγή στο πεδίο της διαχείρισης βιολογικών δεδομένων:

<http://www.dblab.ece.ntua.gr/pubs/uploads/DIPL-2006-5.pdf>

<http://www.dbnet.ece.ntua.gr/pubs/uploads/DIPL-2007-4.pdf>

Γενικές πληροφορίες για το Δένδρο Επιθεμάτων:

<http://www.cise.ufl.edu/~sahni/dsaaj/enrich/c16/suffix.htm>

<http://www.allisons.org/II/AlgDS/Tree/Suffix/>

Γενικές πληροφορίες για τον Πίνακα Επιθεμάτων:

<http://sary.sourceforge.net/docs/suffix-array.html>