



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

DIANA 2.0:

**Προηγμένη Εφαρμογή Ιστού Διαχείρισης Δεδομένων
Βιοεπιστημών**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΓΕΩΡΓΙΑΣ Χ. ΦΩΤΟΠΟΥΛΟΥ

Επιβλέπων : Τιμολέων Σελλής
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2010



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

DIANA 2.0:
Προηγμένη Εφαρμογή Ιστού Διαχείρισης Δεδομένων
Βιοεπιστημών

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΓΕΩΡΓΙΑΣ Χ. ΦΩΤΟΠΟΥΛΟΥ

Επιβλέπων : Τιμολέων Σελλής
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 12^η Ιουλίου 2010.

.....
Τιμολέων Σελλής
Καθηγητής Ε.Μ.Π.

.....
Νεκτάριος Κοζύρης
Αν. Καθηγητής Ε.Μ.Π.

.....
Δημήτριος Φωτάκης
Λέκτορας Ε.Μ.Π.

Αθήνα, Ιούλιος 2010

.....

ΓΕΩΡΓΙΑ ΦΩΤΟΠΟΥΛΟΥ

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Γεωργία Χ. Φωτοπούλου, 2010

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση εργασίας της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς την συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τη συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Στους γονείς μου, Χρήστο και Μαρία.



Περίληψη

Ο σκοπός της διπλωματικής εργασίας ήταν η ανάπτυξη νέων χαρακτηριστικών για την εφαρμογή DIANA. Η DIANA είναι μια προηγμένη εφαρμογή ιστού για διαχείριση βιολογικών δεδομένων, η οποία αναπτύχθηκε σε συνεργασία με το Ε.Κ. Βιοϊατρικών Επιστημών "Αλέξανδρος Φλέμινγκ", στα πλαίσια προηγούμενης διπλωματικής. Η εφαρμογή προσφέρει υπηρεσίες αναζήτησης και παρουσίασης δεδομένων σχετικών με microRNA στόχους σε γονίδια. Τα microRNAs είναι βιομόρια, τα οποία έχουν στόχους πάνω σε γονίδια. Όταν κάποιο microRNA έχει στόχο πάνω σε κάποιο γονίδιο, τότε μπορεί να αποτρέψει την παραγωγή της αντίστοιχης πρωτεΐνης. Υπό αυτή την έννοια, η γνώση των στόχων αυτών είναι ένα σημαντικό εργαλείο στην κατεύθυνση της αντιμετώπισης διαφόρων ασθενειών. Τα δεδομένα που διαχειρίζεται η DIANA είναι προβλέψεις για miRNA στόχους σε γονίδια, οι οποίες έχουν προκύψει με βάση ένα υπολογιστικό μοντέλο που έχει αναπτυχθεί από το Ε.Κ. Βιοϊατρικών Επιστημών "Αλέξανδρος Φλέμινγκ". Η παρούσα διπλωματική κινήθηκε στα πλαίσια του εμπλουτισμού της εφαρμογής με επιπλέον προηγμένα χαρακτηριστικά όπως εύχρηστο περιβάλλον για τους διαχειριστές του συστήματος, απλοποιημένη διεπαφή αναζητήσεων, βελτιωμένο μηχανισμό παροχής προτάσεων για λανθασμένους όρους αναζήτησης και δυνατότητα φόρτωσης και συγκριτικής απεικόνισης για προβλέψεις από μεθόδους που έχουν αναπτυχθεί από διαφορετικές ομάδες. Η ανάπτυξη της εφαρμογής πραγματοποιήθηκε με βάση σύγχρονες αρχές της Τεχνολογίας Λογισμικού όπως το αρχιτεκτονικό πρότυπο Model-View-Controller, η ανάλυση απαιτήσεων με περιπτώσεις χρήσης και ο έλεγχος του συστήματος με σενάρια εκτέλεσης.

Λέξεις Κλειδιά: εφαρμογή ιστού, βιολογία, microRNA, γονίδια, στόχοι microRNA, αναζήτηση, διαχείριση χρηστών, περιβάλλον διαχείρισης, πρότυπο Model-View-Controller

Abstract

The scope of this thesis was the development of new features for the DIANA application. DIANA is an advanced web application for the management of biological data, developed in cooperation with the Biomedical Sciences Research Center "Alexander Fleming", in the course of a previous thesis. The application offers services for the search and presentation of data relating to microRNA targets in genes. MicroRNAs are biomolecules, which have targets on genes. In case a microRNA has a target on a specific gene, it can prevent the production of the corresponding protein. In this sense, knowledge of these targets is an important tool towards the treatment of serious diseases. The data managed by DIANA are predictions of miRNA targets on genes, provided by a computational model developed by the Biomedical Sciences Research Center "Alexander Fleming". The present thesis is focused on enriching the application with additional advanced features, like an easy to use environment for the administrators of the system, a simplified search interface, an improved mechanism for providing suggestions on erroneous search terms and uploading and comparative display of predictions from methods developed by different teams. The development of the application was based on modern Software Engineering principles, such as the Model-View-Controller architectural pattern, requirements analysis based on use cases, and system testing with execution scenarios.

Keywords: web application, biology, microRNA, genes, microRNA targets, search, user management, administration panel, Model-View-Controller pattern

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή κ. Τίμο Σελλή για την ευκαιρία που μου έδωσε να εκπονήσω τη διπλωματική μου εργασία υπό την επίβλεψή του. Ευχαριστώ επίσης τους κ. Θοδωρή Δαλαμάγκα και Θανάση Βεργούλη για την καθοδήγηση και την υποστήριξη που δέχθηκα από αυτούς.

Ευχαριστώ θερμά τους γονείς μου και τα αδέρφια μου Βασιλική και Κοσμά, για την ηθική και υλική υποστήριξη όλα αυτά τα χρόνια, για την στήριξη των επιλογών μου και για όλα όσα μου δίδαξαν και θα μου διδάξουν στη ζωή μου.

Τέλος, ένα μεγάλο ευχαριστώ στον κ. Μιχάλη Πιτίδη για την ηθική συμπαράσταση και τη στήριξη που μου προσφέρει σε κάθε βήμα μου.

Πίνακας περιεχομένων

1	Εισαγωγή.....	5
1.1	Βιολογική έρευνα των μορίων microRNA.....	5
1.2	Αντικείμενο διπλωματικής.....	6
1.2.1	Συνεισφορά.....	7
1.3	Οργάνωση κειμένου.....	8
2	Υπόβαθρο.....	9
2.1	Βασικές αρχές βιολογίας και βιομόρια microRNA.....	9
2.2	Πληροφοριακά συστήματα διαχείρισης microRNA.....	10
2.2.1	DIANA-microT 3.0.....	10
2.2.2	miRanda.....	12
2.2.3	MicroCosm.....	13
2.2.4	EIMMo.....	13
2.2.5	Targetscan 5.0.....	14
2.2.6	PITA.....	14
2.2.7	RNA22.....	15
2.2.8	Σύγκριση δικτυακών εφαρμογών.....	15
3	Ανάλυση Απαιτήσεων Συστήματος.....	17
3.1	Αρχιτεκτονική	17
3.2	Περιγραφή Λειτουργιών.....	21
3.2.1	Διαχείριση χρηστών.....	21
3.2.2	Παροχή προτάσεων για τους όρους αναζήτησης.....	22
3.2.3	Εισαγωγή στοιχείων.....	23
3.2.4	Επαλήθευση στοιχείων.....	23
3.2.5	Αναζήτηση βιολογικών στόχων.....	24
3.2.6	Παρουσίαση αποτελεσμάτων.....	24

3.2.7	Ιστορικό αναζητήσεων.....	24
3.2.8	Τήρηση αγαπημένων αναζητήσεων.....	25
3.3	Περιπτώσεις χρήσης.....	25
3.3.1	Περίπτωση “Λίστα χρηστών”.....	26
3.3.2	Περίπτωση “Εισαγωγή νέων χρηστών”.....	27
3.3.3	Περίπτωση “Διαγραφή χρηστών”.....	29
3.3.4	Περίπτωση “Τροποποίηση στοιχείων χρηστών”.....	30
3.3.5	Περίπτωση “Εισαγωγή αρχείου στόχων με βάση άλλους αλγορίθμους”.....	31
3.3.6	Περίπτωση “Αναζήτηση microRNA”.....	32
3.3.7	Περίπτωση “Παροχή προτάσεων στην αναζήτηση microRNA”.....	33
3.3.8	Περίπτωση “Αναζήτηση γονιδίων”.....	34
3.3.9	Περίπτωση “Παροχή προτάσεων στην αναζήτηση γονιδίων”.....	35
3.3.10	Περίπτωση “Εμφάνιση προβλέψεων με βάση άλλους αλγορίθμους”.....	35
3.4	Μοντέλο Οντοτήτων Συσχετίσεων.....	36
4	Σχεδίαση Συστήματος.....	39
4.1	Αρχιτεκτονική.....	39
4.1.1	Model.....	40
4.1.2	View.....	40
4.1.3	Controller.....	40
4.2	Περιγραφή Κλάσεων.....	41
4.2.1	Κλάσεις Active Record.....	41
4.2.2	Κλάσεις Form Model.....	43
4.2.3	Ελεγκτές.....	47
4.2.4	Widgets.....	48
4.3	Βάση Δεδομένων.....	50
5	Υλοποίηση.....	55
5.1	Επιλογή μηχανής αποθήκευσης.....	55
5.2	Μηχανισμός προτάσεων στις αναζητήσεις χρηστών του συστήματος.....	57
5.2.1	Έλεγχος ομοιότητας με την απόσταση Levenshtein.....	58

5.2.2 Παροχή προτάσεων σε ονομασίες microRNA.....	60
5.2.3 Παροχή προτάσεων σε ονομασίες γονιδίων.....	66
5.3 Πλατφόρμες και προγραμματιστικά εργαλεία	69
5.3.1 Apache HTTP Server	69
5.3.2 MySQL.....	70
5.3.3 PHP.....	70
5.4 Άλλα θέματα.....	71
5.4.1 Τεκμηρίωση.....	71
5.4.2 Αναφορά σφαλμάτων.....	72
6 Έλεγχος και αξιολόγηση.....	73
6.1 Μεθοδολογία ελέγχου.....	73
6.2 Αναλυτική παρουσίαση ελέγχου.....	74
6.2.1 Δημιουργία νέου λογαριασμού χρήστη.....	74
6.2.2 Αναζήτηση με βάση το όνομα του microRNA.....	75
6.2.3 Αναζήτηση με βάση τον κωδικό MIMA.....	79
6.2.4 Αναζήτηση με βάση το όνομα του γονιδίου.....	80
6.2.5 Αναζήτηση με βάση τον κωδικό ENSEMBL	80
6.2.6 Αναζήτηση με βάση τον κωδικό TRANSCRIPT.....	81
6.2.7 Αναζήτηση με βάση τον κωδικό Refseq.....	83
6.2.8 Διαχείριση αγαπημένων αναζητήσεων και ιστορικού.....	83
6.2.9 Παροχή προτάσεων σε λανθασμένο όνομα microRNA	86
6.2.10 Παροχή προτάσεων σε λανθασμένο κωδικό MIMA.....	86
6.2.11 Παροχή προτάσεων σε λανθασμένο όνομα γονιδίου.....	88
6.2.12 Παροχή προτάσεων σε λανθασμένο κωδικό ENSEMBL.....	89
6.2.13 Παροχή προτάσεων σε λανθασμένο κωδικό TRANSCRIPT.....	91
6.2.14 Παροχή προτάσεων σε λανθασμένο κωδικό Refseq.....	93
6.2.15 Σύνδεση απλού χρήστη στο πάνελ διαχείρισης.....	94
6.2.16 Προβολή λίστας χρηστών.....	95
6.2.17 Διαγραφή χρήστη από το σύστημα.....	96

6.2.18 Τροποποίηση στοιχείων χρήστη του συστήματος.....	97
6.2.19 Εισαγωγή χρήστη στην ομάδα διαχειριστών.....	99
6.2.20 Πρόσθεση νέου χρήστη.....	100
6.2.21 Εισαγωγή αρχείου στόχου.....	102
6.3 Αξιολόγηση.....	105
7 Επίλογος.....	107
7.1 Σύνοψη και συμπεράσματα.....	107
7.2 Μελλοντικές εργασίες.....	108
8 Βιβλιογραφία.....	111

1

Εισαγωγή

1.1 Βιολογική έρευνα των μορίων microRNA

Η παρούσα διπλωματική εργασία συνδέεται άμεσα την επιστήμη της Βιολογίας. Στόχος της Βιολογίας είναι η μελέτη και η εξαγωγή συμπερασμάτων που αφορούν στη δομή, τη λειτουργία, την ανάπτυξη, την προέλευση και την εξέλιξη των ζώντων οργανισμών. Η Βιολογία προσπαθεί να απαντήσει σε όλα τα μικρά και μεγάλα ερωτήματα για τη ζωή των οργανισμών με στόχο την επίλυση προβλημάτων που την περιβάλλουν.

Η επιστήμη αυτή γνώρισε μεγάλη άνθηση από το 1953, όπου οι Francis Harry Compton Crick και James D. Watson ανακοίνωσαν το επαναστατικό μοντέλο τους για τη δομή του DNA. Η γνώση που αποκτήθηκε χάρη στην ανακάλυψη αυτή, οδήγησε σταδιακά στην ανάπτυξη πολύτιμων εργαλείων για την καταπολέμηση πολλών από τις ασθένειες του σήμερα.

Το 1993 οι Victor Ambros, Rosalind Lee και Rhonda Feinbaum ανακάλυψαν τα βιομόρια microRNA. Η ανακάλυψη αυτή ήταν ιδιαίτερα σημαντική, γιατί ερχόταν σε αντίθεση με το “Κεντρικό Δόγμα της Βιολογίας” βάσει του οποίου οι βιολόγοι πίστευαν μέχρι τότε ότι πραγματοποιείται η πρωτεϊνοσύνθεση. Τα βιομόρια αυτά προσδένονται στο “αγγελιαφόρο

RNA” (mRNA) που προκύπτει από συγκεκριμένα γονίδια και παρεμποδίζουν την παραγωγή της αντίστοιχης πρωτεΐνης. Τα σημεία πρόσδεσης των microRNAs πάνω στα mRNAs ονομάζονται στόχοι (targets). Αφού υπάρχει αντιστοιχία μεταξύ των mRNAs και των γονιδίων στη συνέχεια μπορούμε να αναφερόμαστε σε στόχους των microRNAs πάνω σε γονίδια.

Μέσω του μηχανισμού που περιγράφηκε προηγουμένως, τα microRNAs λειτουργούν ρυθμιστικά για τα κύτταρα του οργανισμού ελέγχοντας τη σύνθεση πρωτεϊνών. Πολλοί τύποι ασθενειών οφείλονται στη μη παραγωγή μιας συγκεκριμένης πρωτεΐνης. Για το λόγο αυτό, η απόκτηση γνώσης σχετικά ή μη στόχων ενός microRNA σε κάποιο γονίδιο, είναι ιδιαίτερα σημαντική για την κατανόηση διαφόρων ασθενειών. Βιολογικά πειράματα μπορούν να αποκαλύψουν αυτήν την πληροφορία, όμως κοστίζουν πολύ και απαιτείται χρόνος για την περάτωσή τους. Για να επιταχυνθεί η διαδικασία, έχουν προταθεί από ερευνητικές ομάδες υπολογιστικές μέθοδοι για την πρόβλεψη (prediction) των στόχων. Η ομάδα του Ερευνητικού Κέντρου Βιοϊατρικών Επιστημών “Αλέξανδρος Φλέμινγκ” έχει αναπτύξει ένα δικό της αλγόριθμο πρόβλεψης για τον εντοπισμό στόχων. Η εφαρμογή DIANA 2.0, που αναπτύχθηκε στην παρούσα διπλωματική εργασία, έχει ως αντικείμενο τη διαχείριση και παρουσίαση των microRNA στόχων που έχουν εντοπιστεί με βάση τον αλγόριθμο του Ε.Κ “Αλέξανδρος Φλέμινγκ”.

1.2 Αντικείμενο διπλωματικής

Στόχος της παρούσας διπλωματικής εργασίας ήταν η επέκταση του συστήματος DIANA, το οποίο είχε αναπτυχθεί σε συνεργασία με το Ερευνητικό Κέντρο Βιοϊατρικών Επιστημών “Αλέξανδρος Φλέμινγκ” στα πλαίσια προηγούμενης διπλωματικής. Συγκεκριμένα, η νέα εφαρμογή (DIANA 2.0) θα πρέπει να διαχειρίζεται καλύτερα τον μεγάλο όγκο δεδομένων που σχετίζεται με τους microRNA στόχους, βελτιώνοντας ταυτόχρονα την ευχρηστία και τις παρεχόμενες υπηρεσίες εξατομίκευσης. Εκτός αυτού θα πρέπει να προσφέρει επιπλέον χαρακτηριστικά όπως εύχρηστο περιβάλλον για τους διαχειριστές του συστήματος, βελτιωμένο μηχανισμό παροχής προτάσεων για τους λανθασμένους όρους προς αναζήτηση και λειτουργίες συγκριτικής παρουσίασης microRNA στόχων που προέρχονται από άλλες υπολογιστικές μεθόδους.

Στην περίπτωση του περιβάλλοντος διαχείρισης χρηστών οι διαχειριστές του συστήματος θα έχουν τη δυνατότητα να ελέγξουν και να τροποποιήσουν τις πληροφορίες των χρηστών του συστήματος καθώς και να μεταβάλλουν τις ομάδες χρηστών αυτού. Επιπλέον ο διαχειριστής

μπορεί εύκολα και απλά να ανεβάσει στο σύστημα αρχεία στόχων που έχουν προβλεφθεί με παρόμοιους αλγορίθμους με αυτόν της ομάδας του Ε.Κ Βιοϊατρικών Επιστημών “Αλέξανδρος Φλέμινγκ”. Το σύνολο αυτών των δυνατοτήτων θα παρέχεται μέσα από ένα ευέλικτο και εύχρηστο περιβάλλον.

Η νέος μηχανισμός αναζήτησης θα πρέπει να συνοδεύεται από μία απλή και εύχρηστη διεπαφή χρήστη και θα πρέπει να υποστηρίζει δυνατότητες αναζήτησης για όρους διαφόρων τύπων (ονόματα microRNA, ονόματα γονιδίων, κωδικοί microRNA, κωδικοί γονιδίων).

Ο βελτιωμένος μηχανισμός παροχής προτάσεων στους λανθασμένους όρους αναζήτησης οφείλει να είναι σε θέση να αντληφθεί το είδος του όρου προς αναζήτηση που εισήγαγε ο χρήστης και με έξυπνο τρόπο να προτείνει παρόμοιους έγκυρους όρους. Η ανάγκη για υψηλές επιδόσεις, αποτελεσματικότητα και οικονομία πόρων είναι κάτι περισσότερο από επιτακτική.

Τέλος, η συγκριτική παρουσίαση microRNA στόχων που προέρχονται από άλλες υπολογιστικές μεθόδους είναι μία λειτουργία η οποία πρέπει να υπογραμμίζεται από κομψότητα και απλότητα έτσι ώστε ο χρήστης να είναι σε θέση να ενημερωθεί πλήρως για τους συγκεκριμένους στόχους.

1.2.1 Συνεισφορά

Η συνεισφορά της παρούσας διπλωματικής εργασίας συνοψίζεται στα εξής σημεία:

- Δημιουργία εύχρηστου και πολλαπλών δυνατοτήτων περιβάλλοντος διαχείρισης χρηστών του συστήματος της εφαρμογής .
- Νέος απλοποιημένος μηχανισμός αναζήτησης με δυνατότητες υποστήριξης όρων πολλαπλών τύπων.
- Ανάπτυξη βελτιωμένου μηχανισμού παροχής προτάσεων σε λανθασμένους όρους αναζήτησης ο οποίος συνοδεύεται από γρήγορες επιδόσεις.
- Υποστήριξη συγκριτικής παρουσίασης microRNA στόχων που προέρχονται από άλλες υπολογιστικές μεθόδους.
- Τεκμηρίωση αρχείων κώδικα του συστήματος για τη διευκόλυνση μελλοντικών επεκτάσεων αυτού.

1.3 Οργάνωση κειμένου

Η συγγραφή της παρούσας διπλωματικής εργασίας έχει οργανωθεί με τέτοιο τρόπο ώστε να παρουσιάζονται αναλυτικά και ξεχωριστά οι φάσεις της ανάλυσης, της σχεδίασης και της υλοποίησης του συστήματος. Συνοπτικά:

Στο 2^ο κεφάλαιο γίνεται σύντομη περιγραφή των δικτυακών εφαρμογών που έχουν παρόμοιο σκοπό λειτουργίας με το σκοπό της εφαρμογής της παρούσας διπλωματικής εργασίας. Στο 3^ο κεφάλαιο παρουσιάζονται τα επιμέρους υποσυστήματα που απαρτίζουν την εφαρμογή και καθορίζονται οι λειτουργίες που μπορεί να επιτελέσει αυτή.

Στο 4^ο κεφάλαιο πραγματοποιείται η ανάλυση της αρχιτεκτονικής του συστήματος και περιγράφονται οι κυριότερες κλάσεις και μέθοδοι αυτών. Στη συνέχεια αφιερώνουμε μία ενότητα για την περιγραφή της βάσης δεδομένων που χρησιμοποιήθηκε για την εφαρμογή της παρούσας διπλωματικής εργασίας.

Στο 5^ο κεφάλαιο παρουσιάζονται αποφάσεις που πάρθηκαν για την υλοποίηση του συστήματος καθώς και ο νέος βελτιωμένος μηχανισμός παροχής προτάσεων στους όρους αναζήτησης. Επιπλέον γίνεται αναφορά στα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν για την ανάπτυξη της εφαρμογής και παρουσιάζονται ενδιαφέροντα θέματα σχετικά με την φάση υλοποίησης.

Στο 6^ο κεφάλαιο περιγράφεται η μεθοδολογία και η διαδικασία ελέγχου που χρησιμοποιήθηκε για την διαπίστωση της αξιόπιστης λειτουργίας του συστήματος και γίνεται αξιολόγηση του μηχανισμού παροχής προτάσεων σε όρους αναζήτησης.

Τέλος, στο 7^ο κεφάλαιο γίνεται σύνοψη και εξαγωγή συμπερασμάτων που αφορούν την ανάπτυξη της παρούσας εφαρμογής και προτείνονται μελλοντικές επεκτάσεις.

2

Υπόβαθρο

Στο κεφάλαιο αυτό παρουσιάζουμε συνοπτικά πληροφορίες σχετικά με βιομόρια και ειδικότερα για τα βιομόρια τύπου *microRNA* και τις λειτουργίες ανίχνευσης στόχων τους σε γονίδια. Στη συνέχεια παρουσιάζουμε συστήματα διαχείρισης πληροφοριών για τα βιομόρια αυτά.

2.1 Βασικές αρχές βιολογίας και βιομόρια *microRNA*

Η σύνθεση πρωτεϊνών είναι η σημαντικότερη λειτουργία του κυττάρου και σχετίζεται άμεσα με την επιβίωση του. Η πληροφορία για την σύνθεση όλων των απαραίτητων για το κύτταρο πρωτεϊνών είναι καταγεγραμμένη στο γονιδίωμα του και οι δομές στις οποίες οργανώνεται ονομάζονται γονίδια. Κάθε γονίδιο είναι ένα τμήμα DNA, το οποίο μπορεί να μεταγραφεί όποτε είναι απαραίτητο και το προϊόν αυτής της διαδικασίας ονομάζεται αγγελιαφόρο RNA (*mRNA*). Το *mRNA* κατευθύνεται από τον πυρήνα του κυττάρου σε ειδικά οργανίδια στα οποία παράγεται η αντίστοιχη πρωτεΐνη.

Πρόσφατα ανακαλύφθηκε ότι συγκεκριμένα τμήματα DNA, όχι μόνο δεν κωδικοποιούν πληροφορία για κάποια πρωτεΐνη, αλλά αντιθέτως κωδικοποιούν πληροφορία για δομές οι οποίες ευθύνονται για την εμπλοκή της σύνθεσης πρωτεϊνών. Αυτές οι δομές ονομάζονται

microRNA και είναι σχετικά μικρού μήκους (περίπου 30 βάσεις). Τα microRNA προσδένονται πάνω σε διάφορα μόρια mRNA σε συγκεκριμένα σημεία και τα καταστρέφουν. Τα σημεία στα οποία γίνεται η πρόσδεση ονομάζονται **στόχοι (targets)**.

Διάφορες ασθένειες οφείλονται στην παραγωγή ή μη συγκεκριμένων πρωτεϊνών. Για το λόγο αυτό τα microRNA, αποτελούν ένα σημαντικό εργαλείο στα χέρια των βιολόγων για τον έλεγχο της πρωτεϊνοσύνθεσης στην κατεύθυνση της αντιμετώπισης αυτών των ασθενειών.

Ο καθορισμός των στόχων μπορεί να γίνει είτε με βιολογικά είτε με υπολογιστικά πειράματα. Ωστόσο η πρώτη κατηγορία πειραμάτων είναι μεν πολύ ακριβής αλλά αρκετά χρονοβόρα και δαπανηρή διαδικασία. Αντιθέτως η δεύτερη κατηγορία πειραμάτων είναι λιγότερο ακριβής αλλά είναι πολύ πιο γρήγορη από τα βιολογικά πειράματα, λόγω της σύγχρονης προόδου των υπολογιστών. Από τα υπολογιστικά πειράματα προκύπτουν δεδομένα που παρουσιάζουν πιθανούς στόχους συγκεκριμένων microRNA με ένα συγκεκριμένο ποσοστό ακρίβειας πρόβλεψης.

Για το λόγο αυτό έχουν αναπτυχθεί αρκετά πληροφοριακά συστήματα που οργανώνουν τα αποτελέσματα των υπολογιστικών πειραμάτων ανίχνευσης στόχων σε βάσεις δεδομένων. Τα συστήματα αυτά παρέχουν υπηρεσίες αναζήτησης πληροφοριών σε στόχους, δεδομένα microRNA, δεδομένα γονιδίων, κλπ.

Περισσότερα για το βιολογικό υπόβαθρο της εφαρμογής μπορούν να αναζητηθούν στην αναφορά [7].

2.2 Πληροφοριακά συστήματα διαχείρισης microRNA

Στη συνέχεια παρουσιάζονται δικτυακές εφαρμογές οι οποίες σχετίζονται με το αντικείμενο της παρούσας διπλωματικής εργασίας. Για κάθε εφαρμογή παρουσιάζονται συνοπτικά τα κυριότερα χαρακτηριστικά της. Στο τέλος της ενότητας γίνεται και μία συγκριτική παρουσίαση των χαρακτηριστικών κάθε εφαρμογής.

2.2.1 DIANA-microT 3.0

Η εφαρμογή αυτή βρίσκεται στο δικτυακό τόπο <http://diana.cslab.ece.ntua.gr/microT/> και αποτελεί προγενέστερη έκδοση της εφαρμογής που παρουσιάζεται στην παρούσα διπλωματική εργασία. Ο χρήστης έχει τη δυνατότητα να αναζητήσει είτε τους στόχους ενός συγκεκριμένου microRNA, είτε τα ονόματα των microRNA τα οποία προσδένονται στα mRNA που σχετίζονται με ένα συγκεκριμένο γονίδιο, είτε τους στόχους ενός συγκεκριμένου

microRNA στα mRNA που προκύπτουν από ένα συγκεκριμένο γονίδιο. Επιπλέον υπάρχει η δυνατότητα αναζήτησης στόχων ορίζοντας την ακολουθία βάσεων για την οποία θέλει να εντοπιστούν αυτοί. Κάθε επιλογή αναζήτησης γίνεται με τη βοήθεια ξεχωριστής φόρμας στην οποία ο χρήστης εισάγει τους επιθυμητούς όρους. Τα αποτελέσματα παρουσιάζονται σε μορφή πίνακα και υπάρχει η επιλογή προβολής των κυριότερων χαρακτηριστικών κάθε στόχου και ενεργός σύνδεσμος προς τη δικτυακή βιολογική βάση PubMed (<http://www.ncbi.nlm.nih.gov/pubmed/>).



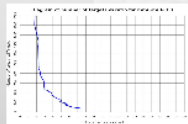
Στην περίπτωση που ο χρήστης πραγματοποιεί αναζήτηση με βάση το όνομα του microRNA, σε κάθε γραμμή του πίνακα αποτελεσμάτων εμφανίζονται με τη σειρά το όνομα του γονιδίου με το οποίο σχετίζεται το συγκεκριμένο microRNA, ο κωδικός ENSEMBL του γονιδίου και τέλος το miTG σκορ, η ακρίβεια πρόβλεψης και η σηματοθορυβική αναλογία (SNR) κάθε στόχου.

Στην περίπτωση που ο χρήστης πραγματοποιεί αναζήτηση με βάση το όνομα ενός γονιδίου ή με βάση των κωδικό ENSEMBL αυτού, σε κάθε γραμμή του πίνακα αποτελεσμάτων εμφανίζονται με τη σειρά το όνομα του microRNA το οποίο προσδένεται σε κάποιο mRNA που μεταγράφεται από το συγκεκριμένο γονίδιο και το miTG σκορ, η ακρίβεια πρόβλεψης και η σηματοθορυβική αναλογία (SNR) κάθε στόχου. Η ίδια μορφή αναζήτησης μπορεί να πραγματοποιηθεί εισάγοντας σε ειδική φόρμα την περιγραφή του γονιδίου, αντί για την ακριβή ονομασία του.

Τέλος, στην περίπτωση που ο χρήστης πραγματοποιεί αναζήτηση στόχων ενός συγκεκριμένου microRNA σε πολλά mRNA που σχετίζονται με ένα συγκεκριμένο γονίδιο, σε κάθε γραμμή του πίνακα αποτελεσμάτων εμφανίζονται το όνομα του γονιδίου και του microRNA που εισήγαγε ο χρήστης στην μπάρα αναζήτησης και τέλος το miTG σκορ, η ακρίβεια πρόβλεψης και η σηματοθορυβική αναλογία (SNR) κάθε στόχου. Όπως και προηγουμένως ο χρήστης μπορεί να εισάγει σε ειδική φόρμα την περιγραφή του γονιδίου αντί για την ακριβή ονομασία του.






Κατά την εισαγωγή των όρων στην αντίστοιχη φόρμα αναζήτησης, ανιχνεύονται τυχόν ορθογραφικά λάθη στις ονομασίες τόσο των microRNA, όσο και των γονιδίων. Για τους λανθασμένους όρους δεν πραγματοποιείται αναζήτηση. Αντιθέτως προτείνονται στο χρήστη έγκυρες ονομασίες αυτών των όρων. Στο σχήμα παρουσιάζεται ένα στιγμιότυπο της εφαρμογής DIANA.

Results for miRNA 'mmu-miR-103' (1497 target sites in 572 genes found) | [New Search](#)

miRNA name: mmu-miR-103   **Figure:** 

Alternative Description: MIMAT0000546

miRNA Sequence: AGCAGCAUUGUACAGGGCUAUGA

Score Threshold  KEGG Filter    Page 1 

Rank	Gene name	Ensembl Gene Id	miTG score	Precision	SNR		
<input type="checkbox"/>	1	Mapkap1	ENSMUSG00000038696	59.48	1	7.34	<input type="checkbox"/>
<input type="checkbox"/>	2	Tnfr6b	ENSMUSG00000047888	40.63	0.99	7.34	<input type="checkbox"/>
<input type="checkbox"/>	3	Med26	ENSMUSG00000045248	39.68	0.99	7.34	<input type="checkbox"/>
<input type="checkbox"/>	4	Ankib1	ENSMUSG00000040351	37.18	0.98	7.34	<input type="checkbox"/>
<input type="checkbox"/>	5	Zbth34	ENSMUSG00000068966	36.27	0.97	7.34	<input type="checkbox"/>
<input type="checkbox"/>	6	Actr2	ENSMUSG00000020152	35.44	0.96	7.34	<input type="checkbox"/>
<input type="checkbox"/>	7	Chd6	ENSMUSG00000057133	35.13	0.96	7.34	<input type="checkbox"/>
<input type="checkbox"/>	8	Mtmr4	ENSMUSG00000018401	34.30	0.96	7.34	<input type="checkbox"/>
<input type="checkbox"/>	9	Dhl1	ENSMUSG00000014773	34.08	0.96	7.34	<input type="checkbox"/>
<input type="checkbox"/>	10	Slc20a2	ENSMUSG00000037656	33.36	0.96	7.34	<input type="checkbox"/>

Εικόνα 1: Στιγμιότυπο της εφαρμογής DIANA-microT 3.0

2.2.2 miRanda

Η εφαρμογή αυτή βρίσκεται στο δικτυακό τόπο <http://www.microrna.org/>. Ο χρήστης έχει τη δυνατότητα να αναζητήσει, είτε τους στόχους ενός ή περισσότερων microRNA για το είδος οργανισμού που επιθυμεί, είτε τους στόχους microRNA που σχετίζονται με κάποιο mRNA με βάση το όνομα αυτού ή του γονιδίου από το οποίο προέρχεται. Ο χρήστης εισάγει τους όρους σε ξεχωριστές φόρμες ανάλογα με το είδος της αναζήτησης που επιθυμεί να πραγματοποιήσει. Τα αποτελέσματα παρουσιάζονται σε μορφή λίστας με βάση το όνομα του mRNA.

Όταν ο χρήστης πραγματοποιεί αναζήτηση με βάση το όνομα ή τα ονόματα των microRNA, αρχικά επιλέγει το είδος οργανισμού που τον ενδιαφέρει. Στη συνέχεια έχει τη δυνατότητα να επιλέξει την προβολή προτάσεων στην περίπτωση που κάποιος ή κάποιοι από τους όρους προς αναζήτηση που εισήγαγε, δεν αντιστοιχούν σε ακριβείς ονομασίες microRNA. Το επόμενο βήμα της διαδικασίας είναι η παρουσίαση μιας νέας φόρμας που περιέχει όλες τις έγκυρες ονομασίες microRNA που σχετίζονται με την αναζήτηση. Αφού επιλέξει τις

επιθυμητές ονομασίες, γίνεται η παρουσίαση των αποτελεσμάτων. Κάθε γραμμή της λίστας αποτελεσμάτων περιέχει το όνομα του mRNA που σχετίζεται με το ή τα συγκεκριμένα microRNA, πληροφορίες σε μορφή κειμένου για το mRNA αυτό και τέλος το είδος οργανισμού στο οποίο ανήκει.

Επιπλέον, στην περίπτωση που ο χρήστης αναζητά στόχους με βάση το όνομα ενός mRNA ή ενός γονιδίου, αφού επιλέξει το είδος οργανισμού που τον ενδιαφέρει, γίνεται η παρουσίαση των αποτελεσμάτων. Όπως και προηγουμένως, σε κάθε γραμμή της λίστας αποτελεσμάτων εμφανίζεται αρχικά το όνομα του mRNA, σύντομες πληροφορίες σε μορφή κειμένου για αυτό και τέλος το είδος οργανισμού στο οποίο εκφράζεται.

2.2.3 MicroCosm

Η εφαρμογή αυτή βρίσκεται στο δικτυακό τόπο <http://www.ebi.ac.uk/enright-srv/microcosm/cgi-bin/targets/v5/search.pl>. Ο χρήστης έχει τη δυνατότητα να πραγματοποιήσει αναζητήσεις στόχων είτε με το όνομα ενός συγκεκριμένου microRNA, είτε με το όνομα του γονιδίου από το οποίο προκύπτουν mRNA στα οποία προσδένονται microRNA, είτε με τον κωδικό ENSEMBL ενός γονιδίου. Επιπλέον, αν το επιθυμεί, μπορεί να καθορίσει το είδος οργανισμού στο οποίο γίνεται η αναζήτηση στόχων. Αυτές οι επιλογές αναζήτησης μπορούν να συνδυαστούν μεταξύ τους με αποτέλεσμα την πραγματοποίηση σύνθετης αναζήτησης. Η εισαγωγή των όρων γίνεται σε μία μόνο φόρμα, συμπληρώνοντας τα πεδία (όνομα microRNA, όνομα γονιδίου, κωδικός ENSEMBL) που είναι επιθυμητά.

Τα αποτελέσματα παρουσιάζονται σε μορφή πίνακα με τα κυριότερα χαρακτηριστικά του κάθε στόχου. Σε κάθε γραμμή του πίνακα εμφανίζονται με τη σειρά και για κάθε στόχο το είδος οργανισμού, το όνομα, ο κωδικός TRANSCRIPT και μια σύντομη περιγραφή του γονιδίου, το σκορ και η ακρίβεια πρόβλεψης και τέλος κάποιες πιο εξειδικευμένες πληροφορίες. Κάθε στόχος του πίνακα περιέχει ενεργό σύνδεσμο σε μια νέα σελίδα στην οποία παρουσιάζονται πιο λεπτομερώς και με γραφικό τρόπο όλα τα χαρακτηριστικά του. Επιπλέον στην περίπτωση που ο χρήστης έχει καθορίσει το είδος οργανισμού το οποίο τον ενδιαφέρει, υπάρχει η δυνατότητα ανάκτησης των αποτελεσμάτων σε μορφή αρχείου.

2.2.4 EIMMO

Η εφαρμογή αυτή βρίσκεται στο δικτυακό τόπο <http://www.mirz.unibas.ch/>. Ο χρήστης, αφού επιλέξει το είδος οργανισμού που επιθυμεί, στη συνέχεια επιλέγει το κριτήριο των mRNA με το οποίο θα γίνει η αναζήτηση και τέλος επιλέγει από μία λίστα με microRNA το

πολύ είκοσι ονόματα. Έτσι δίνεται στο χρήστη η δυνατότητα σύνθετης αναζήτησης, με τον αριθμητικό περιορισμό ωστόσο των ονομάτων microRNA που μπορούν να επιλεγούν. Στην περίπτωση που επιλέξει πάνω από ένα όνομα microRNA, μπορεί να επιλέξει να δει τα αποτελέσματα είτε ξεχωριστά ανά όνομα, είτε συγκεντρωμένα.

Τα αποτελέσματα οργανώνονται σε μορφή πίνακα με βάση τα mRNA στα οποία εντοπίστηκαν στόχοι. Κάθε γραμμή του πίνακα περιέχει σύντομες πληροφορίες σε μορφή κειμένου για το mRNA και έναν ενεργό σύνδεσμο σε μία νέα σελίδα στην οποία προβάλλονται λεπτομερώς τα σημεία πρόσδεσης για κάθε microRNA και έναν ενεργό σύνδεσμο προς τη δικτυακή βιολογική βάση PubMed.

2.2.5 TargetScan 5.0

Η εφαρμογή αυτή βρίσκεται στο δικτυακό τόπο <http://www.targetscan.org/>. Ο χρήστης αφού επιλέξει το είδος οργανισμού που τον ενδιαφέρει, έχει τη δυνατότητα να εισάγει το όνομα ενός γονιδίου. Εναλλακτικά μπορεί να εισάγει είτε το όνομα μιας οικογένειας microRNA ή το όνομα ενός συγκεκριμένου microRNA. Οι δύο επιλογές αναζήτησης μπορούν να συνδυαστούν με σκοπό στην πραγματοποίηση σύνθετης αναζήτησης.

Η λίστα των αποτελεσμάτων οργανώνεται σε μορφή πίνακα με βάση το όνομα του γονιδίου. Για κάθε στόχο αναγράφεται το κωδικό και πλήρες όνομα του γονιδίου, το αντιπροσωπευτικό microRNA για κάθε γονίδιο, το σκορ και η ακρίβεια πρόβλεψης και τέλος κάποιες πιο εξειδικευμένες πληροφορίες για αυτόν. Επιπλέον, στο τέλος κάθε γραμμής του πίνακα περιέχεται ενεργός σύνδεσμος σε μία νέα σελίδα στην οποία προβάλλονται λεπτομερώς και με γραφικό τρόπο τα σημεία πρόσδεσης κάθε microRNA.

2.2.6 PITA

Η εφαρμογή αυτή βρίσκεται στο δικτυακό τόπο <http://genie.weizmann.ac.il/pubs/mir07/>. Ο χρήστης έχει τη δυνατότητα να αναζητήσει είτε τους στόχους ενός ή περισσότερων microRNA, είτε στόχους που σχετίζονται με το όνομα ενός ή περισσότερων γονιδίων στο είδος οργανισμού που επιθυμεί. Επιπλέον οι δύο επιλογές αναζήτησης μπορούν να συνδυαστούν με αποτέλεσμα την πραγματοποίηση σύνθετης αναζήτησης.

Εν αντιθέσει με τις προηγούμενες εφαρμογές, τα αποτελέσματα των αναζητήσεων δεν εμφανίζονται σε μία νέα σελίδα, αλλά οργανώνονται σε μορφή αρχείου. Με το που πραγματοποιηθεί η αναζήτηση και επιλεγούν τα κατάλληλα δεδομένα, ανοίγει μια νέα σελίδα, η οποία περιέχει έναν ενεργό σύνδεσμο για την ανάκτηση του προαναφερθέντος

αρχείου. Με τον τρόπο αυτό ο χρήστης έχει τη δυνατότητα να αποθηκεύσει το αρχείο στον προσωπικό του υπολογιστή. Για κάθε στόχο, αρχικά αναγράφεται το είδος οργανισμού, το όνομα του γονιδίου και του microRNA με τα οποία σχετίζεται η αναζήτηση, το σκορ του στόχου και τέλος κάποιες πιο εξειδικευμένες πληροφορίες για αυτόν.

2.2.7 RNA22

Η εφαρμογή διαφέρει σε σχέση με τις προηγούμενες, ως προς την εισαγωγή των όρων για αναζήτηση. Βρίσκεται στο δικτυακό τόπο http://cbcsrv.watson.ibm.com/rna22_targets.html. Ο χρήστης δεν εισάγει πλέον τον όνομα του microRNA ή το όνομα του mRNA σχετικά με τα όποια αναζητήσει στόχους, αλλά την πλήρη αλληλουχία βάσεων είτε του microRNA, είτε του στόχου σε μορφή FASTA. Όπως εύκολα μπορεί να αντιληφθεί κανείς αυτή είναι η δυσκολία χρήσης της εφαρμογής. Επιπλέον σε αυτήν την εφαρμογή, δίνεται η δυνατότητα στον χρήστη να αναζητήσει στόχους για microRNA, οι οποίοι δεν περιλαμβάνονται σε κάποια βάση δεδομένων. Ο λόγος που αυτό συμβαίνει ότι ο εντοπισμός των στόχων γίνεται τη στιγμή της αναζήτησης με βάση την δοσμένη αλληλουχία βάσεων και συνεπώς δεν είναι απαραίτητο να γίνεται ανάκληση αποτελεσμάτων από τη βάση δεδομένων.

2.2.8 Σύγκριση δικτυακών εφαρμογών

Παρακάτω παρατίθενται συγκριτικοί πίνακες για τα κυριότερα χαρακτηριστικά των εφαρμογών που παρουσιάστηκαν μέχρι τώρα.

Ο Πίνακας 1 αφορά στη σύγκριση των όρων προς αναζήτηση για κάθε εφαρμογή. Τα χαρακτηριστικά που απεικονίζονται στον πίνακα είναι το όνομα ενός ή περισσότερων microRNA, το όνομα ενός ή περισσότερων γονιδίων, ο κωδικός αριθμός ENSEMBL ενός microRNA, το όνομα ενός microRNA που σχετίζεται με το όνομα ενός γονιδίου, το όνομα ενός mRNA, ο καθορισμός μια συγκεκριμένης αλληλουχίας βάσεων και τέλος ο καθορισμός του βιολογικού είδους.

Ο Πίνακας 2 αφορά στη σύγκριση των λειτουργικών χαρακτηριστικών των εφαρμογών. Τα χαρακτηριστικά που απεικονίζονται είναι η εισαγωγή των όρων προς αναζήτηση σε μία ή πολλαπλές φόρμες, η λεπτομερής προβολή των αποτελεσμάτων, η δυνατότητα σύνθετης αναζήτησης, η γραφική απεικόνιση των αποτελεσμάτων, ο καθορισμός κριτηρίων τα οποία πρέπει να πληροί κάποιο συγκεκριμένο mRNA, η προβολή αποτελεσμάτων σε αρχείο και τέλος η δυνατότητα σύνδεσης με τη δικτυακή βιολογική βάση PubMed.

	miRNA	> 1 miRNA	Όνομα γονιδίου	> 1 Όνομα γονιδίου	ENSEMBL	Σύνθετη αναζήτηση	mRNA	Αλληλουχία Βάσεων	Είδος
DIANA	✓		✓		✓	✓		✓	
miRanda	✓	✓	✓				✓		✓
MicroCosm	✓		✓		✓	✓			✓
EIMMo	✓	✓				✓			✓
TargetScan	✓	✓	✓			✓			✓
PITA	✓	✓	✓	✓		✓			✓
RNA22								✓	

Πίνακας 1: Σύγκριση των όρων προς αναζήτηση για κάθε εφαρμογή.

	Κοινή φόρμα	Πολλαπλές φόρμες	Λεπτομερής Προβολή	Γραφική Απεικόνιση	Κριτήρια mRNA	Στόχοι σε ιστοσελίδα	Στόχοι σε αρχείο	Σύνδεση με PubMed
DIANA		✓	✓			✓	✓	✓
miRanda		✓				✓		
MicroCosm	✓		✓	✓		✓	✓	
EIMMo	✓		✓		✓	✓		✓
TargetScan		✓	✓	✓		✓		
PITA		✓					✓	
RNA22		✓				✓		

Πίνακας 2: Σύγκριση επιπλέον χαρακτηριστικών για κάθε εφαρμογή.

3

Ανάλυση Απαιτήσεων Συστήματος

Στο κεφάλαιο αυτό αρχικά παρουσιάζεται η αρχιτεκτονική της εφαρμογής που υλοποιήθηκε με βάση τα υποσυστήματα από τα οποία αποτελείται. Στη συνέχεια γίνεται μία σύντομη περιγραφή της λειτουργίας που επιτελεί καθένα από αυτά τα υποσυστήματα. Στο τέλος του κεφαλαίου αναλύονται οι περιπτώσεις χρήσης όλης της εφαρμογής και παρουσιάζεται το μοντέλο οντοτήτων συσχετίσεων που αναπτύχθηκε για την υλοποίηση αυτής.

3.1 Αρχιτεκτονική

Η υλοποίηση του συστήματος της παρούσας διπλωματικής εργασίας πραγματοποιήθηκε με βάση σύγχρονες αρχές της Τεχνολογίας Λογισμικού. Η παρούσα εφαρμογή αποτελείται από τα εξής υποσυστήματα:

- **Διαχείριση χρηστών:** το συγκεκριμένο υποσύστημα είναι υπεύθυνο για τη διαχείριση των χρηστών της εφαρμογής και το ανέβασμα αρχείων συγκεκριμένου τύπου. Συγκεκριμένα, στην περίπτωση διαχείρισης χρηστών δίνεται η δυνατότητα στο διαχειριστή να εισάγει έναν νέο χρήστη στο σύστημα, να διαγράψει κάποιον από τους υπάρχοντες χρήστες και να τροποποιήσει κάποιο ή κάποια από τα στοιχεία του. Όσον αφορά στην περίπτωση ανεβάσματος αρχείων, ο διαχειριστής μπορεί να

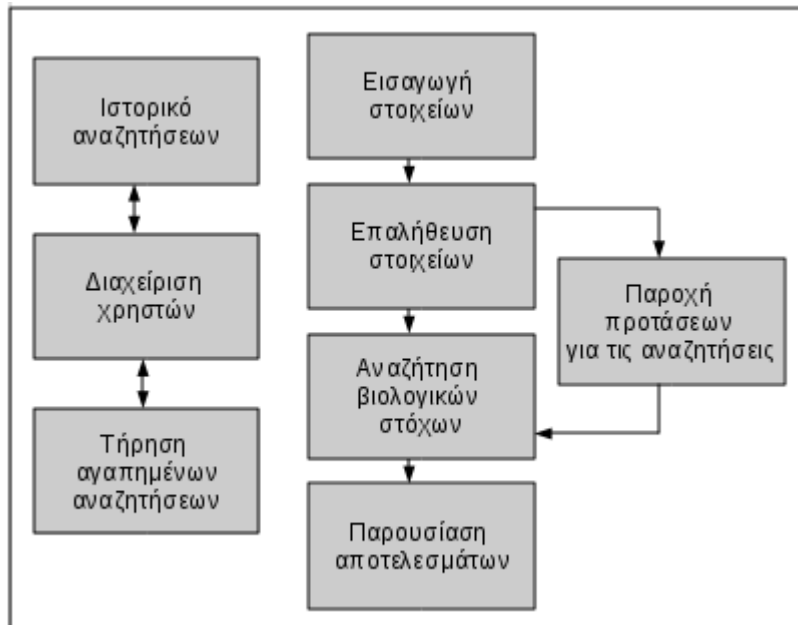
επιλέξει ένα κατάλληλα διαμορφωμένο αρχείο το οποίο περιέχει δεδομένα πρόβλεψης με βάση άλλους αλγορίθμους, και να το ανεβάσει στο σύστημα ενημερώνοντας τον κατάλληλο πίνακα της βάσης. Το συγκεκριμένο υποσύστημα υλοποιήθηκε αποκλειστικά στα πλαίσια της παρούσας διπλωματικής εργασίας.

- **Παροχή προτάσεων σε αναζητήσεις:** το συγκεκριμένο υποσύστημα τίθεται σε λειτουργία στην περίπτωση που ο χρήστης πραγματοποιήσει αναζήτηση με όρους που δεν αντιστοιχούν σε έγκυρες ονομασίες ονομάτων (όνομα microRNA ή όνομα γονιδίου). Σε αυτήν την περίπτωση παρέχονται νέοι προτεινόμενοι όροι προς αναζήτηση στο χρήστη, ανάλογα με το είδος του όρου για τον οποίο αναζητά πληροφορίες. Για κάθε νέο όρο δίνεται δυνατότητα επιλογής για την πραγματοποίηση νέας αναζήτησης με βάση αυτόν τον όρο. Το υποσύστημα αυτό επανασχεδιάστηκε και δημιουργήθηκε αποκλειστικά στα πλαίσια της παρούσας διπλωματικής εργασίας.
- **Εισαγωγή στοιχείων:** το συγκεκριμένο υποσύστημα είναι υπεύθυνο για την εμφάνιση ειδικής φόρμας στο χρήστη, στην οποία μπορεί να εισάγει τους όρους αναζήτησης για τους οποίους θέλει να εντοπίσει βιολογικούς στόχους. Οι όροι αυτοί μπορεί να σχετίζονται είτε με ονομασίες microRNA είτε με ονομασίες γονιδίων. Επιπλέον, παρέχει τη δυνατότητα στο χρήστη να παραμετροποιήσει την αναζήτηση του. Το υποσύστημα αυτό δημιουργήθηκε στην προηγούμενη έκδοση της εφαρμογής αλλά τροποποιήθηκε στην διάρκεια ανάπτυξης της παρούσας έκδοσης προκειμένου να επιτευχθεί μεγαλύτερη απλότητα και ευκολία για το χρήστη.
- **Επαλήθευση στοιχείων:** το συγκεκριμένο υποσύστημα αναλαμβάνει να πιστοποιήσει αν τα στοιχεία που εισήγαγε ένας χρήστης στη φόρμα αναζήτησης, είναι έγκυρα και σε αυτήν την περίπτωση πραγματοποιεί την αναζήτηση με βάση τις ορισμένες παραμέτρους. Σε αντίθετη περίπτωση ενημερώνει το χρήστη με κατάλληλο μήνυμα λάθους, παροτρύνοντας τον να προσπαθήσει ξανά. Το υποσύστημα δημιουργήθηκε στην προηγούμενη έκδοση της εφαρμογής αλλά τροποποιήθηκε κατά τέτοιο τρόπο ώστε να υπάρχει συμβατότητα με τα νέα υποσυστήματα.
- **Αναζήτηση βιολογικών στόχων:** το συγκεκριμένο υποσύστημα πραγματοποιεί την αναζήτηση βιολογικών στόχων ανάλογα με το είδος ονόματος (όνομα microRNA ή όνομα γονιδίου) που εισήγαγε ο χρήστης. Επιπλέον είναι υπεύθυνο για την επιλογή των δεδομένων που πληρούν τα κριτήρια αναζήτησης και παρουσίαση των αποτελεσμάτων στην κατάλληλη μορφή. Ταυτόχρονα, δίνεται η δυνατότητα στο

χρήστη να δει μία λίστα προγραμμάτων που εντόπισαν τους ίδιους βιολογικούς στόχους με το πρόγραμμα της εφαρμογής. Το υποσύστημα δημιουργήθηκε στην προηγούμενη έκδοση της εφαρμογής αλλά επεκτάθηκε κατά τη διάρκεια ανάπτυξης της παρούσας έκδοσης.

- **Ιστορικό αναζητήσεων:** το συγκεκριμένο υποσύστημα είναι υπεύθυνο για τη διατήρηση του ιστορικού των αναζητήσεων που πραγματοποίησε ο κάθε χρήστης καθώς και την εμφάνισή των τελευταίων 10 αναζητήσεων του σε αυτόν. Για κάθε αναζήτηση παρέχεται δυνατότητα επιλογής για την επανάληψη αυτής. Επιπλέον υπάρχει η δυνατότητα διαγραφής του ιστορικού ενός χρήστη από τον ίδιο. Το υποσύστημα αυτό δημιουργήθηκε αποκλειστικά κατά την ανάπτυξη της προηγούμενης έκδοσης της εφαρμογής και τροποποιήθηκε κατά τέτοιο τρόπο έτσι ώστε να υπάρχει συμβατότητα με τα νέα υποσυστήματα.
- **Τήρηση αγαπημένων αναζητήσεων:** το συγκεκριμένο υποσύστημα είναι υπεύθυνο για την διατήρηση και εμφάνιση των αναζητήσεων που ο χρήστης επέλεξε να αποθηκεύσει. Οι αποθηκευμένες αναζητήσεις είναι προσβάσιμες από το χρήστη οποιαδήποτε στιγμή το επιθυμήσει, δίνοντάς του τη δυνατότητα διαγραφής και τροποποίησης αυτών. Το υποσύστημα αυτό δημιουργήθηκε αποκλειστικά κατά την ανάπτυξη της προηγούμενης έκδοσης της εφαρμογής και τροποποιήθηκε κατά τέτοιο τρόπο έτσι ώστε να υπάρχει συμβατότητα με τα νέα υποσυστήματα.

Στην Εικόνα 2 απεικονίζονται τα επιμέρους υποσυστήματα και οι σχέσεις αλληλεπίδρασης μεταξύ αυτών.



Εικόνα 2: Σχηματική απεικόνιση των υποσυστημάτων που απαρτίζουν την εφαρμογή της παρούσας διπλωματικής εργασίας.

Η ανάπτυξη του συστήματος πραγματοποιήθηκε έτσι ώστε να πληροί τα κριτήρια τόσο της υψηλής συνεκτικότητας, όσο και της χαμηλής σύζευξης.

Η συνεκτικότητα ενός συστήματος αποτελεί ένα δείκτη με βάση τον οποίο μπορεί να ελεγχθεί αν τα επιμέρους υποσυστήματα επιτελούν καλά ορισμένες λειτουργίες. Επιπλέον όλες οι λειτουργίες κάθε υποσυστήματος οφείλουν να είναι άμεσα σχετιζόμενες με το σκοπό λειτουργίας αυτού, και όσο το δυνατόν περισσότερο αλληλοεξαρτώμενες. Η εμφάνιση υψηλής συνεκτικότητας σε ένα σύστημα αποτελεί εγγύηση σταθερότητας, αξιοπιστίας, επαναχρησιμοποίησης, εύκολης κατανόησης και καλύτερων δυνατοτήτων ελέγχου αυτού.

Από την άλλη μεριά, η σύζευξη αποτελεί ένα δείκτη ελέγχου για το κατά πόσο τα επιμέρους υποσυστήματα εμφανίζουν εξαρτήσεις μεταξύ τους. Όπως εύκολα μπορεί να συμπεράνει κανείς, η χαμηλή σύζευξη παρέχει δυνατότητες εύκολης προσαρμογής κάθε υποσυστήματος ανάλογα με τις εκάστοτε ανάγκες χωρίς να απαιτούνται μεγάλες αλλαγές στα υπόλοιπα.

Τόσο η υψηλή συνεκτικότητα, όσο και η χαμηλή σύζευξη αποτελούν δείκτη καλής σχεδίασης, και συνεπώς βέλτιστης λειτουργικότητας.

3.2 Περιγραφή Λειτουργιών

Το αντικείμενο της παρούσας εφαρμογής είναι η διευκόλυνση των χρηστών στην αναζήτηση της επιθυμητής πληροφορίας. Για το λόγο αυτό, κάθε επιμέρους υποσύστημα έχει αναπτυχθεί με σκοπό να επιτελεί κάθε λειτουργία του με απλό, αποδοτικό και φιλικό προς το χρήστη τρόπο.

Κάθε χρήστης έχει τη δυνατότητα να προσαρμόσει την αναζήτηση που θέλει να πραγματοποιήσει με βάση τα δικά του προσωπικά κριτήρια. Παράλληλα, του δίνεται η δυνατότητα να “φυλάξει” το υλικό αυτών των αναζητήσεων έτσι ώστε μελλοντικά να μπορεί να το ανακτήσει όποτε χρειαστεί. Όλες αυτές οι δυνατότητες κάνουν το σύστημα εξατομικευμένο και ιδιαίτερα χρηστικό για τον απλό χρήστη.

Επιπλέον οι δυνατότητες διαχείρισης λογαριασμών χρηστών, παρέχουν στους διαχειριστές τους συστήματος ένα απλό, γρήγορο και άμεσο εργαλείο στην κατεύθυνση του ελέγχου όλων των τύπων χρηστών της εφαρμογής.

Από την άλλη μεριά, η δυνατότητα ανεβάσματος αρχείων από τους διαχειριστές επιτρέπει την ενημέρωση των πληροφοριών κατά τέτοιο τρόπο ώστε οι χρήστες να είναι σε θέση να έχουν πρόσβαση στα πιο σύγχρονα αποτελέσματα.

Παρακάτω γίνεται μια προσπάθεια για τη σύντομη περιγραφή των επιμέρους υποσυστημάτων με σαφή καθορισμό του τρόπου λειτουργίας καθενός από αυτά.

3.2.1 Διαχείριση χρηστών

Το συγκεκριμένο υποσύστημα μπορεί να επιτελέσει τις παρακάτω λειτουργίες:

1. Κάθε διαχειριστής έχει τη δυνατότητα να δει, σε μορφή λίστας, όλους τους χρήστες της εφαρμογής χωρισμένους σε δύο ομάδες: μία ομάδα των διαχειριστών και άλλη μία των απλών χρηστών. Για κάθε ομάδα εμφανίζεται ανά χρήστη το όνομα του (username), η ηλεκτρονική διεύθυνσή του (email), η επιλογή διαγραφής λογαριασμού και η επιλογή τροποποίησης στοιχείων αυτού. Κάθε ομάδα είναι αλφαβητικά ταξινομημένη.
2. Κάθε διαχειριστής έχει τη δυνατότητα να τροποποιήσει τα στοιχεία οποιουδήποτε χρήστη, συμπεριλαμβανομένου και του εαυτού του. Τα στοιχεία τα οποία μπορούν να μεταβληθούν είναι το όνομα χρήστη (username), το συνθηματικό (password), η ηλεκτρονική διεύθυνση (email) και ο τύπος χρήστη (απλός ή διαχειριστής).

3. Κάθε διαχειριστής έχει τη δυνατότητα να διαγράψει ένα άλλον οποιονδήποτε χρήστη από χρήστη του συστήματος.
4. Κάθε διαχειριστής έχει τη δυνατότητα να προσθέσει ένα νέο χρήστη στο σύστημα ορίζοντας σε ειδική φόρμα το όνομα του χρήστη (username) που θέλει να εισάγει, το συνθηματικό (password), την ηλεκτρονική διεύθυνση (email) και τον τύπο χρήστη (απλός ή διαχειριστής).
5. Κάθε διαχειριστής έχει τη δυνατότητα να ανεβάσει στο σύστημα ένα κατάλληλα διαμορφωμένο αρχείο το οποίο περιέχει πληροφορίες σχετικά με εντοπισμό βιολογικών στόχων με βάση άλλους αλγορίθμους. Μετά το ανέβασμα του αρχείου ενημερώνεται η βάση δεδομένων με τις νέες πληροφορίες.

3.2.2 Παροχή προτάσεων για τους όρους αναζήτησης

Το συγκεκριμένο υποσύστημα, τίθεται σε λειτουργία εφόσον οι όροι προς αναζήτηση που εισάγει ο χρήστης δεν αντιστοιχούν σε έγκυρες ονομασίες microRNA ή γονιδίων. Οι ονομασίες αυτές μπορεί να ανταποκρίνονται στα πραγματικά ονόματα microRNA ή γονιδίων, ή σε ειδικούς κωδικούς αυτών. Στην περίπτωση αναζήτησης στόχων με βάση ένα microRNA, ο χρήστης έχει τη δυνατότητα να εισάγει είτε το όνομά του είτε τον κωδικό MIMA αυτού. Ο κωδικός MIMA είναι ένας κωδικός που έχει δωθεί σε κάθε microRNA, το ταυτοποιεί μοναδικά και είναι της μορφής “MIMAxxxxxxx” (όπου x ένα ψηφίο). Στην περίπτωση αναζήτησης με βάση ένα γονίδιο, ο χρήστης μπορεί να εισάγει είτε το όνομα αυτού, είτε ένας από τους κωδικούς ENSEMBL, TRANSCRIPT, ή Refseq. Ο κωδικός ENSEMBL είναι ο κωδικός που ταυτοποιεί ένα γονίδιο μοναδικά, αντιστοιχεί σε μία μοναδική εγγραφή στη δικτυακή βάση δεδομένων ENSEMBL και είναι της μορφής “ENSGxxxxxxxxxxxx” ή “ENSMUSGxxxxxxxxxxxx” (όπου x ένα ψηφίο). Ο κωδικός TRANSCRIPT είναι ο κωδικός που ταυτοποιεί ένα μετάγραφο ενός γονιδίου μοναδικά, αντιστοιχεί σε μια μοναδική εγγραφή στη δικτυακή βάση δεδομένων ENSEMBL και είναι της μορφής “ENSTxxxxxxxxxxxx” ή “ENSMUSTxxxxxxxxxxxx” (όπου x ένα ψηφίο). Τέλος, ο κωδικός Refseq είναι κωδικός που ταυτοποιεί ένα γονίδιο μοναδικά, αντιστοιχεί σε μία μοναδική εγγραφή στη δικτυακή βάση δεδομένων Refseq και είναι της μορφής “XM_y” ή “XR_y” ή “NM_y” ή “NR_y” (όπου y ένα μοναδικός αριθμός). Οι λειτουργίες που επιτελεί το σύστημα παροχής προτάσεων για τους όρους αναζήτησης είναι:

1. Αν ο χρήστης εισάγει στην φόρμα αναζήτησης ένα microRNA που δεν αντιστοιχεί σε έγκυρη ονομασία πχ “hsa-let-go”, το σύστημα πρέπει να καταλαβαίνει με κατάλληλη

τεχνική ότι ο χρήστης προσπαθεί να ανακτήσει πληροφορίες για ένα microRNA. Τότε θα καλεί έναν ειδικό αλγόριθμο και θα εντοπίζει προσεγγιστικά έγκυρες ονομασίες microRNA που σχετίζονται στον βέλτιστο δυνατό βαθμό με την αναζήτηση του χρήστη. Τα έγκυρα ονόματα θα παρουσιάζονται στο χρήστη με δυνατότητες επιλογής.

2. Αν ο χρήστης εισάγει στην φόρμα αναζήτησης ένα κωδικό MIMA ενός microRNA ή ένα κωδικό ENSEMBL ή TRANSCRIPT ή Refseq ενός γονιδίου και ο κωδικός αυτός δεν αντιστοιχεί σε έγκυρο κωδικό, τότε το σύστημα πρέπει να καταλαβαίνει τι είδους κωδικό εισήγαγε ο χρήστης. Σε αυτήν την περίπτωση, θα καλεί τον ειδικό αλγόριθμο που εντοπίζει προσεγγιστικά έγκυρους κωδικούς του ίδιου είδους με αυτό που εισήγαγε ο χρήστης. Στη συνέχεια η λίστα με τους έγκυρους κωδικούς παρουσιάζεται στο χρήστη με δυνατότητες επιλογής.

3.2.3 Εισαγωγή στοιχείων

Το συγκεκριμένο υποσύστημα επιτελεί τις εξής λειτουργίες.

1. Ο χρήστης εισάγει στην φόρμα αναζήτησης είτε το όνομα ή τον κωδικό MIMA ενός microRNA, είτε το όνομα ή έναν από τους κωδικούς ENSEMBL και TRANSCRIPT και Refseq ενός γονιδίου. Στη συνέχεια καταλαβαίνει με ειδικά πλην όμως απλά κριτήρια το είδος ονόματος για το οποίο ο χρήστης θέλει να πραγματοποιήσει αναζήτηση και συνεπώς ορίζει τις παραμέτρους της αναζήτησης.
2. Εφόσον πραγματοποιηθεί αναζήτηση για ένα από τα παραπάνω ονόματα, ο χρήστης έχει τη δυνατότητα να ορίσει την παράμετρο KEGG pathway, θέτοντας στην ειδική φόρμα την επιθυμητή τιμή.
3. Εφόσον πραγματοποιηθεί αναζήτηση για ένα από τα ονόματα που αναφέρθηκαν προηγουμένως, ο χρήστης έχει τη δυνατότητα να ορίσει τον ελάχιστο βαθμό βιολογικού στόχου, θέτοντας στην ειδική φόρμα την επιθυμητή τιμή, δηλαδή ένα δεκαδικό αριθμό που κυμαίνεται μεταξύ 0.20 και 1.12.

3.2.4 Επαλήθευση στοιχείων

Το συγκεκριμένο υποσύστημα έχει τη μοναδική εξής λειτουργία:

1. Πιστοποιεί αν ο χρήστης ορίζει με έγκυρο τρόπο τις παραμέτρους αναζήτησης. Σε περίπτωση που οι παράμετροι δεν είναι έγκυρες, το σύστημα καλεί τον χρήστη να δοκιμάσει ξανά με κατάλληλο μήνυμα σφάλματος.

3.2.5 Αναζήτηση βιολογικών στόχων

Το συγκεκριμένο υποσύστημα επιτελεί την εξής λειτουργία:

1. Εντοπίζει τους βιολογικούς στόχους οι οποίοι πληρούν τα κριτήρια που έχουν οριστεί από το χρήστη, και για καθένα από αυτούς αναζητούνται οι περιοχές πρόσδεσης (targets). Στη συνέχεια κατασκευάζει τον πίνακα των αποτελεσμάτων με δομημένο τρόπο, έχοντας στο ένα επίπεδο τους βιολογικούς στόχους και στο άλλο τις περιοχές πρόσδεσης.

3.2.6 Παρουσίαση αποτελεσμάτων

Το συγκεκριμένο υποσύστημα επιτελεί τις εξής λειτουργίες:

1. Αν ο χρήστης πραγματοποιεί αναζήτηση με βάση το όνομα ενός microRNA ή με βάση τον κωδικό MIMA, παρουσιάζει τα αποτελέσματα στην ανάλογη μορφή, τηρώντας τη συσχέτιση βιολογικοί στόχοι-περιοχές πρόσδεσης. Επιπλέον παρέχεται η δυνατότητα στο χρήστη να δει τη λίστα των προγραμμάτων που προέβλεψαν τους ίδιους στόχους με το πρόγραμμα της παρούσας εφαρμογής.
2. Αν ο χρήστης πραγματοποιεί αναζήτηση με βάση το όνομα ενός γονιδίου ή με βάση τον κωδικό ENSEMBL ή TRANSCRIPT ή Refseq, τότε παρουσιάζει τα αποτελέσματα στην ανάλογη μορφή με διατήρηση της συσχέτισης βιολογικοί στόχοι-περιοχές πρόσδεσης. Επιπλέον παρέχεται η δυνατότητα στο χρήστη να δει τη λίστα των προγραμμάτων που προέβλεψαν τους ίδιους στόχους με το πρόγραμμα της παρούσας εφαρμογής.
3. Αν ο χρήστης πραγματοποιεί σύνθετη αναζήτηση, παρουσιάζει τα αποτελέσματα στην ανάλογη μορφή, με διατήρηση της συσχέτισης βιολογικοί στόχοι-περιοχές πρόσδεσης. Επιπλέον παρέχεται η δυνατότητα στο χρήστη να δει τη λίστα των προγραμμάτων που προέβλεψαν τους ίδιους στόχους με το πρόγραμμα της παρούσας εφαρμογής.

3.2.7 Ιστορικό αναζητήσεων

Το συγκεκριμένο υποσύστημα πραγματοποιεί την εξής λειτουργία:

1. Μετά από κάθε είδους αναζήτηση του χρήστη, απλή ή σύνθετη, η αναζήτηση αυτή καταγράφεται. Αυτό γίνεται ανεξάρτητα της επιλογής του χρήστη και του τύπου του και δεν υπάρχουν χρονικοί ή χωρικοί περιορισμοί. Η καταγραφή αυτή γίνεται ανά

λογαριασμό χρήστη, έτσι ώστε κάθε χρήστης να μπορεί να έχει πρόσβαση μόνο στις τελευταίες αναζητήσεις που πραγματοποίησε αυτός.

3.2.8 Τήρηση αγαπημένων αναζητήσεων

Το συγκεκριμένο υποσύστημα επιτελεί τις εξής λειτουργίες:

1. Εφόσον ο χρήστης πραγματοποιήσει μια οποιουδήποτε τύπου αναζήτηση, έχει την επιλογή να αποθηκεύσει στον προσωπικό του χώρο τα αποτελέσματα της αναζήτησης αυτής.
2. Αναλαμβάνει να εμφανίζει στον εκάστοτε χρήστη τα αποθηκευμένα αποτελέσματα σε μορφή λίστας
3. Δίνει στο χρήστη τη δυνατότητα προβολής καθεμίας από τις επιμέρους αναζητήσεις.
4. Δίνει στο χρήστη τη δυνατότητα επεξεργασίας καθεμίας από τις επιμέρους αναζητήσεις.
5. Δίνει στο χρήστη τη δυνατότητα διαγραφής καθεμίας από τις επιμέρους αναζητήσεις.

3.3 Περιπτώσεις χρήσης

Κατά την ανάπτυξη μιας εφαρμογής λογισμικού, είναι απαραίτητο να γίνεται λεπτομερής ανάλυση των απαιτήσεων που αυτή πρέπει να πληροί. Αυτό μπορεί να γίνει με τη βοήθεια μεθόδων της Τεχνολογίας Λογισμικού, και ειδικότερα με τις περιπτώσεις χρήσης.

Η μέθοδος αυτή επιτρέπει την λεπτομερή καταγραφή όλων των ενεργειών που μπορούν να επιτελέσουν οι χρήστες ενός συστήματος. Επιπλέον, καταγράφεται πληροφορία σχετικά με τη συμπεριφορά του συστήματος σε κάθε μία από αυτές τις ενέργειες. Η εφαρμογή της μεθόδου συνίσταται σε δύο μέρη: εύρεση των δραστών (actors) και εύρεση των ενεργειών που μπορεί να επιτελέσει κάθε τύπος δράστη.

Δράστης ονομάζεται κάθε τύπος χειριστή ο οποίος αλληλεπιδρά με το σύστημα. Στην περίπτωση της παρούσας εφαρμογής οι δράστες είναι δύο, ένας για κάθε τύπο χρήστη. Συνεπώς, έχουμε ένα δράστη που αντιπροσωπεύει κάθε απλό χρήστη της εφαρμογής και έναν δράστη που αντιπροσωπεύει κάθε διαχειριστή αυτής.

Οι ενέργειες που μπορούν να επιτελέσουν και οι δύο τύποι δραστών είναι κοινές, με τη διαφορά ότι οι διαχειριστές μπορούν να επιτελέσουν κάποιες επιπλέον ενέργειες όπως είναι φυσικό. Ωστόσο εδώ θα γίνει παρουσίαση μόνο των περιπτώσεων χρήσης της νέας έκδοσης

της εφαρμογής. Για τις περιπτώσεις χρήστης όλης της εφαρμογής πληροφορίες μπορούν να αναζητηθούν και στην αναφορά [7].

Παρακάτω γίνεται συνοπτική παρουσίαση αυτών των ενεργειών και στη συνέχεια ανάλυση καθεμίας από αυτές.

- Αναζήτηση microRNA
- Παροχή προτάσεων στην αναζήτηση microRNA
- Αναζήτηση γονιδίου
- Παροχή προτάσεων στην αναζήτηση γονιδίου
- Εμφάνιση προβλέψεων με βάση άλλους αλγορίθμους
- Προβολή λίστας χρηστών
- Εισαγωγή νέων χρηστών στο σύστημα
- Διαγραφή χρηστών από το σύστημα
- Τροποποίηση στοιχείων χρηστών του συστήματος
- Εισαγωγή αρχείου στόχων με βάση άλλους αλγορίθμους

Από τις ενέργειες αυτές μόνο οι πέντε πρώτες είναι κοινές και για τους δύο τύπους δραστών, ενώ οι πέντε τελευταίες μπορούν να επιτελεστούν αποκλειστικά από τους διαχειριστές του συστήματος.

3.3.1 Περίπτωση χρήσης “Λίστα χρηστών”

<p>Τίτλος περίπτωσης χρήστης: Λίστα Χρηστών</p> <p>Σύντομη περιγραφή: Ο διαχειριστής βλέπει τη λίστα των χρηστών που απαρτίζουν το σύστημα, χωρισμένους σε δύο ομάδες.</p> <p>Ροή γεγονότων</p> <p>Βασική ροή</p> <ol style="list-style-type: none">1. Ο διαχειριστής έχει κάνει σύνδεση στο πάνελ διαχείρισης χρηστών.2. Ο διαχειριστής επιλέγει την προβολή της λίστας χρηστών από τον κατάλληλο σύνδεσμο.3. Εμφανίζεται η λίστα με τις δύο ομάδες χρηστών αλφαβητικά ταξινομημένες, τις
--

κυριότερες πληροφορίες για κάθε χρήστη και επιλογές διαγραφής και τροποποίησης στοιχείων.

Εναλλακτικές ροές

Δεν υπάρχουν εναλλακτικές ροές για τη συγκεκριμένη περίπτωση χρήσης.

Μη λειτουργικές απαιτήσεις

1. Ο τρόπος παρουσίασης των χρηστών οφείλει να είναι καλαίσθητος και καλά δομημένος.
2. Οι πληροφορίες που θα εμφανίζονται για κάθε χρήστη δεν πρέπει να είναι ευαίσθητα προσωπικά δεδομένα, πχ συνηματικό.
3. Η λίστα χρηστών οφείλει να είναι ταξινομημένη για διευκόλυνση του διαχειριστή.

Κατάσταση εισόδου:

1. Ο διαχειριστής πρέπει να έχει κάνει σύνδεση στο πάνελ διαχείρισης χρηστών από τον ειδική φόρμα.
2. Ο χρήστης πρέπει να ανήκει στη λίστα των διαχειριστών που έχουν το δικαίωμα να δουν αυτή τη λίστα χρηστών.

Κατάσταση εξόδου:

Δεν υπάρχει.

3.3.2 Περίπτωση χρήσης “Εισαγωγή νέων χρηστών”

Τίτλος περίπτωσης χρήστης:

Εισαγωγή νέων χρηστών

Σύντομη περιγραφή:

Ο διαχειριστής εισάγει στο σύστημα ένα νέο χρήστη από ειδική φόρμα του πάνελ διαχείρισης χρηστών.

Ροή γεγονότων

Βασική ροή

1. Ο διαχειριστής έχει κάνει σύνδεση στο πάνελ διαχείρισης χρηστών.
2. Ο διαχειριστής επιλέγει την προβολή της λίστας χρηστών.
3. Ο διαχειριστής επιλέγει από τη νέα σελίδα τον σύνδεσμο για την εισαγωγή νέων χρηστών.

4. Ο διαχειριστής συμπληρώνει στη νέα φόρμα τα στοιχεία του νέου χρήστη: όνομα (username), συνθηματικό (password), επανάληψη συνθηματικού, ηλεκτρονική διεύθυνση (email) και τύπο χρήστη (απλός ή διαχειριστής).
5. Ο διαχειριστής πατάει το κουμπί της υποβολής.

Εναλλακτικές ροές

Εναλλακτική ροή 1

1. Ο διαχειριστής εισήγαγε όνομα χρήστη που υπάρχει ήδη στο σύστημα.
2. Εμφανίζεται μήνυμα λάθους που ενημερώνει τον διαχειριστή.

Εναλλακτική ροή 2

1. Ο διαχειριστής επανέλαβε λάθος το συνθηματικό.
2. Εμφανίζεται μήνυμα λάθους που ενημερώνει το χρήστη.

Εναλλακτική ροή 3

1. Ο διαχειριστής εισήγαγε στην ηλεκτρονική διεύθυνση συμβολοσειρά που δεν αντιστοιχεί σε έγκυρη μορφή ηλεκτρονικής διεύθυνσης.
2. Εμφανίζεται μήνυμα λάθους που ενημερώνει το διαχειριστή.

Εναλλακτική ροή 4

1. Ο διαχειριστής έχει αφήσει κενό κάποιο από τα πεδία που πρέπει να συμπληρωθούν.
2. Εμφανίζεται κατάλληλο μήνυμα λάθους που ενημερώνει το διαχειριστή.

Μη λειτουργικές απαιτήσεις

Δεν υπάρχουν μη λειτουργικές απαιτήσεις για τη συγκεκριμένη περίπτωση χρήσης.

Κατάσταση εισόδου:

1. Ο διαχειριστής πρέπει να έχει κάνει σύνδεση στο πάνελ διαχείρισης χρηστών από τον ειδική φόρμα.
2. Ο χρήστης πρέπει να ανήκει στη λίστα των διαχειριστών που έχουν το δικαίωμα να δουν αυτή τη λίστα χρηστών.

Κατάσταση εξόδου:

1. Δημιουργείται ένα νέος λογαριασμός χρήστης με στοιχεία αυτά που εισήγαγε ο διαχειριστής.

3.3.3 Περίπτωση χρήσης “Διαγραφή χρηστών”

Τίτλος περίπτωσης χρήστη:

Διαγραφή χρηστών

Σύντομη περιγραφή:

Ο διαχειριστής εισάγει επιλέγει τη διαγραφή ενός χρήστη από τη λίστα χρηστών του συστήματος.

Ροή γεγονότων

Βασική ροή

1. Ο διαχειριστής πρέπει να έχει κάνει σύνδεση στο πάνελ διαχείρισης χρηστών από τον ειδική φόρμα.
2. Ο διαχειριστής επιλέγει την προβολή της λίστας χρηστών.
3. Ο διαχειριστής επιλέγει το σύνδεσμο “Διαγραφή” δίπλα από το όνομα χρήστη που θέλει να διαγράψει.
4. Εμφανίζεται μια νέα φόρμα που ρωτά τον διαχειριστή αν θέλει σίγουρα να διαγράψει τον χρήστη από το σύστημα.
5. Ο διαχειριστής κάνει κλικ στην επιλογή “ναι”.

Εναλλακτικές ροές

Εναλλακτική ροή 1

1. Ο διαχειριστής κάνει κλικ στην επιλογή “όχι” μετά από την ερώτηση αν είναι σίγουρος για τη διαγραφή του συγκεκριμένου χρήστη.

Μη λειτουργικές απαιτήσεις

1. Το σύστημα οφείλει να ενημερώνει τη βάση δεδομένων σχετικά με το ποιοι χρήστες διαγράφηκαν και συγκεκριμένα να αφαιρεί τις αντίστοιχες εγγραφές από τον κατάλληλο πίνακα.
2. Ο χρήστης που διαγράφηκε δεν θα πρέπει να εμφανίζεται ξανά στη λίστα χρηστών.

Κατάσταση εισόδου:

1. Ο διαχειριστής πρέπει να έχει κάνει σύνδεση στο πάνελ διαχείρισης χρηστών από τον ειδική φόρμα.
2. Ο χρήστης πρέπει να ανήκει στη λίστα των διαχειριστών που έχουν το δικαίωμα να δουν αυτή τη λίστα χρηστών.

Κατάσταση εξόδου:

Ανανεώνεται η λίστα χρηστών.

3.3.4 Περίπτωση χρήσης “Τροποποίηση στοιχείων χρηστών”**Τίτλος περίπτωσης χρήστη:**

Τροποποίηση στοιχείων χρηστών

Σύντομη περιγραφή:

Ο διαχειριστής τροποποιεί τα στοιχεία ενός χρήστη του συστήματος.

Ροή γεγονότων**Βασική ροή**

1. Ο διαχειριστής πρέπει να έχει κάνει σύνδεση στο πάνελ διαχείρισης χρηστών από τον ειδική φόρμα.
2. Ο διαχειριστής επιλέγει την προβολή της λίστας χρηστών.
3. Ο διαχειριστής επιλέγει τον σύνδεσμο “Τροποποίηση” δίπλα από το όνομα χρήστη που επιθυμεί.
4. Ο διαχειριστής τροποποιεί όποιο ή όποια από τα πεδία επιθυμεί.
5. Ο διαχειριστής πατάει το κουμπί της υποβολής.

Εναλλακτικές ροές**Εναλλακτική ροή 1**

1. Ο διαχειριστής τροποποιεί κάποιο πεδίο αφήνοντάς το κενό.
2. Εμφανίζεται μήνυμα σφάλματος που ενημερώνει το διαχειριστή.

Εναλλακτική ροή 2

1. Ο διαχειριστής τροποποιεί το συνθηματικό, χωρίς τα πεδία του συνθηματικού και της επανάληψης αυτού να συμφωνούν.
2. Εμφανίζεται μήνυμα σφάλματος που ενημερώνει το διαχειριστή.

Εναλλακτική ροή 3

1. Ο διαχειριστής εισάγει στο όνομα χρήστη, το όνομα ενός χρήστη που υπάρχει ήδη.
2. Εμφανίζεται μήνυμα σφάλματος που ενημερώνει το διαχειριστή.

Μη λειτουργικές απαιτήσεις

1. Το σύστημα οφείλει να ενημερώνει τη βάση δεδομένων σχετικά με το ποιοι χρήστες

τροποποιήθηκαν και συγκεκριμένα να ενημερώνει τις αντίστοιχες εγγραφές από τον κατάλληλο πίνακα.

2. Τα στοιχεία του χρήστη που τροποποιήθηκαν θα πρέπει να εμφανίζονται ανανεωμένα στη λίστα χρηστών.

Κατάσταση εισόδου:

1. Ο διαχειριστής πρέπει να έχει κάνει σύνδεση στο πάνελ διαχείρισης χρηστών από τον ειδική φόρμα.
2. Ο χρήστης πρέπει να ανήκει στη λίστα των διαχειριστών που έχουν το δικαίωμα να δουν αυτή τη λίστα χρηστών.

Κατάσταση εξόδου:

Ανανεώνεται η λίστα χρηστών.

3.3.5 Περίπτωση χρήσης “Εισαγωγή αρχείου στόχων με βάση άλλους αλγορίθμους”

Τίτλος περίπτωσης χρήστης:

Εισαγωγή αρχείου στόχων με βάση άλλους αλγορίθμους

Σύντομη περιγραφή:

Ο διαχειριστής εισάγει στο σύστημα πληροφορίες σχετικές με αλγορίθμους οι οποίοι εντόπισαν τους ίδιους βιολογικούς στόχους με το πρόγραμμα της παρούσας εφαρμογής.

Ροή γεγονότων

Βασική ροή

1. Ο διαχειριστής πρέπει να έχει κάνει σύνδεση στο πάνελ διαχείρισης χρηστών από τον ειδική φόρμα.
2. Ο διαχειριστής επιλέγει το σύνδεσμο για την εισαγωγή στόχων.
3. Ο διαχειριστής επιλέγει ένα κατάλληλα διαμορφωμένο αρχείο από τον υπολογιστή του.
4. Ο διαχειριστής επιλέγει το κουμπί για το ανέβασμα του αρχείου.
5. Ο διαχειριστής ενημερώνεται για την επιτυχή εισαγωγή του αρχείου.

Εναλλακτικές ροές

Δεν υπάρχουν εναλλακτικές ροές για τη συγκεκριμένη περίπτωση χρήσης.

Μη λειτουργικές απαιτήσεις

1. Το σύστημα οφείλει να ενημερώνει τη βάση δεδομένων σχετικά με τις νέες πληροφορίες και συγκεκριμένα να προσθέσει τις αντίστοιχες εγγραφές στον κατάλληλο πίνακα.

Κατάσταση εισόδου:

1. Ο διαχειριστής πρέπει να έχει κάνει σύνδεση στο πάνελ διαχείρισης χρηστών από τον ειδική φόρμα.
2. Ο χρήστης πρέπει να ανήκει στη λίστα των διαχειριστών που έχουν το δικαίωμα να δουν αυτή τη λίστα χρηστών.

Κατάσταση εξόδου:

Δεν υπάρχει.

3.3.6 Περίπτωση χρήσης “Αναζήτηση microRNA”**Τίτλος περίπτωσης χρήσης:**

Αναζήτηση microRNA

Σύντομη περιγραφή:

Ο χρήστης εισάγει το όνομα ενός microRNA ή τον κωδικό MIMA του και ανακτώνται οι αντίστοιχοι οι βιολογικοί στόχοι.

Ροή γεγονότων**Βασική ροή**

1. Ο χρήστης εισάγει το όνομα η τον κωδικό για τα οποία θέλει να ανακτήσει βιολογικούς στόχους.
2. Ο χρήστης πατάει το κουμπί της υποβολής.
3. Εμφάνιση της σελίδας των αποτελεσμάτων που πληρούν τα κριτήρια της αναζήτησης.

Εναλλακτικές ροές**Εναλλακτική ροή 1**

1. Ο χρήστης δεν εισήγαγε έγκυρη ονομασία ή έγκυρο κωδικό microRNA.
2. Τίθεται σε εφαρμογή το υποσύστημα υποβολής προτάσεων.

Μη λειτουργικές απαιτήσεις

1. Η ανάκτηση των αποτελεσμάτων πρέπει να είναι αποδοτική ως προς το χρόνο.
2. Η εμφάνιση των αποτελεσμάτων πρέπει να είναι καλαίσθητη και δομημένη.

Κατάσταση εισόδου:

Δεν υπάρχουν απαιτήσεις εισόδου για την συγκεκριμένη περίπτωση χρήστης.

Κατάσταση εξόδου:

1. Αν ο χρήστης είναι συνδεδεμένος στο σύστημα με τον προσωπικό του λογαριασμό τότε η αναζήτηση αποθηκεύεται στο ιστορικό αναζητήσεων του συγκεκριμένου χρήστη.

3.3.7 Περίπτωση χρήσης “Παροχή προτάσεων στην αναζήτηση microRNA”

Τίτλος περίπτωσης χρήσης:

Παροχή προτάσεων στην αναζήτηση microRNA

Σύντομη περιγραφή:

Σε περίπτωση που ο χρήστης δεν εισάγει έγκυρη ονομασία ή κωδικό ενός microRNA, παρέχονται σε αυτόν προσεγγιστικές ονομασίες ή κωδικοί αντίστοιχα.

Ροή γεγονότων

Βασική ροή

1. Ο χρήστης εισήγαγε μη έγκυρο όνομα ή κωδικό MIMA ενός microRNA.
2. Με βάση το μη έγκυρο όνομα ή κωδικό εντοπίζονται οι πιο προσεγγιστικές ονομασίες ή κωδικοί.
3. Εμφάνιση λίστας αυτών των ονομασιών ή κωδικών με δυνατότητα επιλογής.

Εναλλακτικές ροές

Δεν υπάρχουν εναλλακτικές ροές για αυτήν την περίπτωση χρήσης.

Μη λειτουργικές απαιτήσεις

1. Ο εντοπισμός των προσεγγιστικών όρων πρέπει να γίνεται με αποδοτικό τρόπο.
2. Οι προτεινόμενοι όροι πρέπει να είναι οι περισσότερο προσεγγιστικοί στον όρο που εισήγαγε ο χρήστης.
3. Η παρουσίαση των προτεινόμενων όρων πρέπει να είναι καλαίσθητη και δομημένη.

Κατάσταση εισόδου:

Δεν υπάρχουν απαιτήσεις εισόδου για τη συγκεκριμένη περίπτωση χρήσης.

Κατάσταση εξόδου:

Δεν υπάρχουν απαιτήσεις εξόδου για τη συγκεκριμένη περίπτωση χρήσης.

3.3.8 Περίπτωση χρήσης “Αναζήτηση γονιδίων”**Τίτλος περίπτωσης χρήσης:**

Αναζήτηση γονιδίων

Σύντομη περιγραφή:

Ο χρήστης εισάγει το όνομα ενός γονιδίου ή τον κωδικό ENSEMBL ή τον κωδικό TRANSCRIPT ή τον κωδικό Refseq του και ανακτώνται οι αντίστοιχοι οι βιολογικοί στόχοι.

Ροή γεγονότων**Βασική ροή**

1. Ο χρήστης εισάγει το όνομα η τον κωδικό για τα οποία θέλει να ανακτήσει βιολογικούς στόχους.
2. Ο χρήστης πατάει το κουμπί της υποβολής.
3. Εμφάνιση της σελίδας των αποτελεσμάτων που πληρούν τα κριτήρια της αναζήτησης.

Εναλλακτικές ροές**Εναλλακτική ροή 1**

1. Ο χρήστης δεν εισήγαγε έγκυρη ονομασία ή έγκυρο κωδικό γονιδίου.
2. Τίθεται σε εφαρμογή το υποσύστημα υποβολής προτάσεων.

Μη λειτουργικές απαιτήσεις

1. Η ανάκτηση των αποτελεσμάτων πρέπει να είναι αποδοτική ως προς το χρόνο.
2. Η εμφάνιση των αποτελεσμάτων πρέπει να είναι καλαίσθητη και δομημένη.

Κατάσταση εισόδου:

Δεν υπάρχουν απαιτήσεις εισόδου για την συγκεκριμένη περίπτωση χρήστης.

Κατάσταση εξόδου:

Αν ο χρήστης είναι συνδεδεμένος στο σύστημα με τον προσωπικό του λογαριασμό τότε η αναζήτηση αποθηκεύεται στο ιστορικό αναζητήσεων του συγκεκριμένου χρήστη.

3.3.9 Περίπτωση χρήσης “Παροχή προτάσεων στην αναζήτηση γονιδίων”

Τίτλος περίπτωσης χρήσης:

Παροχή προτάσεων στην αναζήτηση γονιδίων

Σύντομη περιγραφή:

Σε περίπτωση που ο χρήστης δεν εισάγει έγκυρη ονομασία ή κωδικό ενός γονιδίου, παρέχονται σε αυτόν προσεγγιστικές ονομασίες ή κωδικοί αντίστοιχα.

Ροή γεγονότων**Βασική ροή**

1. Ο χρήστης εισήγαγε μη έγκυρο όνομα ή κωδικό ENSEMBL ή TRANSCRIPT ή Refseq ενός γονιδίου.
2. Με βάση το μη έγκυρο όνομα η κωδικό εντοπίζονται οι πιο προσεγγιστικές ονομασίες ή κωδικοί.
3. Εμφάνιση λίστας αυτών των ονομασιών ή κωδικών με δυνατότητα επιλογής.

Εναλλακτικές ροές

Δεν υπάρχουν εναλλακτικές ροές για αυτήν την περίπτωση χρήσης.

Μη λειτουργικές απαιτήσεις

1. Ο εντοπισμός των προσεγγιστικών όρων πρέπει να γίνεται με αποδοτικό τρόπο.
2. Οι προτεινόμενοι όροι πρέπει να είναι οι περισσότερο προσεγγιστικοί στον όρο που εισήγαγε ο χρήστης.
3. Η παρουσίαση των προτεινόμενων όρων πρέπει να είναι καλαίσθητη και δομημένη.

Κατάσταση εισόδου:

Δεν υπάρχουν απαιτήσεις εισόδου για τη συγκεκριμένη περίπτωση χρήσης.

Κατάσταση εξόδου:

Δεν υπάρχουν απαιτήσεις εξόδου για τη συγκεκριμένη περίπτωση χρήσης.

3.3.10 Περίπτωση χρήσης “Εμφάνιση προβλέψεων με βάση άλλους αλγορίθμους”

Τίτλος περίπτωσης χρήσης:

Εμφάνιση προβλέψεων με βάση άλλους αλγορίθμους

Σύντομη περιγραφή:

Εμφάνιση λίστας προβλέψεων αποτελεσμάτων που έχουν πραγματοποιηθεί με βάση άλλους

αλγορίθμους.

Ροή γεγονότων

Βασική ροή

1. Ο χρήστης έχει πραγματοποιήσει μια αναζήτηση και έχει εμφανιστεί η σελίδα των αποτελεσμάτων.
2. Ο χρήστης επιλέγει το ειδικό σύμβολο που αποτελεί σύνδεσμο προς την παρουσίαση των αποτελεσμάτων.
3. Εμφάνιση νέας σελίδας στην οποία παρουσιάζονται άλλοι αλγόριθμοι που έχουν προβλέψει το συγκεκριμένο στόχο και την ακρίβεια της πρόβλεψης.

Εναλλακτικές ροές

Εναλλακτική ροή 1

1. Ο χρήστης δεν επιλέγει το σύνδεσμο για το άνοιγμα της νέας σελίδας.

Μη λειτουργικές απαιτήσεις

1. Η εμφάνιση των αλγορίθμων και της ακρίβειας τους πρέπει να γίνεται με καλαίσθητο και δομημένο τρόπο.

Κατάσταση εισόδου:

Δεν υπάρχουν απαιτήσεις εισόδου για αυτήν την περίπτωση χρήσης.

Κατάσταση εξόδου:

Δεν υπάρχει.

3.4 Μοντέλο Οντοτήτων Συσχετίσεων

Η φύση της εφαρμογής υποδεικνύει ότι δεν θα μπορούσαμε να προχωρήσουμε στην υλοποίηση της χωρίς τη δημιουργία μιας κατάλληλης βάσης δεδομένων. Η βάση λειτουργεί οργανώνοντας την πληροφορία σε πίνακες με τέτοιο τρόπο ώστε να διευκολύνονται οι αναζητήσεις που πραγματοποιούν οι χρήστες και οι διαχειριστές του συστήματος. Η βάση δεδομένων που χρησιμοποιήθηκε, σχεδιάστηκε αρχικά στην πρώτη έκδοση της παρούσας εφαρμογής, και χρησιμοποιήθηκε με την ίδια μορφή κατά την υλοποίηση της νεότερης έκδοσης αυτής. Για το λόγο αυτό δεν γίνεται αναλυτική περιγραφή των οντοτήτων και συσχετίσεων της βάσης δεδομένων. Για την αναλυτικότερη περιγραφή των οντοτήτων και συσχετίσεων της παρούσας εφαρμογής ο αναγνώστης μπορεί να ανατρέξει στην αναφορά [7].

Gene: Αναπαριστά ένα γονίδιο.

Protein gene: Αναπαριστά ένα γονίδιο το οποίο κωδικοποιεί μια πρωτεΐνη.

microRNA gene: Αναπαριστά ένα γονίδιο το οποίο κωδικοποιεί ένα microRNA.

Transcript. Αναπαριστά ένα μετάγραφο ενός γονιδίου.

Transcription factor: Αναπαριστά ένα μεταγραφικό παράγοντα ενός γονιδίου.

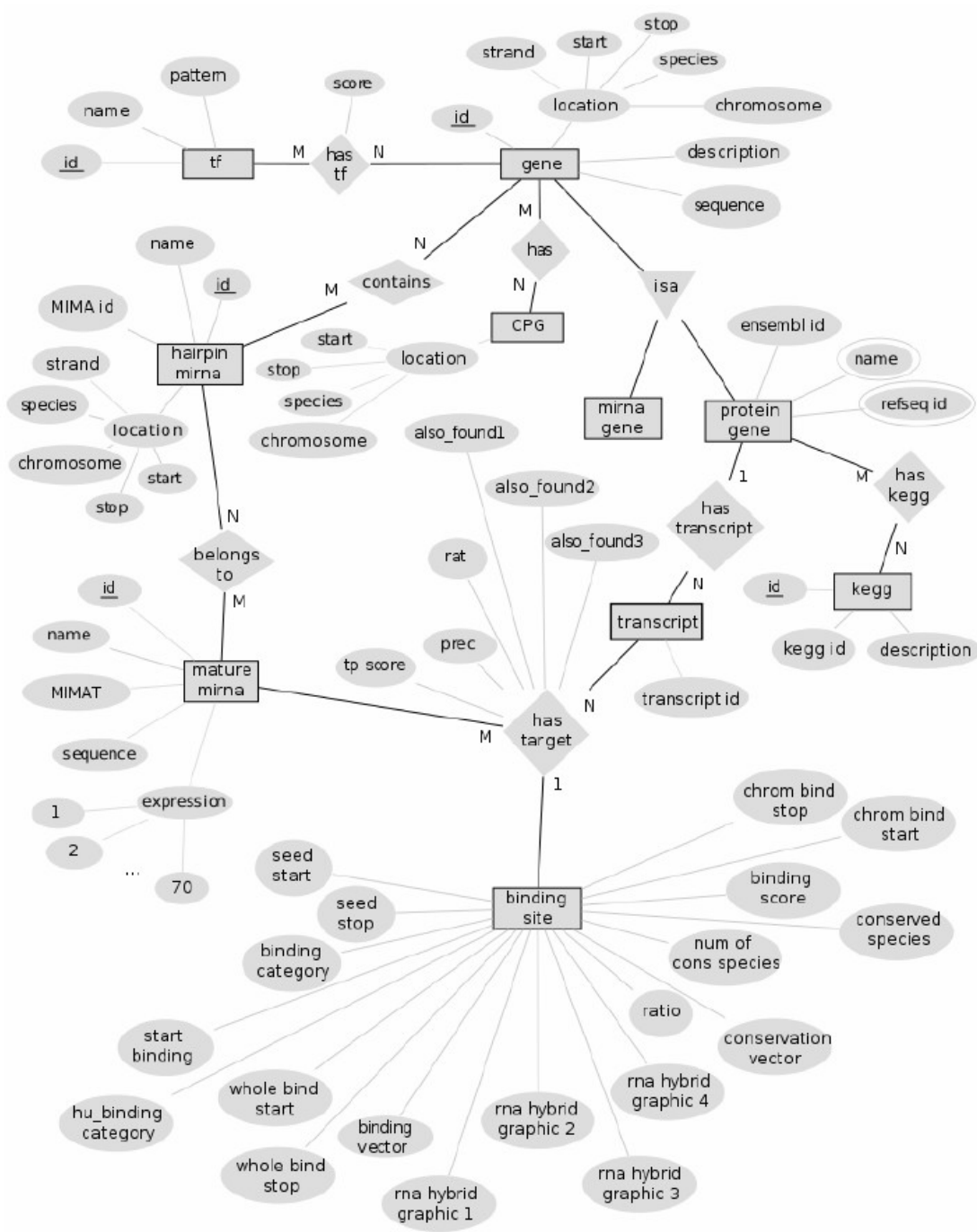
CPG: Αναπαριστά μια περιοχή DNA στην οποία συναντάται η αλληλουχία βάσεων CPG.

Hairpin microRNA: Αναπαριστά ένα microRNA τύπου hairpin.

Mature microRNA: Αναπαριστά ένα microRNA τύπου mature.

KEGG: Αναπαριστά ένα KEGG pathway.

Binding site: Αναπαριστά μια θέση πρόσδεσης ενός microRNA πάνω σε ένα transcript.



Εικόνα 3: Διάγραμμα οντοτήτων-συσχετίσεων της σχεσιακής βάσης δεδομένων της παρούσας εφαρμογής.

4

Σχεδίαση Συστήματος

Στο κεφάλαιο αυτό αρχικά παρουσιάζεται το αρχιτεκτονικό μοντέλο ανάπτυξης της εφαρμογής που επιλέχθηκε στη φάση της σχεδίασης. Στη συνέχεια αναλύονται επιγραμματικά οι κλάσεις του συστήματος και οι μέθοδοι αυτών. Στο τέλος του κεφαλαίου αφιερώνουμε μία ενότητα στην παρουσίαση της δομής της βάσης δεδομένων που σχεδιάστηκε για την εφαρμογή.

4.1 Αρχιτεκτονική

Η σχεδίαση του συστήματος πραγματοποιήθηκε με βάση το αρχιτεκτονικό πρότυπο MVC (Model-View-Controller). Σύμφωνα με το πρότυπο αυτό τα επιμέρους υποσυστήματα διαχωρίζονται με τέτοιο τρόπο ώστε η διαχείριση της πληροφορίας (**Model**) και η οπτική παρουσίαση της εφαρμογής (**View**), να είναι ανεξάρτητα από τη λογική με την οποία λειτουργεί αυτή (**Controller**). Ο διαχωρισμός αυτός οδηγεί σε καλύτερη ανάπτυξη, καλύτερο έλεγχο και καλύτερη συντήρηση της εκάστοτε εφαρμογής.

Η χρήση του αρχιτεκτονικού προτύπου MVC έχει στεφθεί από απόλυτη επιτυχία καθότι λόγω των οφελών που αναφέρθηκαν προηγουμένως, αποτελεί σίγουρη λύση στην ανάπτυξη μεγάλων εφαρμογών λογισμικού. Για το λόγο αυτό έχουν αναπτυχθεί πολλά frameworks σε

διάφορες γλώσσες προγραμματισμού που βοηθούν στην ανάπτυξη εφαρμογών σύμφωνα με το πρότυπο MVC. Η περίπτωση του Yii PHP framework που επιλέχθηκε στην παρούσα διπλωματική εργασία, εντάσσεται στην κατηγορία αυτή.

Στην συνέχεια ακολουθεί μια συνοπτική παρουσίαση των επιμέρους δομικών μονάδων του αρχιτεκτονικού προτύπου MVC.

4.1.1 Model

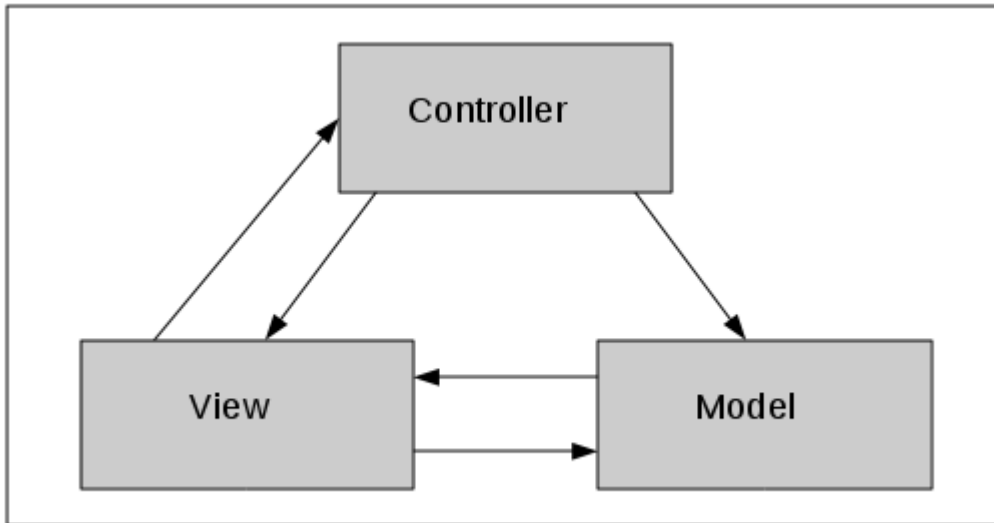
Η δομική μονάδα των Models είναι υπεύθυνη για τη διαχείριση της πληροφορίας και για τη γνωστοποίηση αλλαγών αυτής στις υπόλοιπες δομικές μονάδες. Επιπλέον αποτελεί την αναπαράσταση των δεδομένων πάνω στα οποία επιδρά η εφαρμογή. Για παράδειγμα ένα μοντέλο μπορεί να αντιπροσωπεύει όλες τις εγγραφές σε έναν πίνακα στη βάση δεδομένων μιας εφαρμογής. Με αυτόν τον τρόπο η λογική λειτουργίας αυτής της μονάδας προσθέτει σημασιολογικό περιεχόμενο στα δεδομένα. Όταν κάποιο από τα δεδομένα της εφαρμογής μεταβληθεί τότε το μοντέλο ενημερώνει τα views προκειμένου να ανανεωθούν καταλλήλως.

4.1.2 View

Η δομική μονάδα των Views είναι υπεύθυνη για την παρουσίαση των δεδομένων σε φόρμες. Οι φόρμες αυτές είναι κατάλληλες για αλληλεπίδραση με το χρήστη και στην ουσία αποτελούν μέρος της διεπαφής χρήστη. Κάθε view συνήθως εξυπηρετεί διαφορετικό σκοπό ανάλογα με τις λειτουργίες της εκάστοτε εφαρμογής και έχει δυνατότητες επικοινωνίας τόσο με models όσο και με controllers.

4.1.3 Controller

Η δομική μονάδα των Controllers είναι υπεύθυνη για την υποδοχή των αιτήσεων της εφαρμογής. Ανάλογα με τον τύπο της αίτησης ορίζεται διαφορετικός τρόπος απάντησης με βάση κάποιο συγκεκριμένο μοντέλο και κάποιο συγκεκριμένο view. Συνεπώς ο controller είναι στην ουσία αυτός που πραγματοποιεί την επικοινωνία μεταξύ model και view.



Εικόνα 4: Σχηματικό διάγραμμα του αρχιτεκτονικού προτύπου MVC

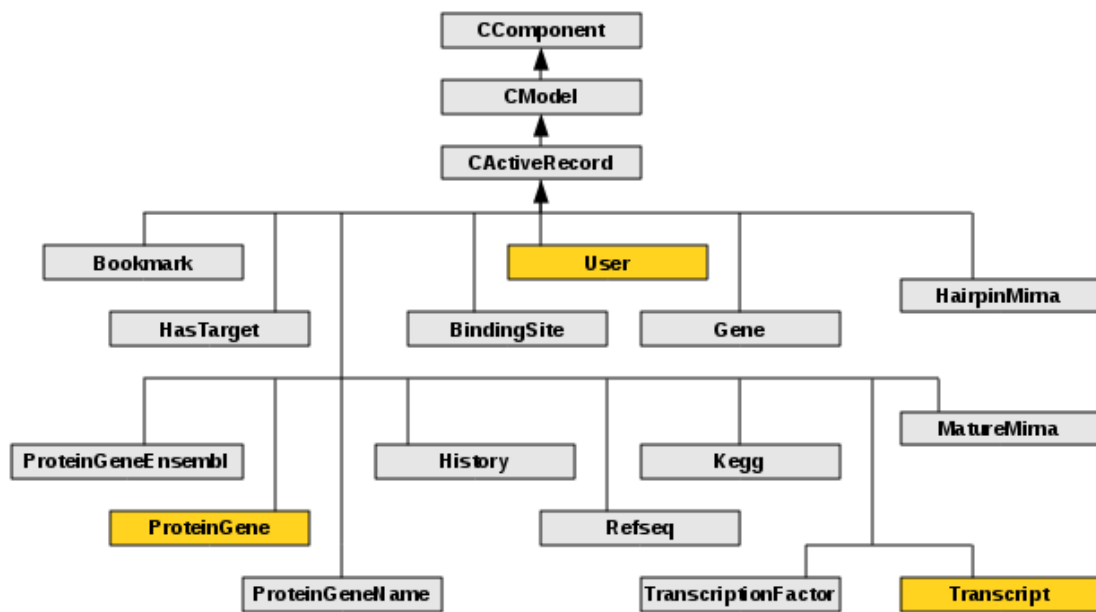
4.2 Περιγραφή κλάσεων

Στο σημείο αυτό θα γίνει συνοπτική παρουσίαση των κλάσεων και των μεθόδων αυτών που απαρτίζουν την εφαρμογή. Η παρουσίαση αφορά μόνο στις κλάσεις που δημιουργήθηκαν εξολοκλήρου ή τροποποιήθηκαν κατά τη διάρκεια της υλοποίησης της εφαρμογής της παρούσας έκδοσης της εφαρμογής. Για τις κλάσεις και τις μεθόδους αυτών της παλιότερης έκδοσης πληροφορίες μπορούν να αναζητηθούν στην αναφορά [7].

4.2.1 Κλάσεις *Active Record*

Το πρότυπο *Active Record* είναι ένα σχεδιαστικό πρότυπο σύμφωνα με το οποίο τα δεδομένα μια εφαρμογής αποθηκεύονται οργανωμένα σε μία βάση δεδομένων. Ουσιαστικά το πρότυπο αυτό ορίζει μια διεπαφή μεταξύ της βάσης και του συστήματος, η οποία περιέχει όλες τις απαραίτητες λειτουργίες για εισαγωγή, διαγραφή και ενημέρωση εγγραφών της βάσης.

Το πρότυπο αυτό τηρεί το μοντέλο του αντικειμενοστραφούς προγραμματισμού με τέτοιο τρόπο ώστε κάθε στιγμιότυπο μιας κλάσης αυτού του τύπου να σχετίζεται με μία μόνο εγγραφή στη βάση δεδομένων.



Εικόνα 5: Διάγραμμα κλάσεων για τις κλάσεις Active Record του συστήματος.

4.2.1.1 Κλάση ProteinGene

Μέθοδος TransIds

1. Δέχεται ως παράμετρο την ονομασία ενός γονιδίου.
2. Επιστρέφει μία λίστα με όλους τους κωδικούς TRANSCRIPT του συγκεκριμένου γονιδίου.

4.2.1.2 Κλάση Transcript

Μέθοδος EnsId

1. Δέχεται ως παράμετρο τον κωδικό-κλειδί ενός γονιδίου.
2. Επιστρέφει τον κωδικό ENSEMBL του αντίστοιχου γονιδίου.

4.2.1.3 Κλάση User

Μέθοδος findAll

1. Δέχεται ως παράμετρο μία συνθήκη σε μορφή συμβολοσειράς και έναν πίνακα που αντιπροσωπεύει τον τρόπο εμφάνισης.
2. Επιλέγει από τον πίνακα των χρηστών όλες τις πληροφορίες των χρηστών με βάση την πρώτη παράμετρο που δέχεται στην είσοδο και τους εμφανίζει με βάση τη δεύτερη παράμετρο.

Μέθοδος `getAdminsOrdered`

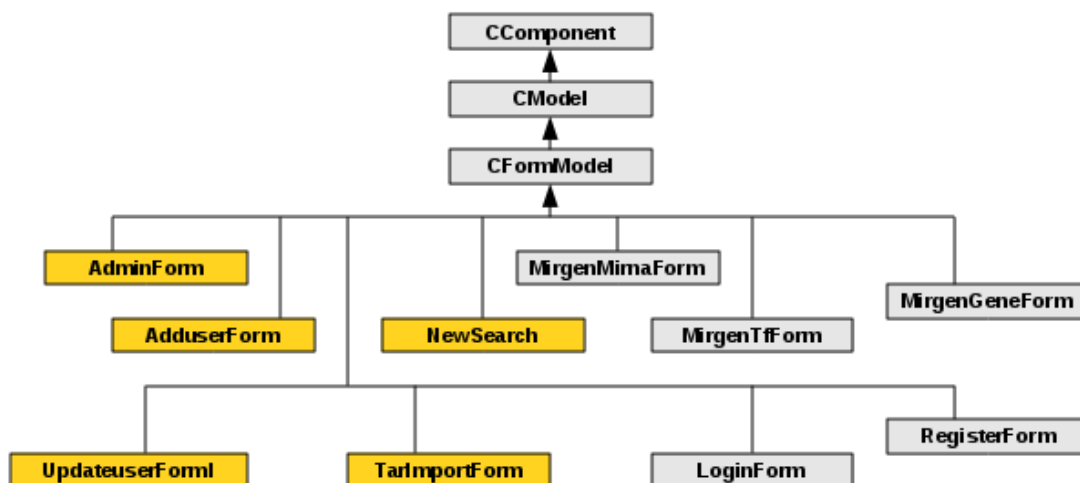
1. Βρίσκει όλους τους διαχειριστές του συστήματος και τους εμφανίζει αλφαβητικά.

Μέθοδος `getSimpleUsersOrdered`

1. Βρίσκει όλους τους απλούς χρήστες του συστήματος και τους εμφανίζει αλφαβητικά.

4.2.2 Κλάσεις *Form Model*

Ο συγκεκριμένος τύπος κλάσεων χρησιμοποιείται από το Yii PHP framework για τη συλλογή των δεδομένων από HTML φόρμες και την επεξεργασία αυτών των δεδομένων. Το σημαντικότερο πλεονέκτημα αυτών των κλάσεων είναι η ευκολία που παρέχουν στο χειρισμό δεδομένων.



Εικόνα 6: Διάγραμμα κλάσεων για τις κλάσεις *Form Model* του συστήματος.

4.2.2.1 Κλάση *AdduserForm*

Μέθοδος `rules`

1. Ορίζει τις προϋποθέσεις που πρέπει να πληρούνται προκειμένου η συμπληρωμένη φόρμα να είναι έγκυρη. Συγκεκριμένα πρέπει όλα τα πεδία της φόρμας να είναι συμπληρωμένα (όνομα χρήστη, συνθηματικό, επανάληψη συνθηματικού, ηλεκτρονική διεύθυνση, τύπος χρήστη), τα πεδία συνθηματικό και επανάληψη συνθηματικού να έχουν συμπληρωθεί με τις ίδιες λέξεις, το όνομα χρήστη να είναι μοναδικό και η ηλεκτρονική διεύθυνση να είναι έγκυρου τύπου.

Μέθοδος `uniqueUsername`

1. Η μέθοδος αυτή ελέγχει αν το όνομα χρήστη που εισήγαγε ο διαχειριστής είναι μοναδικό. Σε διαφορετική περίπτωση εμφανίζεται κατάλληλο μήνυμα λάθους.

Μέθοδος afterValidate

1. Η μέθοδος αυτή καλείται εφόσον τα στοιχεία χρήστη που εισήγαγε ο διαχειριστής πληρούν τις προϋποθέσεις των προηγούμενων μεθόδων.
2. Στη συνέχεια καταχωρεί τα στοιχεία του νέου χρήστη στη βάση δημιουργώντας ένα νέο λογαριασμό χρήστη.

4.2.2.2 Κλάση AdminForm

Μέθοδος rules

1. Ορίζει τις προϋποθέσεις που πρέπει να πληρούνται προκειμένου ο διαχειριστής να κάνει εισαγωγή στο σύστημα. Οι προϋποθέσεις είναι να είναι συμπληρωμένα τόσο τα πεδία του ονόματος χρήστη όσο και του συνθηματικού και ο χρήστης να έχει ταυτοποιηθεί.

Μέθοδος authenticate

1. Η μέθοδος αυτή είναι υπεύθυνη για την ταυτοποίηση του διαχειριστή. Αν ο χρήστης εισάγει λάθος συνδυασμό ονόματος χρήστη και συνθηματικού τότε ενημερώνεται με κατάλληλο μήνυμα λάθους. Σε διαφορετική περίπτωση εισάγεται κανονικά στο σύστημα.

4.2.2.3 Κλάση UpdateuserForm

Μέθοδος rules

1. Ορίζει τις προϋποθέσεις οι οποίες πρέπει να πληρούνται προκειμένου να ενημερωθούν τα στοιχεία ενός συγκεκριμένου χρήστη από το διαχειριστή.
2. Συγκεκριμένα πρέπει τα πεδία όνομα χρήστη, συνθηματικό, ηλεκτρονική διεύθυνση και επανάληψη συνθηματικού να είναι συμπληρωμένα, οι συμβολοσειρές στα πεδία συνθηματικό και επανάληψη συνθηματικού να είναι κοινές, το όνομα χρήστη να είναι μοναδικό, και η ηλεκτρονική διεύθυνση να είναι έγκυρου τύπου.

Μέθοδος uniqueUsername

1. Η μέθοδος ελέγχει αν το νέο όνομα χρήστη είναι μοναδικό, σε περίπτωση που ο διαχειριστής επέλεξε να το ανανεώσει.

Μέθοδος afterValidate

1. Η μέθοδος αυτή αναλαμβάνει την ενημέρωση της κατάλληλης εγγραφής στη βάση δεδομένων με τα νέα στοιχεία.

4.2.2.4 Κλάση *TarImportForm*

Μέθοδος rules

1. Ορίζει τις προϋποθέσεις οι οποίες πρέπει να πληρούνται προκειμένου ο διαχειριστής να ανεβάσει το αρχείο με τους βιολογικούς στόχους που έχουν εντοπιστεί με βάση άλλους αλγορίθμους.
2. Συγκεκριμένα πρέπει το πεδίο για το αρχείο να είναι συμπληρωμένο και το στοιχείο προς ανέβασμα να είναι τύπου “αρχείο”.

4.2.2.5 Κλάση *NewSearch*

Μέθοδος findMirnaHairpinId

1. Η μέθοδος αυτή δέχεται ως παράμετρο το όνομα ενός hairpin microRNA και επιστρέφει τον κωδικό-κλειδί αυτού στην μεταβλητή id του συγκεκριμένου στιγμιότυπου.

Μέθοδος findMirnaMatureId

1. Η μέθοδος αυτή δέχεται ως παράμετρο το όνομα ενός mature microRNA και επιστρέφει τον κωδικό-κλειδί αυτού στην μεταβλητή id του συγκεκριμένου στιγμιότυπου..

Μέθοδος setGeneId

1. Η μέθοδος αυτή δέχεται ως παράμετρο το όνομα ενός γονιδίου και στο στιγμιότυπο της κλάσης θέτει την τιμή του κωδικού-κλειδιού του γονιδίου στην παράμετρο id.

Μέθοδος findMirna

1. Η μέθοδος αυτή δέχεται ως παράμετρο το όνομα ενός πίνακα και το όνομα ενός microRNA.
2. Επιστρέφει όλες τις πληροφορίες για το συγκεκριμένο microRNA από τον πίνακα της βάσης που ορίζεται στην πρώτη παράμετρο.

Μέθοδος findToken

1. Η μέθοδος αυτή χρησιμοποιείται από τη νέα βελτιωμένη έκδοση της αναζήτησης.
2. Δέχεται ως παράμετρο ένα τύπο που ορίζει αν θα επιλέξει ή όχι τα διπλότυπα από τον πίνακα mirna_tokens της βάσης.

3. Σε περίπτωση που ορίζεται και η δεύτερη παράμετρος της, τότε επιλέγει από τον προαναφερθέντα πίνακα όλα εκείνα τα στοιχεία για τα οποία το πεδίο token είναι ίσο με την παράμετρο αυτή.

Μέθοδος checkMima

1. Η μέθοδος αυτή δέχεται ως παράμετρο έναν υποψήφιο κωδικό MIMA και ελέγχει αν αυτός αντιστοιχεί σε κάποιο microRNA.

Μέθοδος checkRefseq

1. Η μέθοδος αυτή δέχεται ως παράμετρο έναν υποψήφιο κωδικό Refseq και ελέγχει αν αυτός αντιστοιχεί σε ένα γονίδιο.

Μέθοδος checkEns

1. Η μέθοδος αυτή δέχεται ως παράμετρο έναν υποψήφιο κωδικό ENSEMBL ή TRANSCRIPT και ελέγχει αν αυτός αντιστοιχεί σε κάποιο γονίδιο.

Μέθοδος checkGene

1. Η μέθοδος αυτή δέχεται ως παράμετρο ένα υποψήφιο όνομα γονιδίου και ελέγχει αν αντιστοιχεί σε κάποιο έγκυρο όνομα.

Μέθοδος fetchMima

1. Η μέθοδος αυτή χρησιμοποιείται από τη βελτιωμένη έκδοση αναζήτησης και σηκώνει από τη βάση όλους τους κωδικούς MIMA για ένα microRNA ανάλογα με τον τύπο που θέλουμε mature ή hairpin.

Μέθοδος fetchEns

1. Η μέθοδος αυτή χρησιμοποιείται από τη βελτιωμένη έκδοση αναζήτησης και σηκώνει από τη βάση όλους τους κωδικούς ENSEMBL.

Μέθοδος fetchTrans

1. Η μέθοδος αυτή χρησιμοποιείται από τη βελτιωμένη έκδοση αναζήτησης και σηκώνει από τη βάση όλους τους κωδικούς TRANSCRIPT.

Μέθοδος fetchRefseq

1. Η μέθοδος αυτή χρησιμοποιείται από τη βελτιωμένη έκδοση αναζήτησης και σηκώνει από τη βάση όλους τους κωδικούς Refseq.

Μέθοδος fetchGeneNames

1. Η μέθοδος αυτή χρησιμοποιείται από τη βελτιωμένη έκδοση αναζήτησης και σηκώνει από τη βάση όλα τα ονόματα γονιδίων.

Μέθοδος ensName

1. Η μέθοδος αυτή δέχεται ως παράμετρο τον κωδικό-κλειδί ενός γονιδίου και επιστρέφει την ονομασία αυτού.

Μέθοδος getEnsId

1. Η μέθοδος αυτή δέχεται ως παράμετρο των κωδικό-κλειδί ενός γονιδίου και επιστρέφει τον κωδικό ENSEMBL αυτού.

Μέθοδος PredictionInfo

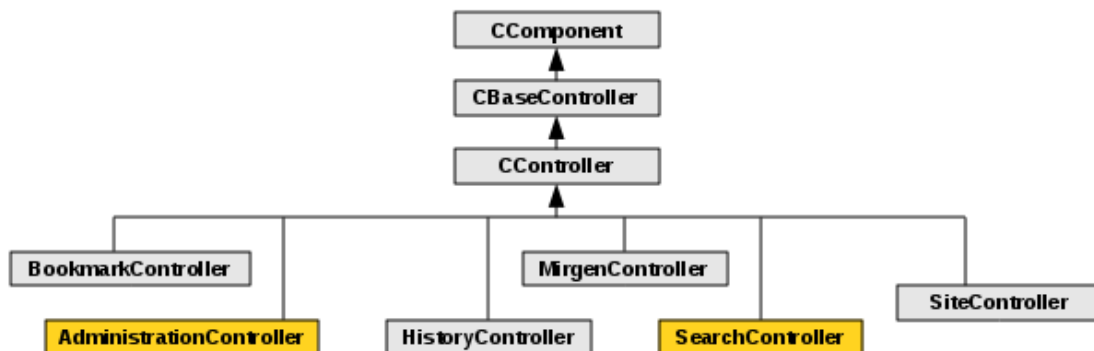
1. Η μέθοδος αυτή δέχεται ως παράμετρο των κωδικό-κλειδί ενός γονιδίου και τον κωδικό κλειδί ενός microRNA και επιστρέφει όλες τις σχετικές πληροφορίες με τα υπόλοιπα προγράμματα που εντόπισαν τον ίδιο βιολογικό στόχο του συγκεκριμένου microRNA στο γονίδιο.

Μέθοδος ProgramNames

1. Η μέθοδος αυτή βρίσκει τη λίστα όλων των προγραμμάτων που έχουν εντοπίσει κοινούς βιολογικούς στόχους με το πρόγραμμα της παρούσας εφαρμογής.

4.2.3 Ελεγκτές

Οι ελεγκτές είναι ίσως η σημαντικότερη κατηγορία κλάσεων του συστήματος, καθότι αναλαμβάνουν να φέρουν εις πέρας την επικοινωνία των άλλων κλάσεων μεταξύ τους και να πετύχουν την εύρυθμη λειτουργία του συστήματος.



Εικόνα 7: Διάγραμμα κλάσεων για τους ελεγκτές του συστήματος.

4.2.3.1 Κλάση AdministrationController

Η συγκεκριμένη κλάση είναι υπεύθυνη για την εκτέλεση των λειτουργιών που σχετίζονται με τη διαχείριση των χρηστών. Οι λειτουργίες που υποστηρίζονται είναι οι εξής:

- Προβολή λίστας χρηστών

- Διαγραφή χρηστών
- Τροποποίηση στοιχείων χρηστών
- Προσθήκη νέου χρήστη
- Εισαγωγή διαχειριστή στο πάνελ διαχείρισης χρηστών
- Έξοδος διαχειριστή από το πάνελ διαχείρισης χρηστών
- Έλεγχος για την εισαγωγή μόνο των διαχειριστών στο πάνελ διαχείρισης χρηστών

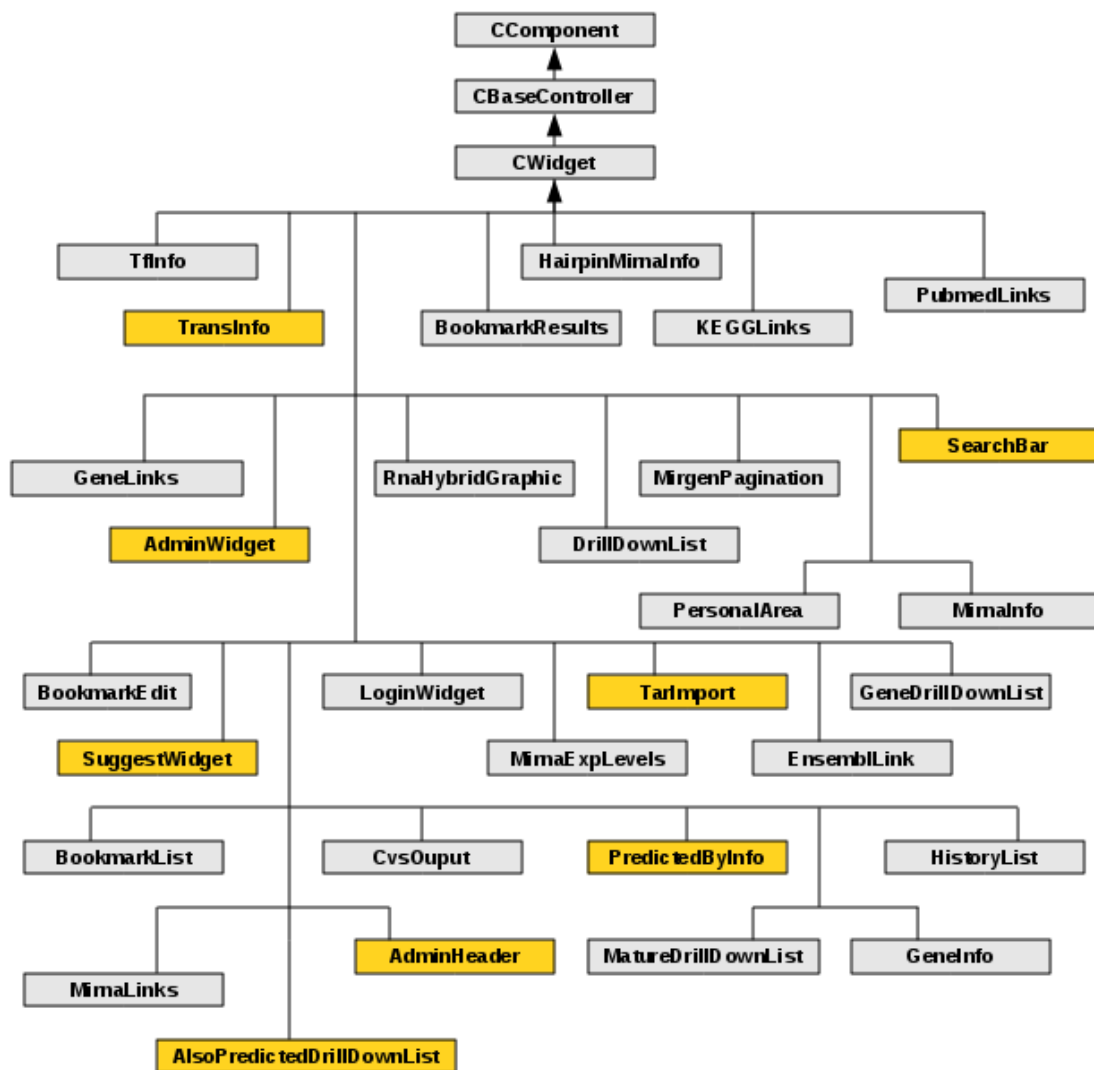
4.2.3.2 Κλάση *SearchController*

Η συγκεκριμένη κλάση είναι υπεύθυνη για την πραγματοποίηση όλων των τύπων αναζητήσεων του συστήματος. Οι λειτουργίες που μπορεί να επιτελέσει είναι οι εξής:

- Παροχή προτάσεων στην περίπτωση που ο χρήστης δεν εισάγει έγκυρους προς αναζήτηση όρους.
- Προβολή αποτελεσμάτων ανάλογα με τον τύπο της αναζήτησης που πραγματοποιεί ένας χρήστης.

4.2.4 *Widgets*

Αυτός ο τύπος κλάσεων χρησιμοποιείται ευρέως σε όλες τις σύγχρονες εφαρμογές καθώς αυξάνει τη δομικότητα αυτών.



Εικόνα 8: Διάγραμμα κλάσεων για τα widget του συστήματος.

4.2.4.1 Κλάση AdminHeader

Η συγκεκριμένη κλάση είναι υπεύθυνη προκειμένου να εμφανίζει στον εκάστοτε διαχειριστή του συστήματος την μπάρα επιλογών στο πάνελ διαχείρισης. Έτσι κάθε διαχειριστής μπορεί από την μπάρα αυτή να μεταβεί είτε στο πάνελ διαχείρισης χρηστών, είτε στην κύρια σελίδα της εφαρμογής είτε και να αποσυνδεθεί από το σύστημα.

4.2.4.2 Κλάση AdminWidget

Η συγκεκριμένη κλάση υλοποιεί τη φόρμα εισαγωγής στοιχείων των διαχειριστών προκειμένου να συνδεθούν στο πάνελ διαχείρισης.

4.2.4.3 Κλάση *AlsoPredictedDrillDownList*

Η συγκεκριμένη κλάση είναι υπεύθυνη προκειμένου να εμφανίζει στη λίστα αποτελεσμάτων τη λίστα με τα προγράμματα που εντόπισαν τους ίδιους βιολογικούς στόχους με το πρόγραμμα της παρούσας εφαρμογής.

4.2.4.4 Κλάση *PredictedByInfo*

Η συγκεκριμένη κλάση είναι υπεύθυνη προκειμένου να εμφανίζει πληροφορίες σχετικά με το στόχο ενός συγκεκριμένου microRNA σε ένα γονίδιο. Οι πληροφορίες που αναγράφονται είναι το όνομα του microRNA, ο κωδικός ENSEMBL του γονιδίου και το σκορ, η ακρίβεια πρόβλεψης και η σηματοθρομβική αναλογία του στόχου.

4.2.4.5 Κλάση *SearchBar*

Η συγκεκριμένη κλάση υλοποιεί τη φόρμα καθορισμού των παραμέτρων στις οποίες μπορεί να δώσει τιμές ο χρήστης μετά από κάθε αναζήτηση προκειμένου να περιορίσει τη λίστα αποτελεσμάτων στο εύρος που επιθυμεί.

4.2.4.6 Κλάση *SuggestWidget*

Η συγκεκριμένη κλάση είναι υπεύθυνη για την παρουσίαση των προτεινόμενων όρων προς αναζήτηση, στην περίπτωση που ο χρήστης δεν εισάγει έγκυρους όρους.

4.2.4.7 Κλάση *TarImport*

Η συγκεκριμένη κλάση είναι υλοποιεί την ειδική φόρμα για τον ανέβασμα αρχείων βιολογικών στόχων με βάση άλλους αλγορίθμους.

4.2.4.8 Κλάση *TransInfo*

Η συγκεκριμένη κλάση είναι υπεύθυνη για την εμφάνιση πληροφοριών σχετικά με τα μετάγραφα ενός γονιδίου. Συγκεκριμένα απεικονίζει τόσο τον κωδικό TRANSCRIPT ενός γονιδίου, όσο και τον κωδικό ENSEMBL αυτού.

4.3 Βάση δεδομένων

Στην ενότητα αυτή θα περιγράψουμε την σχεδίαση της βάσης με βάση το μοντέλο οντοτήτων συσχετίσεων που παρουσιάστηκε στο προηγούμενο κεφάλαιο. Για την ευκολότερη

κατανόηση του σχήματος της βάσης, παραθέτουμε στο τέλος της ενότητας μία σχηματική αναπαράσταση των πινάκων και των συσχετίσεων μεταξύ αυτών. Ειδικότερα, ο σχεδιασμός έγινε με βάση τα εξής κριτήρια:

- Κάθε πίνακας απλού τύπου έχει πρωτεύον κλειδί ώστε να μεγιστοποιείται η ταχύτητα τυχαίας προσπέλασης.
- Το πρωτεύον κλειδί κάθε πίνακα έχει το ελάχιστο δυνατό μήκος για οικονομία χώρου και απόδοση.
- Το μήκος κάθε πεδίου των πινάκων έχει το ελάχιστο δυνατό μήκος τόσο για οικονομία χώρου και απόδοση όσο και για την επιτυχή μελλοντική εισαγωγή στοιχείων.
- Επιλέχθηκε να δημιουργηθούν νέοι πίνακες τύπου MEMORY, προκειμένου να βελτιωθεί ο μηχανισμός της αναζήτησης και να γίνει πιο αποδοτικός.

Οι πίνακες του συστήματος είναι δύο ειδών, MyISAM και MEMORY. Οι πίνακες MyISAM είναι πιο γρήγοροι από τους πίνακες InnoDB. Επιπλέον στην παρούσα εφαρμογή δεν υπάρχουν ανάγκες για transactions και έλεγχοι ακεραιότητας, συνεπώς επιλέχθηκε η χρήση πινάκων MyISAM. Οι πίνακες MEMORY αποθηκεύουν τα δεδομένα τους στη μνήμη RAM του συστήματος με αποτέλεσμα την γρήγορη προσπέλαση αυτών. Για το λόγο αυτό είναι κατάλληλοι να χρησιμοποιηθούν από το σύστημα αναζήτησης. Οι πίνακες περιγράφονται συνοπτικά ακολούθως:

- Πίνακες τύπου MyISAM:

diana_binding_sites: Αναπαριστά ένα σημείο πρόσδεσης ενός microRNA πάνω σε ένα γονίδιο.

diana_genes: Αναπαριστά ένα γονίδιο.

diana_hairpin_mirnas: Αναπαριστά ένα microRNA τύπου hairpin.

diana_has_target: Αναπαριστά ένα στόχο ενός microRNA σε ένα γονίδιο.

diana_keggs: Αναπαριστά ένα KEGG pathway.

diana_mature_mirnas: Αναπαριστά ένα microRNA τύπου mature.

diana_protein_gene_names: Αναπαριστά το όνομα και τα συνώνυμα ενός γονιδίου.

diana_protein_genes: Αναπαριστά ένα γονίδιο το οποίο κωδικοποιεί μια πρωτεΐνη.

diana_refseqs: Αναπαριστά των κωδικό Refseq ενός γονιδίου.

diana_interactions: Αναπαριστά την ακρίβεια πρόβλεψης ενός στόχου με βάση άλλους αλγορίθμους.

diana_users: Αναπαριστά κάθε χρήστη του συστήματος.

diana_history: Αναπαριστά το ιστορικό των αναζητήσεων κάθε χρήστη.

diana_bookmarks: Αναπαριστά τις αγαπημένες αναζητήσεις ενός χρήστη.

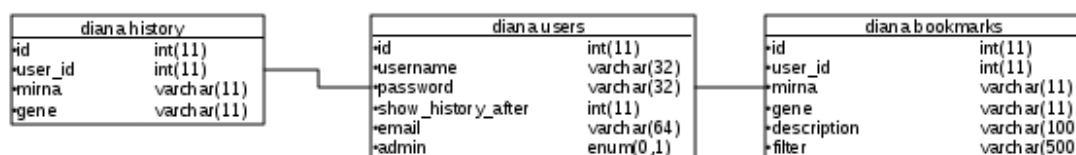
- Πίνακες τύπου MEMORY:

genes_tokens: Σε αυτόν τον πίνακα αποθηκεύεται το όνομα ενός γονιδίου και ο κωδικός-κλειδί του.

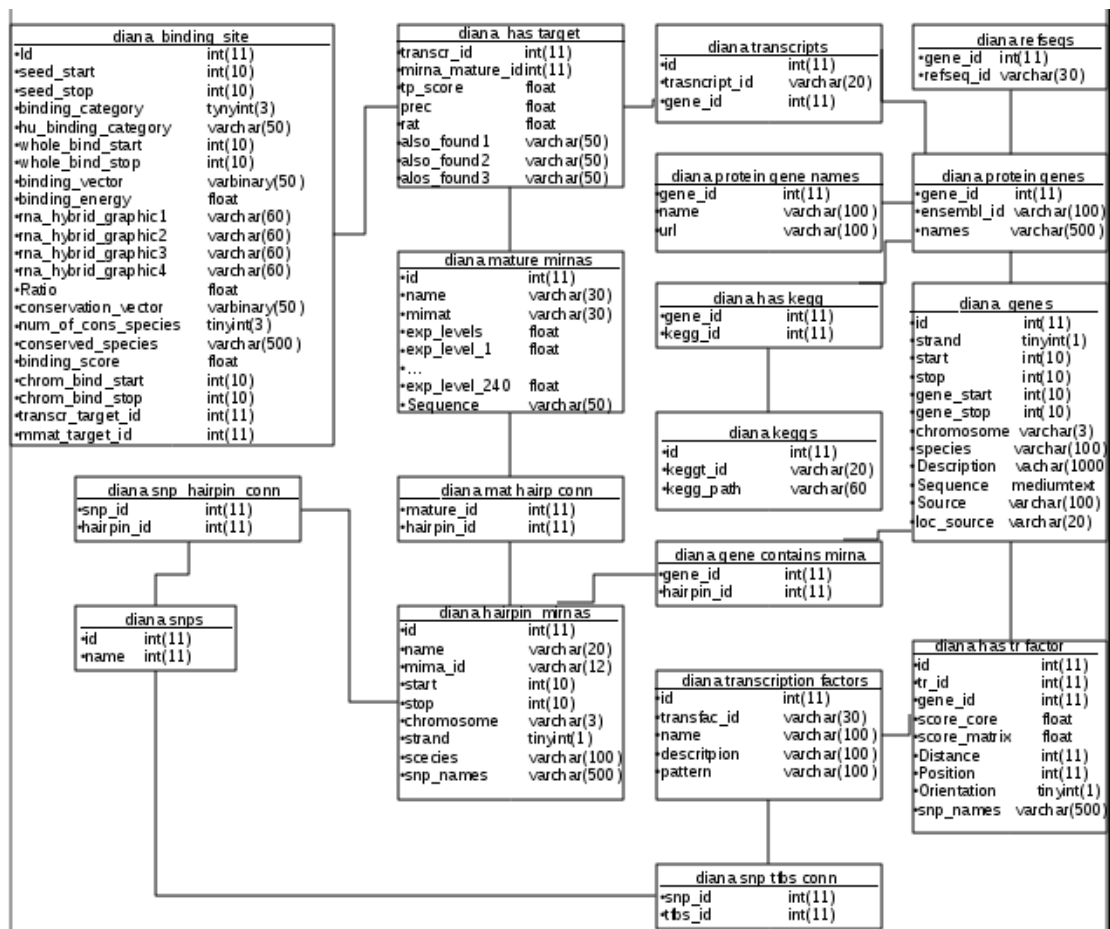
mirna_tokens: Σε αυτόν τον πίνακα αποθηκεύονται όλες οι επιμέρους λέξεις που υπάρχουν ανάμεσα σε παύλες στο όνομα ενός microRNA, και το αντίστοιχο όνομα microRNA που τις περιέχει.

search_tokens: Σε αυτόν τον πίνακα αποθηκεύονται όλοι οι τύποι κωδικών που υπάρχουν (MIMA, ENSEMBL, TRANSCRIPT, XM, XR, NM, NR), ο αριθμός του κάθε κωδικού και τέλος ανάλογα με τον τύπο ο κωδικός-κλειδί που αντιστοιχεί στο αντίστοιχο microRNA ή γονίδιο.

Η σχηματική αναπαράσταση των όσων περιγράφηκαν παραπάνω είναι η εξής:



Εικόνα 9: Σχηματική αναπαράσταση των πινάκων και συσχετίσεων μεταξύ τους που χρησιμοποιούνται για την παροχή προσωπικών υπηρεσιών στο σύστημα της παρούσας εφαρμογής.



Εικόνα 10: Σχηματική αναπαράσταση των πινάκων και των συσχετίσεων μεταξύ τους της βάσης δεδομένων της παρούσας εφαρμογής.

5

Υλοποίηση

Στο κεφάλαιο αυτό εξετάζονται αναλυτικά όλα τα θέματα υλοποίησης της εφαρμογής της παρούσας διπλωματικής εργασίας. Αρχικά, παρουσιάζονται οι δυνατότητες επιλογής μηχανών αποθήκευσης δεδομένων και με τη βοήθεια αυτών δικαιολογούμε τις επιλογές που κάναμε γύρω από αυτό το θέμα. Στη συνέχεια, αναλύεται λεπτομερώς ο μηχανισμός παροχής προτάσεων σε λανθασμένους όρους προς αναζήτηση για κάθε περίπτωση αναζήτησης. Στο τέλος του κεφαλαίου γίνεται αναφορά στα προγραμματιστικά εργαλεία που χρησιμοποιήθηκαν για την ανάπτυξη της εφαρμογής, καθώς και σε ενδιαφέροντα θέματα που προέκυψαν κατά τη διάρκεια αυτής.

5.1 Επιλογή μηχανής αποθήκευσης

Το σύστημα σχεσιακής βάσης δεδομένων MySQL είναι εξοπλισμένο με μια σειρά από μηχανές αποθήκευσης δεδομένων. Ανάλογα με τις ανάγκες μια εφαρμογής, ο προγραμματιστής είναι υπεύθυνος προκειμένου να επιλέξει την ή τις κατάλληλες μηχανές αποθήκευσης δεδομένων. Η MySQL υποστηρίζει τους εξής τύπους μηχανών αποθήκευσης δεδομένων:

- **MyISAM:** Είναι ο προκαθορισμένος τύπος μηχανής αποθήκευσης για την MySQL. Οι περισσότερες εφαρμογές παγκόσμιου ιστού χρησιμοποιούν κατά κύριο λόγο αυτόν τον τύπο μηχανής.
- **InnoDB:** Είναι ένας τύπος μηχανής αποθήκευσης δεδομένων που υποστηρίζει τις ιδιότητες ACID (atomicity, consistency, isolation, durability) και συνεπώς εγγυάται την πραγματοποίηση transactions με αξιόπιστο τρόπο. Επιπλέον, υποστηρίζονται οι λειτουργίες commit και rollback και μπορεί να εγγυηθεί την ακεραιότητα των δεδομένων μετά από απρόσμενο τερματισμό της βάσης δεδομένων. Τέλος υπάρχει η δυνατότητα τοποθέτησης ξένων κλειδιών μεταξύ των πινάκων της βάσης.
- **MEMORY:** Ο συγκεκριμένος τύπος μηχανής αποθήκευσης δεδομένων αποθηκεύει τα δεδομένα στη μνήμη RAM του συστήματος, συνεπώς η προσπέλαση αυτών είναι πολύ γρήγορη.
- **Merge:** Αυτός ο τύπος μηχανής επιτρέπει την προσπέλαση ενός αριθμού πινάκων τύπου MyISAM ως ένα αντικείμενο. Η χρήση αυτής της μηχανής είναι ιδιαίτερα χρήσιμη στην περίπτωση διαχείρισης μεγάλου όγκου δεδομένων.
- **Archive:** Αυτός ο τύπος μηχανής επιτρέπει την αποδοτική πρόσβαση σε μεγάλο όγκο από δεδομένα, τα όποια όμως προσπελάζονται σχετικά σπάνια. Η χρήση αυτής της μηχανής ενδείκνυται στην περίπτωση αποθήκευσης ιστορικών δεδομένων ή δεδομένων που προέρχονται από την καταγραφή ασφάλειας ενός συστήματος.
- **Federated:** Ο συγκεκριμένος τύπος μηχανής αποθήκευσης δεδομένων επιτρέπει τη διασύνδεση πολλών MySQL διακομιστών. Η χρήση του ενδείκνυται στην περίπτωση καταμεμημένων βάσεων δεδομένων.
- **CSV:** Αυτός ο τύπος μηχανής δίνει τη δυνατότητα αποθήκευσης δεδομένων σε αρχεία κειμένου χωρισμένων με ερωτηματικό. Η χρήση του είναι ενδεικτική σε περιπτώσεις όπου είναι απαραίτητη η μεταφορά δεδομένων μεταξύ εφαρμογών που διαχειρίζονται τέτοια αρχεία.
- **Blackhole:** Ο συγκεκριμένος τύπος μηχανής δέχεται δεδομένα χωρίς να τα αποθηκεύει. Η ανάκτηση δεδομένων από μια μηχανή τέτοιου τύπου επιστρέφει πάντα το κενό σύνολο. Χρησιμοποιείται σε ειδικές περιπτώσεις και σχετικά σπάνια.

Η βάση δεδομένων της προηγούμενης έκδοσης της εφαρμογής χρησιμοποιούσε μηχανές τύπου MyISAM για όλους τους πίνακες της βάσης. Η επιλογή αυτή έγινε καθαρά για λόγους επίδοσης και συνεπώς για ταχύτερη προσπέλαση των δεδομένων.

Στην παρούσα έκδοση της εφαρμογής προστέθηκε ένα επιπλέον χαρακτηριστικό που αφορά στη διαχείριση περισσότερων προγραμμάτων πρόβλεψης τα οποία είναι διαφορετικά από το πρόγραμμα της παρούσας διπλωματικής εργασίας. Για το λόγο αυτό προστέθηκε στη βάση δεδομένων ο πίνακας *diana_interactions* ο οποίος περιέχει τα προγράμματα πρόβλεψης και την ακρίβεια αυτών των προβλέψεων ενός *microRNA* πάνω σε ένα γονίδιο.

Κατά τη διάρκεια ανάπτυξης της παρούσας έκδοσης της εφαρμογής δόθηκε έμφαση στον επανασχεδιασμό του συστήματος αναζήτησης βιολογικών στόχων. Ένα ιδιαίτερο και ενδιαφέρον χαρακτηριστικό αυτού του συστήματος είναι και η παροχή προτάσεων στους όρους αναζήτησης στην περίπτωση που αυτοί δεν αντιστοιχούν σε έγκυρες ονομασίες *microRNA* ή γονιδίων. Η παροχή προτάσεων προς το χρήστη είναι μία διαδικασία που διαχειρίζεται μεγάλο όγκο δεδομένων, πλην όμως πρέπει να εκτελείται γρήγορα. Για το λόγο αυτό επιλέχθηκε η χρήση μηχανών τύπου MEMORY μόνο στους πίνακες οι οποίοι προσπελάζονται από το σύστημα παροχής προτάσεων σε όρους αναζήτησης.

Όπως αναφέρθηκε και προηγουμένως, οι μηχανές αποθήκευσης αυτού του τύπου αποθηκεύουν τα δεδομένα στη μνήμη RAM του συστήματος, συνεπώς η προσπέλαση των δεδομένων επιταχύνεται σε σημαντικά μεγαλύτερο βαθμό σε σχέση με κάποια άλλη μηχανή, όπως για παράδειγμα στις μηχανές MyISAM της προηγούμενης έκδοσης του μηχανισμού παροχής προτάσεων. Στην ενότητα που ακολουθεί περιγράφουμε τον τρόπο με τον οποίο λειτουργεί ο νέος βελτιωμένος μηχανισμός.

5.2 Μηχανισμός προτάσεων στις αναζητήσεις χρηστών του συστήματος

Κατά τη διάρκεια της αναζήτησης υπάρχει περίπτωση ο χρήστης να μην εισάγει σωστά τις ονομασίες των όρων για τους οποίους τον ενδιαφέρει να αναζητήσει βιολογικούς στόχους. Σε αυτήν την περίπτωση το σύστημα οφείλει να καταλαβαίνει με σχετικά έξυπνο τρόπο τι είδους αναζήτηση θέλει να πραγματοποιήσει ο χρήστης και να του παρέχει προτάσεις για αυτούς τους λανθασμένους όρους. Οι προτεινόμενοι όροι πρέπει να είσαι όσο το δυνατόν πιο όμοιοι με τον αρχικό όρο αναζήτησης.

Στην προηγούμενη έκδοση της εφαρμογής το σύστημα παροχής προτάσεων γνώριζε εκ των προτέρων τι είδους όρο αναζήτησης εισήγαγε ο χρήστης (*microRNA* ή γονίδιο), καθότι κάθε τύπος όρου προερχόταν από διαφορετική φόρμα. Ωστόσο το συγκεκριμένο σύστημα παρουσίαζε σημαντικές καθυστερήσεις.

Στην παρούσα έκδοση οι όροι προς αναζήτηση προέρχονται από μία μόνο φόρμα, συνεπώς υπάρχει ανάγκη το σύστημα να μαντεύει σωστά το είδος του όρου. Επιπλέον ο μηχανισμός παροχής προτάσεων οφείλει να είναι βελτιωμένος ως προς την ταχύτητα σε σχέση με αυτόν την προηγούμενης έκδοσης. Στην ενότητα αυτή περιγράφεται λεπτομερώς και με παραδείγματα ο τρόπος λειτουργίας του μηχανισμού της παρούσας εφαρμογής.

5.2.1 Έλεγχος ομοιότητας με την απόσταση *Levenshtein*

Στη εφαρμογή της παρούσας διπλωματικής εργασίας ο έλεγχος της ομοιότητας γίνεται κυρίως με τη μετρική της απόστασης *Levenshtein*. Η απόσταση *Levenshtein* μεταξύ δύο ακολουθιών είναι ο ελάχιστος αριθμός πράξεων εισαγωγής, διαγραφής ή αντικατάστασης χαρακτήρων που μπορεί να γίνει έτσι ώστε οι δύο ακολουθίες να γίνουν ταυτόσημες. Για παράδειγμα η απόσταση *Levenshtein* μεταξύ των ακολουθιών *jsad-* και *hsa-7* είναι “3” αφού απαιτείται μία πράξη αντικατάστασης, μία πράξη διαγραφής και μία πράξη εισαγωγής:

jsad- → *hsad-* (αντικατάσταση του “j” με το “h”)

hsad- → *hsa-* (διαγραφή του “d”)

hsa- → *hsa-7* (εισαγωγή του “7”)

Ο αλγόριθμος για τον υπολογισμό της απόστασης *Levenshtein* περιγράφεται σε μορφή ψευδοκώδικα ακολούθως:

```
int LevenshteinDistance(char s[1..m], char t[1..n])
{
    // d is a table with m+1 rows and n+1 columns
    declare int d[0..m, 0..n]

    for i from 0 to m
        d[i, 0] := i // deletion
    for j from 0 to n
        d[0, j] := j // insertion

    for j from 1 to n
    {
        for i from 1 to m
        {
            if s[i] = t[j] then
                d[i, j] := d[i-1, j-1]
            else
                d[i, j] := minimum
                    (
                        d[i-1, j] + 1, // deletion
```

```

        d[i, j-1] + 1, // insertion
        d[i-1, j-1] + 1 // substitution
    )
}
}

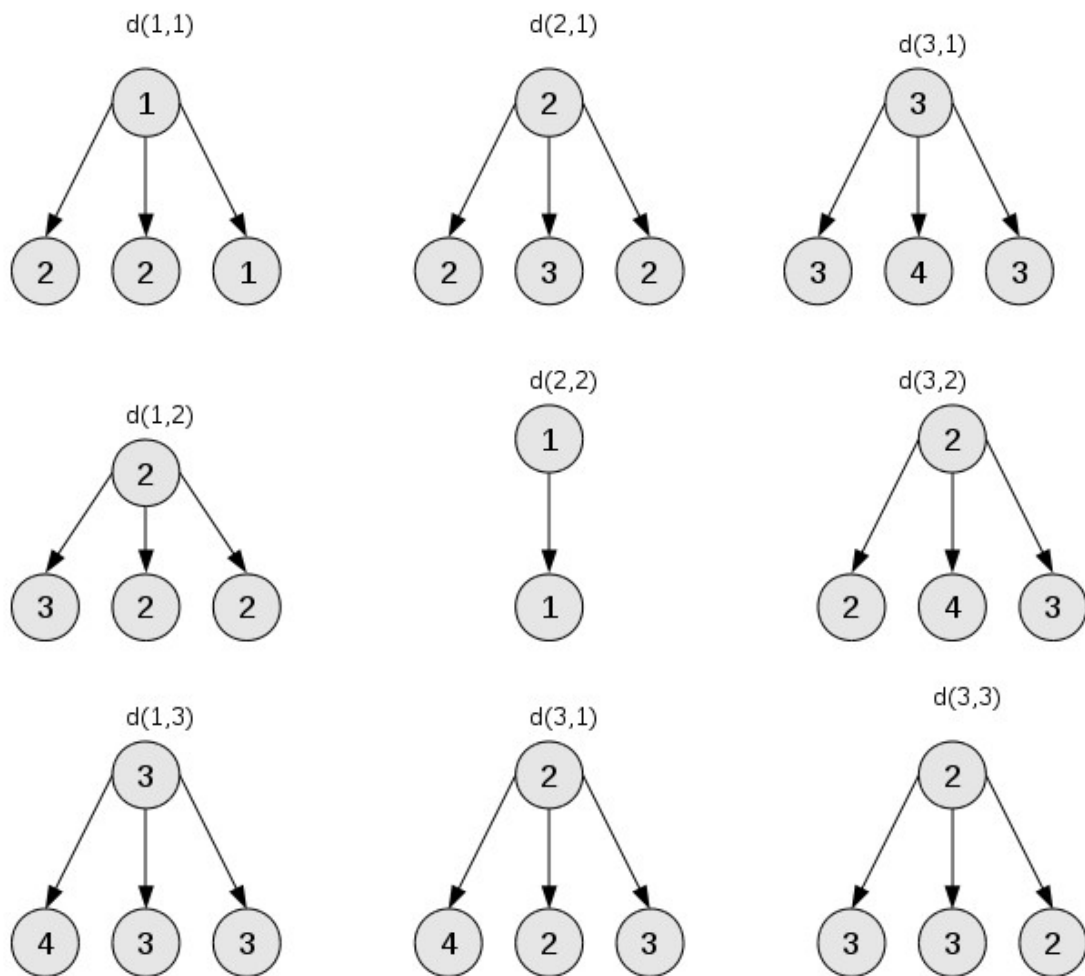
return d[m,n]
}

```

Ας υποθέσουμε ότι εφαρμόζουμε τον αλγόριθμο στις ακολουθίες hsa και jse. Ο διαμορφωμένος πίνακας d του αλγορίθμου φαίνεται ακολούθως:

		j	s	e
	0	1	2	3
h	1	1	2	3
s	2	2	1	2
a	3	3	2	<u>2</u>

Η ζητούμενη απόσταση είναι το στοιχείο d(3,3) δηλαδή 2. Πράγματι χρειάζονται δύο αντικαταστάσεις χαρακτήρων προκειμένου οι ακολουθίες να ταυτιστούν. Στην Εικόνα 9 παρουσιάζεται πως υπολογίζεται κάθε στοιχείο του πίνακα, καθώς και ποια στοιχεία συμμετέχουν σε κάθε βήμα αυτού. Στο στοιχείο d(2,2) είναι το μοναδικό για το οποίο ισχύει s[2]=t[2].



Εικόνα 11: Υπολογισμός των επιμέρους βημάτων του αλγορίθμου Levenshtein.

Ανάλογα με τον τύπο του όρου αναζήτησης (microRNA ή γονίδιο) αναπτύχθηκαν αλγόριθμοι στους οποίους γίνεται χρήση της απόστασης Levenshtein, πλην όμως είναι εξειδικευμένοι για κάθε περίπτωση αναζήτησης. Στην συνέχεια γίνεται παρουσίαση αυτών των αλγορίθμων για κάθε περίπτωση ξεχωριστά.

5.2.2 Παροχή προτάσεων σε ονομασίες microRNA

Ονομασία ενός microRNA μπορεί να είναι είτε το όνομα ή ο κωδικός MIMA αυτού. Το όνομα ενός microRNA συνήθως αποτελείται από μερικές λεκτικές μονάδες (tokens) χωρισμένες με παύλα (-) μεταξύ τους. Ο κωδικός MIMA είναι μία λέξη της μορφής MIMATxxxxxxx (όπου x ένα ψηφίο) για τα microRNA τύπου mature, ή μία λέξη της μορφής

MIxxxxxxx (όπου x ένα ψηφίο) για τα microRNA τύπου hairpin. Κάθε περίπτωση ονομασίας αντιμετωπίζεται ξεχωριστά:

I. Κωδικός MIMA

Όπως αναφέρθηκε προηγουμένως ο κωδικός MIMA ξεκινάει με τη συμβολοσειρά MI ή MIMAT ακολουθούμενη από 7 ψηφία. Συνήθως τα πρώτα ψηφία του 7ψήφιου αριθμού είναι μηδενικά, συνεπώς ένα συχνό λάθος ενός χρήστη είναι η παράλειψη κάποιων από αυτών των μηδενικών ή η τοποθέτηση περισσοτέρων μηδενικών. Παραδείγματα κωδικών MIMA είναι οι συμβολοσειρές MIMAT0000070 και MI0000800.

Ο αλγόριθμος αρχικά ελέγχει τα πέντε πρώτα γράμματα και αν αυτά είναι η συμβολοσειρά MIMAT τότε αντιλαμβάνεται ότι ο χρήστης αναζητά βιολογικούς στόχους για ένα microRNA τύπου mature. Αν αυτό δεν ισχύει ελέγχει τα δύο πρώτα γράμματα και αν αυτά είναι η συμβολοσειρά MI τότε αντιλαμβάνεται ότι ο χρήστης πραγματοποιεί αναζήτηση με βάση των κωδικό MIMA ενός microRNA τύπου hairpin. Για παράδειγμα αν ο χρήστης εισάγει τον όρο “**MIMAT005**” ο αλγόριθμος θα εξετάσει τα πέντε πρώτα γράμματα του όρου και θα αντιληφθεί ότι πρόκειται για όρο που σχετίζεται με ένα microRNA τύπου mature. Αντίθετα αν εισάγει τον όρο “**MI00000003**” θα εξετάσει στην αρχή τα πέντε πρώτα γράμματα και επειδή αυτά δεν αντιστοιχούν στη συμβολοσειρά MIMAT, θα εξετάσει τα δύο πρώτα τα οποία αντιστοιχούν επιτυχώς στη συμβολοσειρά “MI”. Με αυτόν τον τρόπο θα αντιληφθεί ότι πρόκειται για microRNA τύπου hairpin. Η συνέχεια από αυτό το σημείο είναι κοινή και για τις δύο περιπτώσεις, με τη διαφορά ότι με τον όρο prefix θα εννοούμε τη συμβολοσειρά MIMAT στην πρώτη περίπτωση (mature) και τη συμβολοσειρά MI (hairpin) στη δεύτερη περίπτωση.

Ο αλγόριθμος ελέγχει το πλήθος των ψηφίων μετά το prefix και αν αυτό είναι εφτά ελέγχει αν ο αρχικός όρος αντιστοιχεί σε έγκυρη ονομασία κωδικού MIMA. Σε περίπτωση που αντιστοιχεί σε έγκυρη ονομασία παρουσιάζει τα αντίστοιχα αποτελέσματα. Στην αντίθετη περίπτωση παίρνει διαδοχικά από τον πίνακα search_tokens όλα τα token με τύπο prefix και με χρήση της απόστασης Levenshtein βρίσκει τα 20 πιο όμοια token με τον 7ψήφιο αριθμό του κωδικού MIMA (πχ για το token “**000005**” τα είκοσι πιο όμοια tokens είναι “**0000705**”, “**0000095**” κτλ). Η συμβολοσειρά prefix συνενωμένη με κάθε ένα από τα είκοσι πιο πιθανά token αποτελεί έγκυρες ονομασίες microRNA τις οποίες ο χρήστης μπορεί να επιλέξει για την πραγματοποίηση αναζήτησης. Η λίστα των έγκυρων ονομασιών θα είναι “**MIMAT0000705**”, “**MIMAT0000095**” κτλ.

Αν το πλήθος ψηφίων που εξετάστηκε στο προηγούμενο βήμα είναι μικρότερο από επτά, ο αλγόριθμος παίρνει από το τέλος προς την αρχή τον μεγαλύτερο δυνατό μη μηδενικό αριθμό της συμβολοσειράς μετά το prefix και προσθέτει σε αυτόν τόσα μηδενικά ώστε να δημιουργηθεί ένας νέος επταψήφιος αριθμός. Η συμβολοσειρά prefix συνενώνεται με τον αριθμό που δημιουργήθηκε και επαναλαμβάνεται το προηγούμενο βήμα. Για παράδειγμα στην περίπτωση του όρου “**MIMAT005**” επειδή το πλήθος των ψηφίων του αριθμού “005” είναι τρία, θα συμπληρωθούν τα υπόλοιπα μηδενικά και ο νέος επταψήφιος αριθμός θα γίνει “0000005”. Η τελική συμβολοσειρά που θα δημιουργηθεί είναι η “**MIMAT0000005**”.

Τέλος, αν το πλήθος των ψηφίων που εξετάστηκε είναι μεγαλύτερο από επτά, ο αλγόριθμος αφαιρεί διαδοχικά από την αρχή προς το τέλος μηδενικά μέχρις ότου δημιουργηθεί ένας νέος επταψήφιος αριθμός. Η συμβολοσειρά prefix συνενώνεται με τον αριθμό που δημιουργήθηκε και επαναλαμβάνεται το αρχικό βήμα. Για παράδειγμα όταν ο όρος που εισήγαγε ο χρήστης είναι ο “**MI00000003**” και επειδή το πλήθος των ψηφίων του αριθμού “00000003” είναι οκτώ, θα αφαιρεθεί το επιπλέον μηδενικό. Ο νέος όρος προς αναζήτηση που θα δημιουργηθεί είναι ο “**MI0000003**”.

Ακολουθεί ο ψευδοκώδικας της περιγραφής που έγινε προηγουμένως.

```
search(term):
if term like MIMATn
    if n > 7 digit
        if term valid
            return results
        else
            similar_tokens = find from search_tokens 20 most
similar tokens to n, with type MIMAT
            foreach item in similar_tokens
                new_term = MIMAT.item
                add(item,suggestions)
            return suggestions
    else
        if n < 7 digit
            non_zero = find last non-zero number
            seven_digit = add zeros to non_zero to make a 7digit
number
            new_term = MIMAT.seven_digit
```

```

        search(new_term)
    else if n > 7 digit
        new_number = n
        while new_number > 7 digit
            new_number = remove zero from beginning of
new_number

            new_term = MIMAT.new_number
            search(new_term)
else if term like MIn
    if n 7 digit
        if term valid
            return results
        else
            similar_tokens = find from search_tokens 20 most
similar tokens to n, with type MI
            foreach item in similar_tokens
                new_term = MI.item
                add(item,suggestions)
            return suggestions
    else
        if n < 7 digit
            non_zero = find last non-zero number
            seven_digit = add zeros to non_zero to make a
7digit number

            new_term = MI.seven_digit
            search(new_term)
        else if n > 7 digit
            new_number = n
            while new_number > 7 digit
                new_number = remove zero from beginning of
new_number

            new_term = MI.new_number
            search(new_term)

```

Στη συνέχεια δίνεται ένα πιο αντιπροσωπευτικό παράδειγμα για την κατανόηση των όσων περιγράφηκαν μέχρι τώρα:

Έστω ότι ο πίνακας `search_tokens` της βάσης είναι ως εξής:

type	token
MIMAT	“0000042”
MI	“0000067”
MIMAT	“0000093”
MIMAT	“0000147”
MIMAT	“0000087”
MIMAT	“0000097”

Για παράδειγμα, αν ο χρήστης εισάγει την ονομασία `MIMAT00047`, τότε ο αλγόριθμος θα σχηματίσει την ονομασία `MIMAT000047` και αν αυτή αντιστοιχεί σε έγκυρη ονομασία θα εμφανίσει τα αντίστοιχα αποτελέσματα του συγκεκριμένου `microRNA`. Αν όχι, θα φέρει από τη βάση όλα `tokens` τύπου “MIMAT” από τη βάση, δηλαδή τα `tokens` “0000042”, “0000093”, “0000147”, “0000087” και “0000097”. Από αυτά τα `tokens` θα εντοπίσει τα πιο όμοια με τη συμβολοσειρά “0000047”, δηλαδή τα “0000147”, “0000087” και “0000097”. Σε κάθε ένα από αυτά τα `token` θα προσθέσει τη συμβολοσειρά “MIMAT” οπότε οι τελικοί όροι που θα προταθούν στο χρήστη θα είναι οι “MIMAT0000147”, “MIMAT0000087” και “MIMAT0000097”.

II. Όνομα *microRNA*

Στην περίπτωση που ο χρήστης αναζητά βιολογικούς στόχους με βάση το όνομα ενός `microRNA` αρχικά ελέγχεται αν ο όρος που εισήγαγε αντιστοιχεί σε έγκυρη ονομασία. Αν ναι γίνεται η παρουσίαση των αντίστοιχων αποτελεσμάτων. Αν όχι, ο αλγόριθμος χωρίζει το όνομα σε `tokens` με βάση τις παύλες που τα διαχωρίζουν.

Στη συνέχεια, για κάθε ένα από αυτά τα `tokens` ελέγχει αν αρχίζουν από κάποιο αριθμό. Αν ναι σπάει το `token` αυτό στον αριθμό αυτό και στο υπόλοιπο κομμάτι του `token` και έτσι δημιουργείται ένα επιπλέον `token`. Αν δεν ξεκινάει από κάποιον αριθμό δεν κάνει τίποτα.

Για κάθε στοιχείο της νέας λίστας με τα `token`, αν υπάρχει ο τύπος `token` στον πίνακα `mirna_tokens` της βάσης παίρνει όλα τα στοιχεία αυτού και τα εισάγει στην λίστα των υποψήφιων προτάσεων (λίστα A). Αν δεν υπάρχει αυτός ο τύπος παίρνει από τον πίνακα `mirna_tokens` όλους τους διαφορετικούς τύπους `token` και για κάθε ένα από αυτούς τους τύπους ο αλγόριθμος ελέγχει την απόσταση Levenshtein μεταξύ του τύπου και του `token`. Αν η απόσταση αυτή είναι μικρότερη από ένα κατώφλι, φέρνει από τη βάση όλα τα στοιχεία του συγκεκριμένου τύπου και τα προσθέτει στη λίστα των υποψήφιων προτάσεων (λίστα B).

Ταυτόχρονα, με όλη αυτή τη διαδικασία και για όλα τα tokens συνολικά, ο αλγόριθμός επισημαίνει την ελάχιστη απόσταση μεταξύ του token και του τύπου (d). Με βάση των αριθμό των tokens που υπάρχει αντίστοιχος τύπος στη βάση ορίζουμε ένα αριθμό (x). Στη συνέχεια κανονικοποιούμε τη λίστα A διαιρώντας κάθε στοιχείο αυτής με τον αριθμό x και πολλαπλασιάζοντας το αποτέλεσμα με τον αριθμό 0.5. Αντίστοιχα, κανονικοποιούμε τη λίστα B διαιρώντας κάθε στοιχείο αυτής με τον αριθμό d και πολλαπλασιάζοντας το αποτέλεσμα με τον αριθμό 0.5. Τέλος συγχωνεύουμε τις δύο λίστες, και όπου έχουμε κοινά στοιχεία προσθέτουμε τους δύο αριθμούς που βρήκαμε προηγουμένως. Από την τελική κανονικοποιημένη λίστα επιλέγουμε τα είκοσι στοιχεία με τους υψηλότερους αριθμούς και τα προτείνουμε στο χρήστη.

Αξίζει να σημειωθεί ότι οι προαναφερθέντες λίστες στην πραγματικότητα υλοποιήθηκαν με τη βοήθεια hash πινάκων, στους οποίους ισχύει ότι η αναζήτηση με βάση ένα κλειδί είναι πιο γρήγορη από κάθε άλλη δομή (πχ απλή λίστα).

Ακολουθεί ο αλγόριθμος σε μορφή ψευδοκώδικα.

```

search(term):
if term valid mirna_name
    return results;
else
    list_of_tokens = split term by -
    foreach item in list_of_tokens
        if item begins with number
            new_token1 = number
            new_token2 = token - {number}
            add(new_token1,list_of_tokens)
            add(new_token2,list_of_tokens)
x = number of elements in list_of_tokens
foreach item in list_of_tokens
    if item exists in mirna_tokens(token)
        listA = find from mirna_tokens all mirna_names with type =
token
    else
        types = find all tokens from mirna_tokens
        foreach type in types
            dis = levenshtein(item,type)
            if dis < min_dis
                min_dis = dis
            if dis < threshold
                listB = find from mirna_tokens all mirna_names with
type = type
foreach itemA in listA
    new_itemA = (itemA / x)*0.5
foreach itemB in listB
    new_itemB = (itemB / min_dis)*0.5
    if itemB in listA
        new_item = new_itemA + new_itemB
list = append(listA,listB)
suggestions = select 20 highest scored elements from list
return suggestions

```

Για παράδειγμα αν ο χρήστης εισάγει προς αναζήτηση τον όρο “**hsa-let-100go**”, τότε ο αλγόριθμος θα σχηματίσει μία λίστα με τα token (hsa, let, go,100). Στη συνέχεια, για τα token “hsa”, “let” και “100” που υπάρχουν στη βάση, θα φέρει όλα τα έγκυρα ονόματα microRNA που περιέχουν τα συγκεκριμένα token (πχ **hsa-mir-454**, **hsa-mir-19b**, **hsa-let-7a**, **mmu-let-7c**, **hsa-let-7a**, **hsa-mir-100** κτλ) και θα σχηματίσει τη λίστα Α. Για το token “go” που δεν υπάρχει στη βάση, θα βρει τα πιο όμοια σε αυτά token πχ (hsa, mir) και για αυτά τα token θα κάνει ότι έκανε στο προηγούμενο βήμα. Έτσι θα σχηματιστεί η λίστα Β (**hsa-mir-454**, **hsa-let-7a**, **mmu-mir-10b** κτλ). Με βάση τη φόρμουλα που χρησιμοποιεί ο αλγόριθμος η λίστα Α θα γίνει [(**hsa-let-7a,0.25**),(**hsa-mir-19b,0.125**),(**mmu-let-7c,0.125**),(**hsa-mir-100,0.125**) κτλ] και η λίστα Β (έστω ότι η ελάχιστη απόσταση είναι 1) θα γίνει [(**hsa-mir-454,0.5**),(**hsa-let-7a,0.5**),(**mmu-mir-10b,0.5**) κτλ]. Η συγχωνευμένη λίστα θα είναι [(**hsa-let-7a,0.75**),(**hsa-mir-454,0.5**),(**mmu-mir-10b,0.5**),(**hsa-mir-19b,0.125**),(**mmu-let-7c,0.125**),(**hsa-mir-100,0.125**), κτλ]. Από την τελική λίστα θα επιλεγούν οι είκοσι έγκυρες ονομασίες με το πιο υψηλό σκορ και θα παρουσιαστούν στον χρήστη ως προτεινόμενες.

5.2.3 Παροχή προτάσεων σε ονομασίες γονιδίων

Ονομασία ενός γονιδίου μπορεί να είναι είτε το όνομα του γονιδίου ή κάποιος από τους κωδικούς ENSEMBL ή TRANSCRIPT ή Refseq αυτού. Ο κωδικός ENSEMBL είναι μια λέξη που ξεκινά με το πρόθεμα ENSG ή ENSMUSG ακολουθούμενου από 11 ψηφία. Ο κωδικός TRANSCRIPT είναι μία λέξη που ξεκινά με το πρόθεμα ENST ή ENSMUST ακολουθούμενου από 11 ψηφία. Από τα 11 αυτά ψηφία, στην αρχή συναντώνται κυρίως μηδενικά. Τέλος ο κωδικός Refseq είναι μία λέξη που ξεκινά με NM_, ή NR_, ή XM_, ή XR_.

Στην περίπτωση αναζήτηση βιολογικών στόχων με βάση κάποιο κωδικό ο μηχανισμός παροχής προτάσεων μοιάζει σε μεγάλο βαθμό με αυτόν για τον κωδικό MIMA. Στη συνέχεια παρουσιάζονται οι μηχανισμοί παροχής προτάσεων για κάθε περίπτωση.

I. Κωδικός ENSEMBL ή TRANSCRIPT

Προκειμένου να ανιχνευθεί η περίπτωση αυτή ελέγχονται τα πρώτα τέσσερα και τα πρώτα επτά γράμματα του όρου προς αναζήτηση. Επειδή η διαδικασία είναι κοινή, με τη διαφορά ότι όπου έχουμε πρόθεμα ENSG ή ENSMUSG για τον κωδικό ENSEMBL έχουμε αντίστοιχα πρόθεμα ENST ή ENSMUST για τον κωδικό TRANSCRIPT, περιγράφουμε την περίπτωση αναζήτησης με βάση τον κωδικό ENSEMBL. Η τεχνική του αλγορίθμου είναι κοινή με την

τεχνική που χρησιμοποιήθηκε στην περίπτωση κωδικών MIMA με τη διαφορά ότι το πλήθος των ψηφίων σε αυτή την περίπτωση πρέπει να είναι έντεκα αντί για εφτά.

Ο αλγόριθμος αρχικά διαβάζει τα τέσσερα πρώτα γράμματα του όρου. Αν αυτά αντιστοιχούν στη συμβολοσειρά ENSG τότε ελέγχει το πλήθος των ψηφίων του αριθμού που ακολουθεί τη συμβολοσειρά. Αν το πλήθος είναι 11 ελέγχεται αν ο κωδικός αντιστοιχεί σε έγκυρο κωδικό ENSEMBL και στην περίπτωση που αυτό ισχύει παρουσιάζονται τα ανάλογα αποτελέσματα. Σε αντίθετη περίπτωση βρίσκουμε με όμοιο τρόπο όπως στην περίπτωση του κωδικού MIMA, τις είκοσι πιο όμοιες αριθμητικές συμβολοσειρές τύπου ENSG από τον πίνακα search_tokens και συνενώνουμε κάθε μία από αυτές τις συμβολοσειρές με το πρόθεμα ENSG. Αν το πλήθος των ψηφίων του αριθμού που ακολουθούν το πρόθεμα είναι διαφορετικό του έντεκα, τότε συμπληρώνονται ή αφαιρούνται τα επιπλέον μηδενικά.

Στην περίπτωση που τα τέσσερα πρώτα γράμματα δεν αντιστοιχούν στη συμβολοσειρά ENSG, ελέγχεται αν τα εφτά πρώτα γράμματα αντιστοιχούν στη συμβολοσειρά ENSMUSG, οπότε και πράττουμε δυσαναλογίας με τα όσα περιγράφηκαν μέχρι τώρα.

Στη συνέχεια ακολουθεί ο αλγόριθμος που περιγράφηκε σε μορφή ψευδοκώδικα.

```
search(term):
if term like ENSGn
    if n 11 digit
        if term valid
            return results
        else
            similar_tokens = find from search_tokens 20 most
similar tokens to n, with type ENSG
            foreach item in similar_tokens
                new_term = ENSG.item
                add(item,suggestions)
            return suggestions
    else
        if n < 11 digit
            non_zero = find last non-zero number
            eleven_digit = add zeros to non_zero to make a 11digit
number
            new_term = ENSG.eleven_digit
            search(new_term)
        else if n > 11 digit
            new_number = n
            while new_number > 11 digit
                new_number = remove zero from beginning of
new_number
            new_term = ENSG.new_number
            search(new_term)
else if term like ENSMUSGn
    if n 11 digit
        if term valid
            return results
        else
            similar_tokens = find from search_tokens 20 most
```

```

similar tokens to n, with type ENSMUSG
    foreach item in similar_tokens
        new_term = ENSMUSGM.item
        add(item,suggestions)
    return suggestions
else
    if n < 11 digit
        non_zero = find last non-zero number
        eleven_digit = add zeros to non_zero to make a
11digit number
        new_term = ENSMUSG.eleven_digit
        search(new_term)
    else if n > 11 digit
        new_number = n
        while new_number > 11 digit
            new_number = remove zero from beginning of
new_number
        new_term = ENSMUSG.new_number
        search(new_term)

```

Για παράδειγμα έστω ότι ο χρήστης εισάγει στην μπάρα αναζήτησης τον όρο ENSG0001. Τότε ο αλγόριθμος θα καταλάβει ότι ο χρήστης κάνει αναζήτηση με βάση τον κωδικό ENSEMBL και θα σχηματίσει τον αριθμό 0000000001, και μετέπειτα τον όρο ENS0000000001. Αν ο νέος όρος αντιστοιχεί σε έγκυρο κωδικό ENSEMBL κάποιου γονιδίου, θα παρουσιαστούν στο χρήστη τα αντίστοιχα αποτελέσματα. Αν όχι, θα αναζητήσει τα είκοσι πιο όμοια token τύπου ENSG από τον πίνακα search_tokens της βάσης, και θα προτείνει τους όρους ENSG{token}. Η διαδικασία αυτή πραγματοποιείται όπως ακριβώς στο παράδειγμα που περιγράφηκε για τους κωδικούς MIMA.

Αντίθετα, αν ο χρήστης εισάγει τον όρο ENSG0000000000000001, ο αλγόριθμος θα σχηματίσει τον όρο ENSG0000000001 και θα συνεχίσει όπως περιγράφηκε προηγουμένως.

II. Κωδικός Refseq

Προκειμένου να ανιχνευθεί η περίπτωση αυτή ελέγχονται τα τρία πρώτα γράμματα του όρου και αν αυτά αντιστοιχούν σε μία από τις συμβολοσειρές NM_, NR_, XM_ ή XR_ λειτουργεί όπως περιγράφηκε προηγουμένως. Η διαφορά σε αυτήν την περίπτωση είναι ότι το πλήθος ψηφίων είναι μεταβλητό, οπότε το στάδιο για την τροποποίηση του πλήθους των ψηφίων παρακάμπτεται.

III. Όνομα γονιδίου

Η περίπτωση αυτή ανιχνεύεται όταν δεν ισχύει καμία από τις προηγούμενες περιπτώσεις. Όταν ο χρήστης εισάγει ένα μη έγκυρο όνομα γονιδίου, ο αλγόριθμος απλά εντοπίζει τα είκοσι πιο όμοια ονόματα γονιδίων από τον πίνακα genes_tokens και τα προτείνει στο χρήστη. Και πάλι ο έλεγχος της ομοιότητας γίνεται με χρήση της απόστασης Levenshtein.

Για παράδειγμα έστω ότι ο πίνακας της βάσης περιέχει τα εξής δεδομένα:

token
RNASEL
FIGN
MIER1
FIGF
FIGNL2

Αν ο χρήστης εισάγει τον όρο “**FIGNG**” τότε ο αλγόριθμος θα προτείνει στο χρήστη μόνο τα ονόματα “**FIGN**”, “**FIGF**” και “**FIGNL2**”.

5.3 Πλατφόρμες και προγραμματιστικά εργαλεία

Για την ανάπτυξη της εφαρμογής της παρούσας διπλωματικής εργασίας επιλέχθηκε μια σειρά από πλατφόρμες ανάπτυξης και εκτέλεσης λογισμικού και προγραμματιστικά εργαλεία. Η εφαρμογή ανήκει στην κατηγορία των δικτυακών εφαρμογών και για αυτό το λόγο τα εργαλεία που αναλύονται στην συνέχεια σχετίζονται άμεσα με την παροχή υπηρεσιών παγκόσμιου ιστού.

5.3.1 Apache HTTP Server

Ο Apache HTTP Server είναι ο πιο διαδεδομένος διακομιστής σελίδων παγκόσμιου ιστού. Τα χαρακτηριστικά που διαθέτει είναι ιδιαίτερα ελκυστικά και είναι ένας από τους πιο γρήγορους διακομιστές σελίδων. Μπορεί να λειτουργήσει σε διάφορα λειτουργικά συστήματα, με πιο διαδεδομένο το συνδυασμό του με το λειτουργικό σύστημα GNU/Linux. Ο Apache ανήκει στην κατηγορία του ανοιχτού λογισμικού υπό την άδεια Apache License. Επιπλέον παρέχεται υποστήριξη και ανάπτυξη από μια μεγάλη και ανοιχτή κοινότητα προγραμματιστών του Apache Software Foundation.

Τα περισσότερα χαρακτηριστικά του Apache HTTP Server υλοποιούνται ως ξεχωριστές δομικές μονάδες (modules). Τα κυριότερα χαρακτηριστικά του είναι:

- Υποστήριξη server side γλωσσών προγραμματισμού (πχ PHP, Perl, Python, κτλ).
- Ταυτοποίηση χρηστών με τη βοήθεια διαδεδομένων συστημάτων ταυτοποίησης.
- Ασφαλή μεταφορά πληροφορίας μέσω του πρωτοκόλλου TLS.
- Συμπύεση της μεταφερόμενης πληροφορίας για γρήγορες επιδόσεις.

Τα χαρακτηριστικά αυτά αποτελούν και το λόγο για τον οποίο επιλέχθηκε η χρήση του ως διακομιστή στην ανάπτυξη της παρούσας διπλωματικής εργασίας.

5.3.2 MySQL

Η MySQL είναι ένα διαδεδωμένο σύστημα σχεσιακής βάσης δεδομένων που χρησιμοποιείται σε πληθώρα σύγχρονων εφαρμογών. Όπως και στην περίπτωση του Apache χρησιμοποιείται πολύ συχνά συνδυασμένη με το λειτουργικό σύστημα GNU/Linux και είναι λογισμικό ανοιχτού κώδικα υπό την άδεια GNU General Public License (GPL). Επιπλέον, συναντάται συχνά ο συνδυασμός της με τον Apache HTTP Server. Τα κυριότερα χαρακτηριστικά της είναι:

- Πρόσβαση πολλαπλών χρηστών σε πολλαπλές βάσεις δεδομένων ταυτόχρονα.
- Υποστήριξη ANSI SQL 99.
- Υποστήριξη για stored procedures, cursors και triggers.
- Υποστήριξη για updatable views.
- Μεγάλη πληθώρα διαφορετικών τύπων μηχανών αποθήκευσης δεδομένων (MyISAM, InnoDB, MEMORY κτλ).
- Υποστήριξη full-text indexing.

Τα χαρακτηριστικά αυτά αποτελούν ενδεδειγμένη την χρήση της στην ανάπτυξη της παρούσας εφαρμογής με σκοπό την διαχείριση των δεδομένων αυτής.

5.3.3 PHP

Η PHP είναι μία scripting γλώσσα προγραμματισμού και χρησιμοποιείται κατά κύριο λόγο στην ανάπτυξη δικτυακών εφαρμογών. Μπορεί να συνδυαστεί με διάφορα λειτουργικά συστήματα και ανήκει και αυτή στα εργαλεία ανοιχτού λογισμικού. Διανέμεται υπό την άδεια PHP License.

Η PHP υποστηρίζει τόσο τη δημιουργία στατικών σελίδων όσο και την ανάπτυξη εφαρμογών γραμμής εντολών ή γραφικού interface. Οι επιδόσεις της είναι ιδιαίτερα καλές, γεγονός που καθιστούν τη χρήση της ενδεδειγμένη σε δικτυακές εφαρμογές. Επιπρόσθετα (από την έκδοση 5.0 και μετά) υποστηρίζεται πλήρως το αντικειμενοστρεφές μοντέλο προγραμματισμού με αποτέλεσμα να είναι ιδανική για την ανάπτυξη μεγάλων εφαρμογών λογισμικού. Για το λόγο αυτό μπορεί να γίνει χρήση όλων των σύγχρονων τεχνικών της Τεχνολογίας Λογισμικού.

Τα χαρακτηριστικά που συνοπτικά περιγράψαμε προηγουμένως, καθιστούν τη γλώσσα PHP ιδανική επιλογή για την ανάπτυξη της παρούσας διπλωματικής εργασίας. Επιπλέον ο συνδυασμός της τόσο με τον Apache HTTP Server όσο και με την MySQL σχεσιακή βάση δεδομένων είναι ιδιαίτερα διαδεδομένος και χρήζει μεγάλης προγραμματιστικής ευκολίας. Τέλος, η προηγούμενη έκδοση της εφαρμογής είχε αναπτυχθεί με τη βοήθεια της PHP, συνεπώς η επιλογή αυτής αποτέλεσε μονόδρομο.

5.4 Άλλα θέματα

Σε αυτήν την ενότητα γίνεται αναφορά σε δύο σημεία στα οποία δόθηκε ιδιαίτερη έμφαση κατά την ανάπτυξη της παρούσας έκδοσης της εφαρμογής, την τεκμηρίωση του κώδικα και την διόρθωση σφαλμάτων της προηγούμενης έκδοσης.

5.4.1 Τεκμηρίωση

Η προηγούμενη έκδοση της εφαρμογής αποτελούνταν από περίπου 7.000 γραμμές κώδικα PHP. Από όλο το σύνολο των αρχείων κώδικα, ελάχιστα περιείχαν επαρκή τεκμηρίωση για τη λειτουργία που επιτελούν. Οι δυνατότητες που παρέχονται για επιπλέον επέκταση αυτής, καθιστούν απαραίτητη την τεκμηρίωση αυτής.

Για το λόγο αυτό κρίθηκε αναγκαίο να τεκμηριωθούν σωστά με βάση το πρότυπο Doxygen τόσο τα αρχεία της προηγούμενης έκδοσης όσο και τα αρχεία που δημιουργήθηκαν για την ανάπτυξη της νεότερης έκδοσης.

Η παρούσα έκδοση αριθμεί περίπου 11.000 γραμμές κώδικα PHP. Σε κάθε αρχείο αυτής αναγράφονται οι συγγραφείς του αρχείου καθώς και το ρόλο που επιτελεί το κάθε το αρχείο. Για κάθε μέθοδο δίνεται μια σύντομη, πλην όμως περιεκτική περιγραφή της λειτουργίας της , των παραμέτρων που δέχεται ως είσοδο και του επιστρεφόμενου αποτελέσματος.

Η τεκμηρίωση της εφαρμογής θα αποτελέσει ένα σημαντικό εργαλείο για τους προγραμματιστές που μελλοντικά θα αναλάβουν την επέκταση αυτής, επιταχύνοντας την διαδικασία εξοικείωσης και κατανόησης της λειτουργίας της.

5.4.2 Αναφορά σφαλμάτων

Η ανάπτυξη της προηγούμενης έκδοσης της εφαρμογής πραγματοποιήθηκε έχοντας ορίσει για την PHP, να πραγματοποιεί αναφορά σφαλμάτων (error reporting) μόνο για τις

περιπτώσεις όπου αυτά τα λάθη δεν οφείλονταν σε warnings ή notices. Ωστόσο, αυτό έκρυβε κινδύνους καθότι η εφαρμογή συνέχιζε να λειτουργεί σε περιπτώσεις όπου δεν θα έπρεπε.

Για το λόγο αυτό κρίθηκε απαραίτητο η ανάπτυξη της παρούσας έκδοσης της εφαρμογής να πραγματοποιηθεί ορίζοντας για την PHP καθολική αναφορά σφαλμάτων. Αυτό πρακτικά σημαίνει ότι η ύπαρξη ενός μόνο warning ή notice, αρκεί για την έγερση σφάλματος. Η ρύθμιση αυτή εγγυάται την αξιόπιστη λειτουργία της εφαρμογής σε συμφωνία με τα όσα ορίζουν οι λειτουργικές απαιτήσεις αυτής.

6

Έλεγχος και αξιολόγηση

Στο κεφάλαιο αυτό παρουσιάζεται η μεθοδολογία ελέγχου που επιλέχθηκε για την διαπίστωση της αξιοπιστίας του συστήματος. Ειδικότερα, αναλύονται με τη χρήση των σεναρίων εκτέλεσης και με τη βοήθεια εικόνων όλες οι επιμέρους λειτουργίες που μπορεί να επιτελέσει το σύστημα. Στο τέλος του κεφαλαίου αξιολογείται ο νέος μηχανισμός παραχής προτάσεων σε σχέση με τον μηχανισμό της προηγούμενης έκδοσης της εφαρμογής.

6.1 Μεθοδολογία ελέγχου

Ο έλεγχος του συστήματος επιλέχθηκε να γίνει με τη βοήθεια των σεναρίων εκτέλεσης. Για κάθε διαφορετικό σενάριο εκτέλεσης, δηλαδή για κάθε διαφορετική λειτουργία που θέλει να πραγματοποιήσει ο χρήστης, παρουσιάζονται με εικόνες τα επιμέρους βήματα και γίνεται λεκτική περιγραφή όλης της διαδικασίας.

Συνολικά παρουσιάζονται τα εξής σενάρια εκτέλεσης:

- Δημιουργία νέου λογαριασμού χρήστη
- Αναζήτηση με βάση το όνομα του microRNA
- Αναζήτηση με βάση τον κωδικό MIMA

- Αναζήτηση με βάση το όνομα του γονιδίου
- Αναζήτηση με βάση τον κωδικό ENSEMBL
- Αναζήτηση με βάση τον κωδικό TRANSCRIPT
- Αναζήτηση με βάση τον κωδικό Refseq
- Διαχείριση αγαπημένων αναζητήσεων και ιστορικού
- Παροχή προτάσεων σε λανθασμένο όνομα microRNA
- Παροχή προτάσεων σε λανθασμένο κωδικό MIMA
- Παροχή προτάσεων σε λανθασμένο όνομα γονιδίου
- Παροχή προτάσεων σε λανθασμένο κωδικό ENSEMBL
- Παροχή προτάσεων σε λανθασμένο κωδικό TRANSCRIPT
- Παροχή προτάσεων σε λανθασμένο κωδικό Refseq
- Σύνδεση απλού χρήστη στο πάνελ διαχείρισης
- Προβολή λίστας χρηστών
- Διαγραφή χρήστη από το σύστημα
- Τροποποίηση στοιχείων χρήστη του συστήματος
- Εισαγωγή χρήστη στην ομάδα διαχειριστών
- Πρόσθεση νέου χρήστη
- Ανέβασμα αρχείου στόχων

6.2 Αναλυτική παρουσίαση ελέγχου

Σε αυτήν την ενότητα περιγράφουμε τόσο με τη χρήση εικόνων όσο και λεκτικά τα διάφορα σενάρια εκτέλεσης.

6.2.1 Δημιουργία νέου λογαριασμού χρήστη

Εστω ότι θέλουμε να δημιουργήσουμε έναν νέο λογαριασμό χρήστη με το όνομα “georgia”. Τότε πραγματοποιούμε εγγραφή πατώντας το σύνδεσμο “**New account**” για τη δημιουργία νέου λογαριασμού. Στη συνέχεια παρουσιάζεται η ακόλουθη φόρμα:

A screenshot of a web registration form. It contains four input fields: 'Username' with the text 'georgia', 'Password' with four dots, 'Password Repeat' with four dots, and 'Email' with the text 'gf@c.com'. Below the fields is a 'Register' button.

Εικόνα 12: Συμπλήρωση στοιχείων χρήστη για τη δημιουργία νέου λογαριασμού

Αφού συμπληρώσουμε τα στοιχεία μας, πατάμε το κουμπί **“Register”** για την εγγραφή μας στο σύστημα.

Στη βάση εμφανίζεται η ακόλουθη η εγγραφή:

```
| 49 | georgia | r00t | 0 | gf@c.com | 0 |
```

Εικόνα 13: Εγγραφή του νέου λογαριασμού χρήστη στη βάση δεδομένων του συστήματος.

Όπως μπορούμε να δούμε ο χρήστης “georgia” προστέθηκε επιτυχώς στο σύστημα.

6.2.2 Αναζήτηση με βάση το όνομα του *microRNA*

Στο συγκεκριμένο σενάριο εκτέλεσης ο χρήστης “georgia” θέλει να πραγματοποιήσει αναζήτηση με βάση το όνομα του *microRNA*.

A screenshot of a search interface titled 'DIANA - microT v5.0 (beta version)'. It features a search input field containing the text 'hsa-let-7e'.

Εικόνα 14: Εισαγωγή του ονόματος *microRNA* στην μπάρα αναζήτησης.

Αφού εισάγει στο όνομα του microRNA για το οποίο ενδιαφέρεται να ανακτήσει αποτελέσματα βιολογικών στόχων (“**hsa-let-7e**”), πατάει το κουμπί της υποβολής “**VIEW RESULTS**”. Στη συνέχεια εμφανίζεται η φόρμα των αποτελεσμάτων:

Results for miRNA hsa-let-7e (3366 target sites in 2488 genes found) | [New Search](#)

miRNA name: **hsa-let-7e**

Alternative Description: MIMAT0000066

miRNA Sequence: UGAGGUAGGAGGUUGUAUAGUU

Related hairpins: hsa-let-7e

Score Threshold: KEGG Filter:

miRNA name: **hsa-let-7e**

Alternative Description: MIMAT0000066

miRNA Sequence: UGAGGUAGGAGGUUGUAUAGUU

Related hairpins: hsa-let-7e

Rank	Gene name	Ensembl Gene Id	miTG score	Precision	SNR	Also predicted by
1	FIGN	ENS00000182263	0.928784	1	18.9684	
2	ARID3B	ENS00000179361	0.927275	1	18.9684	>>
3	CLCN5	ENS00000171365	0.923375	1	18.9684	
4	HMG2	ENS00000149948	0.871348	0.993333	18.9684	>>
5	C14ORF28	ENS00000179476	0.867339	0.993333	18.9684	
6	FOXO2	ENS00000126773	0.832781	0.985263	18.9684	
7	TRMT1	ENS00000780757	0.819961	0.885263	18.9684	
8	TMAT1	ENS00000164400	0.810768	0.885263	18.9684	
9	PAPPAS,PAPPA	ENS00000187750	0.788098	0.838095	18.9684	>>
10	IGDC3	ENS00000171980	0.786540	0.838095	18.9684	>>



Εικόνα 15: Εμφάνιση αποτελεσμάτων αναζήτησης με βάση το όνομα microRNA

Στη φόρμα των αποτελεσμάτων αρχικά εμφανίζεται ένας πίνακας με το όνομα του microRNA, ο κωδικός MIMA αυτού και κάποιες πρόσθετες πληροφορίες. Επιπλέον υπάρχει

η δυνατότητα ορισμού των παραμέτρων όπως το σκορ ή το KEGG pathway. Στο κάτω μέρος της φόρμα εμφανίζονται οργανωμένα τα αποτελέσματα με βάση το όνομα του γονιδίου. Ο χρήστης μπορεί να πλοηγηθεί σε όλες τις σελίδες αποτελεσμάτων απλά ορίζοντας τη σελίδα που επιθυμεί με το βοήθεια του κατάλληλου πλήκτρου.

Ας υποθέσουμε σε αυτό το σημείο ότι ο χρήστης θέλει να εντοπίσει όλους τους βιολογικούς στόχους για το συγκεκριμένο microRNA με σκορ μεγαλύτερο του 0.6 (strict) και ότι θέλει να ορίσει την παράμετρο KEGG pathway ίση με “cancer”. Αφού εισάγει στην ειδική φόρμα κάθε παραμέτρου την τιμή που επιθυμεί πατάει το κουμπί υποβολής μιας από αυτές. Τα νέα αποτελέσματα εμφανίζονται και ο χρήστης επιλέγει να δει την σελίδα τέσσερα όπως παρουσιάζεται ακολούθως:

Results for miRNA hsa-let-7e (256 target sites in 100 genes found) | [New Search](#)






miRNA name: **hsa-let-7e**  

Alternative Description: MIMAT0000066

miRNA Sequence: UGAGGUAGGAGGUUGUAUGUU

Related hairpins: hsa-let-7e

[Show](#) / [Hide](#) expression level graph

Score Threshold  KEGG Filter    Page 4 

Rank	Gene name	Ensembl Gene Id	miTG score	Precision	SNR	Also predicted by
1	CCND2	ENSG00000118971	0.712707	0.971212	18.9684	>>
2	ACVR2A	ENSG00000121989	0.707492	0.971212	18.9684	
3	GATM	ENSG00000171766	0.705093	0.971212	18.9684	>>
4	PRTG	ENSG00000166450	0.698526	0.969192	18.9684	
5	ZNF644	ENSG00000122482	0.695215	0.969192	18.9684	>>
6	ARHGAP28	ENSG00000088756	0.690693	0.969192	18.9684	
7	LRIG3	ENSG00000139263	0.689928	0.969192	18.9684	>>
8	COL1A2	ENSG00000164692	0.688929	0.969192	18.9684	>>
9	MEIS3	ENSG00000105419	0.688858	0.969192	18.9684	
10	DLC1	ENSG00000164741	0.688039	0.969192	18.9684	>>

[expand all](#) | [collapse all](#)

Εικόνα 16: Εμφάνιση αποτελεσμάτων της σελίδας 4 με ορισμό των παραμέτρων *threshold* και *KEGG pathway*

Ο χρήστης έχει τη δυνατότητα να δει τη λίστα προγραμμάτων που εντόπισαν τον ίδιο βιολογικό στόχο με το πρόγραμμα της παρούσας εφαρμογής, απλά πατώντας το σύνδεσμο

6.2.3 Αναζήτηση με βάση τον κωδικό MIMA

Στο συγκεκριμένο σενάριο εκτέλεσης ο χρήστης αναζητά βιολογικούς στόχους με βάση τον κωδικό MIMA. Αφού εισάγει τον κωδικό MIMA του miRNA που τον ενδιαφέρει στην μπάρα αναζήτησης (“MIMAT0000063”), εμφανίζεται η λίστα των αποτελεσμάτων:

DIANA - microT v5.0 (beta version)

MIMAT0000063

VIEW RESULTS

Advanced Search Help

Results for miRNA hsa-let-7b (3181 target sites in 2344 genes found) | New Search

miRNA name: hsa-let-7b

Alternative Description: MIMAT0000063

miRNA Sequence: UGAGGUAGUAGGUUGUGUGGUU

Related hairpins: hsa-let-7b

Show / Hide expression level graph

Score Threshold: medium 0

KEGG Filter

Page 1

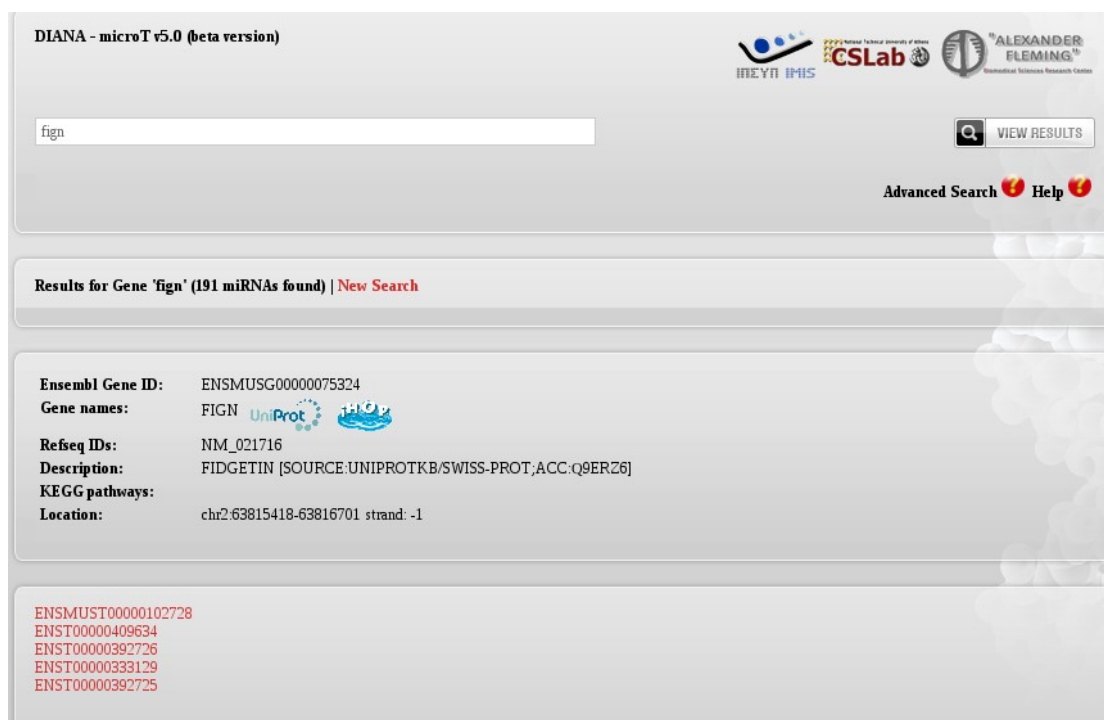
Rank	Gene name	Ensembl Gene Id	miTG score	Precision	SNR	Also predicted by
1	FIGN	ENSG00000182263	0.948241	1	19.6279	
2	ARID3B	ENSG00000179361	0.919733	1	19.6279	>>
3	FOXP2	ENSG00000128573	0.872364	0.993333	19.6279	
4	HMGA2	ENSG00000149948	0.856157	0.993333	19.6279	>>
5	BACH1	ENSG00000156273	0.853783	0.993333	19.6279	>>
6	C14ORF28	ENSG00000179476	0.831538	0.991667	19.6279	
7	PRTG	ENSG00000166450	0.815932	0.991667	19.6279	
8	COL1A2	ENSG00000164692	0.814001	0.991667	19.6279	>>
9	ARHGAP28	ENSG00000088756	0.810749	0.991667	19.6279	
10	C11ORF90	ENSG00000214376	0.801366	0.991667	19.6279	

Εικόνα 18: Εμφάνιση αποτελεσμάτων για την αναζήτηση με βάση τον κωδικό MIMA

Οι δυνατότητες που έχει ο χρήστης από αυτό το σημείο και μετά είναι κοινές με το προηγούμενο σενάριο εκτέλεσης και για αυτό δεν περιγράφονται εκ νέου.

6.2.4 Αναζήτηση με βάση το όνομα του γονιδίου

Στο συγκεκριμένο σενάριο εκτέλεσης ο χρήστης αναζητά βιολογικούς στόχους με βάση το όνομα ενός γονιδίου. Αφού εισάγει το όνομα του γονιδίου που τον ενδιαφέρει στην μπάρα αναζήτησης (“**FIGN**”), εμφανίζεται η λίστα των αποτελεσμάτων:



The screenshot shows the DIANA-microT v5.0 (beta version) search interface. At the top, there are logos for IRIE YTI IHS, CSLab, and ALEXANDER FLEMING. A search bar contains the text 'fign'. To the right of the search bar is a 'VIEW RESULTS' button. Below the search bar are links for 'Advanced Search' and 'Help'. The main content area displays the search results for the gene 'fign', indicating that 191 miRNAs were found. The results section includes the following information:

- Ensembl Gene ID:** ENSMUSG00000075324
- Gene names:** FIGN (with UniProt and iH3 logos)
- Refseq IDs:** NM_021716
- Description:** FIDGETIN [SOURCE:UNIPROTKB/SWISS-PROT;ACC:Q9ERZ6]
- KEGG pathways:**
- Location:** chr2:63815418-63816701 strand: -1

Below the gene information, a list of miRNA IDs is shown in red text:

- ENSMUST00000102728
- ENST00000409634
- ENST00000392726
- ENST00000333129
- ENST00000392725




Εικόνα 19: Εμφάνιση αποτελεσμάτων με βάση το όνομα του γονιδίου.

Στη λίστα αποτελεσμάτων αρχικά εμφανίζεται ένα πίνακας ο οποίος περιέχει τον κωδικό ENSEMBL του γονιδίου, το όνομα του και κάποιες επιπρόσθετες πληροφορίες. Στη συνέχεια εμφανίζεται ένας δεύτερος πίνακας που περιέχει μία λίστα από όλα τα μετάγραφα του συγκεκριμένου γονιδίου με βάση τον κωδικό TRANSCRIPT αυτών και με τη δυνατότητα επιλογής καθενός από αυτούς τους κωδικούς.



6.2.5 Αναζήτηση με βάση τον κωδικό ENSEMBL

Στο συγκεκριμένο σενάριο εκτέλεσης ο χρήστης αναζητά βιολογικούς στόχους με βάση τον κωδικό ENSEMBL ενός γονιδίου. Αφού εισάγει τον κωδικό που τον ενδιαφέρει στην μπάρα αναζήτησης (“**ENSG00000000005**”), εμφανίζεται η λίστα των αποτελεσμάτων:



DIANA - microT v5.0 (beta version)

ENSG00000000005

Advanced Search  Help 

Results for Gene 'TNMD' (23 miRNAs found) | [New Search](#)

Ensembl Gene ID: ENSG00000000005
Gene names: TNMD  
Refseq IDs: NM_022144
Description: TENOMODULIN (TEM)(HTEM)(CHONDROMODULIN-I-LIKE PROTEIN)(CHM1L)(HCHM1L)(MYODULIN)(TENDIN)
 [SOURCE:UNIPROT;KB/SWISS-PROT;ACC:Q9H2S6]
KEGG pathways:
Location: chrX:- strand: 1

ENSMUST00000033602
 ENST00000373031

Εικόνα 20: Εμφάνιση αποτελεσμάτων με βάση τον κωδικό ENSEMBL ενός γονιδίου.

Όπως και πριν στη λίστα αποτελεσμάτων αρχικά εμφανίζεται ένα πίνακας ο οποίος περιέχει τον κωδικό ENSEMBL του γονιδίου, το όνομα του και κάποιες επιπρόσθετες πληροφορίες. Στη συνέχεια εμφανίζεται ένας δεύτερος πίνακας που περιέχει μία λίστα από όλα τα μεταγράφα του συγκεκριμένου γονιδίου με βάση τον κωδικό TRANSCRIPT αυτών και με τη δυνατότητα επιλογής καθενός από αυτούς τους κωδικούς.

6.2.6 Αναζήτηση με βάση τον κωδικό TRANSCRIPT

Στο συγκεκριμένο σενάριο εκτέλεσης ο χρήστης αναζητά βιολογικούς στόχους με βάση τον κωδικό TRANSCRIPT ενός γονιδίου. Αφού εισάγει τον κωδικό που τον ενδιαφέρει στην μπάρα αναζήτησης (“ENSMUST0000004829”), εμφανίζεται η λίστα των αποτελεσμάτων:

DIANA - microT v5.0 (beta version)

HEVA IHS CSLab ALEXANDER FLEMING

ENSMUST0000004829

VIEW RESULTS

Advanced Search Help

Results for Transcript 'ENSMUST0000004829' (27 miRNAs found) | New Search

Transcript ID: ENSMUST0000004829
Ensembl Gene ID: ENSG00000104848

Score Threshold: medium 0 KEGG Filter

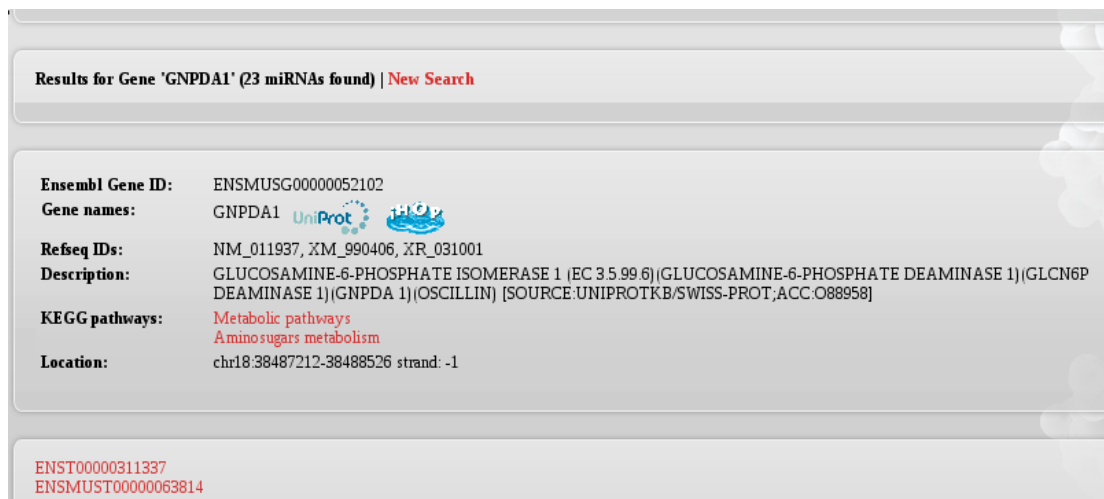
Rank	miRNA name	Alternative indicator	miTG score	Precision	SNR	Also predicted by
1	mmu-mir-300-star	MIMAT0004578	0.43357	0.0916981	1.10298	
2	mmu-mir-302c-star	MIMAT0003375	0.399274	0.337379	1.23114	
3	mmu-mir-1897-5p	MIMAT0007864	0.382845	0.316771	1.52274	
4	mmu-mir-877-star	MIMAT0004862	0.365428	0.00985663	0.934116	
5	mmu-mir-302b-star	MIMAT0003373	0.247154	0.388666	1.41217	
6	mmu-mir-299-star	MIMAT0000377	0.242045	0.401647	1.76259	
7	mmu-mir-218-2-star	MIMAT0005444	0.241428	0	0.818368	
8	mmu-mir-380-5p	MIMAT0000744	0.239982	0	0.741514	
9	mmu-mir-188-3p	MIMAT0004541	0.234383	0.351426	0.970333	
10	mmu-mir-881-star	MIMAT0004845	0.230613	0.143888	0.86829	

Εικόνα 21: Εμφάνιση αποτελεσμάτων με βάση τον κωδικό TRANSCRIPT ενός γονιδίου.



Στη λίστα αποτελεσμάτων αρχικά παρουσιάζεται ένα πίνακας που περιέχει τόσο τον κωδικό TRANSCRIPT του μετάγραφου που σχετίζεται με το γονίδιο όσο και τον κωδικό ENSEMBL του γονιδίου. Στη συνέχεια εμφανίζεται μία λίστα με ονόματα microRNA τα οποία έχουν βιολογικό στόχο πάνω στο συγκεκριμένο μετάγραφο και πληροφορίες σχετικές με τον κάθε στόχο. Οι παράμετροι για το σκορ και το KEGG pathway μπορούν να οριστούν όπως περιγράφηκε στο δεύτερο σενάριο εκτέλεσης και για αυτό δεν παρουσιάζουμε τη διαδικασία αυτή ξανά.

6.2.7 Αναζήτηση με βάση τον κωδικό Refseq

Στο συγκεκριμένο σενάριο εκτέλεσης ο χρήστης αναζητά βιολογικούς στόχους με βάση τον κωδικό Refseq ενός γονιδίου. Αφού εισάγει τον κωδικό που τον ενδιαφέρει στην μπάρα αναζήτησης (πχ “**xr_031001**”), εμφανίζεται η λίστα των αποτελεσμάτων:



Results for Gene 'GNPDA1' (23 miRNAs found) | [New Search](#)

Ensembl Gene ID: ENSMUSG00000052102
Gene names: GNPDA1  
Refseq IDs: NM_011937, XM_990406, XR_031001
Description: GLUCOSAMINE-6-PHOSPHATE ISOMERASE 1 (EC 3.5.99.6)|GLUCOSAMINE-6-PHOSPHATE DEAMINASE 1|(GLCN6P DEAMINASE 1)|(GNPDA 1)|(OSCILLIN) [SOURCE:UNIPROT/KB/SWISS-PROT;ACC:O88958]
KEGG pathways: [Metabolic pathways](#)
[Amino sugars metabolism](#)
Location: chr18:38487212-38488526 strand: -1

ENST00000311337
ENSMUST00000063814

Εικόνα 22: Εμφάνιση των αποτελεσμάτων με βάση τον κωδικό Refseq ενός γονιδίου.

Η μορφή των αποτελεσμάτων είναι κοινή με την αναζήτηση με βάση τον κωδικό ENSEMBL και για αυτό δεν αναλύεται περαιτέρω.

6.2.8 Διαχείριση αγαπημένων αναζητήσεων και ιστορικού

Στο συγκεκριμένο σενάριο εκτέλεσης ο χρήστης επιθυμεί να αποθηκεύσει μία αναζήτηση που πραγματοποίησε. Για το λόγο αυτό επιλέγει το σύνδεσμο “**Bookmark results page**”. Η εικόνα της εφαρμογής εκείνη τη στιγμή είναι η ακόλουθη στα αριστερά:



Εικόνα 23: Λίστα αγαπημένων αναζητήσεων και ιστορικού



Εικόνα 24: Τροποποίηση αγαπημένης αναζήτησης

Ο χρήστης έχει τη δυνατότητα τώρα είτε να δει, είτε να τροποποιήσει, είτε να διαγράψει μία από τις αγαπημένες του αναζητήσεις. Έστω ότι επιλέγει να δει την πρώτη από αυτές. Η εικόνα που εμφανίζεται είναι η πάνω δεξιά.

Στη συνέχεια ο χρήστης μπορεί να επιλέξει να ορίσει την περιγραφή της συγκεκριμένης αναζήτησης. Στο συγκεκριμένο παράδειγμα, επέλεξε να ορίσει ως περιγραφή το όνομα του microRNA για το οποίο πραγματοποίησε αναζήτηση. Πατώντας το κουμπί “**Update**” ή το κουμπί “**Save**” ενημερώνονται τα στοιχεία της αναζήτησης. Η μόνη διαφορά μεταξύ των δύο κουμπιών είναι ότι το κουμπί “**Save**” ανακατευθύνει το χρήστη στην φόρμα αναζήτησης.

Η εικόνα που βλέπει ο χρήστης στη συνέχεια είναι η ακόλουθη:



Εικόνα 25: Ενημέρωση της αναζήτησης μετά από την τροποποίηση της περιγραφής της.

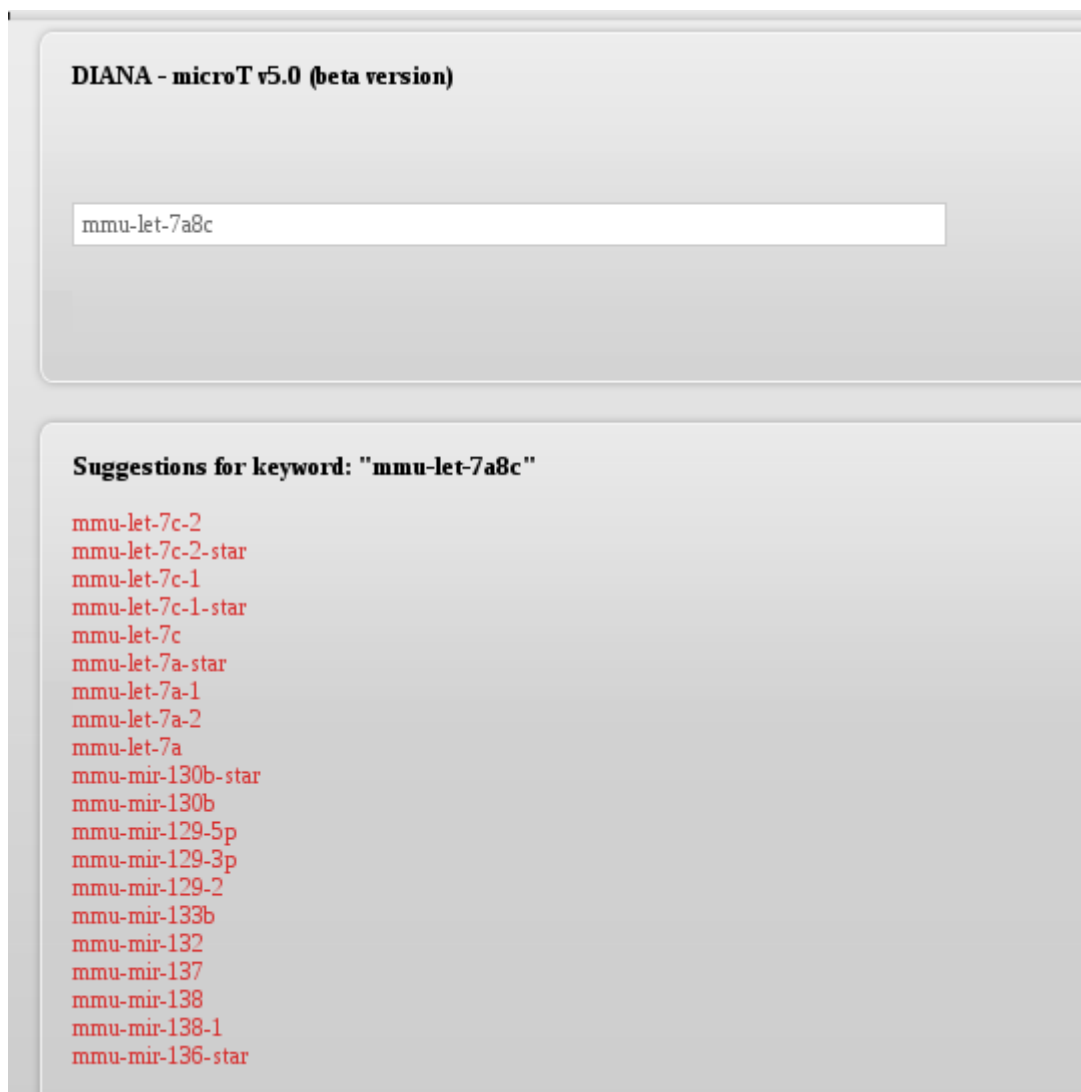
Από αυτό το σημείο και μετά ο χρήστης μπορεί επιπλέον να διαγράψει κάποια ή κάποιες από τις αγαπημένες του αναζητήσεις απλά επιλέγοντας το σύνδεσμο “Delete” των συγκεκριμένων αναζητήσεων. Η νέα εικόνα που παρουσιάζεται στην οθόνη του υπολογιστή είναι η ακόλουθη:



Εικόνα 26: Διαγραφή της αναζήτησης "fign gene".

6.2.9 Παροχή προτάσεων σε λανθασμένο όνομα *microRNA*

Στο συγκεκριμένο σενάριο εκτέλεσης ο χρήστης επιθυμεί να αναζητήσει βιολογικούς στόχους με βάση το όνομα ενός *microRNA*, αλλά εισάγει τον όρο λανθασμένα (“**mmu-let-7a8c**”). Σε αυτήν την περίπτωση το σύστημα αποκρίνεται προτείνοντας στο χρήστη τους παρακάτω όρους με δυνατότητα επιλογής καθενός από αυτούς:

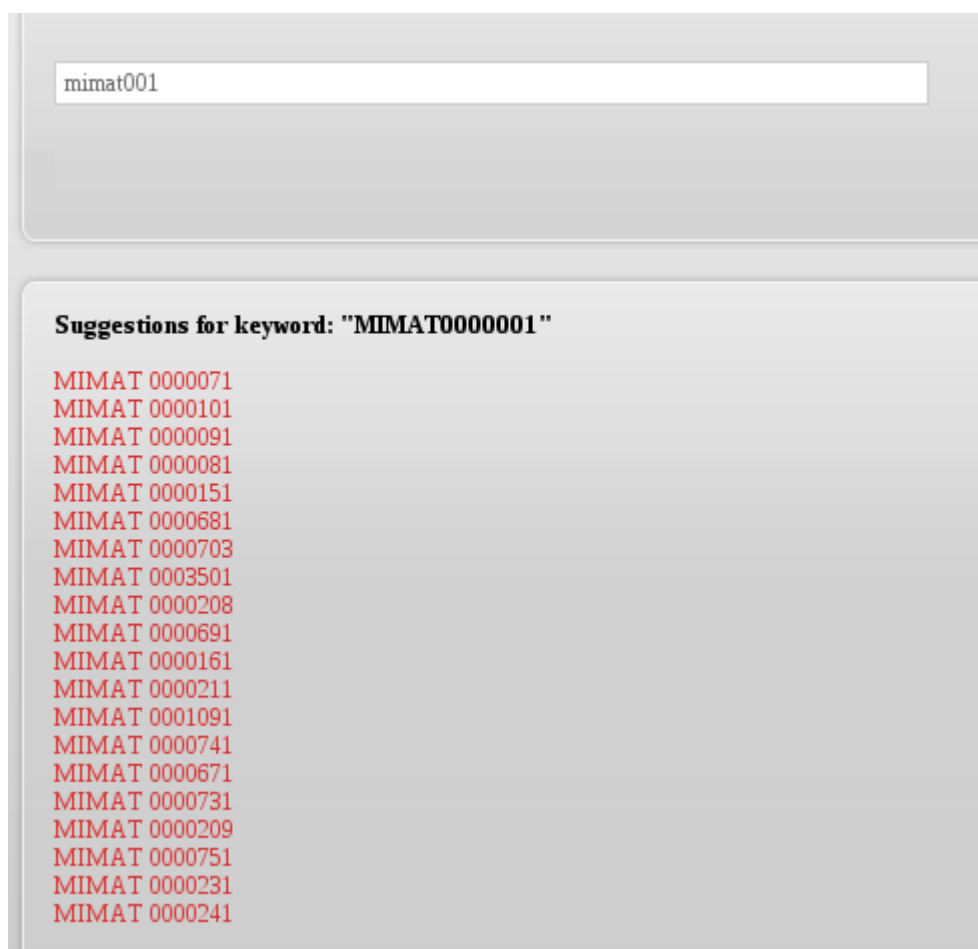


Εικόνα 27: Παροχή προτάσεων σε λανθασμένο όνομα *microRNA*.

6.2.10 Παροχή προτάσεων σε λανθασμένο κωδικό *MIMA*

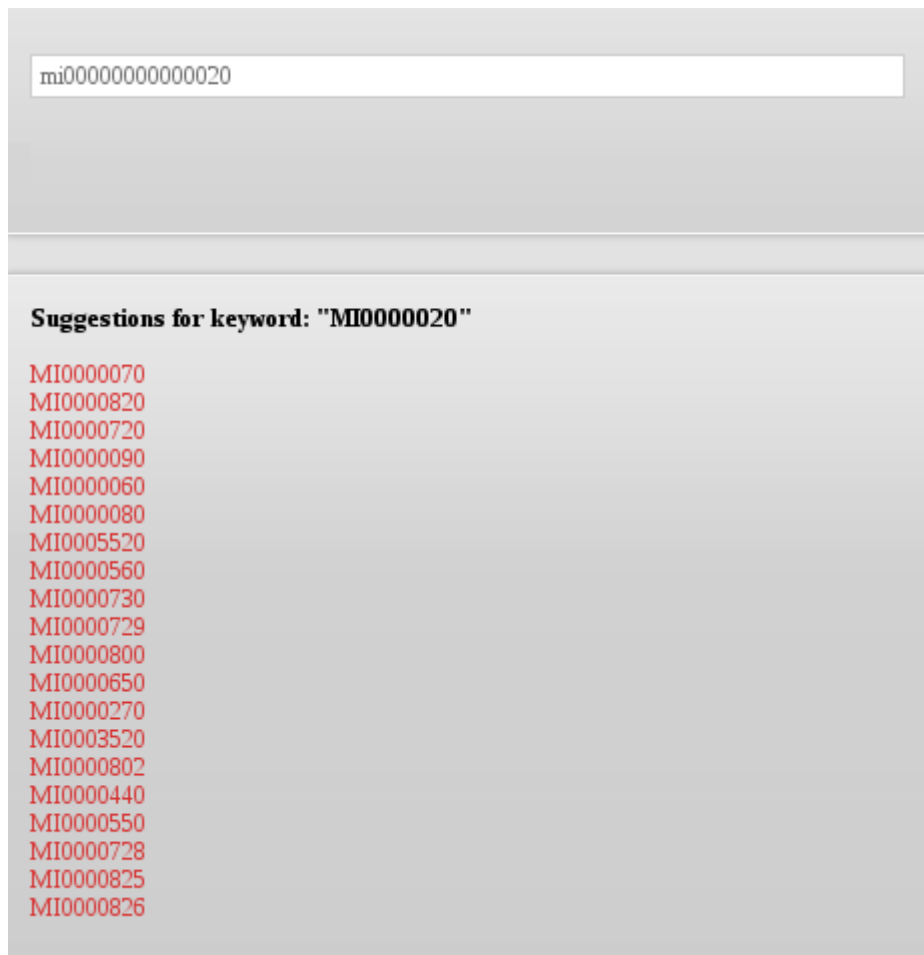
Στο συγκεκριμένο σενάριο εκτέλεσης ο χρήστης επιθυμεί να αναζητήσει βιολογικούς στόχους με βάση τον κωδικό *MIMA* ενός mature *microRNA*, αλλά εισάγει τον όρο

λανθασμένα (“**mimat001**”). Σε αυτήν την περίπτωση το σύστημα αποκρίνεται προτείνοντας στο χρήστη τους παρακάτω όρους με δυνατότητα επιλογής καθενός από αυτούς:



Εικόνα 28: Παροχή προτάσεων σε λανθασμένο κωδικό MIMA ενός *mature microRNA*

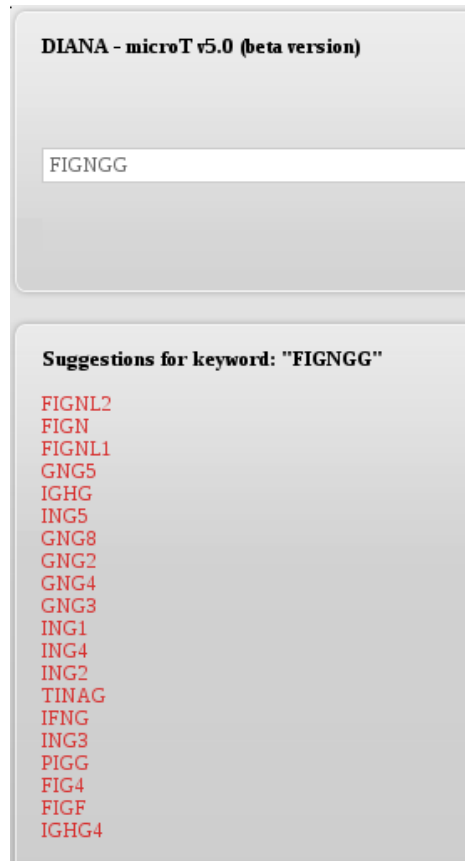
Αντίστοιχα αν εισάγει λανθασμένα τον κωδικό MIMA ενός *hairpin microRNA* (“**mi00000000000020**”) εμφανίζονται προς επιλογή οι ακόλουθοι όροι:



Εικόνα 29: Παροχή προτάσεων σε λανθασμένο κωδικό MIMA ενός *hairpin microRNA*

6.2.11 Παροχή προτάσεων σε λανθασμένο όνομα γονιδίου

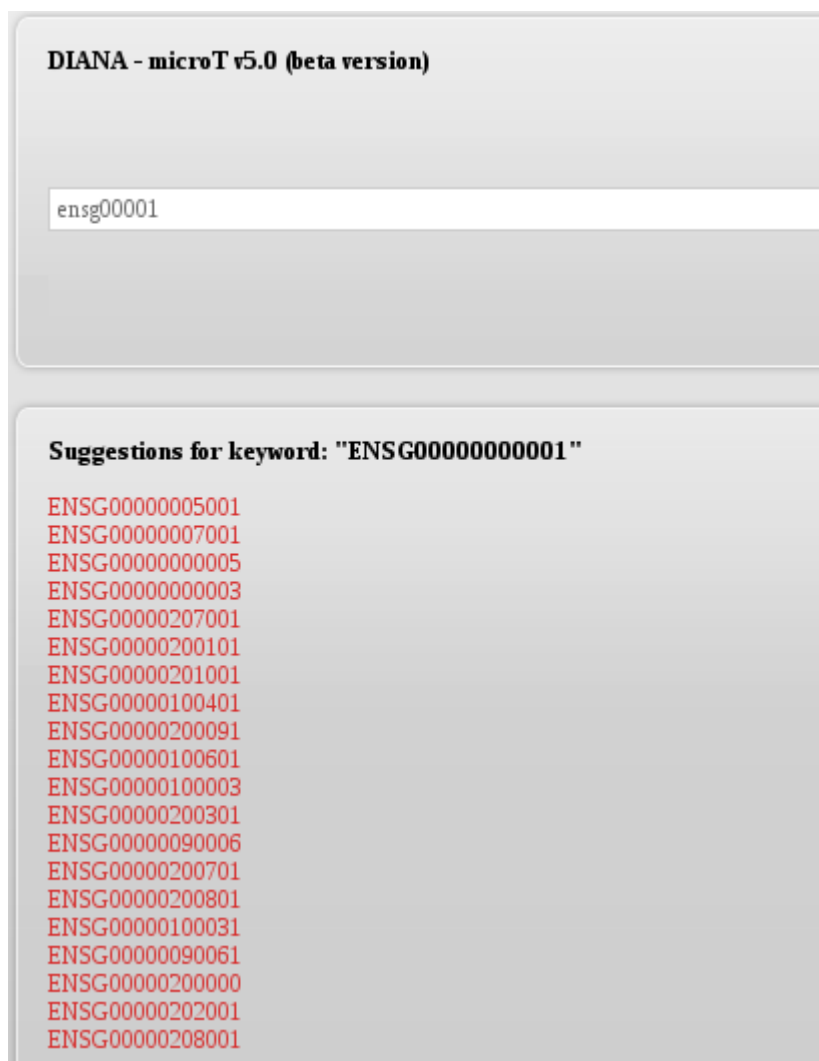
Στο συγκεκριμένο σενάριο εκτέλεσης ο χρήστης επιθυμεί να πραγματοποιήσει αναζήτηση με βάση το όνομα ενός γονιδίου, αλλά ο όρος που εισάγει δεν είναι έγκυρος ("FIGNGG"). Σε αυτήν την περίπτωση εμφανίζεται η ακόλουθη φόρμα με τους εξής όρους προς επιλογή:



Εικόνα 30: Παροχή προτάσεων σε λανθασμένο όνομα γονιδίου.

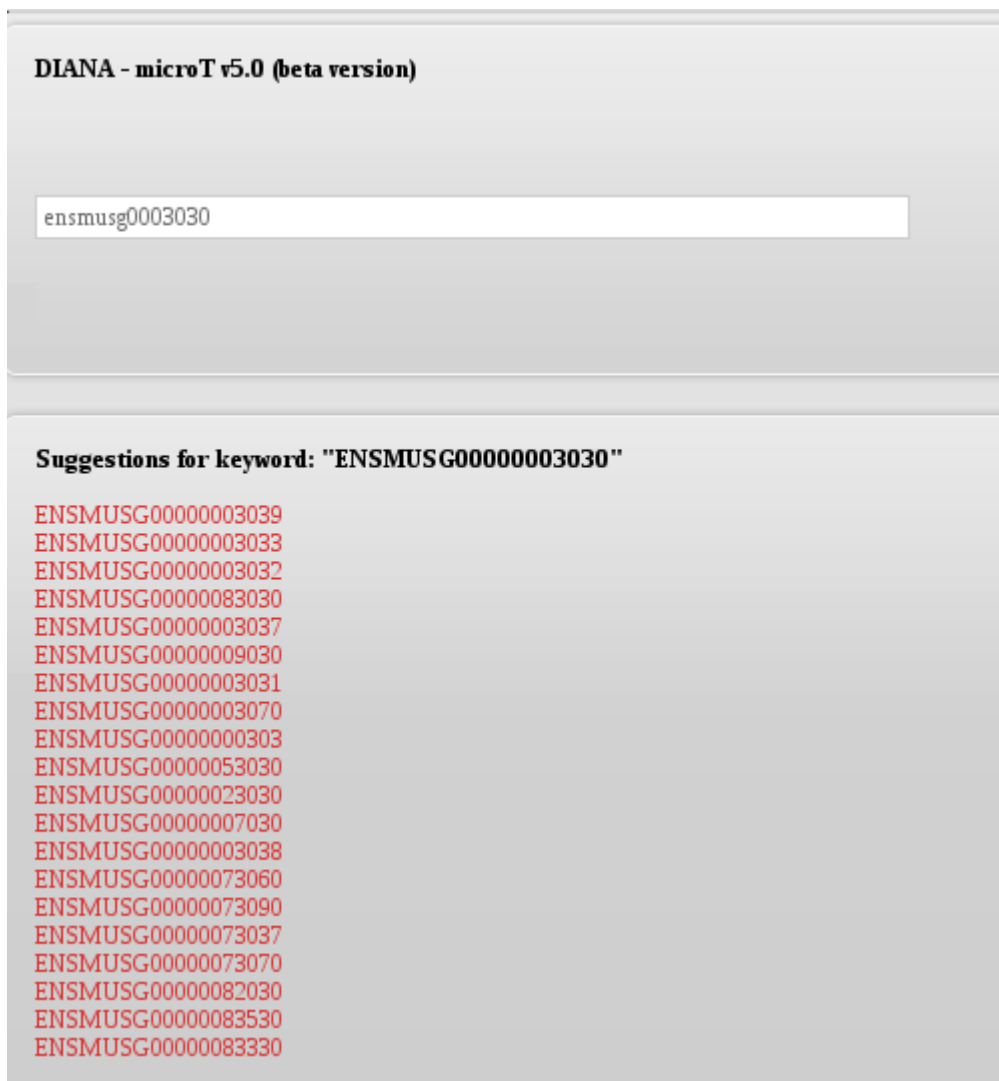
6.2.12 Παροχή προτάσεων σε λανθασμένο κωδικό *ENSEMBL*

Στο συγκεκριμένο σενάριο εκτέλεσης ο χρήστης εισάγει λάθος τον κωδικό *ENSEMBL* ενός γονιδίου (“**ensg000001**”). Σε αυτήν την περίπτωση εμφανίζεται μια φόρμα με τους ακόλουθους όρους προς επιλογή:



Εικόνα 31: Παροχή προτάσεων σε λανθασμένο κωδικό *ENSEMBL* ενός γονιδίου

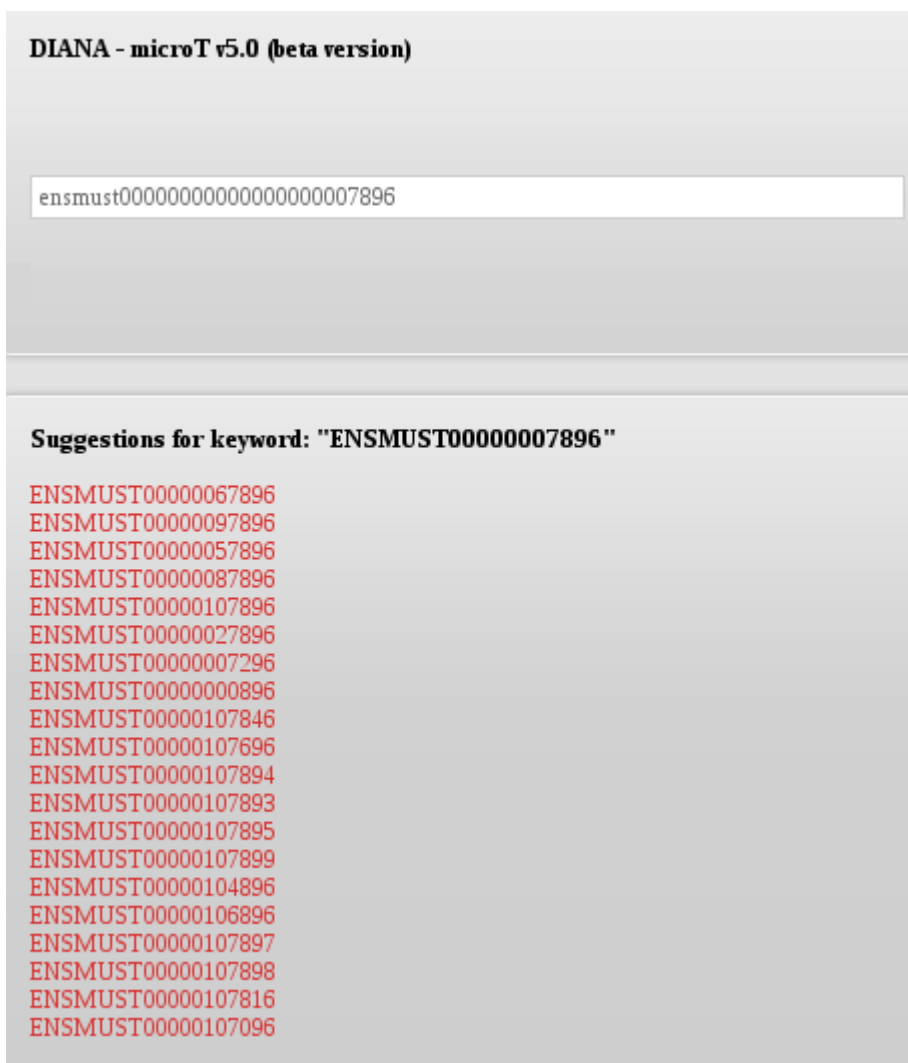
Εναλλακτικά, αν ο όρος που εισάγει είναι ο “**ensmusg0003030**” εμφανίζονται οι εξής όροι προς επιλογή:



Εικόνα 32: Παροχή προτάσεων σε λανθασμένο κωδικό ENSEMBL ενός γονιδίου

6.2.13 Παροχή προτάσεων σε λανθασμένο κωδικό TRANSCRIPT

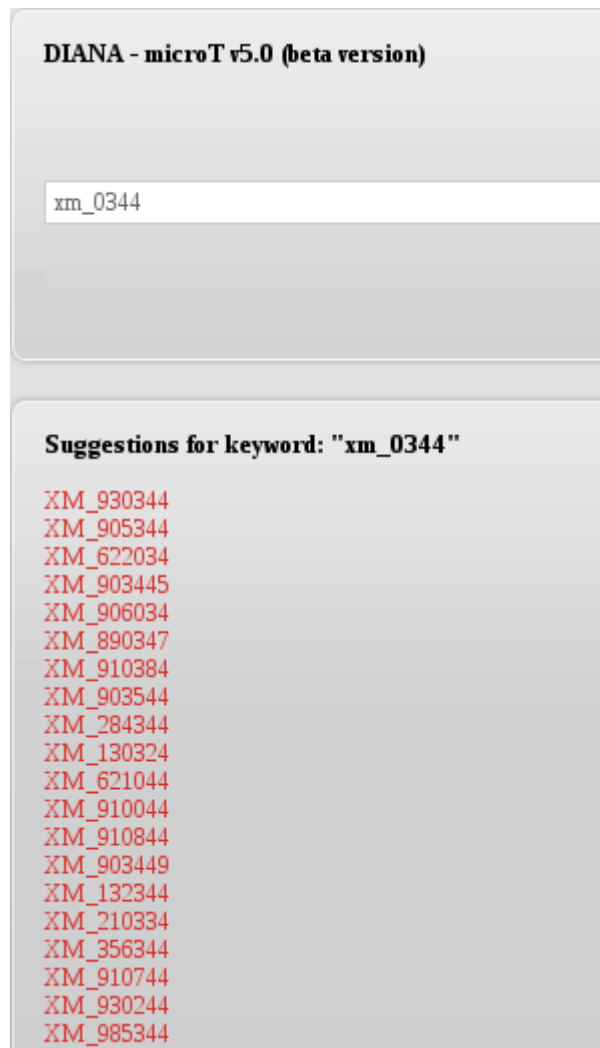
Στο συγκεκριμένο σενάριο εκτέλεσης ο χρήστης εισάγει λάθος τον κωδικό TRANSCRIPT ενός γονιδίου (“**enst003047**”). Σε αυτήν την περίπτωση εμφανίζεται μια φόρμα με τους ακόλουθους όρους προς επιλογή:



Εικόνα 34: Παροχή προτάσεων σε λανθασμένο κωδικό TRANSCRIPT

6.2.14 Παροχή προτάσεων σε λανθασμένο κωδικό Refseq

Στο συγκεκριμένο σενάριο εκτέλεσης ο χρήστης πραγματοποιεί αναζήτηση με βάση τον κωδικό Refseq ενός γονιδίου, ο οποίος όμως είναι λανθασμένος ("xm_0344"). Σε αυτήν την περίπτωση εμφανίζεται μια νέα φόρμα προς αναζήτηση με τους εξής προτεινόμενους όρους:



Εικόνα 35: Παροχή προτάσεων σε λανθασμένο κωδικό Refseq

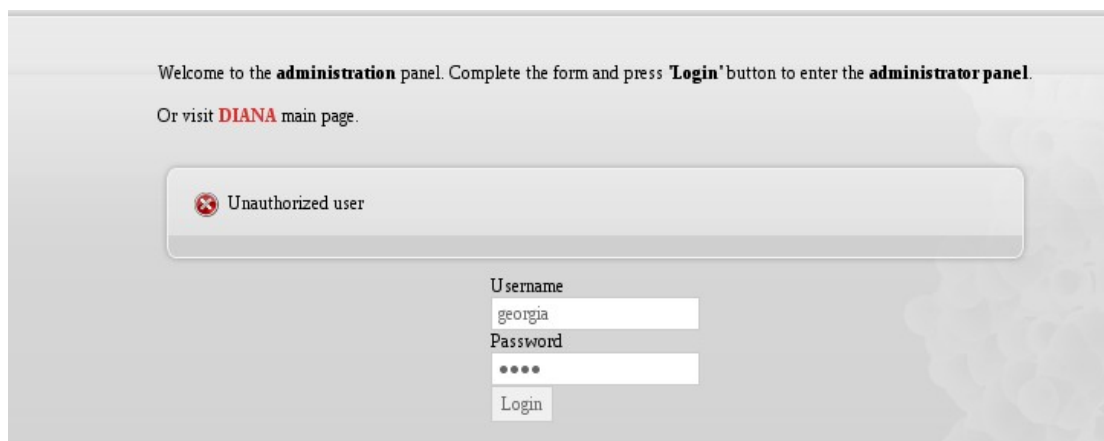
Ανάλογοι είναι και οι προτεινόμενοι όροι στην περίπτωση που ο χρήστης εισάγει λανθασμένο κωδικό Refseq μιας από τις μορφές “**nr_**”, “**xr_**”, και “**nm_**”.

6.2.15 Σύνδεση απλού χρήστη στο πάνελ διαχείρισης

Στο συγκεκριμένο σενάριο εκτέλεσης ο απλό χρήστης “georgia” επιχειρεί να συνδεθεί στον πάνελ διαχείρισης της εφαρμογής.

Όπως είναι φυσικό, κάτι τέτοιο δεν θα έπρεπε αν είναι επιτρεπτό για έναν απλό χρήστη, παρά μόνο για τους διαχειριστές. Οι διαχειριστές πραγματοποιούν τη σύνδεση τους στο πάνελ διαχείρισης από διαφορετική και μυστική τοποθεσία από ότι οι οι χρήστες του συστήματος.

Παρόλα αυτά, έστω ότι ο απλός χρήστης “georgia” εντοπίζει αυτή την τοποθεσία και συμπληρώνει τα στοιχεία του στην κατάλληλη φόρμα. Σε αυτήν την περίπτωση εμφανίζεται στην οθόνη του υπολογιστή η εξής εικόνα:

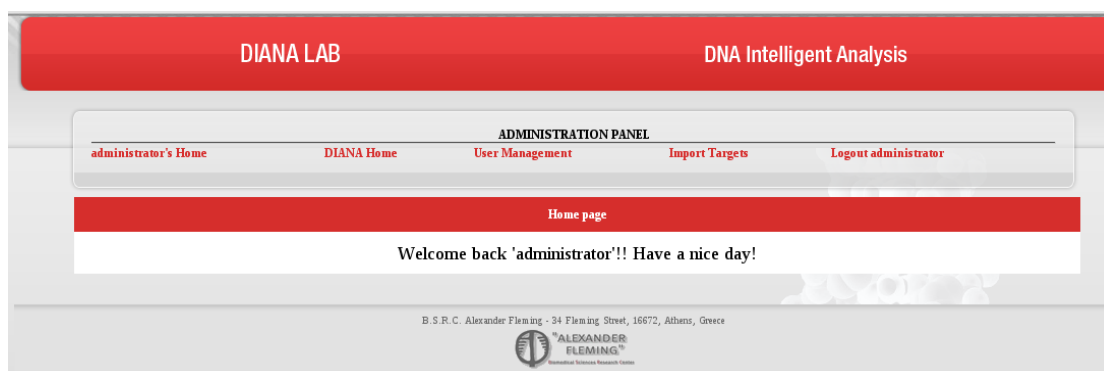


Εικόνα 36: Απαγόρευση της εισόδου του απλού χρήστη "georgia" στο πάνελ διαχείρισης.

6.2.16 Προβολή λίστας χρηστών

Έστω ο χρήστης “administrator” ο οποίος ανήκει στην ομάδα διαχειριστών του συστήματος. Ο χρήστης “administrator” πραγματοποιεί τη σύνδεση του στο πάνελ διαχείρισης της εφαρμογής από τη φόρμα που παρουσιάστηκε στο προηγούμενο σενάριο εκτέλεσης.

Μετά την εισαγωγή του στο πάνελ διαχείρισης η πρώτη εικόνα που βλέπει ο “administrator” είναι η εξής:



Εικόνα 37: Αρχική σελίδα του πάνελ διαχείρισης.

Οι δυνατότητες που έχει ο διαχειριστής είναι η μετάβασή του στην αρχική σελίδα της εφαρμογής, η διαχείριση των χρηστών, το ανέβασμα ενός αρχείου στόχων ή αποσύνδεση του από το σύστημα.

Ο χρήστης επιλέγει να μεταβεί στην σελίδας διαχείρισης χρηστών, και η εικόνα που βλέπει είναι η εξής:

Add new user

DIANA ADMINISTRATORS GROUP
Update administrators info or delete users from administrators group.

Username	Email	Edit	Delete
administrator	noemail@example.com	Edit	Delete
coma	e@c.com	Edit	Delete
georgia	gf@c.com	Edit	Delete

DIANA USERS GROUP
Update users info or delete users from simple users group.

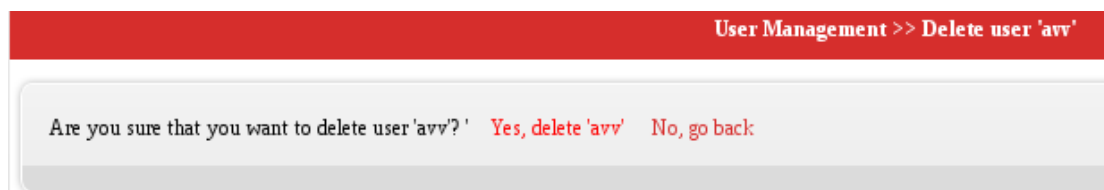
Username	Email	Edit	Delete
abad	ako@c.com	Edit	Delete
alanbridge	alan.bridge@isb-sib.ch	Edit	Delete
avv	ako@c.com	Edit	Delete
bob	ako@c.com	Edit	Delete
callie	ako@c.com	Edit	Delete
doughowe	dhowe@cs.uoregon.edu	Edit	Delete
fleming		Edit	Delete
messi	lionel@c.com	Edit	Delete
new_adddda	akom@c.com	Edit	Delete
prekageo	a@b.com	Edit	Delete
testv	bergceid@yahoo.com	Edit	Delete
thanos	thanos@example.com	Edit	Delete

Εικόνα 38: Προβολή της λίστας χρηστών του συστήματος της εφαρμογής.

Όπως φαίνεται οι χρήστες χωρίζονται σε δύο ομάδες (απλοί και διαχειριστές) και κάθε ομάδα προβάλλεται με αλφαβητική σειρά. Ορατές είναι μόνο οι πληροφορίες του ονόματος χρήστη (username) και της ηλεκτρονικής διεύθυνσης (email). Επιπλέον υπάρχουν οι επιλογές διαγραφής και τροποποίησης στοιχείων χρηστών για κάθε χρήστη ξεχωριστά. Τέλος, υπάρχει η δυνατότητα δημιουργίας νέου χρήστη στο σύστημα επιλέγοντας το σύνδεσμο “**Add new user**”.

6.2.17 Διαγραφή χρηστών από το σύστημα

Στο συγκεκριμένο σενάριο εκτέλεσης ο διαχειριστής “administrator” επιθυμεί να διαγράψει έναν χρήστη του συστήματος. Αρχικά, επιλέγει από το πάνελ διαχείρισης χρηστών την επιλογή “**Delete**” δίπλα από το όνομα του χρήστη που επιθυμεί. Έστω ότι σκοπός του διαχειριστή είναι να διαγράψει τον χρήστη “**avv**”. Αφού επιλέξει τον κατάλληλο σύνδεσμο εμφανίζεται η εξής φόρμα που ρωτά τον διαχειριστή για επιβεβαίωση της ενέργειας που θέλει να πραγματοποιήσει:



Εικόνα 39: Φόρμα για επιβεβαίωση της διαγραφής του χρήστη από το σύστημα.

Ο διαχειριστής επιλέγει την επιλογή “**Ναι**”, συνεπώς ο χρήστης “**avn**” διαγράφεται από το σύστημα και εμφανίζεται η ανανεωμένη λίστα χρηστών με ταυτόχρονη γνωστοποίηση προς τον διαχειριστή της διαγραφής που μόλις πραγματοποιήθηκε.

User 'avn' has been successfully deleted from the system.

Add new user

DIANA ADMINISTRATORS GROUP
Update administrators info or delete users from administrators group.

Username	Email	Edit	Delete
administrator	noemail@example.com	Edit	Delete
coma	e@c.com	Edit	Delete
georgia	gf@c.com	Edit	Delete

DIANA USERS GROUP
Update users info or delete users from simple users group.

Username	Email	Edit	Delete
abad	ako@c.com	Edit	Delete
alanbridge	alan.bridge@isb-sib.ch	Edit	Delete
bob	ako@c.com	Edit	Delete
callie	ako@c.com	Edit	Delete
doughowe	dhowe@cs.uoregon.edu	Edit	Delete
fleming		Edit	Delete
messi	lionel@c.com	Edit	Delete
new_adddda	akom@c.com	Edit	Delete
prekageo	a@b.com	Edit	Delete
testv	bergceid@yahoo.com	Edit	Delete
thanos	thanos@example.com	Edit	Delete

Εικόνα 40: Η ανανεωμένη λίστα των χρηστών του συστήματος μετά τη διαγραφή.

6.2.18 Τροποποίηση στοιχείων χρήστη του συστήματος

Στο συγκεκριμένο σενάριο εκτέλεσης ο διαχειριστής “administrator” έχει σκοπό να τροποποιήσει τα στοιχεία του χρήστη “**adddda**”. Αφού επιλέξει από την λίστα προβολής χρηστών την επιλογή “**Edit**” για τον χρήστη αυτόν, εμφανίζεται η ακόλουθη φόρμα η οποία περιέχει όλα στοιχεία του χρήστη συμπληρωμένα.

Update user info
Insert into the form fields the data of the user to be updated. Then, Press submit button to save the changes.

Username

Password

Password Repeat

Email

Type Of User

Εικόνα 41: Φόρμα για την τροποποίηση στοιχείων ενός χρήστη.

Στη συνέχεια ο διαχειριστής μπορεί να τροποποιήσει όποιο από τα στοιχεία αυτά επιθυμεί. Εστω ότι επιλέγει να τροποποιήσει το όνομα χρήστη και το συνθηματικό του χρήστη. Τότε συμπληρώνει με τα επιθυμητά ονόματα μόνο τα πεδία που τον τον ενδιαφέρουν, δηλαδή το πεδίο **“username”** και τα πεδία **“password”** και **“password repeat”**. Αυτό παρουσιάζεται στην ακόλουθη εικόνα:

User Management >> Edit user add

Update user info
Insert into the form fields the data of the user to be updated. Then, Press submit button to save the changes.

Username

Password

Password Repeat

Email

Type Of User

Εικόνα 42: Συμπλήρωση νέων στοιχείων χρήστη.

Ο διαχειριστής εισήγαγε στην ειδική φόρμα το νέο όνομα χρήστη **“new_adddda”** και το νέο συνθηματικό. Μόλις πατήσει του κουμπί **“Save”** εμφανίζεται η νέα ανανεωμένη λίστα χρηστών, με τα νέα στοιχεία του χρήστη:

Add new user			
DIANA ADMINISTRATORS GROUP			
Update administrators info or delete users from administrators group.			
Username	Email		
administrator	noemail@example.com	Edit	Delete
coma	e@c.com	Edit	Delete
DIANA USERS GROUP			
Update users info or delete users from simple users group.			
Username	Email		
abad	ako@c.com	Edit	Delete
alanbridge	alan.bridge@isb-sib.ch	Edit	Delete
avv	ako@c.com	Edit	Delete
bob	ako@c.com	Edit	Delete
callie	ako@c.com	Edit	Delete
doughowe	dhowe@cs.uoregon.edu	Edit	Delete
fleming		Edit	Delete
georgia	gf@c.com	Edit	Delete
new_adddda	akom@c.com	Edit	Delete
prekageo	a@b.com	Edit	Delete
testv	bergceid@yahoo.com	Edit	Delete
thanos	thanos@example.com	Edit	Delete

Εικόνα 43: Ενημερωμένη λίστα χρηστών.

Αξίζει να σημειωθεί πως αν ο διαχειριστής αφήσει κάποιο πεδίο κενό, ενημερώνεται με κατάλληλο μήνυμα σφάλματος.

6.2.19 Εισαγωγή χρήστη στην ομάδα διαχειριστών

Στο συγκεκριμένο σενάριο εκτέλεσης ο διαχειριστής θέλει να εισάγει το χρήστη “georgia” στην ομάδα των διαχειριστών. Αυτό γίνεται απλά τροποποιώντας τα στοιχεία του χρήστη “georgia” και συγκεκριμένα το πεδίο “**Type of User**” επιλέγοντας την ομάδα χρηστών “**Admin**”. Αυτό φαίνεται στην ακόλουθη εικόνα:

User Management >>> Edit User	
Update user info	
Insert into the form fields the data of the user to be updated. Then, Press submit button to save the changes.	
Username	<input type="text" value="georgia"/>
Password	<input type="password" value="••••"/>
Password Repeat	<input type="password" value="••••"/>
Email	<input type="text" value="gf@c.com"/>
Type Of User	<input type="text" value="Admin"/> ▼
<input type="button" value="Save"/>	

Εικόνα 44: Τροποποίηση των στοιχείων του χρήστη "georgia" έτσι ώστε να ανήκει στην ομάδα των διαχειριστών.

Add new user			
DIANA ADMINISTRATORS GROUP			
Update administrators info or delete users from administrators group.			
Username	Email		
administrator	noemail@example.com	Edit	Delete
coma	e@c.com	Edit	Delete
georgia	gf@c.com	Edit	Delete
DIANA USERS GROUP			
Update users info or delete users from simple users group.			
Username	Email		
abad	ako@c.com	Edit	Delete
alanbridge	alan.bridge@isb-sib.ch	Edit	Delete
avv	ako@c.com	Edit	Delete
bob	ako@c.com	Edit	Delete
callie	ako@c.com	Edit	Delete
doughowe	dhowe@cs.uoregon.edu	Edit	Delete
fleming		Edit	Delete
new_adddda	akom@c.com	Edit	Delete
prekageo	a@b.com	Edit	Delete
testv	bergceid@yahoo.com	Edit	Delete
thanos	thanos@example.com	Edit	Delete

Εικόνα 45: Ανανεωμένη λίστα των χρηστών του συστήματος.

6.2.20 Πρόσθεση νέου χρήστη

Στο συγκεκριμένο σενάριο εκτέλεσης ο διαχειριστής επιθυμεί να προσθέσει ένα νέο χρήστη στο σύστημα. Για το λόγο αυτό επιλέγει το σύνδεσμο “**Add new user**” από το πάνελ διαχείρισης χρηστών και του εμφανίζεται η μία φόρμα της οποίας συμπληρώνει όλα τα πεδία εκτός από την ηλεκτρονική διεύθυνση. Τότε ενημερώνεται με κατάλληλο μήνυμα σφάλματος για το πεδίο που ξέχασε να συμπληρώσει. Η ακόλουθη εικόνα δείχνει τα όσα περιγράψαμε μέχρι τώρα:

Add a new user
Insert into the form fields the data of the new user. Then, Press submit button to insert the new user.

Username

Password

Password Repeat

Email ❌
Email cannot be blank.

Type Of User

Εικόνα 46: Συμπλήρωση των στοιχείων του νέου χρήστη. Μήνυμα σφάλματος για την παράλειψη συμπλήρωσης του πεδίου ηλεκτρονικής διεύθυνσης.

Στη συνέχεια συμπληρώνει το πεδίο που λείπει και πατάει του κουμπί “Save” όπως φαίνεται ακολούθως:

Add a new user
Insert into the form fields the data of the new user. Then, Press submit button to insert the new user.

Username

Password

Password Repeat

Email

Type Of User

Εικόνα 47: Συμπλήρωση του πεδίου της ηλεκτρονικής διεύθυνσης που έλειπε.

Η λίστα χρηστών ανανεώνεται και παρουσιάζεται ακολούθως:

Successful creation of user 'messi'

Add new user

DIANA ADMINISTRATORS GROUP
Update administrators info or delete users from administrators group.

Username	Email		
administrator	noemail@example.com	Edit	Delete
coma	e@c.com	Edit	Delete
georgia	gf@c.com	Edit	Delete

DIANA USERS GROUP
Update users info or delete users from simple users group.

Username	Email		
abad	ako@c.com	Edit	Delete
alanbridge	alan.bridge@isb-sib.ch	Edit	Delete
avv	ako@c.com	Edit	Delete
bob	ako@c.com	Edit	Delete
callie	ako@c.com	Edit	Delete
doughowe	dhowe@cs.uoregon.edu	Edit	Delete
fleming		Edit	Delete
messi	lionel@c.com	Edit	Delete
new_adddda	akom@c.com	Edit	Delete
prekageo	a@b.com	Edit	Delete
testv	bergceid@yahoo.com	Edit	Delete
thanos	thanos@example.com	Edit	Delete

Εικόνα 48: Ανανεωμένη λίστα των χρηστών.

6.2.21 Ανέβασμα αρχείου στόχων

Στο συγκεκριμένο σενάριο εκτέλεσης ο διαχειριστής επιθυμεί να ανεβάσει στο σύστημα ένα αρχείο το οποίο περιέχει τα ονόματα προγραμμάτων που εντόπισαν κοινούς βιολογικούς στόχους με το πρόγραμμα της παρούσας εφαρμογής. Ο διαχειριστής αφού επιλέξει από την αρχική σελίδα του πάνελ διαχείρισης τον σύνδεσμο “**Import targets**”, βλέπει την ακόλουθη φόρμα:

Use this form in order to upload data for a specific program.

Upload File

Εικόνα 49: Ειδική φόρμα για το ανέβασμα του αρχείου στόχων στη βάση δεδομένων.

Από τη φόρμα αυτή μπορεί να επιλέξει ένα κατάλληλα διαμορφωμένο αρχείο από τον προσωπικό υπολογιστή του και να το ανεβάσει στο σύστημα. Τα δεδομένα της βάσης πριν το ανέβασμα του αρχείου είναι τα εξής:

```
mysql> select * from diana_interactions;
+-----+-----+-----+-----+
| gene_id | mirna_mature_id | program | score |
+-----+-----+-----+-----+
| 20164 | 118 | targetscan | 0,9 |
| 20164 | 118 | pictar | 0,8 |
| 20164 | 118 | pro1 | 0,833 |
| 3103 | 1538 | pro2 | 0,33 |
| 3103 | 1538 | pictar | 0,23 |
| 3103 | 1538 | pro3 | 0,9393 |
| 5014 | 138 | proggn | 0,99 |
| 5014 | 138 | proggn1 | 0,9 |
| 5014 | 138 | proggn2 | 0,9 |
| 33856 | 1738 | proggn | 0,8 |
| 33856 | 1738 | proggnn | 0,7 |
| 85014 | 1738 | proggn1 | 0,9 |
| 85014 | 1738 | proggn2 | 0,9 |
| 85014 | 1738 | proggn | 0,99 |
| 3356 | 138 | proggnn | 0,7 |
| 3356 | 138 | proggn | 0,8 |
+-----+-----+-----+-----+
16 rows in set (0,01 sec)
```

Εικόνα 50: Εικόνα της βάσης πριν το ανέβασμα του αρχείου στόχων.

Έστω ότι ο χρήστης επιλέγει το αρχείο testfile.txt από τον υπολογιστή του το οποίο έχει τη μορφή:

```
25014 7738 proggn1 0,9
25014 7738 proggn2 0,9
25014 7738 proggn 0,99
98256 7738 proggn 0,8
98256 7738 proggnn 0,7
```

Εικόνα 51: Μορφή του αρχείου στόχων προς ανέβασμα.

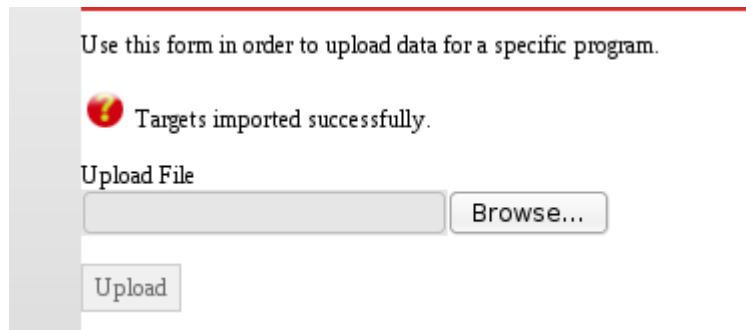
Ο διαχειριστής με τη βοήθεια του πλήκτρου “**Browse**” επιλέγει το αρχείο που επιθυμεί και πατάει το κουμπί “**Upload**”.

Use this form in order to upload data for a specific program.

Upload File

Εικόνα 52: Επιλογή του αρχείου testfile.txt για ανέβασμα.

Στη συνέχεια εμφανίζεται κατάλληλο μήνυμα που ενημερώνει το διαχειριστή ότι το ανέβασμα του αρχείου πραγματοποιήθηκε επιτυχώς.



Εικόνα 53: Ενημέρωση του διαχειριστή για το επιτυχή ανέβασμα του αρχείου.

Η εικόνα του πίνακα diana_interactions της βάσης μετά το ανέβασμα είναι η εξής:

```
mysql> select * from diana_interactions;
+-----+-----+-----+-----+
| gene_id | mirna_mature_id | program | score |
+-----+-----+-----+-----+
| 20164 | 118 | targetscan | 0.9 |
| 20164 | 118 | pictar | 0.8 |
| 20164 | 118 | pro1 | 0.833 |
| 3103 | 1538 | pro2 | 0.33 |
| 3103 | 1538 | pictar | 0.23 |
| 3103 | 1538 | pro3 | 0.9393 |
| 5014 | 138 | proggn | 0.99 |
| 5014 | 138 | proggn1 | 0.9 |
| 5014 | 138 | proggn2 | 0.9 |
| 98256 | 7738 | proggn | 0.8 |
| 98256 | 7738 | proggn | 0.7 |
| 25014 | 7738 | proggn | 0.99 |
| 33856 | 1738 | proggn | 0.8 |
| 33856 | 1738 | proggn1 | 0.7 |
| 25014 | 7738 | proggn1 | 0.9 |
| 25014 | 7738 | proggn2 | 0.9 |
| 85014 | 1738 | proggn1 | 0.9 |
| 85014 | 1738 | proggn2 | 0.9 |
| 85014 | 1738 | proggn | 0.99 |
| 3356 | 138 | proggn1 | 0.7 |
| 3356 | 138 | proggn | 0.8 |
+-----+-----+-----+-----+
21 rows in set (0.00 sec)
```

Εικόνα 54: Εικόνα της βάσης μετά το ανέβασμα του αρχείου στόχων.

Όπως εύκολα μπορούμε να διαπιστώσουμε η βάση έχει ενημερωθεί σωστά.

6.3 Αξιολόγηση

Η κυριότερη λειτουργία της εφαρμογής είναι η αναζήτηση βιολογικών στόχων. Ωστόσο είναι αρκετά συχνό φαινόμενο οι χρήστες να μην εισάγουν σωστά τους όρους προς αναζήτηση είτε λόγο απροσεξίας, είτε λόγο του ότι δεν θυμούνται σωστά τη σωστή ονομασία του όρου. Για το λόγο αυτό δόθηκε ιδιαίτερη έμφαση στον μηχανισμό παροχής προτάσεων κατά τη διάρκεια ανάπτυξης της παρούσας έκδοσης της εφαρμογής.

Συγκεκριμένα ο νέος μηχανισμός παροχής προτάσεων στους όρους αναζήτησης επανασχεδιάστηκε και υλοποιήθηκε εξολοκλήρου από την αρχή. Οι αλγόριθμοι του μηχανισμού δεν δίνουν απαραίτητα καλύτερα ως προς την ποιότητα αποτελέσματα. Αυτό που κάνει τον μηχανισμό να ξεχωρίζει από την προηγούμενη έκδοση αυτού είναι ότι λειτουργεί πολύ πιο έξυπνα και απλά, και κυρίως έχει καλύτερες επιδόσεις.

Ο προηγούμενος μηχανισμός μπορούσε να γνωρίζει ανά πάσα στιγμή τι είδους όρο αναζήτησης (microRNA ή γονίδιο) εισήγαγε ο χρήστης, καθότι υπήρχε διαφορετική φόρμα εισαγωγής για κάθε είδος όρου. Ο νέος μηχανισμός διαβάζει τους όρους προς αναζήτηση από μία μόνο φόρμα και προσπαθεί με έξυπνο τρόπο να μαντέψει το είδος του όρου (microRNA ή γονίδιο). Η ύπαρξη μιας μόνο φόρμας εγγυάται ευκολία για το χρήστη της εφαρμογής και δημιουργεί ένα καλαίσθητο αποτέλεσμα.

Για την παροχή των προτάσεων, ο προηγούμενος μηχανισμός χρησιμοποιούσε μεν την μετρική της απόσταση Levenshtein, αλλά σήκωνε όλα τα δεδομένα από τη μνήμη και συνέκρινε ένα προς ένα από αυτά με τον όρο που εισήγαγε ο χρήστης. Ο νέος μηχανισμός καταρχάς χρησιμοποιεί την μετρική της απόστασης Levenshtein μόνο όπου κρίνεται απαραίτητο και επιπλέον σηκώνει τα δεδομένα από τη μνήμη κατά ομάδες (buffering) γεγονός που εγγυάται καλύτερες ταχύτητες στην εξαγωγή προτάσεων.

Σε ορισμένες κατηγορίες όρων (κωδικοί MIMA, ENSEMBL, TRANSCRIPT, Refseq) ο νέος μηχανισμός εκμεταλλεύεται τη μορφή τους, και δεν χρειάζεται καν να πραγματοποιήσει αναζήτηση για την παροχή προτάσεων. Για παράδειγμα αν λείπουν κάποια μηδενικά από αυτούς τους όρους ο μηχανισμός τα συμπληρώνει κατάλληλα και αποφεύγεται τελείως η χρήση της απόστασης Levenshtein.

Τέλος, η επιλογή χρήσης πινάκων της βάσης τύπου MEMORY στους πίνακες που έχει πρόσβαση ο μηχανισμός, επιτάχυνε τη διαδικασία παροχής προτάσεων καθότι τα δεδομένα προέρχονταν από τη μνήμη RAM του συστήματος και συνεπώς ήταν σχεδόν άμεσα προσπελάσιμα.

7

Επίλογος

Στο κεφάλαιο αυτό συνοψίζεται η παρούσα διπλωματική εργασία. Αρχικά αναφέρονται οι στόχοι που επιτεύχθηκαν και δίνεται έμφαση στα κυριότερα θετικά και αρνητικά στοιχεία της διαδικασίας ανάπτυξης της εφαρμογής. Στο τέλος του κεφαλαίου παρουσιάζονται ορισμένες προτάσεις για μελλοντικές επεκτάσεις της παρούσας έκδοσης της εφαρμογής.

7.1 Σύνοψη και συμπεράσματα

Η παρούσα διπλωματική εργασία έχει ως θέμα την ανάπτυξη πρόσθετων λειτουργιών και τη βελτίωση υπαρχόντων χαρακτηριστικών στην προηγμένη εφαρμογή ιστού για διαχείριση βιολογικών δεδομένων DIANA. Αποτέλεσμα είναι το σύστημα DIANA 2.0. Η ανάπτυξη των χαρακτηριστικών πραγματοποιήθηκε σύμφωνα με σύγχρονες αρχές της Τεχνολογίας Λογισμικού, τόσο για την διευκόλυνση και την επιτάχυνση της διαδικασίας προγραμματισμού αλλά και για τη δημιουργία μιας ποιοτικής εφαρμογής που μπορεί να ανταγωνιστεί άμεσα επιτυχημένες εφαρμογές που επιτελούν τους ίδιους σκοπούς και λειτουργίες.

Οι εφαρμογές DIANA και DIANA 2.0, έχουν ως στόχο τη διευκόλυνση των χρηστών-βιολόγων στις αναζητήσεις που πραγματοποιούν για τον εντοπισμό microRNA στόχων σε

γονίδια. Το περιβάλλον χρήσης τους είναι σχεδιασμένο ώστε να προσφέρει ευκολία χρήσης, ενώ οι αναζητήσεις πραγματοποιούνται με ταχύτητα και τα αποτελέσματά τους παρουσιάζονται μέσω καλοσχεδιασμένων και εύχρηστων φορμών. Η ποικιλομορφία των υποστηριζόμενων αναζητήσεων έχει επιλεγεί ώστε να ανταποκρίνεται σε όλες τις ανάγκες που μπορεί να παρουσιαστούν στον ερευνητή βιολόγο.

Οι στόχοι που τέθηκαν για το DIANA 2.0 στα πλαίσια της παρούσας διπλωματικής ήταν οι εξής:

- Δημιουργία εύχρηστου περιβάλλοντος για τις λειτουργίες διαχείρισης (π.χ. διαχείριση χρηστών) της εφαρμογής.
- Απλοποίηση του μηχανισμού αναζητήσεων.
- Σχεδίαση νέου, βελτιωμένου μηχανισμού παροχής προτάσεων για λανθασμένους όρους αναζητήσεων.
- Υλοποίηση υποδομής για τη συγκριτική παρουσίαση προβλέψεων microRNA στόχων που προέρχονται από άλλες υπολογιστικές μεθόδους.
- Βελτίωση τεκμηρίωσης του κώδικα του συστήματος για τη διευκόλυνση μελλοντικών επεκτάσεων.

Τα πρόσθετα αυτά χαρακτηριστικά αναμένεται να βοηθήσουν την ομάδα διαχείρισης της εφαρμογής, τους χρήστες της εφαρμογής καθώς επίσης και τους προγραμματιστές που αναμένεται να υλοποιήσουν επόμενες επεκτάσεις στον κώδικα. Η επίτευξη των πρόσθετων στόχων που μόλις περιγράφηκε, στα πλαίσια της διπλωματικής εργασίας, μας γεμίζει με ικανοποίηση και αισιοδοξία για το μέλλον της εφαρμογής

Προβλήματα που παρουσιάστηκαν, πλην όμως αντιμετωπίστηκαν, ήταν τόσο η έλλειψη τεκμηρίωσης των αρχείων κώδικα της προηγούμενης έκδοσης όσο και η επαναλειτουργία της εφαρμογής με καθολική αναφορά σφαλμάτων από τη μεριά του συστήματος. Επιπλέον σε κάθε φάση ανάπτυξης έπρεπε να διασφαλίζεται η συμβατότητα των υποσυστημάτων της προηγούμενης έκδοσης με τα νέα υποσυστήματα.

7.2 Μελλοντικές εργασίες

Το σύστημα που αναπτύχθηκε στα πλαίσια της παρούσας διπλωματικής εργασίας θα αποτελέσει τη βάση για μελλοντικές επεκτάσεις και προσθήκες. Η εφαρμογή DIANA 2.0 είναι μια πραγματική εφαρμογή, η οποία προβλέπεται να εξυπηρετεί καθημερινά εκατοντάδες

βιολόγους-ερευνητές και της οποίας οι ανάγκες συνεχώς μεταβάλλονται λόγω των μεταβολών που υφίσταται το πεδίο της υπολογιστικής βιολογίας.

Ο εμπλουτισμός του συστήματος καταγραφής με επιπλέον πληροφορία που θα αφορά στον τρόπο πλοήγησης του χρήστη μέσα στην εφαρμογή θα μπορούσε να είναι ένα από αυτούς τους τρόπους. Οι τιμές των παραμέτρων που ορίζει ένας χρήστης, οι κωδικοί και τα ονόματα των μορίων microRNA ή των γονιδίων που επιλέγει θα μπορούσαν να καταγράφονται και να χρησιμοποιούνται για την εξαγωγή πληροφορίας που θα αφορά στο ενδιαφέρον του χρήστη γύρω από τα δεδομένα της εφαρμογής. Έτσι η εφαρμογή θα μπορούσε να είναι όλο και πιο προσαρμοσμένη στον κάθε χρήστη εξατομικεύοντας τις υπηρεσίες που παρέχει σε μεγαλύτερο βαθμό. Εκτός αυτού, η καταγραφή της πληροφορίας που αναφέρθηκε προηγουμένως, θα μπορούσε να μελετηθεί περαιτέρω με στόχο την παροχή προτάσεων στο χρήστη για δημοσιεύσεις, άρθρα ή βιβλία γύρω από τα προσωπικά του ερευνητικά ενδιαφέροντα.

Επιπλέον η πλατφόρμα διαχείρισης θα μπορούσε να επεκταθεί κατά τέτοιο τρόπο ώστε να παρέχονται στους διαχειριστές πρόσθετες λειτουργίες. Για παράδειγμα, χρήσιμο θα ήταν να δοθεί η δυνατότητα πραγματοποίησης στατιστικών ερωτημάτων πάνω στα ίδια τα βιολογικά δεδομένα ή πάνω στις συμπεριφορές των χρηστών της εφαρμογής. Επίσης, η ανανέωση των δεδομένων της βάσης, έτσι ώστε να περιλαμβάνει κάθε στιγμή τις πιο πρόσφατες πληροφορίες (ονόματα και ακολουθίες γονιδίων και μορίων microRNA) είναι μια διαδικασία η οποία βαραίνει τους διαχειριστές και η οποία θα μπορούσε να απλοποιηθεί και να αυτοματοποιηθεί μέσω της πλατφόρμας διαχείρισης.

Τέλος, ενδιαφέρον θα παρουσίαζε η πιο εκτεταμένη εκμετάλλευση των προβλέψεων από εναλλακτικά προγράμματα πρόβλεψης προκειμένου να παραχθούν ενοποιημένες λίστες. Τέτοιου είδους λίστες θα μπορούσαν να οδηγήσουν σε ενδιαφέροντα συμπεράσματα για τη βελτίωση της ακρίβειας στις προβλέψεις των διαφόρων αλγορίθμων.

8

Βιβλιογραφία

- [1] Apache HTTP Server Version 2.2 Documentation, <http://httpd.apache.org/docs/2.2/en/>
- [2] Aslam, Javed and Montague, Mark, Models for Metasearch, Annual ACM Conference on Research and Development in Information Retrieval, 2001
- [3] Dwork, Cynthia and Kumar, Ravi and Naor, Moni and Sivakumar, D, Rand Aggregation Methods for the Web, WWW10, 2001
- [4] Liu, Yu-Ting and Liu, Tie-Yan and Qin, Tao and Ma, Zhi-Ming and Li, Hang, Supervised Rank Aggregation, International World Wide Web Conference, 2007
- [5] MySQL 5.1 Reference Manual, <http://dev.mysql.com/doc/refman/5.1/en/index.html>
- [6] PHP Manual, <http://www.php.net/manual/en/>

- [7] Προηγμένη εφαρμογή ιστού διαχείρισης δεδομένων βιοεπιστημών, Γεώργιος Πρέκας, Εργαστήριο Συστημάτων Βάσεων Γνώσεων και Δεδομένων, Δεκέμβριος 2009