



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

ΤΟΜΕΑΣ ΜΗΧΑΝΟΛΟΓΙΚΩΝ ΚΑΤΑΣΚΕΥΩΝ ΚΑΙ ΑΥΤΟΜΑΤΟΥ ΕΛΕΓΧΟΥ

ΣΧΕΔΙΑΣΜΟΣ ΚΑΙ ΑΝΑΠΤΥΞΗ ΤΗΣ ΕΦΑΡΜΟΓΗΣ ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ «SIGNAL TO RESPONSE CORRELATOR» ΓΙΑ ΤΗΝ ΣΥΣΧΕΤΙΣΗ ΠΡΩΤΕΪΝΩΝ ΜΕΣΩ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ ΤΟΥ
ΠΑΠΑΚΩΝΣΤΑΝΤΙΝΟΥ ΑΝΑΣΤΑΣΙΟΥ

ΕΠΙΒΛΕΠΩΝ:
Αλεξόπουλος Λεωνίδας
Λέκτορας ΕΜΠ

Αθήνα, Οκτώβριος 2011



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΜΗΧΑΝΟΛΟΓΙΚΩΝ ΚΑΤΑΣΚΕΥΩΝ ΚΑΙ ΑΥΤΟΜΑΤΟΥ ΕΛΕΓΧΟΥ

ΣΧΕΔΙΑΣΜΟΣ ΚΑΙ ΑΝΑΠΤΥΞΗ ΤΗΣ ΕΦΑΡΜΟΓΗΣ
ΒΙΟΠΛΗΡΟΦΟΡΙΚΗΣ «SIGNAL TO RESPONSE CORRELATOR» ΓΙΑ
ΤΗΝ ΣΥΣΧΕΤΙΣΗ ΠΡΩΤΕΪΝΩΝ ΜΕΣΩ ΕΞΟΥΥΞΗΣ ΔΕΔΟΜΕΝΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ ΤΟΥ
ΠΑΠΑΚΩΝΣΤΑΝΤΙΝΟΥ ΑΝΑΣΤΑΣΙΟΥ

ΕΠΙΒΛΕΠΩΝ:
Αλεξόπουλος Λεωνίδας
Λέκτορας ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 14η Οκτωβρίου 2011.

Αλεξόπουλος Λεωνίδας
Λέκτορας ΕΜΠ

Προβατίδης Χριστόφορος
Καθηγητής ΕΜΠ

Τσαγκάρης Σωκράτης
Καθηγητής ΕΜΠ

Αθήνα, Οκτώβριος 2011

ΠΑΠΑΚΩΝΣΤΑΝΤΙΝΟΥ ΑΝΑΣΤΑΣΙΟΣ

Διπλωματούχος Μηχανολόγος Μηχανικός Ε.Μ.Π.

© 2011 – All rights reserved

Περίληψη

Η παρούσα Διπλωματική Εργασία πραγματοποιήθηκε στο πλαίσιο του μαθήματος «Βιοϊατρικής Τεχνολογίας» του Τομέα Κατασκευών και Αυτομάτου Ελέγχου της Σχολής Μηχανολόγων Μηχανικών και προτείνει μια μεθοδολογία για την εύρεση του συντελεστή συσχέτισης ανάμεσα σε σηματοδότες και σε ενεργούσες πρωτεΐνες εντός και εκτός κυττάρου, μέσω της εξόρυξης δεδομένων και επεξεργασίας των από τις περιλήψεις των χιλιάδων δημοσιεύσεων του Endnote. Η μεθοδολογία αυτή μετουσιώθηκε στον προγραμματισμό της εφαρμογής «Signal to Response Correlator».

Η δομή της είναι χωρισμένη στις ενότητες της Εισαγωγής, των Μεθόδων, των Αποτελεσμάτων, της Συζήτησης, των Μελλοντικών Βελτιώσεων και τέλος της Βιβλιογραφίας.

Στην ενότητα των Μεθόδων αναλύονται οι βασικές αρχές της τεχνολογίας εξόρυξης δεδομένων, η διαδικασία εύρεσης συσχέτισεων στην βιοπληροφορική, η μεθοδολογία που ακολουθούμε μέχρι την εύρεση του συντελεστή συσχέτισης των πρωτεϊνών, η ροή δεδομένων, η ροή εργασιών, επεξήγηση των βάσεων δεδομένων και της εφαρμογής που κατασκεύασα, παράμετροι που χρησιμοποιεί το σύστημα και τέλος ανάλυση της γραφικής διεπιφάνειας Graphic User Interface της εφαρμογής.

Στην ενότητα των Αποτελεσμάτων αναλύονται οι τιμές των παραμέτρων που χρησιμοποιούνται και δίνεται μια μελέτη περιπτώσης για ένα ζευγάρι πρωτεϊνών.

Στην ενότητα της Συζήτησης αναλύονται τα πλεονεκτήματα και μειονεκτήματα της εφαρμογής έναντι άλλων εφαρμογών Βιοϊατρικής τεχνολογίας.

Στην ενότητα των Μελλοντικών Βελτιώσεων αναλύονται όλες οι μελλοντικές προσθήκες, αλλαγές και βελτιώσεις που θα μπορούν να γίνουν στην εφαρμογή.

Τέλος παρουσιάζεται η Βιβλιογραφία που χρησιμοποιήθηκε για την συγγραφή της Διπλωματικής Εργασίας.

Λέξεις κλειδιά: Εξόρυξη κειμένου, βιοπληροφορική, συντελεστής συσχέτισης πρωτεϊνών, σηματοδότες πρωτεΐνες, ενεργούσες πρωτεΐνες

Abstract

This thesis titled “Design and development of the data mining-bioinformatics application “Signal To Response Correlator””, was written in association with the class "Biomedical Engineering" of the Systems Biology & Bioengineering Research Laboratory, Dept. of Mechanical Engineering, National Technical University of Athens. We propose a methodology for assigning the correlation coefficient among signaling proteins inside the cell and response proteins outside the cell, using data mining techniques while processing thousands of abstracts from Endnote. The methodology was implicated in the software «Signal to Response Correlator».

The structure of the thesis is divided into the following sections: Introduction, Methods, Results, Discussion, Future Improvements and finally the References.

The Methods section includes the basic principles of data mining technology, the process of finding correlations in bioinformatics, the methodology which was followed in order to assign the correlation coefficient, the data flow, workflow, explanation of the database of the software that I built, the parameters that the system is using and finally analyzes the Graphic User Interface.

The Results section includes the parameters used and given a case study of a pair of proteins.

The Discussion section includes the advantages and disadvantages of the application.

The Future Improvements section includes all future additions, changes and improvements can be made on implementation.

Finally the References used for writing the thesis are presented.

Keywords: data mining, bioinformatics, correlation coefficient, signaling proteins, response proteins

Περιεχόμενα

Κεφάλαιο 1ο ΕΙΣΑΓΩΓΗ	1
Κεφάλαιο 2ο ΜΕΘΟΔΟΙ	2
2.1 Τεχνολογία εξόρυξης δεδομένων κειμένου	2
2.1.1 Γενικά	2
2.1.2 Εννοιες	3
2.1.3 Διαδικασία εξόρυξης δεδομένων	3
2.1.4 Απαιτήσεις υλικού	6
2.2 Βιοπληροφορική	8
2.2.1 Γενικά	8
2.2.2 Online εργαλεία	8
2.3 Η μεθοδολογία μας	11
2.3.1 Γενική αρχή της μεθοδολογίας μας	11
2.3.2 Διαγράμματα ροής	12
2.3.2.1 Γενικό διάγραμμα ροής	13
2.3.2.2 Διάγραμμα ροής εισαγωγής στοιχείων στις βάσεις δεδομένων	15
2.3.2.3 Διάγραμμα ροής αναζήτησης	23
2.3.2.4 Διάγραμμα ροής επεξεργασίας προτάσεων από τον χρήστη	26
2.3.3 Επεξήγηση του GUI	28
2.3.3.1 Είσοδος στην εφαρμογή	28
2.3.3.2 Μενού αναζήτησης	29
2.3.3.3 Μενού κέντρου ελέγχου	36
2.3.3.4 Αλλαγή προτιμήσεων χρήστη	49
2.3.3.5 Βοήθεια	50
2.3.3.6 Επιπρόσθετα δικαιώματα διαχειριστή	51
Κεφάλαιο 3ο ΑΠΟΤΕΛΕΣΜΑΤΑ	55
Κεφάλαιο 4ο ΣΥΖΗΤΗΣΗ	64
4.1 Πλεονεκτήματα	64
4.1.1 Σύγκριση με τεχνικές εξόρυξης δεδομένων	64
4.1.2 Σύγκριση με αντίστοιχα προγράμματα	65
4.1.3 Πλεονεκτήματα εφαρμογής	65
4.2 Μειονεκτήματα	69
4.3 Ελάχιστες απαιτήσεις	69
Κεφάλαιο 5ο ΜΕΛΛΟΝΤΙΚΕΣ ΒΕΛΤΙΩΣΕΙΣ	71
Παράρτημα ΒΙΒΛΙΟΓΡΑΦΙΑ	72

Κεφάλαιο 1^ο

ΕΙΣΑΓΩΓΗ

Η τεχνολογία εξόρυξης δεδομένων έχει χρησιμοποιηθεί ευρέως σε εργασίες βιοπληροφορικής και βιοϊατρικής τεχνολογίας. Η τεχνολογία εξόρυξη δεδομένων είναι η διαδικασία της εξόρυξης προτύπων από μεγάλα σύνολα δεδομένων με το συνδυασμό των μεθόδων από την στατιστική και τη διαχείριση βάσεων δεδομένων. Ο πρωταρχικός στόχος της βιοπληροφορικής είναι να αυξηθεί η κατανόηση των βιολογικών διαδικασιών.

Το γενικότερο αντικείμενο του διεπιστημονικού κλάδου που μας αφορά είναι οι συσχετίσεις-αλληλεπιδράσεις μεταξύ πρωτεϊνών.

Η μια περίπτωση είναι η εκκίνηση της έρευνας να γίνει με βάση την ταυτοποίηση της συσχέτισης μεταξύ αγνώστων μεταξύ τους πρωτεϊνών. Εδώ αναφέρουμε την χρήση προγραμμάτων όπως τα Ingenuity, Pathway studio, Reactome, τα οποία χρησιμοποιούνται για την ανάπτυξη συνδέσεων-διαδρόμων (“pathways”) ανάμεσα σε σηματοδότες και σε ενεργούσες πρωτεΐνες, για τις οποίες δεν γνωρίζουμε την σχέση τους.

Σε αυτή την διπλωματική προτείνουμε μια (αντίστροφη ως προς την πρώτη περίπτωση) μεθοδολογία για την εύρεση του συντελεστή συσχέτισης ανάμεσα σε σηματοδότες (δίνουν το σήμα) πρωτεΐνες εντός κυττάρου και σε ενεργούσες (αυτές που ενεργοποιούνται) πρωτεΐνες εκτός κυττάρου, δηλαδή σε πρωτεΐνες που γνωρίζουμε ότι συσχετίζονται. Αυτή η μεθοδολογία έχει ως αρχικό στάδιο την εισαγωγή δεδομένων από τις περιλήψεις χιλιάδων δημοσιεύσεων του Endnote, συνώνυμα πρωτεϊνών, λέξεις συσχετίσεων, στάδιο επεξεργασίας την εξόρυξη δεδομένων και προτύπων από αυτά και στάδιο αποτελεσμάτων την παρουσίαση τους στον χρήστη, για να γίνει η τελική επιλογή τους και η εξεύρεση-επιλογή του συντελεστή συσχέτισης.

Στην συνέχεια αυτές οι συσχετίσεις πρωτεϊνών μαζί με τους συντελεστές μπορούν να συνθέσουν μια τοπολογία πρωτεϊνών όπου θα φαίνονται το πώς η μία επιρρεάζει την άλλη.

Η μεθοδολογία αυτή μετουσιώθηκε στον προγραμματισμό της εφαρμογής «Signal to Response Correlator» <http://www.jojo.gr/thesis>, για την οποία θα γίνει εκτενής αναφορά.

Κεφάλαιο 2^ο

ΜΕΘΟΔΟΙ

2.1 Τεχνολογία εξόρυξης δεδομένων κειμένου

2.1.1 Γενικά

Η *εξόρυξη δεδομένων*, ένα σχετικά νέο διεπιστημονικό πεδίο στην επιστήμη των υπολογιστών, είναι η διαδικασία της εξόρυξης προτύπων (ονομάζονται και συσχετίσεις, συσχετισμοί, μοτίβα) από μεγάλα σύνολα δεδομένων με το συνδυασμό των μεθόδων από την Στατιστική και Τεχνητή Νοημοσύνη με τη διαχείριση βάσεων δεδομένων. Επιτρέπει στους χρήστες να αναλύουν δεδομένα από πολλές διαφορετικές διαστάσεις ή οπτικές γωνίες, να ταξινομούν, και να συνοψίσουν τις σχέσεις που εντοπίζουν. Τεχνικά, η εξόρυξη δεδομένων είναι η διαδικασία της εύρεσης προτύπων ανάμεσα σε δεκάδες τομείς με μεγάλες σχεσιακές βάσεις δεδομένων.¹

Η *εξόρυξη κειμένου*, μερικές φορές αναφέρεται ως εξόρυξη δεδομένων κειμένου, είναι η διαδικασία που απορρέουν υψηλής ποιότητας πληροφορίες από το κείμενο. Η εξόρυξη κειμένου συνήθως περιλαμβάνει τη διαδικασία της κατάλληλης δόμησης/επεξεργασίας διαμέσου φίλτρων του κειμένου εισόδου, την επακόλουθη εισαγωγή σε μια βάση δεδομένων, την επεξεργασία και απόρροια συσχετισμών πληροφοριών και τέλος την αξιολόγηση και ερμηνεία των συσχετισμών δεδομένων εξόδου. «Υψηλής ποιότητας» στην εξόρυξη κειμένου αναφέρεται συνήθως σε κάποιο συνδυασμό συνάφειας, καινοτομίας και ενδιαφέροντος. Τυπικές εφαρμογές εξόρυξης κειμένου μπορούμε να βρούμε στην Λογοτεχνία, σε διαδικασίες Μαρκετινγκ, στην Βιοπληροφορική κα.²

Με τις πρόσφατες τεχνικές προόδους στην επεξεργαστική ισχύ και χωρητικότητα αποθήκευσης, η εξόρυξη δεδομένων έχει θεωρηθεί ως ένα όλο και πιο σημαντικό εργαλείο από τις σύγχρονες επιχειρήσεις να μετατρέψουν τις άνευ προηγουμένου ποσοτήτες ψηφιακών δεδομένων σε επιχειρηματική ευφυΐα.

2.1.2 Έννοιες

Θα πρέπει εδώ να αποσαφηνίσουμε κάποιες έννοιες.

Δεδομένα: είναι οποιαδήποτε στοιχεία, αριθμοί ή κείμενο που μπορεί να υποστούν επεξεργασία από έναν υπολογιστή. Σήμερα, οργανισμοί συσσωρεύουν τεράστιες και αυξανόμενες ποσότητες δεδομένων σε διαφορετικές μορφές και διαφορετικές βάσεις δεδομένων. Αυτό περιλαμβάνει δεδομένα και μετα-δεδομένα (στοιχεία σχετικά με τα δεδομένα ή ίδια, όπως λογικό σχεδιασμό βάσεων δεδομένων ή ορισμούς λεξικό δεδομένων)

Πρότυπο: Μια έκφραση σε μια γλώσσα που χαρακτηρίζει ένα υποσύνολο των δεδομένων (π.χ. ένας κανόνας).

Εγκυρότητα: Το πρότυπο να είναι συνεπές σε νέα δεδομένα.

Πληροφορίες: είναι συσχετίσεις μεταξύ όλων αυτών των δεδομένων.

Γνώση: είναι οι πληροφορίες που μπορούν να μετατραπούν σε γνώση σχετικά με παλαιότερους συσχετισμούς αλλά και με μελλοντικά πρότυπα.

Αποθήκευση δεδομένων: ορίζεται ως μια διαδικασία της κεντρικής διαχείρισης των δεδομένων και της ανάκτηση τους. Αποθήκευση δεδομένων αποτελεί ένα ιδανικό όραμα διατήρησης μιας κεντρικής αποθήκης όλων των δεδομένων. Κεντροποιημένη συγκέντρωση των δεδομένων είναι απαραίτητη για να μεγιστοποιηθεί η πρόσβαση των χρηστών και η ανάλυση των δεδομένων.

Φίλτρα: κομμάτια κώδικα ή οντότητες προγραμματισμού που εντοπίζουν και απομακρύνουν σε ένα σύνολο δεδομένων άχρηστες πληροφορίες

2.1.3 Διαδικασία εξόρυξης δεδομένων

Η διαδικασία της εξόρυξης δεδομένων κειμένου περιλαμβάνει την προεπεξεργασία, την εξόρυξη και τα αποτελέσματα.³

Προ-επεξεργασία: Πριν εφαρμοστούν οι αλγόριθμοι εξόρυξης δεδομένων πρέπει να γίνει η συλλογή των δεδομένων. Δεδομένου ότι η εξόρυξη δεδομένων μπορεί να αποκαλύψει μόνο τα πρότυπα που υπάρχουν ήδη στα δεδομένα, το σύνολο δεδομένων που θα χρησιμοποιήσουμε πρέπει να είναι αρκετά μεγάλο ώστε να περιέχει αυτά τα πρότυπα και να μπορούν να εξάγονται σε ένα αποδεκτό χρονικό πλαίσιο. Τα δεδομένα στη συνέχεια φιλτράρονται, απομακρύνονται οι «θόρυβοι», δηλαδή άχρηστες πληροφορίες και δομούνται κατάλληλα για

την εισαγωγή τους στην βάση δεδομένων. Η προ-επεξεργασία είναι απαραίτητη για την ανάλυση των πολυδιάστατων σύνολων δεδομένων πριν από την εξόρυξη δεδομένων.

Στην διαδικασία *εξόρυξης δεδομένων κειμένου*, παρέχεται πρόσβασης στα δεδομένα από τους αναλυτές και το λογισμικό αναλύει τους συσχετισμούς και τα πρότυπα στα αποθηκευμένα δεδομένα, που βασίζονται στα ανοικτού τύπου ερωτήματα των χρηστών.

Η εξόρυξη δεδομένων περιλαμβάνει συνήθως την εφαρμογή των παρακάτω αλγορίθμων για παραγωγή συσχετίσεων:^{4 5}

Προβλεπτικού τύπου, που άνει πρόβλεψη συμπεριφοράς κάποιων μεταβλητών που παρουσιάζουν ενδιαφέρον. Εδώ υπάρχουν οι περιπτώσεις:

- Κατηγοριοποίηση (Classification). Βασίζεται στην εξέταση των χαρακτηριστικών ενός νέου αντικειμένου το οποίο με βάση τα χαρακτηριστικά αυτά αντιστοιχίζεται σε ένα προκαθορισμένο σύνολο κλάσεων. Τα προς κατηγοριοποίηση αντικείμενα αναπαριστάνονται γενικά από τις εγγραφές της βάσης δεδομένων. Η διαδικασία της κατηγοριοποίησης αποτελείται από την ανάθεση κάθε εγγραφής σε κάποιες από τις προκαθορισμένες κατηγορίες.
- Παλινδρόμηση ή Παρεμβολή (Regression). Αξιολογεί τις σχέσεις μιας μεταβλητής, εξαρτημένη μεταβλητή, με μία ή περισσότερες άλλες μεταβλητές, που ονομάζονται ανεξάρτητες μεταβλητές. Κάθε ανεξάρτητη μεταβλητή ιστορικά έχει σχέση με την εξαρτημένη μεταβλητή, η μελλοντική αξία της οποίας μπορεί να εκτιμηθεί.
- Ανάλυση χρονοσειρών (Time series analysis). Ακολουθία αριθμών κάθε ένας από τους οποίους έχει ένα timestamp (ετικέτα χρόνου). Χαρακτηριστικά υποθέτουμε ότι οι διαδοχικοί αριθμοί χωρίζονται από ένα σταθερό χρονικό διάστημα, και η πραγματική ετικέτα χρόνου παραλείπεται.
- Πρόβλεψη (Prediction). Προβλέπει την κατηγορία, κατηγοριοποιεί δεδομένα (κατασκευάζοντας ένα μοντέλο) βασισμένο σε training set και τις τιμές του χαρακτηριστικού κατηγορίας και το χρησιμοποιεί για κατηγοριοποίηση νέων δεδομένων

Περιγραφικού τύπου, που βρίσκει συσχετίσεις που υπάρχουν στα δεδομένα. Εδώ υπάρχουν οι περιπτώσεις:

- Συσταδοποίηση (Clustering). Η συσταδοποίηση δεδομένων (data clustering) είναι η εργασία του καταμερισμού ενός ετερογενούς πληθυσμού σε ένα σύνολο συστάδων

(clusters). Τα αντικείμενα σε μια ομάδα θα είναι παρόμοια το ένα με ένα άλλο και διαφορετικό από τα αντικείμενα σε άλλες ομάδες. Επίσης γνωστό ως επιβλεπόμενη ταξινόμηση. Στην συσταδοποίηση δεν υπάρχουν προκαθορισμένες κατηγορίες. Οι εγγραφές ομαδοποιούνται σε σύνολα με βάση την ομοιότητα που παρουσιάζουν μεταξύ τους. Επαφίεται σε εμάς να καθορίσουμε την σημασία που θα έχει κάθε μία από τις ομάδες που προκύπτουν.

- Σύνοψη (Summarization) ή Γενίκευση (Generalization)
- Εύρεση Κανόνων Συσχέτισης (Association Rules). Οι κανόνες συσχέτισης ανακαλύπτουν κρυμμένες «συσχετίσεις» μεταξύ των γνωρισμάτων ενός συνόλου των δεδομένων. Αυτοί οι συσχετισμοί παρουσιάζονται στην ακόλουθη μορφή: $A \rightarrow B$ όπου το **A** και το **B** αναφέρονται στα σύνολα γνωρισμάτων που υπάρχουν στα υπό ανάλυση δεδομένα.
- Ανακάλυψη Συσχετίσεων σε Ακολουθίες (Pattern Discovery in Sequences)

Αποτελέσματα: είναι το τελικό στάδιο της ανακάλυψης γνώσης από τα δεδομένα για την επαλήθευση των συσχετίσεων που παράγονται από τους αλγόριθμους εξόρυξης δεδομένων. Αναφέρουμε ότι δεν είναι όλοι οι συσχετισμοί που βρέθηκαν από τους αλγόριθμους εξόρυξης δεδομένων απαραίτητα έγκυροι. Είναι κοινό για τους αλγόριθμους εξόρυξης δεδομένων να βρουν συσχετισμούς στο σύνολο δεδομένων εκμάθησης που δεν είναι παρόντα στο γενικό σύνολο δεδομένων. Για να ξεπεραστεί αυτό, η αξιολόγηση των αλγορίθμων χρησιμοποιεί για δοκιμή ένα σύνολο δεδομένων κατά την οποία ο αλγόριθμος εξόρυξης δεδομένων δεν είναι εκπαιδευμένος. Οι συσχετισμοί που έχουν βρεθεί εφαρμόζονται σε αυτό το σύνολο δεδομένων και τα αποτελέσματα συγκρίνονται με το επιθυμητό αποτέλεσμα. Εάν οι συσχετίσεις που θα βρεθούν δεν πληρούν τις επιθυμητές προδιαγραφές, τότε θα πρέπει να αξιολογηθεί εκ νέου το στάδιο προ-επεξεργασίας και εξόρυξης δεδομένων. Στην άλλη περίπτωση οι συσχετίσεις μετατρέπονται σε γνώση και παρουσιάζονται σε μια χρήσιμη μορφή, όπως ένα γράφημα, έναν πίνακα ή μια αναφορά.

Διαφορετικοί τρόποι ανάλυσης είναι διαθέσιμοι:

- Τεχνητά νευρωνικά δίκτυα: Μη-γραμμικά προβλεπτικά μοντέλα που μαθαίνουν μέσω της κατάρτισης και μοιάζουν στη δομή με βιολογικά νευρωνικά δίκτυα.
- Γενετικοί αλγόριθμοι: τεχνικές βελτιστοποίησης που χρησιμοποιούν διαδικασίες όπως ο γενετικός συνδυασμός, μετάλλαξη και η φυσική επιλογή σε ένα σχέδιο που βασίζεται στις έννοιες της φυσικής εξέλιξης.

- Δέντρα απόφασεων: Δομές σχήματος δέντρου που αντιπροσωπεύουν το σετ των αποφάσεων. Οι αποφάσεις αυτές παράγουν κανόνες για την ταξινόμηση ενός συνόλου δεδομένων.
- Μέθοδος κοντινότερου: Μια τεχνική που ταξινομεί κάθε εγγραφή σε ένα σύνολο δεδομένων με βάση ένα συνδυασμό των ομάδων της εγγραφής k που μοιάζουν περισσότερο σε ένα ιστορικό σύνολο δεδομένων.
- Επαγωγικός κανόνας: Η εξαγωγή χρήσιμων AN-TOTE κανόνων από δεδομένα που βασίζονται σε στατιστική.
- Οπτικοποίηση δεδομένων: Η οπτική ερμηνεία των πολύπλοκων σχέσεων σε πολυδιάστατα δεδομένα. Γραφικά εργαλεία που χρησιμοποιούνται για να απεικονίζουν τις σχέσεις των δεδομένων.

Σχηματικά μπορούμε να παρουσιάσουμε το παρακάτω αναλυτικό διάγραμμα για την διαδικασία λειτουργίας της εξόρυξης κειμένου.

ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ → ΕΠΙΛΟΓΗ ΔΕΔΟΜΕΝΩΝ → ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ → ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ ΔΕΔΟΜΕΝΩΝ → ΕΞΟΡΥΞΗ ΣΥΣΧΕΤΙΣΜΩΝ → ΑΞΙΟΛΟΓΗΣΗ ΣΥΣΧΕΤΙΣΜΩΝ → ΠΛΗΡΟΦΟΡΙΕΣ → ΠΑΡΟΥΣΙΑΣΗ ΓΝΩΣΗΣ

Υπάρχουν αρκετοί γνωστοί αλγόριθμοι εξόρυξης δεδομένων. Οι πιο κοινοί είναι οι C4.5, The k -Means algorithm, Support Vector Machines, The Apriori algorithm, Expectation-Maximization, PageRank, AdaBoost, k -Nearest Neighbor Classification, Naive Bayes, CART (Classification and Regression Trees)⁶

2.1.4 Απαιτήσεις υλικού

Οι απαιτήσεις για την εξόρυξη δεδομένων είναι ο χειρισμός διαφορετικών τύπων δεδομένων, η απόδοση και εξελισιμότητα των αλγορίθμων εξόρυξης δεδομένων, η χρησιμότητα, εγκυρότητα και εκφραστικότητα των αποτελεσμάτων εξόρυξης δεδομένων από διαφορετικές πηγές δεδομένων.

Σχετικά με την τεχνολογική υποδομή υλικού που απαιτούν οι εφαρμογές εξόρυξης δεδομένων, έχουμε διαθέσιμα συστήματα μέγεθους mainframe, client / server και πλατφόρμες PC. Οι τιμές κυμαίνονται από χιλιάδες ευρώ για τις μικρότερες εφαρμογές μέχρι

και 1 εκατ. ευρώ ανά terabyte για τις μεγαλύτερες. Υπάρχουν δύο κρίσιμα τεχνολογικά κατευθυντήρια κριτήρια:

- Το μέγεθος της βάσης δεδομένων: ο αριθμός των δεδομένων που υποβάλλονται σε επεξεργασία και διατήρηση, τόσο πιο ισχυρό το σύστημα που απαιτείται.
- Η πολυπλοκότητα των ερωτημάτων: όσο πιο πολύπλοκα τα ερωτήματα στην βάση και όσο μεγαλύτερος ο αριθμός των ερωτήσεων που υποβάλλονται σε επεξεργασία, τόσο πιο ισχυρό το σύστημα που απαιτείται.

2.2 Βιοπληροφορική

2.2.1 Γενικά

Βιοπληροφορική είναι η εφαρμογή της επιστήμης των υπολογιστών και της τεχνολογίας των πληροφοριών στον τομέα της βιολογίας. Ο πρωταρχικός στόχος της βιοπληροφορικής είναι να αυξηθεί η κατανόηση των βιολογικών διαδικασιών. Αυτό που ξεχωρίζει από άλλες προσεγγίσεις, είναι εστίασή της στην ανάπτυξη και εφαρμογή υπολογιστικών τεχνικών (π.χ., αναγνώριση προτύπων, εξόρυξη δεδομένων, αλγόριθμοι μηχανικής μάθησης, και οπτικοποίηση) για την επίτευξη αυτού του στόχου. Σημαντικές ερευνητικές προσπάθειες στον τομέα περιλαμβάνουν εύρεση γονιδίων, συναρμολόγηση γονιδιώματος, τον σχεδιασμό των φαρμάκων, την ανακάλυψη φαρμάκων, πρόβλεψη δομής πρωτεϊνών, η πρόβλεψη της γονιδιακής έκφρασης και των αλληλεπιδράσεων μεταξύ πρωτεϊνών.⁷

Η βιοπληροφορική περιλαμβάνει την δημιουργία και την εξέλιξη των βάσεων δεδομένων, αλγόριθμων, υπολογιστικές και στατιστικές τεχνικές και θεωρία για την επίλυση τυπικών και πρακτικών προβλημάτων που προκύπτουν από τη διαχείριση και την ανάλυση των βιολογικών δεδομένων. Κατά τις τελευταίες δεκαετίες ραγδαίες εξελίξεις σε γονιδιακές και άλλες μοριακές τεχνολογικές έρευνες και της ανάπτυξης στον τομέα των τεχνολογιών πληροφοριών, έχουν δημιουργήσει ένα τεράστιο ποσό των πληροφοριών που σχετίζονται με τη μοριακή βιολογία.

Κοινές δραστηριότητες στη Βιοπληροφορική περιλαμβάνουν τη χαρτογράφηση και την ανάλυση του DNA και τις ακολουθίες πρωτεϊνών, σύγκριση ακολουθιών και προβολή 3-D μοντέλων των πρωτεϊνικών δομών.

Η *ανάλυση των δημοσιεύσεων* είναι ένας από τους τομείς της βιοπληροφορικής. Η αύξηση του αριθμού των δημοσιεύσεων καθιστά σχεδόν αδύνατο σε κάποιον να διαβάσει κάθε έγγραφο. Η ανάλυση των δημοσιεύσεων έχει ως στόχο χρησιμοποιώντας εξόρυξη δεδομένων κειμένου, υπολογιστική γλωσσολογία και τη στατιστική να εξορύσει πληροφορίες από τα κείμενα. Μία από τις μεγάλες χρήσεις είναι η εύρεση αλληλεπίδρασης μεταξύ πρωτεϊνών.

Υπάρχει ένα αυξανόμενο ενδιαφέρον για την εξόρυξη κειμένου στην βιολογική λογοτεχνία, λόγω του αυξανόμενου αριθμού των ηλεκτρονικά διαθέσιμων δημοσιεύσεων που αποθηκεύονται σε βάσεις δεδομένων όπως το PubMed.⁸

2.2.2 Online εργαλεία

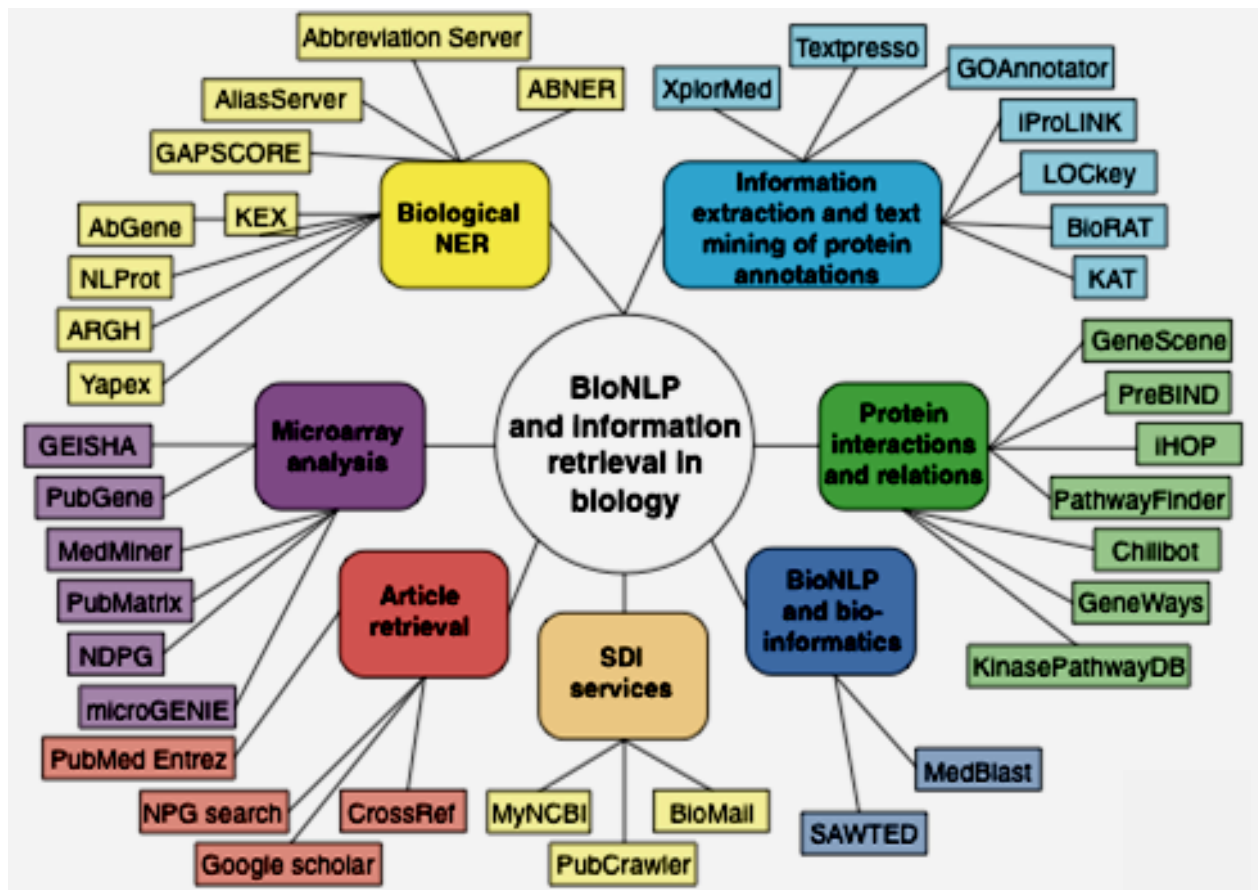
Μερικά από τα online εργαλεία που υπάρχουν είναι:

- KLEIO: ένα σύστημα ανάκτησης πληροφοριών βιοϊατρικής
- FACTA+: μια μηχανή αναζήτησης της Medline για την εύρεση συσχετίσεων μεταξύ βιοϊατρικών εννοιών. Η FACTA + Visualizer βοηθά στην κατανόηση των αποτελεσμάτων αναζήτησης FACTA+ μέσω της γραφικής απεικόνισης των αποτελεσμάτων.
- TerMine: ένα σύστημα διαχείρισης όρων και αναγνώρισής τους σε κείμενα βιοϊατρικής.
- AcroMine: Ένα αρκτικόλεξο λεξικό το οποίο μπορεί να χρησιμοποιηθεί για να βρούμε διαφορετικά ακρωνύμια από το Medline
- Chilibot: Ένα εργαλείο για την εύρεση σχέσεων μεταξύ των γονιδίων και των γονιδιακών προϊόντων.
- EBIMed: είναι μια διαδικτυακή εφαρμογή που συνδυάζει ανάκτηση πληροφορίας και εξόρυξη κειμένων από το Medline
- GOAnnotator: ένα online εργαλείο που χρησιμοποιεί τη σημασιολογική ομοιότητα για την επαλήθευση των πρωτεϊνών χρησιμοποιώντας GO όρους από τη βιβλιογραφία.
- GoPubMed: ανακτά περιλήψεις δημοσιεύσεων από το PubMed, μετά ανιχνεύει όρους οντολογίας και επιτρέπει στο χρήστη να περιηγηθεί στα αποτελέσματα αναζήτησης. Εμφανίζει μόνο τα έγγραφα που παραπέμπουν σε επιλεγμένους όρους ή συνώνυμα.
- IHOP: «Ένα δίκτυο γονιδίων και πρωτεϊνών που εκτείνεται μέσα από την επιστημονική βιβλιογραφία για φαινότυπους και λειτουργία των γονιδίων. Το IHOP παρέχει αυτό το δίκτυο ως ένα φυσικό τρόπο πρόσβασης σε εκατομμύρια περιλήψεις του PubMed. Με τη χρήση των γονιδίων και των πρωτεϊνών ως υπερ-συνδέσεις μεταξύ των προτάσεων και περιλήψεων, οι πληροφορίες στο PubMed μπορεί να μετατραπούν σε ένα διαδραστικό πόρο, φέρνοντας όλα τα πλεονεκτήματα του διαδικτύου στην επιστημονική έρευνα βιβλιογραφίας.»
- NextBio-Life: μηχανή αναζήτησης, με λειτουργικότητα εξόρυξης κειμένου που χρησιμοποιεί περιλήψεις του PubMed.
- PubAnatomy: μια αλληλεπιδραστική οπτική μηχανή αναζήτησης που προσφέρει νέους τρόπους για να διερευνήσει τις σχέσεις μεταξύ δημοσιεύσεων στο Medline, τα

αποτελέσματα εξόρυξης κειμένου, την έκφραση των γονιδίων και άλλες βασικές πληροφορίες.

- PubGene: συνύπαρξη δικτύων απεικόνισης των γονιδίων και πρωτεϊνών.

Παραθέτω ένα διάγραμμα για το πώς συνδέονται οι πιο γνωστές εφαρμογές βιοπληροφορικής.



2.3 Η μεθοδολογία μας

Θα αναλύσουμε την δομή με την πλήρη ανάλυση της μεθοδολογίας που ακολουθούμε για την εύρεση του συντελεστή συσχέτισης ανάμεσα σε σηματοδότες (δίνουν το σήμα) πρωτεΐνες εντός κυττάρου και σε ενεργούσες (αυτές που ενεργοποιούνται) πρωτεΐνες εκτός κυττάρου, που δηλαδή γνωρίζουμε ότι έχουν κάποια συσχέτιση.

2.3.1 Γενική αρχή της μεθοδολογίας μας.

Η γενική αρχή είναι ότι τα άρθρα της βιολογίας που αφορούν μια ομάδα πρωτεϊνών, θα τις περιέχουν περιέχουν ως σημαντική αναφορά στην περίληψή τους.⁹ Πρωτεΐνες που έχουν σημαντική σχέση μεταξύ τους θα βρίσκονται στην ίδια πρόταση των περιλήψεων των άρθρων.

Άρα εμείς αναζητούμε με το Endnote άρθρα τα οποία θα έχουν αναφορά στο ζευγάρι πρωτεϊνών που μας ενδιαφέρει να ορίσουμε τον συντελεστή συσχέτισής των. Από την πλευρά της εξόρυξης δεδομένων, αναπτύσσουμε μια εφαρμογή που διασπάει την περίληψη κάθε άρθρου σε προτάσεις και αναζητεί σε κάθε πρόταση τις λέξεις-κλειδιά δηλαδή της πρωτεΐνες. Τώρα αν αυτές οι πρωτεΐνες βρίσκονται σε μια σχετική εγγύτητα και υπάρχουν στην γειτνίασή τους λέξεις που υποδηλώνουν συσχέτιση τους, τότε μπορούμε να θεωρήσουμε ότι αυτή την πρόταση υποδηλώνει συσχέτιση για τις λέξεις-κλειδιά (πρωτεΐνες).¹⁰

Για παράδειγμα μετά από την εφαρμογή του παραπάνω τμήματος της μεθοδολογίας, βρίσκουμε την εξής πρόταση: “Protein A induces Protein B”. Σε αυτή την πρόταση οι λέξεις κλειδιά απέχουν 2 λέξεις (σχετική εγγύτητα) και παρεμβάλεται ανάμεσά τους μια λέξη συσχέτισης. Αυτή η πρόταση υποδηλώνει θετική συσχέτιση μεταξύ των A και B.

Το σύνολο των επεξεργασμένων προτάσεων τις παρουσιάζει στον χρήστη για να αποφανθεί της πληρότητας και της ορθότητάς τους. Ελέγχει δηλαδή αν είναι θετική, αρνητική ή καμία συσχέτιση μεταξύ των λέξεων-κλειδιών. Μετά από το πέρας του ελέγχου του συνόλου των προτάσεων, ο χρήστης βλέπει μια συνοπτική αναφορά με αυτές και αποφένεται για τον συντελεστή συσχέτισης (συντελεστής βαρύτητας) των δυο λέξεων κλειδιών.

Σημείωση: μετά από ενδελεχή αναζήτηση βρήκαμε ότι οι λέξεις που υποδηλώνουν συσχέτιση είναι οι : *activate, affect, associate, bind, block, cause, characterize, compare, contain, decrease, detect, develop, effect, expose, express, find, include, increase, incubate, induce, inhibit, involve, measure, mediate, obtain, occur, play, prevent, produce, reduce, regulate, result, reveal, secrete, show, specific, stimulate, treat*. Επίσης συνήθως η Protein A και η Protein B δεν απέχουν περισσότερο από 4 λέξεις.

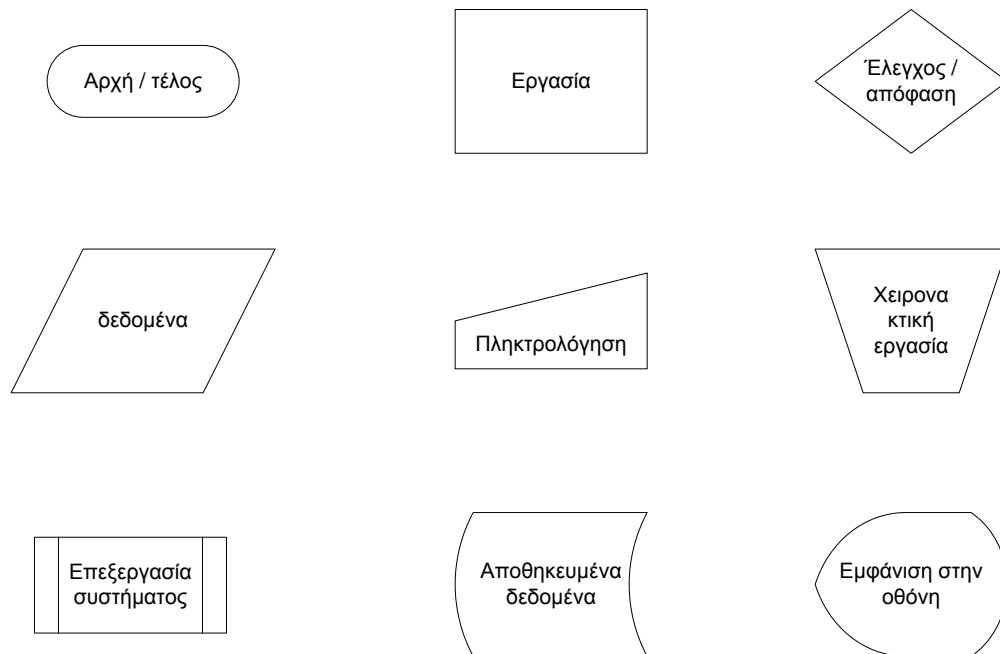
Ένα σχηματικό διάγραμμα της μεθοδολογίας μας είναι το παρακάτω.

ΑΡΘΡΑ → ΑΝΑΓΝΩΡΙΣΗ ΠΡΟΤΑΣΕΩΝ → ΕΥΡΕΣΗ ΛΕΞΕΩΝ ΚΛΕΙΔΙΩΝ → ΕΛΕΓΧΟΣ ΕΓΓΥΤΗΤΑΣ ΛΕΞΕΩΝ ΚΛΕΙΔΙΩΝ ΚΑΙ ΛΕΞΗΣ ΣΥΣΧΕΤΙΣΗΣ → ΕΛΕΓΧΟΣ ΣΥΣΧΕΤΙΣΗΣ ΑΠΟ ΤΟΝ ΧΡΗΣΤΗ → ΑΘΡΟΙΣΤΙΚΑ ΑΠΟΤΕΛΕΣΜΑΤΑ → ΑΠΟΔΟΣΗ ΣΥΝΤΕΛΕΣΤΗ ΣΥΣΧΕΤΙΣΗΣ (ΒΑΡΟΥΣ)

Στις παρακάτω ενότητες θα αναλύσουμε σε βάθος της παραπάνω γενική αρχή της μεθοδολογίας μας και θα δώσουμε έμφαση στις παραμέτρους και στους περιορισμούς της.

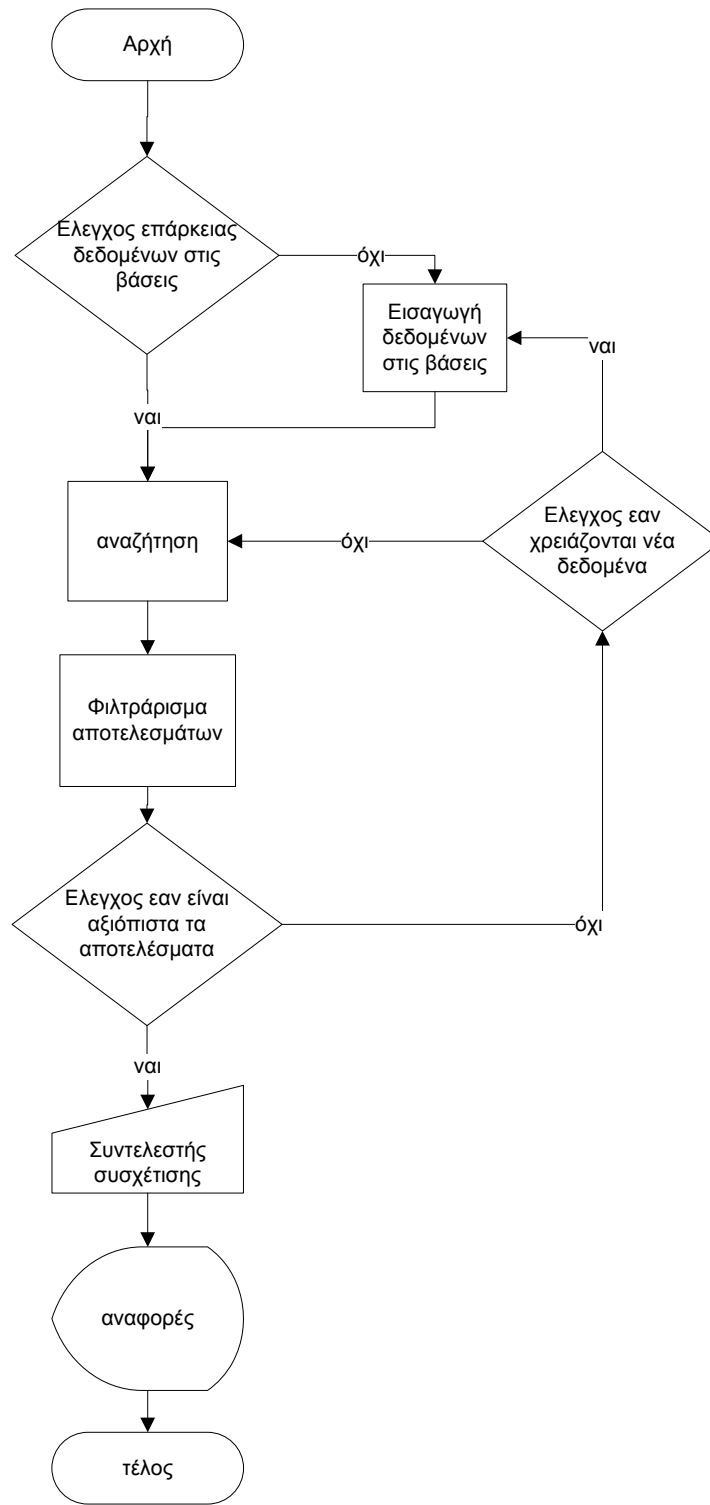
2.3.2 Διαγράμματα ροής

Θα ξεκινήσουμε με το διάγραμμα ροής και θα παράθεσουμε το λεξικό και την σημασία των διαφόρων σχημάτων που θα χρησιμοποιήσουμε.¹¹



2.3.2.1 Γενικό διάγραμμα ροής

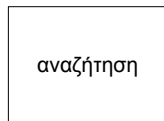
Ξεκινούμε με το γενικό διάγραμμα ροής της μεθοδολογίας μας.



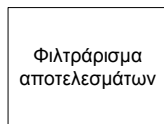


πληθαίνονται τα δεδομένα στις βάσεις

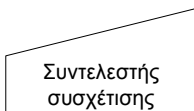
Με την εκκίνηση της μεθοδολογίας μας ο χρήστης θα καλεστεί να κρίνει εάν τα δεδομένα στις βάσεις είναι επαρκή. Εάν δεν είναι τότε γίνεται η εισαγωγή επιπρόσθετων δεδομένων. Οι παράμετροι ελέγχου των βάσεων, μετά από κάποια χρήση της εφαρμογής είναι τα προηγούμενα αποτελέσματα και η ποιότητά των, για ένα ζευγάρι πρωτεϊνών. Για κάθε νέο ζευγάρι πρωτεϊνών ή στην αρχή χρήσης της εφαρμογής πρέπει να



Εαν τα δεδομένα στις βάσεις είναι επαρκή, τότε μπορούμε να προχωρήσουμε στην αναζήτηση. Η αναζήτηση γίνεται με διάφορους τρόπους και εμπεριέχει παραμέτρους στην μεθοδολογία της. Θα γίνει εκτενή ανάλυση της αναζήτησης παρακάτω.



Τα αποτελέσματα επεξεργάζονται από το σύστημα μέσω των φίλτρων και εμφανίζονται στην οθόνη του χρήστη. Ο χρήστης μετά θα αποφανθεί για την εγκυρότητα και πληρότητα των αποτελεσμάτων. Ελέγχει δηλαδή αν είναι θετική, αρνητική ή καμία συσχέτιση μεταξύ των λέξεων-κλειδιών. Εάν είναι ανεπαρκή θα πρέπει ή/και να προσθέσει νέα σχετικά δεδομένα στο σύστημα είτε να αλλάξει τις παραμέτρους αναζήτησης.



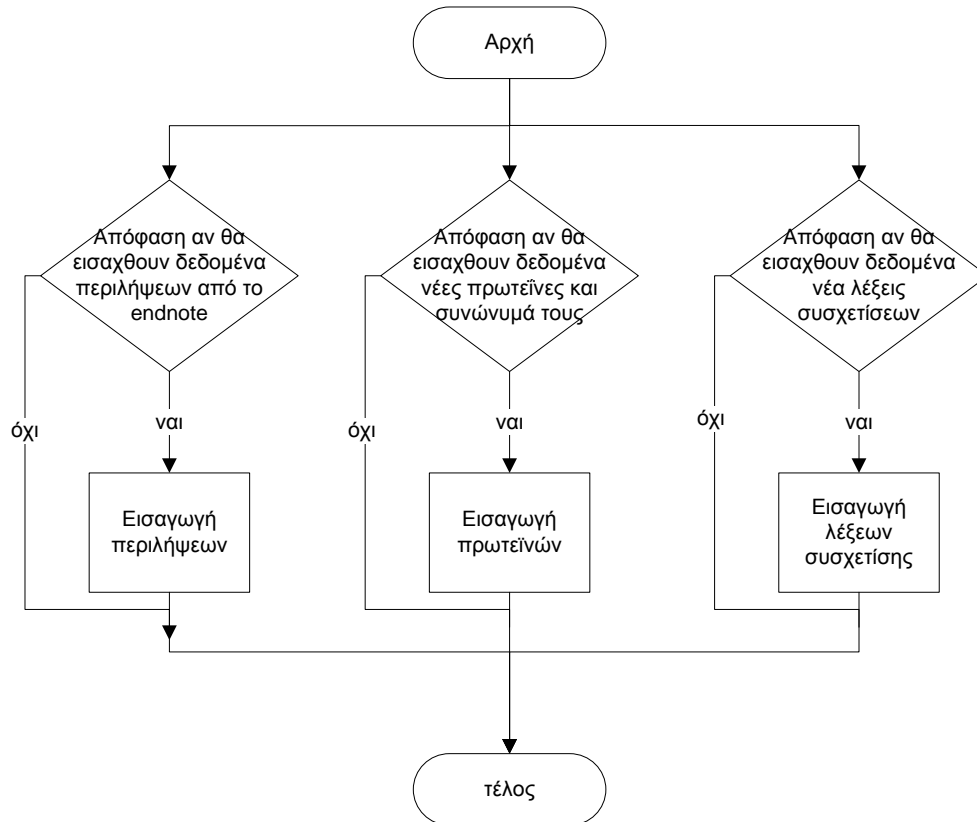
Από την στιγμή που ο χρήστης αποφάινεται για την πληρότητα, εγκυρότητα και αξιοπιστία των αποτελεσμάτων, αποδίδει τον συντελεστή συσχέτισης του ζεύγους πρωτεϊνών της αναζήτησης.



Τελικά οδηγούμαστε στο τομέα των αναφορών όπου μπορούμε να συγκρίνουμε διαφορετικούς συντελεστές συσχέτισης διαφορετικών ή ιδίων ζευγών πρωτεϊνών.

Εισαγωγή
δεδομένων
στις βάσεις

2.3.2.2 Διάγραμμα ροής εισαγωγής στοιχείων στις βάσεις δεδομένων



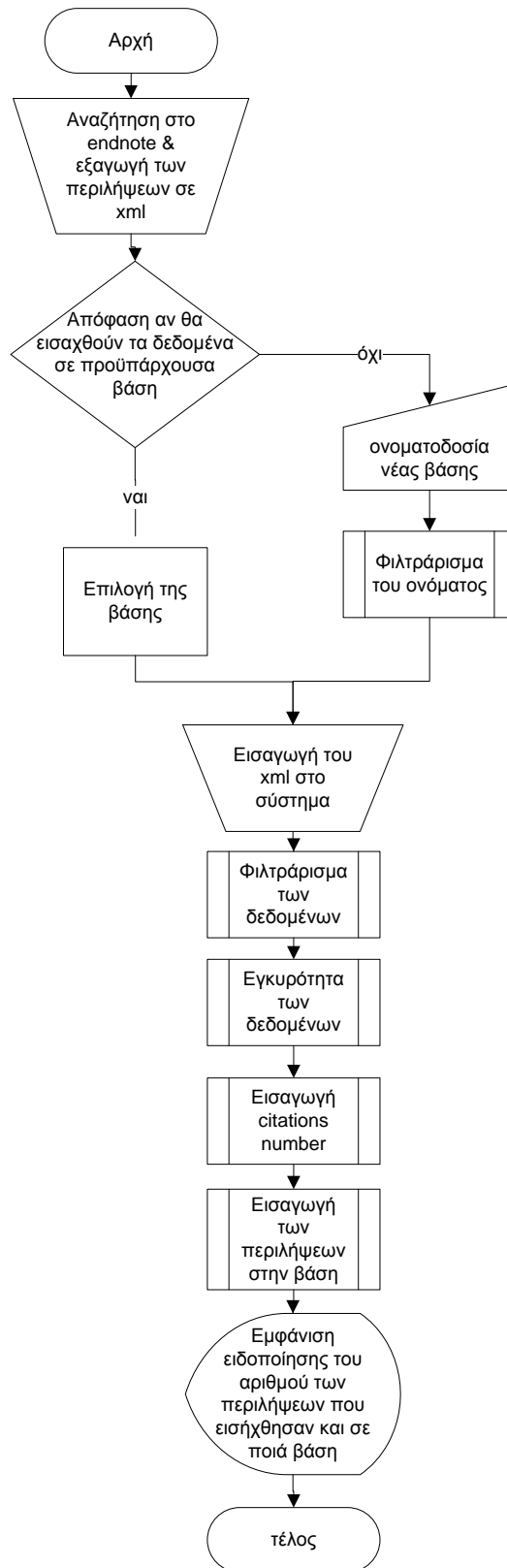
Το παραπάνω διάγραμμα αφορά την εισαγωγή νέων δεδομένων στις βάσεις από τον χρήστη. Η σχέση του με το συνολικό διάγραμμα φαίνεται παραπάνω.

Η εισαγωγή δεδομένων αφορά 3 διαφορετικά είδη δεδομένων στις βάσεις: περιλήψεις από το endnote, πρωτεΐνες και συνώνυμά τους και λέξεις συσχετίσεων.

Αυτές οι τρεις διαδικασίες είναι ανεξάρτητες η μία από την άλλη και μπορούν να γίνουν σε οποιοδήποτε χρόνο. Ξεχωριστή περίπτωση αποτελεί η έναρξη νέου ζεύγους πρωτεϊνών όπου θα πρέπει να προστεθούν περιλήψεις αλλά και συνώνυμα για το νέο ζευγάρι λέξεων-κλειδιών.

Εισαγωγή
περιλήψεων

Διάγραμμα ροής βάσης δεδομένων του περιλήψεων (Endnote database).



Αναζήτηση στο
endnote &
εξαγωγή των
περιλήψεων σε
xml

Το διάγραμμα ξεκινάει με την χειρονακτική εργασία της αναζήτησης περιλήψεων στο endnote. Αναζητούμε άρθρα τα οποία θα έχουν στην περίληψή τους και τις δύο λέξεις-κλειδιά πρωτεΐνες. Άρα με αυτό τον τρόπο υποδηλώνεται ότι οι δύο πρωτεΐνες θα έχουν κάποια συσχέτιση, όπως αναφέραμε στην εισαγωγή και στην βασική δομή της μεθοδολογίας μας. Αποσκοπούμε να εξαγάγουμε ένα μεγάλο αριθμό περιλήψεων (>500). Το σύνολο των περιλήψεων εξαγάγονται σε μορφή xml.

Απόφαση αν θα
εισαχθούν τα δεδομένα
σε προϋπάρχουσα
βάση

Μετά θα επιλέξουμε αν θα εισάγουμε τις περιλήψεις του xml σε μια προϋπάρχουσα βάση ή σε μια καινούργια. Στην δεύτερη περίπτωση θα πρέπει να πληκτρολογήσουμε το όνομα της βάσης και αυτό να φιλτραριστεί κατάλληλα, ώστε να μην υπάρχουν μη αποδεκτοί χαρακτήρες και να είναι μοναδική η ονομασία της βάσης.

Φιλτράρισμα
του ονόματος

Το όνομα της βάσης φιλτράρεται ώστε να απομακρύνονται οι μη αποδεκτοί χαρακτήρες και να ελέγχεται η μοναδικότητα του ονόματος.

Εισαγωγή του
xml στο
σύστημα

Αφού έχουμε επιλέξει την βάση, μπορούμε να εισαγάγουμε το αρχείο xml με τις περιλήψεις. Το σύστημα θα δημιουργήσει στο σύστημα αρχείων (filesystem) έναν φάκελο όπου θα εισαχθεί το xml. Η ονομασία του φακέλου αποτελείται από το username του χρήστη και την ημερομηνία που γίνεται η εισαγωγή.

Φιλτράρισμα
των
δεδομένων

Τα δεδομένα από το αρχείο xml επεξεργάζονται μέσω φίλτρων ώστε να απομακρύνονται οι μη αποδεκτοί χαρακτήρες, σχόλια και άλλες άχρηστες πληροφορίες από το xml.

Εγκυρότητα
των
δεδομένων

Η κάθε περίληψη με τα πλήρη στοιχεία και ιδιότητες της, ελέγχονται ως προς την πληρότητά και την μοναδικότητά της. Περιλήψεις άρθρων με ελλιπή στοιχεία δεν θα μπορέσουν να εισαχθούν στην βάση.

Εισαγωγή citations number

Για κάθε περίληψη, καλείται ένα web crawl που αναζητεί στο Google Scholar τον αριθμό των citations, δηλαδή των αριθμό που το κάθε άρθρο έχει αναφερθεί ως βιβλιογραφία. Αυτό μας δείχνει την εγκυρότητα του κάθε άρθρου.

Εισαγωγή των περιλήψεων στην βάση
--

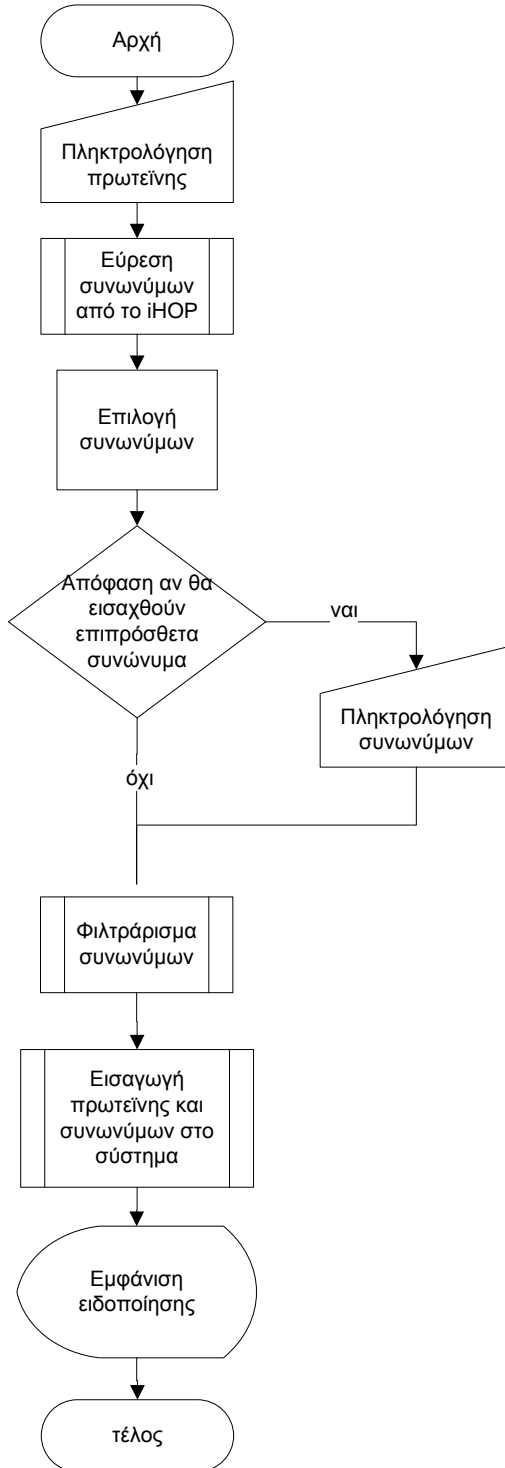
Μετά εισάγονται στην βάση δεδεμένων των περιλήψεων, οι περιλήψεις των άρθρων, με όλα τους τα προαναφερόμενα δεδομένα, για μελλοντικές αναζητήσεις.

Εμφάνιση ειδοποίησης του αριθμού των περιλήψεων που εισήχθησαν και σε ποιά βάση
--

Παρουσιάζεται στον χρήστη, η επιτυχία ή αποτυχία της διαδικασίας και ο αριθμών των περιλήψεων που έχουν εισαχθεί στην βάση δεδομένων.

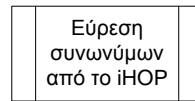
Εισαγωγή πρωτεϊνών

Διάγραμμα ροής βάσης δεδομένων πρωτεϊνών-συνωνύμων (Keywords database)

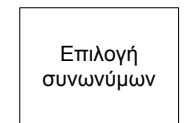




Το διάγραμμα εισαγωγής δεδομένων στην βάση πρωτεϊνών-συνωνύμων ξεκινά με την πληκτρολόγηση από τον χρήστη της πρωτεΐνης που τον ενδιαφέρει.



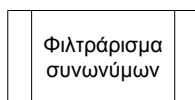
Το σύστημα καλεί ένα web crawl που αναζητεί την συγκεκριμένη πρωτεΐνη αναγνωρισμένα συνώνυμα από το iHOP. Αναλυτικότερα, τα αποτελέσματα του web crawl που φαίνονται στον χρήστη προέρχονται από τις στήλες συμβόλου, ονόματος, συνωνύμου και οργανισμού κάθε συνωνύμου στο iHOP.



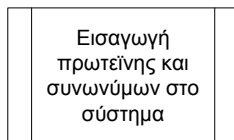
Ο χρήστης επιλέγει τα συνώνυμα που τον ενδιαφέρουν κλικάροντας τα. Συνώνυμα που ήδη υπάρχουν για την συγκεκριμένη πρωτεΐνη, θα αναφέρονται ξεχωριστά.



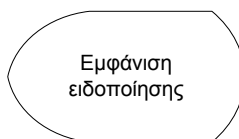
Μετά το πέρας της παραπάνω διαδικασίας ο χρήστης καλείται να προσθέσει επιπλέον συνώνυμα. Αυτά τα πληκτρολογεί σε συγκεκριμένο πεδίο.



Το σύστημα επεξεργάζεται όλα τα συνώνυμα με φίλτρα μοναδικότητας και απομάκρυνσης των μη αποδεκτών χαρακτήρων.



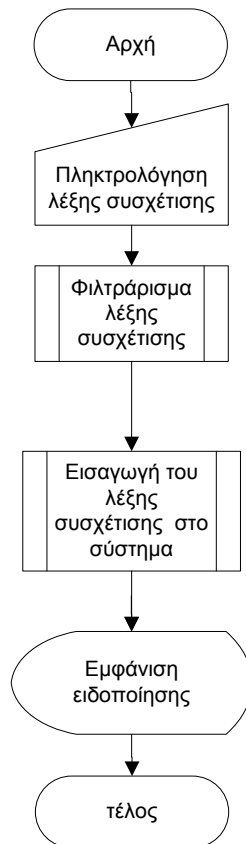
Μετά το πέρας της παραπάνω διαδικασίας, το σύστημα εισάγει όλα συνώνυμα για την προεπιλεγθήσα-πληκτρολογημένη πρωτεΐνη στην βάση δεδομένων.



Η διαδικασία τελειώνει με την ένδειξη αναφοράς στην οθόνη του χρήστη.

Εισαγωγή
λέξεων
συσχετίσης

Διάγραμμα ροής βάσης δεδομένων λέξεων συσχετίσεων (relationship database)



Πληκτρολόγηση
λέξης συσχέτισης

Το διάγραμμα εισαγωγής δεδομένων στην βάση λέξεων συσχετίσεων ξεκινά με την πληκτρολόγηση από τον χρήστη της λέξης συσχέτισης που τον ενδιαφέρει.

Φιλτράρισμα
λέξης
συσχέτισης

Το σύστημα επεξεργάζεται όλα την λέξη με φίλτρα μοναδικότητας και απομάκρυνσης των μη αποδεκτών χαρακτήρων.

Εισαγωγή του
λέξης συσχέτισης
στο σύστημα

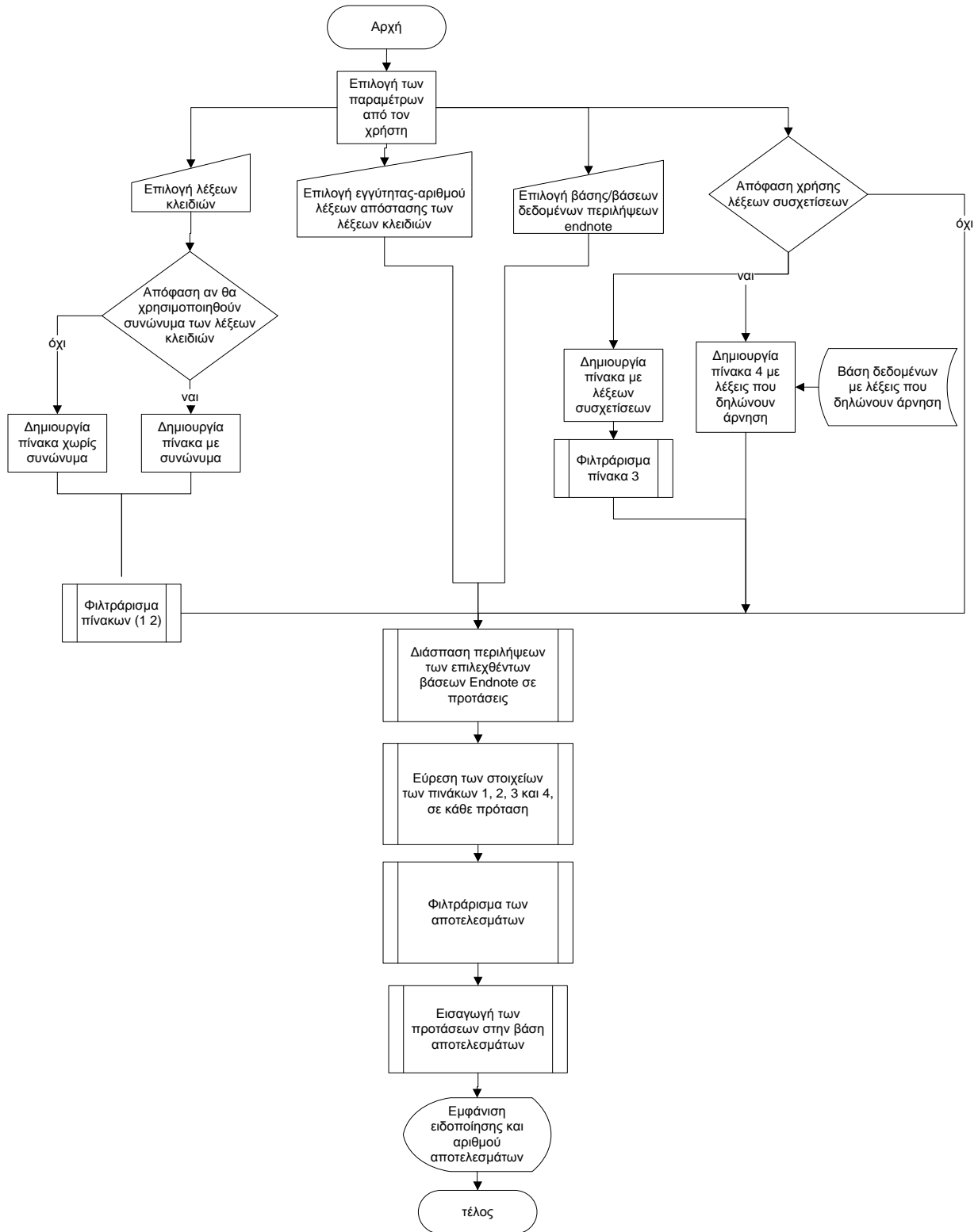
Η λέξη εισάγεται στην βάση δεδομένων λέξεων συσχετίσεων.

Εμφάνιση
ειδοποίησης

Η διαδικασία τελειώνει με την ένδειξη αναφοράς στην οθόνη του χρήστη.

αναζήτηση

2.3.2.3 Διάγραμμα ροής αναζήτησης (correlations search)



Επιλογή των
παραμέτρων
από τον
χρήστη

Η αναζήτηση ξεκινά με την επιλογή των παραμέτρων από τον χρήστη.

Επιλογή λέξεων
κλειδιών

Επιλέγει τις δύο λέξεις κλειδιά – πρωτεΐνες και αν θα χρησιμοποιήσει συνώνυμα για καθεμία από αυτές. Αυτές προέρχονται από την βάση δεδομένων των συνονύμων που προαναφέραμε.

Φιλτράρισμα
πίνακων (1 2)

Οι πρωτεΐνες (και τα συνώνυμά τους, εάν έχει αποφασίσει ο χρήστης να τα συμπεριλάβει) εισάγονται σε δύο πίνακες που δημιουργούνται. Το σύστημα επεξεργάζεται τους πίνακες με φίλτρα διαμόρφωσης των στοιχείων τους, όπου σε κάθε στοιχείο η χαρακτήρων, τοποθετείται μια παύλα σε κάθε σημείο ενδιάμεσα των τελευταίων τριών χαρακτήρων. Άρα για κάθε στοιχείο, ο κάθε πίνακας πληθαίνει κατά επιπρόσθετα δύο στοιχεία (εφόσον το αρχικό στοιχείο έχει περισσότερα από τρεις χαρακτήρες). Αυτό το κάνουμε καθώς πολλοί συγγραφείς αναφέρονται στις πρωτεΐνες που χρησιμοποιούν με παύλες. Πχ την πρωτεΐνη nfkB μπορεί να την δούμε γραμμένη nf-kb.

Επιλογή εγγύτητας-αριθμού
λέξεων απόστασης των
λέξεων κλειδιών

Ο χρήστης θα επιλέξει τον μέγιστο αριθμό λέξεων που θα απέχουν τα στοιχεία των πίνακων 1 και 2 (πρωτεΐνες και συνώνυμα)

Επιλογή βάσης/βάσεων
δεδομένων περιλήψεων
endnote

Ο χρήστης θα επιλέξει τις βάσεις δεδομένων των περιλήψεων που έχει ήδη δημιουργήσει, για να χρησιμοποιηθούν για την εξόρυξη δεδομένων.

Απόφαση χρήσης
λέξεων συσχέτισης

Ο χρήστης θα επιλέξει εάν θα χρησιμοποιήσει λέξεις συσχέτισης. Σε περίπτωση που απαντήσει καταφατικά, τότε θα δημιουργηθούν δύο επιπρόσθετοι πίνακες on-the-fly.

Φιλτράρισμα
πίνακα 3

Ο πίνακας 3 θα εμπλουτιστεί με τις λέξεις συσχέτισης από την βάση δεδομένων λέξεων συσχέτισης. Αυτές οι λέξεις όμως θα έχουν υποστεί την διαδικασία stemming. Δηλαδή θα έχει αφαιρεθεί το τελευταίο γράμμα τους ώστε οι λέξεις να έχουν επανέλθει στην ρίζα τους. Πχ η λέξη συσχέτισης induce θα γίνει induc και με αυτή την λέξη θα πληθύνει ο πίνακας 2. Το stemming το κάνουμε γιατί δεν ξέρουμε την

μορφή που θα βρούμε την κάθε λέξη συσχέτισης. Μπορεί να βρεθεί με την μορφή induced, inducing, inducer κτλ

Δημιουργία πίνακα 4 με λέξεις που δηλώνουν άρνηση

Ο πίνακας 4 θα εμπλουτιστεί με τις λέξεις που δηλώνουν άρνηση από την βάση δεδομένων λέξεων άρνησης. Αυτή η βάση δεδομένων είναι φανερή μόνο στον διαχειριστή του συστήματος και αποτελείται από τις λέξεις not, barely, nor.

Διάσπαση περιλήψεων των επιλεχθέντων βάσεων Endnote σε προτάσεις

Αφού έχουν συλλεχθεί όλοι οι προηγούμενοι παράμετροι, το σύστημα επιλέγει όλα τα άρθρα που ανήκουν στις προεπιλεγμένες βάσεις δεδομένων περιλήψεων και τις διασπάει σε προτάσεις. Η διάσπαση σε κάθε πρόταση γίνεται με την ανεύρεση τελείας-κενού, γιατί την αποφυγή εμφάνισης λανθασμένων θετικών αποτελεσμάτων.

Εύρεση των στοιχείων των πινάκων 1, 2, 3 και 4, σε κάθε πρόταση

Σε κάθε λέξη κάθε πρότασης, το σύστημα εντοπίζει τα στοιχεία των πινάκων 1, 2, 3 και 4. Τα στοιχεία των πινάκων 1,2 και 3 εντοπίζονται εσωτερικά των λέξεων κάθε πρότασης ή/και ταυτίζονται με αυτές, ενώ τα στοιχεία του πίνακα 4 εντοπίζονται μόνο όταν ταυτίζονται με τις λέξεις. Τα στοιχεία των πινάκων 3 και 4 εμφανίζονται μόνο αφού ο χρήστης έχει επιλέξει να χρησιμοποιήσει λέξεις συσχέτισης στην αναζήτηση του. Για την πλήρη ανάλυση του διαγράμματος ροής θα θεωρήσουμε ότι έχει επιλέξει την χρήση τους.

Φιλτράρισμα των αποτελεσμάτων

Το σύστημα επεξεργάζεται τις προτάσεις που εμπεριέχουν τα στοιχεία των πινάκων 1, 2, 3 και 4. Πρώτο φίλτρο είναι η μέγιστη απόσταση των συνδυασμών ευρεθέντων στοιχείων των πινάκων 1 και 2. Εδώ δεν μας απασχολεί ποιο είναι πρώτο και ποιο δεύτερο, στα αποτελέσματα συμπεριλαμβάνεται η ενεργητική και η παθητική φωνή της εύρεσης συσχέτισης ανάμεσα στα στοιχεία των πινάκων 1 και 2. Δεύτερο φίλτρο είναι η εύρεση στοιχείων του πίνακα 3 ανάμεσα στα στοιχεία των πινάκων 1 και 2.

Εισαγωγή των προτάσεων στην βάση αποτελεσμάτων

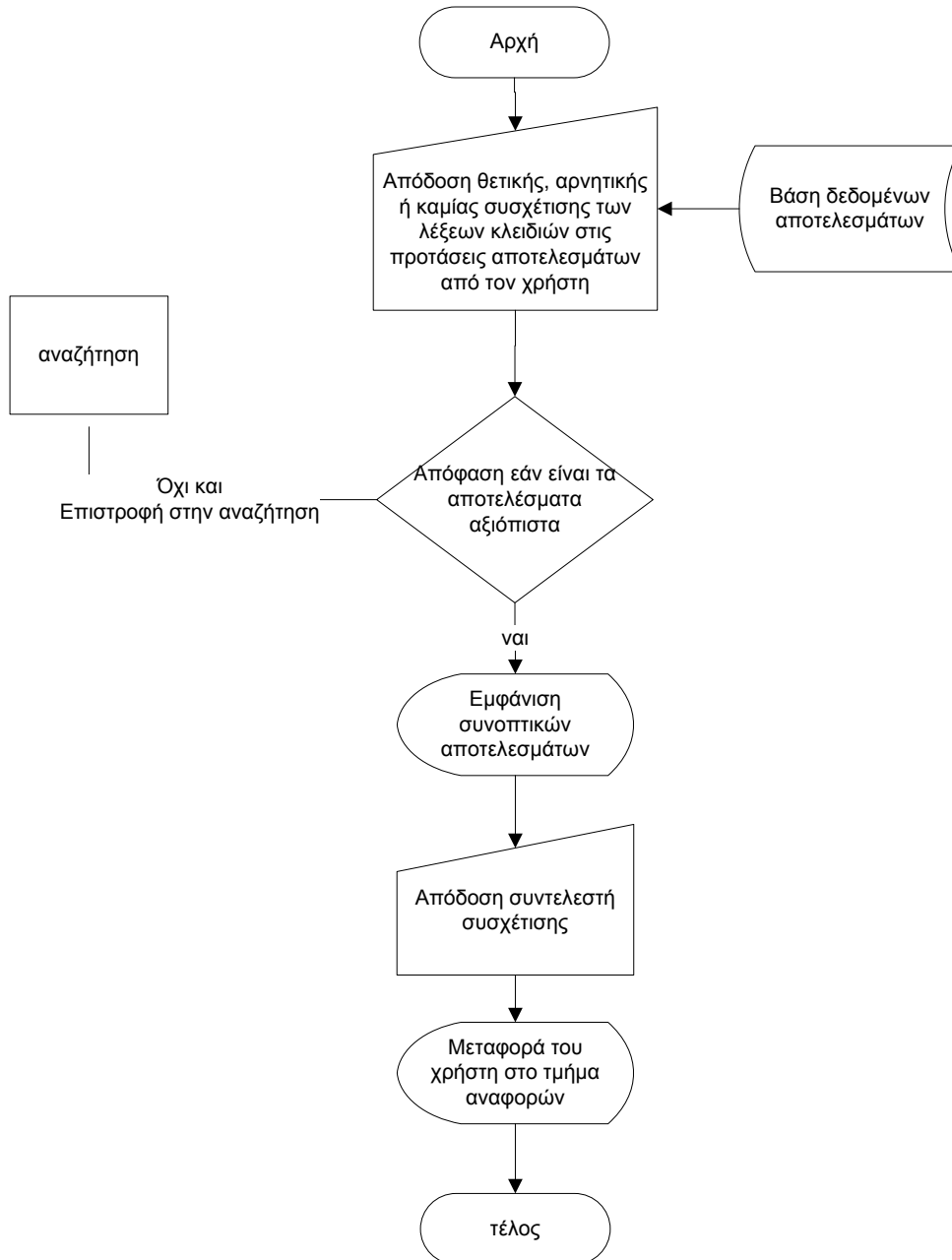
Αυτές οι προτάσεις που πληρούν τα παραπάνω κριτήρια θα εισαχθούν στην βάση δεδομένων αποτελεσμάτων.

Εμφάνιση ειδοποίησης και αριθμού αποτελεσμάτων

Η διαδικασία τελειώνει με την εμφάνιση ειδοποίησης στον χρήστη του αριθμού των αποτελεσμάτων που ευρέθηκαν.

Φιλτράρισμα αποτελεσμάτων

2.3.2.4 Διάγραμμα ροής επεξεργασίας προτάσεων από τον χρήστη (user selection)



Απόδοση θετικής, αρνητικής ή καμίας συσχέτισης των λέξεων κλειδιών στις προτάσεις αποτελεσμάτων από τον χρήστη

Η διαδικασία ξεκινά με την εμφάνιση στον χρήστη όλων των προτάσεων που εντόπισε το σύστημα στην προηγούμενη διαδικασία. Εκείνος αποφαινεται για την θετική, αρνητική ή καμία συσχέτιση που εμφανίζουν οι προτάσεις-αποτελέσματα.

Απόφαση εάν είναι τα αποτελέσματα αξιόπιστα

Από την στιγμή που έχει τελειώσει την διαδικασία επιλογής, ο χρήστης αποφαινεται για το αν τα αποτελέσματά που επεξεργάζεται είναι αξιόπιστα και επαρκή. Σε περίπτωση που δεν είναι επιστρέφει να αλλάξει τις παραμέτρους αναζήτησης ή/και προσθέτει νέα δεδομένα στις βάσεις (πρώτο διάγραμμα ροής)

Εμφάνιση συνοπτικών αποτελεσμάτων

Εάν είναι αξιόπιστα και επαρκή τα αποτελέσματα τότε εμφανίζονται συνοπτικά πόσες προτάσεις βρέθηκαν με θετική/αρνητική συσχέτιση και παρατίθεται οι αθροιστικές βιβλιογραφικές αναφορές των.

Απόδοση συντελεστή συσχέτισης

Ο χρήστης αποφαινεται για τον συντελεστή συσχέτισης (weight) με βάση την προηγούμενη εμφάνιση συνοπτικών αποτελεσμάτων. Ο συντελεστής συσχέτισης μπορεί να είναι από -1 για πλήρη αρνητική συσχέτιση μέχρι +1 για πλήρη θετική συσχέτιση.

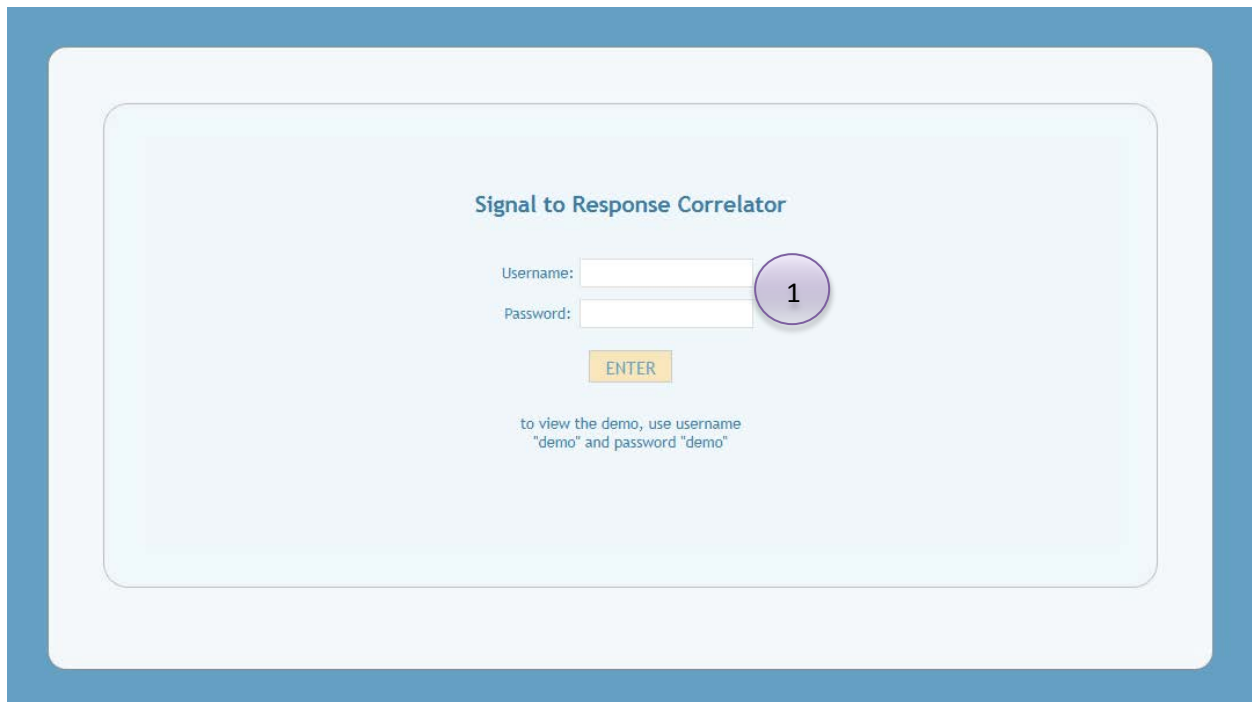
Μεταφορά του χρήστη στο τμήμα αναφορών

Ο χρήστης μεταφέρεται στο τμήμα αναφορών όπου παρουσιάζονται όλοι οι προηγούμενοι συντελεστές συσχέτισης.

2.3.3 Επεξήγηση του GUI

Μετά από την μετουσίωση των διαγραμμάτων ροής σε προγραμμαστική αρχιτεκτονική και ανάπτυξη, καταλήξαμε στον προγραμματισμό της εφαρμογή μας. Παρακάτω αναρτούνται και μετά επεξηγούνται τα διάφορα screenshots από την λειτουργία της εφαρμογής.

2.3.3.1 Είσοδος στην εφαρμογή



Signal to Response Correlator

Username:

Password:

ENTER

to view the demo, use username: "demo" and password "demo"

- 1 Ο χρήστης εισάγει τα στοιχεία του (ψευδώνυμο χρήστη και συνθηματικό) και πατάει το Enter. Για την demo χρήση του συστήματος εισάγει ως ψευδώνυμο χρήστη *demo* και συνθηματικό *demo*.


2.3.3.2 Μενού αναζήτησης

The screenshot shows the 'Signal to Response Correlator' web application. At the top, the title 'Signal to Response Correlator' is followed by a subtitle 'Search your endnote abstracts and define correlation among proteins of interest' and a user status 'You are logged in as "demo user" Logout'. Below this are two tabs: 'Correlations' and 'Control Panel'. A large blue bar contains a '3' callout. The main content area is titled 'Define correlations' with a '4' callout. It features a 'How to get started?' link, a 'Use advanced search' button, and two input fields for 'Signal keyword' and 'Response keyword', each with an 'add new' link. A '5' callout points to the 'Endnote databases' section, which includes a list of databases: 'All databases', 'p38-nfkb (499)', and 'test (199)'. A 'Search' button is located below the list. At the bottom left, a '6' callout points to a 'Logs' section displaying a list of system events. At the bottom right, there are 'Setup' and 'Help' buttons, with a '7' callout pointing to the 'Help' button.

Με το που εισάγεται ο χρήστης στην Γραφική Διεπιφάνεια Χρήσης (GUI) της Εφαρμογής, βλέπει την δομή της, που έχει ως παρακάτω:

- 1 Τίτλος και περιγραφή συστήματος. Επίσης φαίνεται το όνομα του χρήστη και το logout.
- 2 Οι δύο καρτέλες χρήσης του συστήματος: αναζήτηση συσχετίσεων (correlations) και ο πίνακας ελέγχου (control panel)
- 3 Υπομενού για την κάθε καρτέλα (φαίνεται μόνο στον πίνακα ελέγχου)
- 4 Τίτλος τρέχουσας διαδικασίας
- 5 Επιλογές διαδικασίας

- 6 Ιστορικό με τις τελευταίες κινήσεις του χρήστη (logs)
- 7 Μενού αλλαγών προτιμήσεων χρήστη και βοήθεια με διαδραστικά εγχειρίδια χρήσης

Αν επικεντρωθούμε στο 5, θα δούμε ότι υπάρχει ένα σύμβολο . Αυτό μας δείχνει μικρές κοινοποιήσεις βοήθειας για το πώς μπορεί να γίνει η εκκίνηση της διαδικασίας. Ενεργοποιείται με απλή μετακίνηση του ποντικιού πάνω από το σύμβολο. Τέτοια σύμβολα υπάρχουν σε όλες τις διαδικασίες, διαμέσου του GUI.

Στην απλή μορφή (simple search) ο χρήστης μπορεί να επιλέγει τις δύο πρωτεΐνες και τις βάσεις δεδομένων περιλήψεων που θέλει το σύστημα να επεξεργαστεί.

Signal to Response Correlator
Search your endnote abstracts and define correlation among proteins of interest
You are logged in as "demo user" [Logout](#)

[Correlations](#) [Control Panel](#)

Define correlations

"How to get started"

[Use simple search](#) **1**

Signal keyword [add new](#)

 Enable synonyms: Yes No

Response keyword [add new](#)

 Enable synonyms? Yes No

Endnote databases: [add new](#)

All databases
 p38-nfkb (499)
 test (199)

Keyword proximity

Enable relationship words?
 Yes No

[Search](#)

Logs:
 Fri 29 Jul 2011, 05:56:21 - demo user logged in
 Thu 28 Jul 2011, 17:29:34 - demo user added 499 articles in database p38-nfkb
 Thu 28 Jul 2011, 17:24:23 - demo user logged in
 Thu 28 Jul 2011, 16:50:36 - demo user deleted synonym nfkbib for protein nfkb

[Setup](#) [Help](#)

1 Εάν ο χρήστης πατήσει το κουμπί για προχωρημένη αναζήτηση (advanced search), τότε θα έχει τις προστιθέμενες επιλογές για να επιλέξει την χρήση συνωνύμων για κάθε πρωτεΐνη, τον αριθμό λέξεων απόστασης των λέξεων κλειδιών (εγγύτητα) και την χρήση λέξεων συσχετίσεων (μαζί με την χρήση λέξεων που δηλώνουν άρνηση).

Το σύστημα παρέχει default-προεπιλεγμένη τιμή για τον αριθμό λέξεων απόστασης των λέξεων-κλειδιών στοιχείων πινάκων 1 & 2 (εγγύτητα). Όπως προαναφέραμε αυτή η τιμή είναι 4, αλλά μπορεί ο χρήστης να την αλλάξει.

Signal to Response Correlator
Search your endnote abstracts and define correlation among proteins of interest
You are logged in as "demo user" [Logout](#)

Correlations Control Panel

Define correlations

"How to get started" ⓘ

Use simple search

Signal keyword [add new](#)
p38
Enable synonyms: Yes No

Response keyword [add new](#)
nkfb
Enable synonyms? Yes No

Endnote databases: [add new](#)

- All databases
- p38 nkfb (499)
- test (199)

Keyword proximity
8

Enable relationship words?
yes No

[Search](#)

Logs:
Mon 01 Aug 2011, 16:29:05 - demo user logged in
Mon 01 Aug 2011, 16:26:51 - demo user logged in
Fri 29 Jul 2011, 06:23:08 - demo user assigned weight for job demou110729-001
Fri 29 Jul 2011, 06:23:05 - demo user finished positive/negative correlation for the results of job demou110729-001

[Login](#) [Help](#)

Εδώ βλέπουμε ένα screenshot αφού έχουμε προσθέσει τις πληροφορίες που θέλουμε για να γίνει η αναζήτηση. Πατούμε το κουμπί search.

Signal to Response Correlator
Search your endnote abstracts and define correlation among proteins of interest
You are logged in as "demo user" Logout

Signal keyword: p38
Response keyword: nfkb
Job: demou110801-001
Db: p38-nfkb,test

Correlations Control Panel

Define correlations

There are 4 results in 698 abstracts that matched the criteria

"How to get started"

Use simple search

Signal keyword: add new
p38
Enable synonyms? Yes No

Response keyword: add new
nfkb
Enable synonyms? Yes No

Endnote databases: add new
 All databases
 p38+nfkb (499)
 test (199)

Keyword proximity
8

Enable relationship words?
Yes No

Search again

Continue to view the results

Logs:
Mon 01 Aug 2011, 16:38:02 - demo user logged in
Mon 01 Aug 2011, 16:29:05 - demo user logged in
Mon 01 Aug 2011, 16:26:51 - demo user logged in
Fri 29 Jul 2011, 06:23:08 - demo user assigned weight for job demou110729-001

Setup Help

Αφού έχει γίνει η αναζήτηση του συστήματος μας δείχνει:

- 1 Περιοχή όπου φαίνονται οι βασικές πληροφορίες για την αναζητήτή μας και το όνομα της τρέχουσας/παρούσας εργασίας, όπως προκύπτει από το σύστημα με βάση το όνομα του χρήστη, την ημερομηνία και τον αύξοντα αριθμό της εργασίας.
- 2 Ειδοποίηση για το αποτέλεσμα της διαδικασίας του συστήματος
- 3 Κουμπί εμφάνισης των αποτελεσμάτων «Continue to view the results». Ο χρήστης μπορεί επίσης να αλλάξει τις προτιμήσεις αναζήτησης και να πατήσει το κουμπί «search again».

Signal to Response Correlator
 Search your endnote abstracts and define correlation among proteins of interest
 You are logged in as "demo user" [Logout](#)

Signal keyword: p38
 Response keyword: nfkb
 Job: demou110801-001
 Db: p38-nfkb,test

[Correlations](#) [Control Panel](#)




Finalizing selections or you can go [back to the settings page](#)

Sentence	Journal	Year	Cited			
The aim of this study was to investigate p38-mitogene-activated protein kinase (p38-MAPK), nuclear factor kappa B (p65-NF-κB) and inducible nitric oxide synthase (iNOS) expression in an experimental model of varicocele in the rat testis.	Andrologia	2011	0	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
Malondialdehyde (MDA), nitric oxide (NO) and reduced glutathione (GSH) levels were biochemically assessed, and the p38-MAPK and NF-κB activity, and iNOS expression were immunohistochemically studied in the right and left testicles of rats from each group.	Andrologia	2011	0	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
There was a marked staining for iNOS, p38-MAPK and p65-NF-κB expression in rats of varicocele group compared with the sham group.	Andrologia	2011	0	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>
p38-MAPK and p65-NF-κB activation, and iNOS expression have a significant role in varicocele-induced testicular dysfunction.	Andrologia	2011	0	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

Total sentences left: 4

Logs:
 Mon 01 Aug 2011, 16:26:51 - demo user logged in
 Fri 29 Jul 2011, 06:23:08 - demo user assigned weight for job demou110729-001
 Fri 29 Jul 2011, 06:23:09 - demo user finished positive/negative correlation for the results of job demou110729-001
 Fri 29 Jul 2011, 05:56:21 - demo user logged in

[Setup](#) [Help](#)

Αφού έχουμε πατήσει το κουμπί εμφάνισης των αποτελεσμάτων «Continue to view the results», βλέποντάς τα, ο χρήστης επιλέγει την θετική , αρνητική  ή καμία  συσχέτιση για κάθε μια πρόταση των αποτελεσμάτων. Επίσης βλέπουμε ότι οι πρωτεΐνες (στοιχείων πινάκων 1 και 2) παρουσιάζονται με highlight χρώματα, για ευκολότερη ανεύρεση.

Signal to Response Correlator
Search your endnote abstracts and define correlation among proteins of interest
You are logged in as "demo user" [Logout](#)

Signal keyword: p38
Response keyword: nfb
Job: demou110801-001
Db: p38-nfb,test

Correlations Control Panel

Finalizing selections

Results:
3 positive sentences with 0 article citations
1 negative sentences with 0 article citations

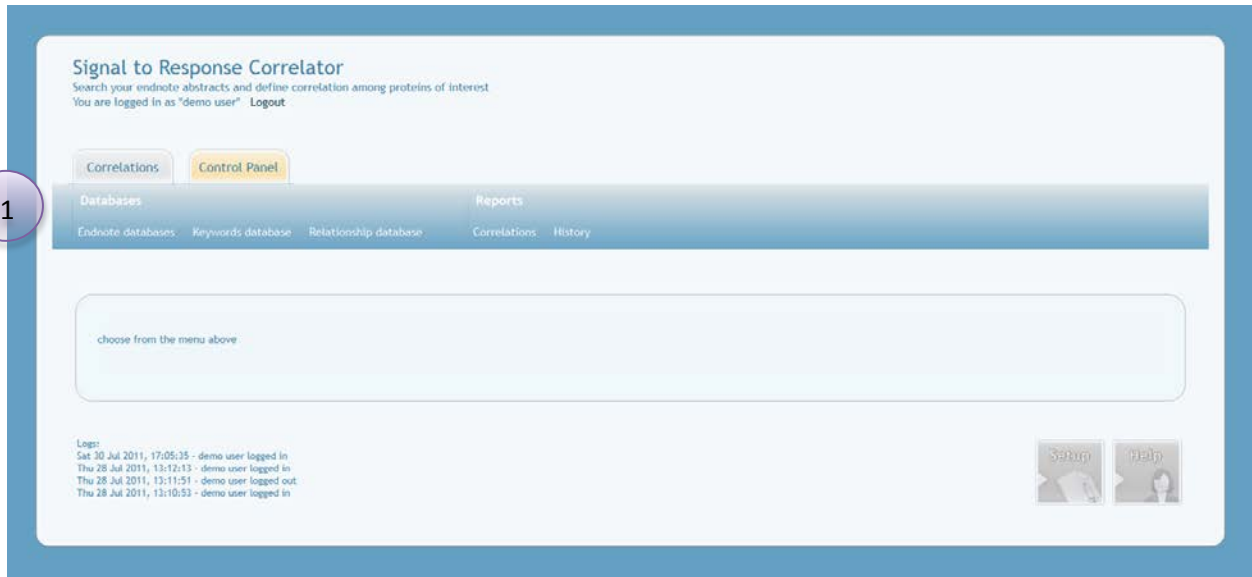
Assign weight: [Save](#)

Logs:
Mon 01 Aug 2011, 16:42:01 - demo user finished positive/negative correlation for the results of job demou110801-001
Mon 01 Aug 2011, 16:42:53 - demo user logged in
Mon 01 Aug 2011, 16:38:02 - demo user logged in
Mon 01 Aug 2011, 16:29:05 - demo user logged in

Help Logout

1 Στο τέλος της επιλογής των συσχετίσεων, ο χρήστης θα κληθεί να προσθέσει τον συντελεστή συσχέτισης (assigning weight) στο ζεύγος των δύο πρωτεϊνών. Πατώντας «save», θα μεταφερθεί στο μενού των αναφορών.

2.3.3.3 Μενού κέντρου ελέγχου



1 Εδώ βλέπουμε το κεντρικό μενού του πίνακα ελέγχου. Στην περιοχή του υπομενού βλέπουμε τις διάφορες επιλογές που θα αναλύσουμε παρακάτω.

Βάσεις δεδομένων περιλήψεων (Endnote databases)

Signal to Response Correlator
Search your endnote abstracts and define correlation among proteins of interest
You are logged in as "demo user" [Logout](#)

Correlations **Control Panel**

Databases **Reports**

Endnote databases Keywords database Relationship database **Correlations** History

Endnote databases of 698 total abstracts

1 **Search database(s)** All databases p38-nfkb (499) test (199)

Authors Title Year Journal Abstract

2 **Add abstracts**

Database create new OR add to existing

xml file

3 **Delete databases**

p38-nfkb (499) test (199)


Logs:
Fri 29 Jul 2011, 05:56:21 - demo user logged in
Thu 28 Jul 2011, 17:29:34 - demo user added 499 articles in database p38-nfkb
Thu 28 Jul 2011, 17:24:23 - demo user logged in
Thu 28 Jul 2011, 16:50:36 - demo user deleted synonym nfkbib for protein nfkb

Διακρίνονται 3 σημαντικές περιοχές.

1

Η περιοχή αναζήτησης στις προϋπάρχουσες βάσεις δεδομένων περιλήψεων. Μπορεί να είναι συνδυαστική αναζήτηση σε συγγραφείς, τίτλο, έτος, περιοδικό και κείμενο περίληψης κάθε περίληψης και να χρησιμοποιούνται τελεστές AND και OR σε κάθε πεδίο, πχ εισάγουμε για authors «Young AND Chen», σε πολλαπλές βάσεις δεδομένων περιλήψεων. Δίπλα από κάθε βάση περιλήψεων φαίνεται ο αριθμός περιλήψεων που περιέχει.

2 Γίνεται η προσθήκη xml αρχείων από το endpoint σε νέα ή προϋπάρχουσα βάση δεδομένων περιλήψεων.

3 Μπορεί ο χρήστης να διαγράψει βάση. Μόλις πατήσει το κουμπί , μια τελευταία ειδοποίηση αναδύεται ώστε ο χρήστης να ερωτηθεί αν είναι σίγουρος για την διαγραφή της βάσης. Μόλις διαγραφεί η βάση θα αφαιρεθούν τα στοιχεία της που χρησιμοποιήθηκαν στο μενού των αναφορών.

Signal to Response Correlator

Search your endnote abstracts and define correlation among proteins of interest
 You are logged in as "demo user" [Logout](#)

Correlations

Control Panel

Databases

Reports

Endnote databases

Keywords database

Relationship database

Correlations

History

Endnote databases of 698 total abstracts

Search database(s) All databases
 p38-nfkb (499) test (199)

Authors	Title	Year	Journal	Abstract
				cortical neural progenitor: <input type="text" value="search"/>

1

Author(s): Rochira, J. A., Matluk, N. N., Adams, T. L., Karaczyn, A. A., Oxburgh, L., Hess, S. T., Verdi, J. M.
 Title: A Small Peptide Modeled after the NRAGE Repeat Domain Inhibits XIAP-TAK1-TAK1 Signaling for NF-kappaB Activation and Apoptosis in P19 Cells
 Year: 2011
 Journal: PLoS One
 Abstract: In normal growth and development, apoptosis is necessary to shape the central nervous system and to eliminate excess neurons which are not required for innervation. In some diseases, however, apoptosis can be either overactive as in some neurodegenerative disorders or severely attenuated as in the spread of certain cancers. Bone morphogenetic proteins (BMPs) transmit signals for regulating cell growth, differentiation, and apoptosis. Responding to BMP receptors stimulated from BMP ligands, neurotrophin receptor-mediated IJAGE homolog (NRAGE) binds and functions with the XIAP-TAK1-TAK1 complex to activate p38(MAPK) and induces apoptosis in [cortical neural progenitors](#). NRAGE contains a unique repeat domain that is only found in human, mouse, and rat homologs that we theorize is pivotal in its BMP MAPK role. Previously, we showed that deletion of the repeat domain inhibits apoptosis, p38(MAPK) phosphorylation, and caspase-3 cleavage in P19 neural progenitor cells. We also showed that the XIAP-TAK1-TAK1 complex is dependent on NRAGE for IKK-alpha/beta phosphorylation and NF-kappaB activation. XIAP is a major inhibitor of caspases, the main executioners of apoptosis. Although it has been shown previously that NRAGE binds to the RING domain of XIAP, it has not been determined which NRAGE domain binds to XIAP. Here, we used fluorescence resonance energy transfer (FRET) to determine that there is a strong likelihood of a direct interaction between NRAGE and XIAP occurring at NRAGE's unique repeat domain which we also attribute to be the domain responsible for downstream signaling of NF-kappaB and activating IKK subunits. From these results, we designed a small peptide modeled after the NRAGE repeat domain which we have determined inhibits NF-kappaB activation and apoptosis in P19 cells. These intriguing results illustrate that the paradigm of the NRAGE repeat domain may hold promising therapeutic strategies in developing pharmaceutical solutions for combating harmful diseases involving excessive downstream BMP signaling, including apoptosis.
 Uri: <http://www.ncbi.nlm.nih.gov/pubmed/21789165>
 Cited: 0

2

Total abstracts found: 1

Add abstracts

Database create new OR add to existing

xml file

Delete databases

p38-nfkb (499) test (199)

Logs:
 Fri 29 Jul 2011, 05:56:21 - demo user logged in
 Thu 28 Jul 2011, 17:29:34 - demo user added 499 articles in database p38-nfkb
 Thu 28 Jul 2011, 17:24:23 - demo user logged in
 Thu 28 Jul 2011, 16:50:36 - demo user deleted synonym nfkbib for protein nfkb



- 1 Φαίνεται η παρουσίαση των αποτελεσμάτων αναζήτησης στην βάση «p38-nfkb» για τον όρο «cortical neural progenitors» στην περίληψη.
- 2 Ο χρήστης μπορεί να διαγράψει μια περίληψη, μόνο στην περίπτωση που δεν έχει χρησιμοποιηθεί από την εφαρμογή για την απόδοση συντελεστή συσχέτισης σε ζεύγος πρωτεϊνών.

Η διαδικασία της αναζήτησης μπορεί να ολοκληρωθεί με την επιλογή του μενού.

Βάση δεδομένων πρωτεϊνών-συνωνύμων (Keywords databases)

Signal to Response Correlator
Search your endnote abstracts and define correlation among proteins of interest
You are logged in as "demo user" [Logout](#)

Correlations **Control Panel**

Databases: Endnote databases **Keywords database** Relationship database Reports: Correlations History

Keywords & synonyms database

Keyword	Synonym	ID	Update	Delete
nfkB		72	Update	Delete
nfkB	nfkB1	73	Update	Delete
p38		62	Update	Delete
p38	mapk1	66	Update	Delete
p38	p38b	65	Update	Delete

1

Add keyword

Keyword

2

Logs:
 Fri 29 Jul 2011, 05:56:21 - demo user logged in
 Thu 28 Jul 2011, 17:29:34 - demo user added 499 articles in database p38-nfkB
 Thu 28 Jul 2011, 17:24:23 - demo user logged in
 Thu 28 Jul 2011, 16:50:36 - demo user deleted synonym nfkb1 for protein nfkb

- 1 Περιοχή όπου φαίνεται η λέξη κλειδί και με τα συνώνυμά της. Μπορεί να γίνει αναζήτηση, αλφαβητική διαλογή, αλλαγή ή διαγραφή στην λέξη κλειδί και συνώνυμο.
- 2 Περιοχή όπου γίνεται προσθήκη συνδυασμού λέξης-κλειδί και συνωνύμων. Στο πεδίο πληκτρολογεί ο χρήστης την πρωτεΐνη που τον ενδιαφέρει και το σύστημα βρίσκει τα συνώνυμά της μέσω του iHOP.

Signal to Response Correlator

Search your endnote abstracts and define correlation among proteins of interest
 You are logged in as "demo user" [Logout](#)

Correlations **Control Panel**

Databases Reports

Endnote databases **Keywords database** Relationship database Correlations History

select synonyms for "p38"

Symbol	Name (description)	Synonym	Organism
p38	p38 kinase		clona intestinalis
p38	coat protein		pna stem necrosis virus
p38	p38		drosophila melanogaster
cp	p38		carnation mottle virus
pap2_gp38	hypothetical protein		nsuwtmshoaat aeruainosa bacteriophage pap2
mapk1	mitogen-activated protein kinase 1	p38	homo sapiens
mapk14	mitogen-activated protein kinase 14	p38	homo sapiens
crk	v-crk sarcoma virus ct10 oncogene homolog [avian]	p38	homo sapiens
grap2	grb2-related adaptor protein 2	p38	homo sapiens
rnf19a	ring finger protein 19a	p38	homo sapiens
poldip2	polynemase (dna-directed), delta interacting protein 2	p38	homo sapiens
ahat1	ahat1, activator of heat shock 90kda protein atpase homolog 1 (yeast)	p38	homo sapiens
jtvl	jtvl gene	p38	homo sapiens
mapk14	mitogen-activated protein kinase 14	p38	mus musculus
syb	synaptophysin	p38	mus musculus
crk	v-crk sarcoma virus ct10 oncogene homolog [avian]	p38	mus musculus
grap2	grb2-related adaptor protein 2	p38	mus musculus
ahat1	ahat1, activator of heat shock protein atpase homolog 1 (yeast)	p38	mus musculus
nurf-2B	nucleosome remodeling factor - 28kd	p38	drosophila melanogaster
hrb67f	heterogeneous nuclear ribonucleoprotein at 87f	p38	drosophila melanogaster
fic	licorne	p38	drosophila melanogaster
mpk2	mpk2	p38	drosophila melanogaster
peb	protein ejaculatory bulb	p38	drosophila melanogaster
at3g15520	rom1-like protein crk trans. homologue 1b p38, chloroplast / thylakoid lumen ppia...		arabidopsis thaliana
at4g08780	peroxidase, putative	p38	arabidopsis thaliana
mapk14a	mitogen-activated protein kinase 14a	p38	danio rerio
crk	v-crk sarcoma virus ct10 oncogene	p38	gallus gallus
mapk14	mitogen-activated protein kinase 14	p38	xenopus tropicalis
mapk14	mitogen activated protein kinase 14	p38	rattus norvegicus
crk	v-crk sarcoma virus ct10 oncogene homolog [avian]	p38	rattus norvegicus
spcv2gp2	p38 protein		indian peacock clump virus
mpv1p32	p38 protein		mycoplasma virus p1
ap38-1_38	putative integrase	p38	acretthesinhan rhusii bacteriophage ap38-1

1

Add more synonyms for "p38" separated by a comma

2

Save

Logs:
 Fri 29 Jul 2011, 05:56:21 - demo user logged in
 Thu 28 Jul 2011, 17:29:34 - demo user added 499 articles in database p38-nfb
 Thu 28 Jul 2011, 17:24:23 - demo user logged in
 Thu 28 Jul 2011, 16:50:36 - demo user deleted synonym rnf6ib for protein rnf6



Παραπάνω η αναζήτηση συνωνύμων γίνεται για την πρωτεΐνη «p38».

- 1 Το σύστημα έχει βρει τα συνώνυμά της πρωτεΐνης μέσω του iHOP και ο χρήστης μπορεί να τα προσθέσει κλικάροντάς τα. Η διαγραφή τους γίνεται ξανακλικάροντάς τα.
- 2 Μπορεί ο χρήστης να προσθέσει επιπρόσθετα συνώνυμα τα οποία δεν έχουν ευρεθεί από το iHOP.

Η διαδικασία τελειώνει πατώντας «Save».

Βάση δεδομένων λέξεων συσχέτισης (Relationship databases)

Signal to Response Correlator
Search your endnote abstracts and define correlation among proteins of interest
You are logged in as "demo user" [Logout](#)

Correlations **Control Panel**

Databases **Reports**
Endnote databases Keywords database **Relationship database** Correlations History

Relationship database

Word	ID	Update	Delete
activate	197	Update	Delete
affect	198	Update	Delete
associate	199	Update	Delete
bind	200	Update	Delete
block	201	Update	Delete
cause	202	Update	Delete
characterize	203	Update	Delete
compare	204	Update	Delete
contain	205	Update	Delete
decrease	227	Update	Delete
database	206	Update	Delete

1

2 Add word

Word

Save **Cancel**

Logs:
Fri 29 Jul 2011, 05:56:21 - demo user logged in
Thu 28 Jul 2011, 17:29:34 - demo user added 499 articles in database p38-nfkb
Thu 28 Jul 2011, 17:24:23 - demo user logged in
Thu 28 Jul 2011, 16:50:36 - demo user deleted synonym nfkbib for protein nfkb

Setup **Help**

1 Περιοχή όπου φαίνονται οι λέξεις συσχέτισης (τις έχουμε ήδη αναφέρει και προεγκαταστήθηκαν στην εφαρμογή). Μπορεί να γίνει αναζήτηση, αλφαβητική διαλογή, αλλαγή ή διαγραφή στην κάθε λέξη συσχέτισης.

2 Περιοχή όπου γίνεται η προσθήκη λέξης συσχέτισης. Την πληκτρολογούμε και πατάμε «Save».

Αναφορές

Signal to Response Correlator
Search your endnote abstracts and define correlation among proteins of interest
You are logged in as "demo user" [Logout](#)

Correlations Control Panel

Databases: Endnote databases, Keywords database, Relationship database
Reports: Correlations, History

Reports on previous correlations

Signal keyword	Response keyword	Database(s)	Weight	Positive/Negative	Job	Date
p38	nfb	p38-nfb,test	0.85	3/1	demou110729-001	2011-07-29

Logs:
 Fri 29 Jul 2011, 06:23:08 - demo user assigned weight for job demou110729-001
 Fri 29 Jul 2011, 06:23:05 - demo user finished positive/negative correlation for the results of job demou110729-001
 Fri 29 Jul 2011, 05:56:21 - demo user logged in
 Thu 28 Jul 2011, 17:29:34 - demo user added 499 articles in database p38-nfb

Setup Help

1 Φαίνεται ο συντελεστής συσχέτισης κάθε ζεύγους πρωτεϊνών και επιπρόσθετες πληροφορίες για τις βάσεις δεδομένων περιλήψεων που χρησιμοποιήθηκαν, το άθροισμα θετικών και αρνητικών αποτελεσμάτων, η ονομασία και ο χρόνος που πραγματοποιήθηκε η αναζήτηση-εργασία. Όπως και προηγουμένως μπορεί να γίνει συνδυαστική αναζήτηση και αλφαβητική διαλογή στα στοιχεία του πίνακα.

Πατώντας στο όνομα της αναζήτησης-εργασίας (job), μας παρουσιάζεται το επόμενο screenshot.

Signal to Response Correlator

Search your endnote abstracts and define correlation among proteins of interest
 You are logged in as "demo user" [Logout](#)

Correlations

Control Panel

Databases

Reports

Endnote databases

Keywords database

Relationship database

Correlations

History

View sentences for the job demou110729-001

Settings for this job

1 Keyword proximity:4, Relationship database used:no, Synonyms for the Signal keyword used:no, Synonyms for the Response keyword used:no

Sentence	Journal	Abstract	Year	
<p>2</p> <p>p38-MAPK and p65-NF-κB activation, and iNOS expression have a significant role in varicocele-induced testicular dysfunction.</p>	Andrologia	Potential role of p38-mitogene-activated protein kinase and nuclear factor-kappa B expression in testicular dysfunction associated with varicocele: an experimental study	2011	
<p>There was a marked staining for iNOS, p38-MAPK and p65-NF-κB expression in rats of varicocele group compared with the sham group</p>	Andrologia	Potential role of p38-mitogene-activated protein kinase and nuclear factor-kappa B expression in testicular dysfunction associated with varicocele: an experimental study	2011	
<p>Malondialdehyde (MDA), nitric oxide (NO) and reduced glutathione (GSH) levels were biochemically assessed, and the p38-MAPK and NF-κB activity, and iNOS expression were immunohistochemically studied in the right and left testicles of rats from each group</p>	Andrologia	Potential role of p38-mitogene-activated protein kinase and nuclear factor-kappa B expression in testicular dysfunction associated with varicocele: an experimental study	2011	
<p>The aim of this study was to investigate p38-mitogene-activated protein kinase (p38-MAPK), nuclear factor-kappa B (p65-NF-κB) and inducible nitric oxide synthase (iNOS) expression in an experimental model of varicocele in the rat testis</p>	Andrologia	Potential role of p38-mitogene-activated protein kinase and nuclear factor-kappa B expression in testicular dysfunction associated with varicocele: an experimental study	2011	

Logs:

Fri 29 Jul 2011, 06:23:08 - demo user assigned weight for job demou110729-001
 Fri 29 Jul 2011, 06:23:05 - demo user finished positive/negative correlation for the results of job demou110729-001
 Fri 29 Jul 2011, 05:56:21 - demo user logged in
 Thu 28 Jul 2011, 17:29:34 - demo user added 499 articles in database p38-nfkb



- 1 Περιοχή όπου φαίνονται οι προτιμήσεις για την αναζήτηση που έχουμε επιλέξει από τις αναφορές.
- 2 Περιοχή όπου φαίνονται οι προτάσεις όπως ήταν κατά την διαδικασία αναζήτησης-επιλογής συντελεστή συσχέτισης και η θετική-αρνητική συσχέτισή τους που επέλεξε ο χρήστης.

Ιστορικό

Signal to Response Correlator
 Search your endnote abstracts and define correlation among proteins of interest
 You are logged in as "demo user" [Logout](#)

Correlations **Control Panel**

Databases: Endnote databases, Keywords database, Relationship database | Reports: Correlations, **History**

History

ID	Date	Action
272	Fri 29 Jul 2011, 06:23:08	demo user assigned weight for job demou110729-001
271	Fri 29 Jul 2011, 06:23:05	demo user finished positive/negative correlation for the results of job demou110729-001
270	Fri 29 Jul 2011, 05:56:21	demo user logged in
269	Thu 28 Jul 2011, 17:29:34	demo user added 499 articles in database p38-nfkb
268	Thu 28 Jul 2011, 17:24:23	demo user logged in
267	Thu 28 Jul 2011, 16:50:36	demo user deleted synonym nfkbib for protein nfkb
266	Thu 28 Jul 2011, 16:50:34	demo user added synonym nfkbib for protein nfkb
265	Thu 28 Jul 2011, 15:36:14	demo user logged in
241	Tue 19 Jul 2011, 11:33:51	demo user logged in
233	Sat 16 Jul 2011, 22:28:43	demo user logged out
232	Sat 16 Jul 2011, 22:27:12	demo user logged in

Logs:
 Fri 29 Jul 2011, 06:23:08 - demo user assigned weight for job demou110729-001
 Fri 29 Jul 2011, 06:23:05 - demo user finished positive/negative correlation for the results of job demou110729-001
 Fri 29 Jul 2011, 05:56:21 - demo user logged in
 Thu 28 Jul 2011, 17:29:34 - demo user added 499 articles in database p38-nfkb

Setup Help

1 Φαίνεται το ιστορικό κινήσεων του χρήστη (logs). Αναφέρεται το όνομα του χρήστη, ο χρόνος και η ενέργεια/κίνηση που πραγματοποίησε. Όπως και προηγουμένως, μπορεί να γίνει αναζήτηση και αλφαβητική διαλογή στην περιοχή μέσω των πεδίων του πίνακα.

2 Όπως ήδη έχουμε αναφέρει, στο κάτω μέρος του GUI υπάρχουν πάντα οι τέσσερις τελευταίες κινήσεις του χρήστη.

2.3.3.4 Αλλαγή προτιμήσεων χρήστη

The screenshot displays the 'Signal to Response Correlator' web application. At the top, it says 'Search your endnote abstracts and define correlation among proteins of interest' and 'You are logged in as "demo user" Logout'. Below this are two tabs: 'Correlations' and 'Control Panel'. The main content area is titled 'Change your personal details' and contains a form titled 'Edit your details'. The form has four input fields: 'Name' (containing 'demo user'), 'Username' (containing 'demo'), 'Password' (containing 'demo'), and 'Signature' (empty). Below the fields are 'Save' and 'Cancel' buttons. A purple circle with the number '1' is overlaid on the form. At the bottom left, there is a 'Logs' section with four entries: 'Fri 29 Jul 2011, 05:56:21 - demo user logged in', 'Thu 28 Jul 2011, 17:29:34 - demo user added 499 articles in database p38-nfkb', 'Thu 28 Jul 2011, 17:24:23 - demo user logged in', and 'Thu 28 Jul 2011, 16:50:36 - demo user deleted synonym nfkbib for protein nfkb'. At the bottom right, there are 'Setup' and 'Help' buttons.

1

Εδώ ο χρήστης μπορεί να αλλάζει όλες τις προσωπικές του πληροφορίες εκτός από το ψευδώνυμό του.

2.3.3.5 Βοήθεια

Signal to Response Correlator
Search your endnote abstracts and define correlation among proteins of interest
You are logged in as "demo user" [Logout](#)

Correlations Control Panel

Help tutorials

- How to get started
- Correlations
- Control Panel
 - Endnote database
 - Keywords database
 - Relationship database
- Contact Programmer

Logs:
Fri 29 Jul 2011, 05:56:21 - demo user logged in
Thu 28 Jul 2011, 17:29:34 - demo user added 499 articles in database p38-nfkb
Thu 28 Jul 2011, 17:24:23 - demo user logged in
Thu 28 Jul 2011, 16:50:36 - demo user deleted synonym nfkbib for protein nfkb

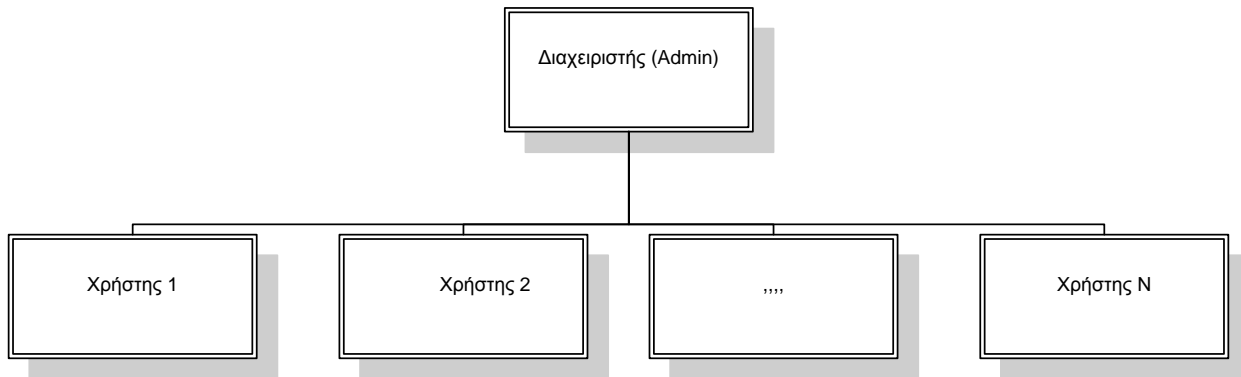
Setup Help

1

Εδώ ο χρήστης μπορεί να διαβάσει τα διαδραστικά εγχειρίδια χρήσεως. Το «how to get started» δείχνει μια μελέτη περίπτωσης χρήσεως της εφαρμογής.

2.3.3.6 Επιπρόσθετα δικαιώματα διαχειριστή

Στο παρακάτω διάγραμμα φαίνεται το οργανόγραμμα της εφαρμογής.



Παρακάτω φαίνονται screenshots από την εφαρμογή που είναι διαθέσιμες μόνο στον διαχειριστή.

Βάση δεδομένων λέξεων που δηλώνουν άρνηση

Signal to Response Correlator
Search your endnote abstracts and define correlation among proteins of interest
You are logged in as "Papakonstantinou Anastasios" [Logout](#)

Correlations **Control Panel**

Databases **Reports** **Personnel**
Endnote databases Keywords database Relationship database **Negative database** Correlations History Users

Negative database

Word	ID	Update	Delete
barely	76	Update	Delete
nor	77	Update	Delete
not	75	Update	Delete

1

Add word

Word

[Save](#) [Cancel](#)

2

Logs:
 Fri 29 Jul 2011, 06:57:41 - Papakonstantinou Anastasios logged in
 Tue 26 Jul 2011, 09:47:18 - Papakonstantinou Anastasios logged in
 Sun 24 Jul 2011, 20:34:10 - Papakonstantinou Anastasios assigned weight for job PapaA110724-002
 Sun 24 Jul 2011, 20:34:02 - Papakonstantinou Anastasios deleted database: p38-nfkb

[Setup](#) [Help](#)

1 Περιοχή όπου στον διαχειριστή φαίνονται οι λέξεις που δηλώνουν άρνηση. Έχει την ίδια χρήση με την λειτουργία προσθήκης λέξεων συσχέτισης (εμπεριέχονται φίλτρα μοναδικότητας και επεξεργασίας).

2 Περιοχή που ο διαχειριστής εισάγει νέα λέξη που δηλώνη άρνηση.

Σημείωση: έχει αλλάξει το υπομενού του κέντρου έλεγχου μόνο για τον διαχειριστή. Οι λέξεις άρνησης αφορούν και χρησιμοποιούνται όλους τους χρήστες που δημιουργούνται από τον διαχειριστή.

Βάση δεδομένων χρηστών

Signal to Response Correlator

Search your endnote abstracts and define correlation among proteins of interest
You are logged in as "Papakonstantinou Anastasios" [Logout](#)

Correlations
Control Panel

Databases

Reports

Personnel

Endnote databases
Keywords database
Relationship database
Negative database

Correlations
History
Users

View users

1

Name	ID	Update	Delete
<input style="width: 90%;" type="text"/>	<input style="width: 10%;" type="text"/>		
demo user	83	Update	Delete
test	87	Update	Delete

2

Add User

Name

Username

Password

Signature

Save
Cancel

Logs:

Fri 29 Jul 2011, 06:57:41 - Papakonstantinou Anastasios logged in

Tue 26 Jul 2011, 09:47:18 - Papakonstantinou Anastasios logged in

Sun 24 Jul 2011, 20:34:10 - Papakonstantinou Anastasios assigned weight for job PapaA110724-002

Sun 24 Jul 2011, 20:34:02 - Papakonstantinou Anastasios deleted database: p38-nfkb

Setup

Help

- 1 Περιοχή όπου στον διαχειριστή φαίνονται οι χρήστες. Έχει την ίδια χρήση με την λειτουργία προσθήκης λέξεων συσχέτισης (εμπεριέχονται φίλτρα μοναδικότητας και επεξεργασίας).
- 2 Περιοχή που ο διαχειριστής εισάγει νέο χρήστη. Μόλις εισάγεται ένας νέος χρήστης τότε αυτόματα προστίθενται στην βάση του οι προεγκατεστημένες λέξεις συσχέτισης.

Σημείωση: έχει αλλάξει το υπομενού του κέντρου έλεγχου μόνο για τον διαχειριστή.

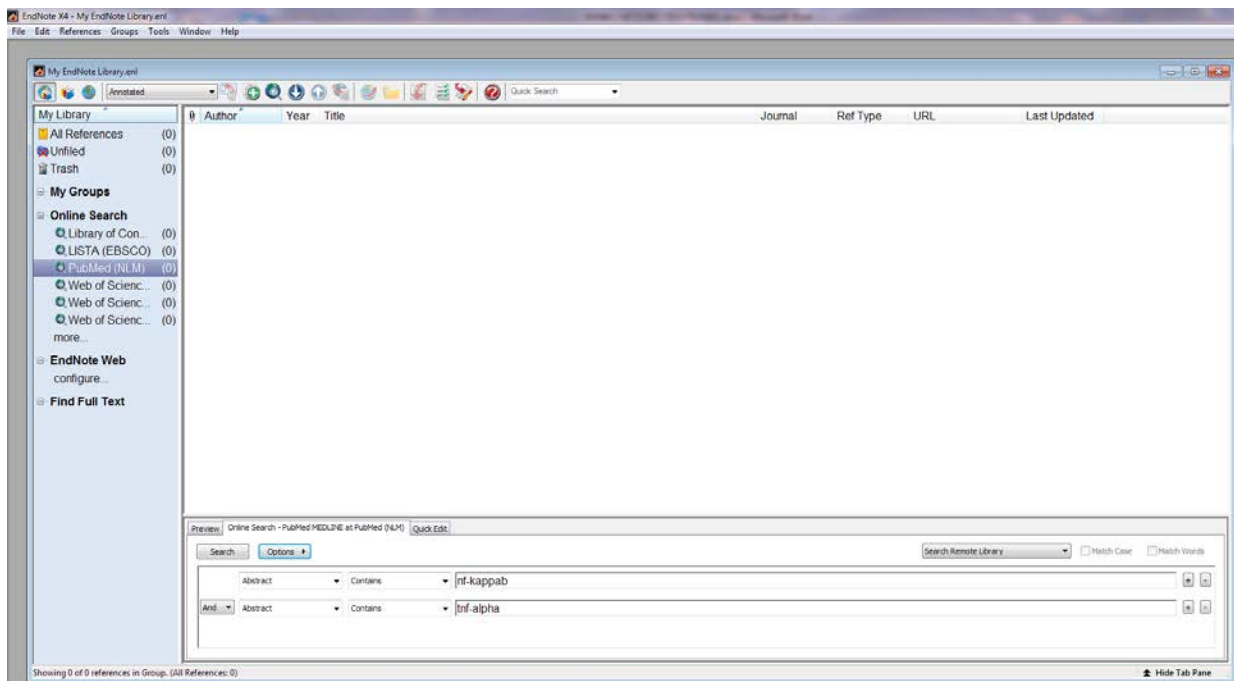
Κεφάλαιο 3^ο

ΑΠΟΤΕΛΕΣΜΑΤΑ

Σε αυτό το κεφάλαιο θα πραγματοποιήσουμε την μελέτη-περίπτωσης μας.

Θα μελετήσω την σχέση ανάμεσα στο ζευγάρι πρωτεϊνών: **nf-kappab** και **tnf-alpha**.

Ξεκινάω να εισαγάγω περιλήψεις από το Endnote.



Αναζητώ από το PubMed άρθρα που θα έχουν στην περίληψή τους τις πρωτεΐνες **nf-kappab** και **tnf-alpha**.

The screenshot displays the EndNote X4 interface. The main window shows a list of references with columns for Author, Year, Title, Journal, Ref Type, URL, and Last Updated. A search dialog box is open at the bottom, showing search criteria for 'nf-kappaB' and 'tnf-alpha'.

Author	Year	Title	Journal	Ref Type	URL	Last Updated
Hristova	2011	The Tobacco Smoke Component Acrolein Suppresses Innate Macrophage Respons...	Am J Respir...	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Chen	2011	ZLJ-6, a novel COX/5-LOX inhibitor, attenuates TNF-alpha-induced endothelial E-sel	Vascul Pha...	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Song	2011	Suppression of TNF-alpha-induced MMP-9 expression by a cell-permeable superoxid...	BMB Rep	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Chanput	2010	Transcription profiles of LPS-stimulated THP-1 monocytes and macrophages: a tool	Food Funct	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Mo	2011	Altered gene expression by low-dose arsenic exposure in humans and cultured cardi...	Int J Environ...	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Suzuki	2011	Inhibitory effect of aminoimidazole carboxamide ribonucleotide (AICAR) on endotoxi...	Invest Ophth...	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Xu	2011	An immunomodulatory beta-glucan from Lentinus edodes activates mitogen activate...	J Biol Chem	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Su	2011	Beneficial Effects of Ethyl Pyruvate through Inhibiting High-Mobility Group Box 1 Expr	Mediators I...	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Sriram	2011	Modulation of reactive oxygen species in skeletal muscle by myostatin is mediated th...	Aging Cell	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Hu	2011	Class A scavenger receptor attenuates myocardial infarction-induced cardiomyocyte...	Basic Res ...	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Liang	2011	Viscolin reduces VCAM-1 expression in TNF-alpha-treated endothelial cells via the J...	Free Radic ...	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
da Silva	2011	Neurotensin downregulates the pro-inflammatory properties of skin-dendritic cells an...	Biochim Bi...	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Dai	2011	Gastrodin Inhibits Expression of Inducible NO Synthase, Cyclooxygenase-2 and Proi...	PLoS One	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Lakhdari	2011	Identification of NF-kappaB Modulation Capabilities within Human Intestinal Comme...	J Biomed B...	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Chen	2011	Classical swine fever virus suppresses maturation and modulates functions of mono...	Res Vet Sci	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Kim	2011	Compound C independent of AMPK inhibits ICAM-1 and VCAM-1 expression in infl...	Atheroscler...	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Kim	2011	T-Cell Receptor Signaling Enhances Transcriptional Elongation from Latent HIV Prov...	J Mol Biol	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Mohamed	2011	Epicatechin attenuates doxorubicin-induced brain toxicity: Critical role of TNF-alpha...	Brain Res ...	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Chen	2011	Chitinase 3-Like-1 (CH3L1/YKL-40) Expression in Colonic Epithelial Cells as a Pot...	Am J Pathol	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Jain	2011	Differential expression of TNF-alpha signaling molecules and ERK1 in distal and pro...	Tumour Biol	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Jiang	2011	Curcúogside A attenuates experimental cerebral ischemia injury in vitro and vivo	Neuroscien...	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Impellizzeri	2011	CGS 21680, an agonist of the adenosine (A2A) receptor, decreases acute lung infl...	Eur J Phar...	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Liu	2011	Small compound 6-o-angeloylplenolin induces mitotic arrest and exhibits therapeutic...	PLoS One	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11
Perrot-Appia	2011	Similar NF-kappaB Gene Signatures in TNF-alpha Treated Human Endothelial Cells...	PLoS One	Journal Arti...	http://www.ncbi.nlm...	14-Aug-11

The search dialog box at the bottom shows the following search criteria:

- Search Remote Library: [Dropdown]
- Match Case:
- Match Words:
- Abstract: [Dropdown]
- Contains: [Dropdown]
- nf-kappaB
- And: [Dropdown]
- Abstract: [Dropdown]
- Contains: [Dropdown]
- tnf-alpha

Βρίσκω πάνω από 4000 αποτελέσματα που εξάγω ως xml και τα εισάγω στην εφαρμογή μου σε μια καινούργια βάση που ονομάζω *nfkb-tnfalpa*.

Signal to Response Correlator
Search your endnote abstracts and define correlation among proteins of Interest
You are logged in as "demo user" [Logout](#)

[Correlations](#) [Control Panel](#)

Databases [Endnote databases](#) [Keywords database](#) [Relationship database](#) **Reports** [Correlations](#) [History](#)

Endnote databases of 4012 total abstracts

Search database(s) All databases
 nfkb-tnfalpha (4018)

Authors	Title	Year	Journal	Abstract	
<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="button" value="search"/>

Add abstracts

Database: create new OR add to existing

xml file:

Delete databases

nfkb-tnfalpha (4018)

Logs:
Mon 15 Aug 2011, 01:14:06 - demo user logged in
Sun 14 Aug 2011, 23:01:47 - demo user deleted synonym p38b1 for p38
Sun 14 Aug 2011, 23:01:45 - demo user deleted synonym mask1 for p38
Sun 14 Aug 2011, 23:01:42 - demo user deleted synonym nfkb1 for nfkb

Οι περιλήψεις έχουν εισαχθεί στην βάση όπως φαίνονται.

Προχωρούμε στην εισαγωγή των πρωτεϊνών και των συνώνυμών τους.

Signal to Response Correlator
Search your endnote abstracts and define correlation among proteins of interest
You are logged in as "demo user" [Logout](#)

Correlations **Control Panel**

Databases Reports

Endnote databases **Keywords database** Relationship database Correlations History

select synonyms for "nf-kappab"

Symbol	Name (description)	Synonym	Organism
nkrf	nfkb repressing factor	nf-kappab repressing factor	homo sapiens
nkaj	nfkb activating protein	nf-kappab activating protein	homo sapiens
nfkb1	nuclear factor of kappa light polypeptide gene enhancer in b-cells 1, p105	nf-kappab	mus musculus
nfkb2	nuclear factor of kappa light polypeptide gene enhancer in b-cells 2, p49/p10...	nf-kappab2	mus musculus
nkrf	nf-kappab repressing factor		mus musculus
rela	v-rel reticuloendotheliosis viral oncogene homolog a (avian)	nf-kappab transcription factor p65 subunit	bos taurus
nkrf	nf-kappab repressing factor		gallus gallus
nkrf	nf-kappab repressing factor		rattus norvegicus

Add more synonyms for "nf-kappab" separated by a comma

Save

Logs:
 Mon 15 Aug 2011, 01:22:46 - demo user added synonym nfkb2 for protein nf-kappab
 Mon 15 Aug 2011, 01:22:46 - demo user added synonym nfkb1 for protein nf-kappab
 Mon 15 Aug 2011, 01:22:44 - demo user added synonym nkao for protein nf-kappab
 Mon 15 Aug 2011, 01:22:25 - demo user added keyword nf-kappab

[Setup](#) [Help](#)

Για την nf-kappab βάζουμε τα συνώνυμα nkao, nfkb1, nfkb2.

Signal to Response Correlator
 Search your endnote abstracts and define correlation among proteins of interest
 You are logged in as "demo user" [Logout](#)

Correlations **Control Panel**

Databases: Endnote databases **Keywords database** Relationship database Reports: Correlations History

select synonyms for "tnf-alpha"

Symbol	Name (description)	Synonym	Organism
tnf-alpha	putative tumour necrosis factor alpha		oncorhynchus mykiss
tnf-alpha	tumor necrosis factor alpha		salmo salar
tnf-alpha	tumor necrosis factor alpha		macaca mulatta
tnf-alpha-1	tumor necrosis factor alpha-1 precursor		salmo salar
tnf-alpha-2	tumor necrosis factor alpha-2 precursor		salmo salar
tnf	tumor necrosis factor (tnf superfamily, member 2)	tnf-alpha	homo sapiens
adam17	adams metalloproteinase domain 17	tnf-alpha-converting enzyme	homo sapiens
abc1	atp-binding cassette, sub-family f (gcn20), member 1	tnf-alpha-stimulated abc protein	homo sapiens
tnf	tumor necrosis factor	tnf-alpha	mus musculus
tnfrsf1a	tumor necrosis factor receptor superfamily, member 1a	tnf-alpha1	mus musculus
tnfrsf1b	tumor necrosis factor receptor superfamily, member 1b	tnf-alpha-2	mus musculus
adam17	a disintegrin and metalloproteinase domain 17	tnf-alpha-converting enzyme	mus musculus
adam17	adams metalloproteinase domain 17	tnf-alpha converting enzyme	sus scrofa
tnfb	tumor necrosis factor b (tnf superfamily, member 2)	tnf-alpha	danio rerio
tnf	tumor necrosis factor (tnf superfamily, member 2)	tnf-alpha	pan troglodytes
tnf	tumor necrosis factor (tnf superfamily, member 2)	tnf-alpha	bos taurus
b2br	tnf-alpha-receptor-like protein		vaccinia virus
c22l	tnf-alpha-receptor-like		vaccinia virus
vacwr002	tnf-alpha-receptor-like protein		vaccinia virus
vacwr217	tnf-alpha-receptor-like protein		vaccinia virus
tnf	tumor necrosis factor (tnf superfamily, member 2)	tnf-alpha	ovis aries
tnf	tumor necrosis factor (tnf superfamily, member 2)	tnf-alpha	rattus norvegicus
adam17	a disintegrin and metalloproteinase domain 17	tnf-alpha-converting enzyme	rattus norvegicus
tatv_dah68_004	host defense modulator	tnf-alpha-receptor-like protein	taterapox virus
tatv_dah68_005	host defense modulator	tnf-alpha-receptor-like protein	taterapox virus
tatv_dah68_221	host defense modulator	tnf-alpha-receptor-like	taterapox virus
tatv_dah68_222	host defense modulator	tnf-alpha-receptor-like protein	taterapox virus

Add more synonyms for "tnf-alpha" separated by a comma

Save

Logs:
 Mon 15 Aug 2011, 01:23:11 - demo user added synonym tnf-alpha-2 for protein tnf-alpha
 Mon 15 Aug 2011, 01:23:09 - demo user added synonym tnf-alpha-1 for protein tnf-alpha
 Mon 15 Aug 2011, 01:23:06 - demo user added keyword tnf-alpha
 Mon 15 Aug 2011, 01:22:46 - demo user added synonym nfkb2 for protein nf-kappaB

[Setup](#) [Help](#)

Για την tnf-alpha βάζουμε τα συνώνυμα tnf-alpha-1, tnf-alpha-2.


Signal to Response Correlator
 Search your endnote abstracts and define correlation among proteins of interest
 You are logged in as "demo user" [Logout](#)

Correlations **Control Panel**

Databases: Endnote databases | **Keywords database** | Relationship database | Reports: Correlations | History

Keywords & synonyms database

Keyword	Synonym	ID	Update	Delete
rf-kappab		80	Update	Delete
rf-kappab	rfkb1	82	Update	Delete
rf-kappab	rfkb2	83	Update	Delete
rf-kappab	rikap	81	Update	Delete
trf-alpha		84	Update	Delete
trf-alpha	trf-alpha-1	85	Update	Delete
trf-alpha	trf-alpha-2	86	Update	Delete

Add keyword 

Keyword

[Retrieve synonyms](#) [Cancel](#)

Loggs:
 Mon 15 Aug 2011, 01:23:11 - demo user added synonym trf-alpha-2 for protein trf-alpha
 Mon 15 Aug 2011, 01:23:09 - demo user added synonym trf-alpha-1 for protein trf-alpha
 Mon 15 Aug 2011, 01:23:06 - demo user added keyword trf-alpha
 Mon 15 Aug 2011, 01:22:46 - demo user added synonym rfkb2 for protein rf-kappab

[Setup](#) [Help](#)

Άρα καταλήγουμε στο παραπάνω screenshot.

Δεν εισαγάγουμε επιπρόσθετες λέξεις που δηλώνουν συσχέτιση μιας και στο σύστημα υπάρχουν προεγκατεστημένες.

Προχωρούμε στην διαδικασία αναζήτησης για τον καθορισμό του συντελεστή συσχέτισης.

Signal to Response Correlator
Search your endnote abstracts and define correlation among proteins of interest
You are logged in as "demo user" [Logout](#)

Signal keyword: nf-kappab
Response keyword: tnf-alpha
Job: demou110815-001
Db: nfkb-tnfalpa

Correlations Control Panel

Define correlations

"How to get started"

Use simple search

Signal keyword [add new](#)
nf-kappab
Enable synonyms: Yes No

Response keyword [add new](#)
tnf-alpha
Enable synonyms? Yes No

Endnote databases: [add new](#)
 All databases
 nfkb-tnfalpa (4018)

Keyword proximity
3

Enable relationship words?
Yes No

[Search](#)

Logs:
Mon 15 Aug 2011, 01:23:11 - demo user added synonym tnf-alpha-2 for protein tnf-alpha
Mon 15 Aug 2011, 01:23:09 - demo user added synonym tnf-alpha-1 for protein tnf-alpha
Mon 15 Aug 2011, 01:23:08 - demo user added keyword tnf-alpha
Mon 15 Aug 2011, 01:22:46 - demo user added synonym nfkb2 for protein nf-kappab

[Setup](#) [Help](#)

Επιλέγουμε advanced settings, τις πρωτεΐνες χωρίς συνώνυμα, απόσταση λέξεων 3 και την χρήση λέξεων συσχέτισης και λέξεων άρνησης.

Signal to Response Correlator
 Search your endnote abstracts and define correlation among proteins of interest
 You are logged in as "demo user" [Logout](#)

Signal keyword: nf-kappab
 Response keyword: tnf-alpha
 Job: demou110815-001
 Db: nfkb-tnfalpaha

[Correlations](#) [Control Panel](#)

Define correlations

There are 618 results in 4012 abstracts that matched the criteria.

"How to get started"

[Use simple search](#)

Signal keyword: [add new](#)
 nf-kappab
 Enable synonyms: Yes No

Response keyword: [add new](#)
 tnf-alpha
 Enable synonyms? Yes No

Endnote databases: [add new](#)
 All databases
 nfkb-tnfalpaha (4018)

Keyword proximity
 3

Enable relationship words?
 Yes No

[Search again](#)

[Continue to view the results](#)

Logs:
 Mon 15 Aug 2011, 01:23:11 - demo user added synonym tnf-alpha-2 for protein tnf-alpha
 Mon 15 Aug 2011, 01:23:09 - demo user added synonym tnf-alpha-1 for protein tnf-alpha
 Mon 15 Aug 2011, 01:23:06 - demo user added keyword tnf-alpha
 Mon 15 Aug 2011, 01:22:46 - demo user added synonym nfkb2 for protein nf-kappab

[Setup](#) [Help](#)

Βρίσκουμε 618 προτάσεις-αποτελέσματα που πρέπει να εξετάσουμε. Αλλάζουμε τις ρυθμίσεις και επιλέγουμε απόσταση λέξεων 2. Βρίσκουμε 336 προτάσεις.

Τα αποτελέσματα είναι 143 προτάσεις με θετική συσχέτιση από άρθρα που έχουν αναφερθεί ως βιβλιογραφία 105 φορές και 1 πρόταση με αρνητική συσχέτιση και καμία βιβλιογραφική αναφορά.

Άρα βάζω συντελεστή συσχέτισης (+) 0.999, όπως βλέπουμε και παρακάτω.

Signal keyword	Response keyword	Database(s)	Weight	Positive/Negative	Job	Date
nf-kappab	tnf-alpha	nfkb-tnfalpa	0.999	143/1	demou110815-001	2011-08-15

Κεφάλαιο 4^ο

ΣΥΖΗΤΗΣΗ

Σε αυτή την διπλωματική προτείνουμε μια μεθοδολογία για την εύρεση του συντελεστή συσχέτισης ανάμεσα σε σηματοδότες (δίνουν το σήμα) πρωτεΐνες εντός κυττάρου και σε ενεργούσες (αυτές που ενεργοποιούνται) πρωτεΐνες εκτός κυττάρου, δηλαδή σε πρωτεΐνες που γνωρίζουμε ότι συσχετίζονται. Αυτή η μεθοδολογία έχει ως αρχικό στάδιο την εισαγωγή δεδομένων από τις περιλήψεις χιλιάδων δημοσιεύσεων του Endnote, συνώνυμα πρωτεϊνών, λέξεις συσχέτισεων, στάδιο επεξεργασίας την εξόρυξη δεδομένων και προτύπων από αυτά και στάδιο αποτελεσμάτων την παρουσίαση τους στον χρήστη, για να γίνει η τελική επιλογή τους και η εξεύρεση-επιλογή του συντελεστή συσχέτισης.

Η μεθοδολογία αυτή μετουσιώθηκε στον προγραμματισμό της εφαρμογής «Signal to Response Correlator» <http://www.jojo.gr/thesis>.

Παρακάτω θα αναλύσουμε διάφορες πτυχές της εφαρμογής.

4.1 Πλεονεκτήματα

Θα αναλύσουμε τα πλεονεκτήματα της εφαρμογής σε σύγκριση με τις τεχνικές της εξόρυξης δεδομένων, με αντίστοιχα προγράμματα και αυτούσια μόνη της.

4.1.1 Σύγκριση με τεχνικές εξόρυξης δεδομένων

Οι τεχνικές εξόρυξης δεδομένων σε συνδυασμό με την βιοπληροφορική έχουν μερικές προκλήσεις να αντιμετωπίσουν. Αυτές είναι η απεραντοσύνη (vastness) των άρθρων-περιλήψεων, η ασάφεια (vagueness) σχετικά με την χρήση των πρωτεϊνών, η αβεβαιότητα (uncertainty) σχετικά με διαφορετικές περιλήψεις με παρόμοια λεκτική συμπτωματολογία, η ασυνέπεια (inconsistency) δηλαδή οι λογικές συγκρούσεις που παρουσιάζονται από περίληψη σε περίληψη και η εξαπάτηση (deceit) από αρθρογράφους.¹²

Η απεραντοσύνη των άρθρων μειώνεται καθώς η διαδικασία ξεκινά από την χρήση του endnote όπου φιλτράρουμε τις περιλήψεις και εξάγουμε σε xml εκείνες που μας ενδιαφέρουν.

Η ασάφεια της χρήσης των πρωτεϊνών ελαχιστοποιείται καθώς το σύστημα καλεί ένα web crawl που φιλτράρει το iHOP για συνώνυμα πρωτεϊνών και δίνεται η δυνατότητα στον χρήστη αν θέλει να τα χρησιμοποιήσει στην αναζήτησή του.

Η αβεβαιότητα και η ασυνέπεια μειώνεται καθώς στο στάδιο πριν από την απόδοση του συντελεστή συσχέτισης, ο χρήστης καλείται να ελέγξει το επίπεδο της αξιοπιστίας των προτάσεων-αποτελεσμάτων της αναζήτησης.

Η εξαπάτηση δεν υφίσταται στον συγκεκριμένο τομέα της βιοπληροφορικής καθώς οι περιλήψεις εξάγονται από άρθρα που για την συγγραφή τους έχουν γίνει πειράματα.

4.1.2 Σύγκριση με αντίστοιχα προγράμματα

Σε σύγκριση με άλλα προγράμματα, η εφαρμογή εφαρμογής «Signal to Response Correlator» χαρακτηρίζεται από απλότητα (simplicity) στην χρήση και συμβατότητα (compatibility) με άλλα προγράμματα.

Η απλότητα φαίνεται από την κατανοητή Γραφική Διεπιφάνεια Χρήσης (GUI).

Η συμβατότητα αφορά την χρήση xml για την σύνδεση με το endnote και την χρήση web crawls όπου συνδέεται η εφαρμογή με το Google Scholar και το iHOP.

4.1.3 Πλεονεκτήματα εφαρμογής

Κάθε εφαρμογή θα πρέπει να χαρακτηρίζεται από συγκεκριμένες ιδιότητες, όπως ορθότητα (correctness), αξιοπιστία (reliability), αποδοτικότητα (performance), ευελιξία-συντηρησιμότητα (maintainability), ελεγχιμότητα (verifiability), δυνατότητα επαναχρησιμοποίησης (reusability), φορητότητα (portability) και ευχρηστία (usability).¹³

Στην θεωρητική επιστήμη των υπολογιστών, η ορθότητα ενός αλγορίθμου αποδίδεται, όταν λέγεται ότι ο αλγόριθμος είναι σωστός σε σχέση με τις προδιαγραφές. Λειτουργική ακρίβεια αναφέρεται στη συμπεριφορά εισόδου-εξόδου του αλγόριθμου (δηλαδή, για κάθε είσοδο που παράγει τη σωστή έξοδο).¹⁴ Η ορθότητα της εφαρμογής κρίνεται από την ανάπτυξή της σύμφωνα με τις λειτουργικές απαιτήσεις και στην προκειμμένη περίπτωση καλύπτονται.

Στην δικτύωση υπολογιστών, ένα αξιόπιστο πρωτόκολλο είναι αυτό που παρέχει τις ιδιότητες αξιοπιστία όσον αφορά την παράδοση των δεδομένων με τον δέκτη (εξ), σε αντίθεση με ένα αναξιόπιστο πρωτόκολλο, το οποίο δεν παρέχει ειδοποιήσεις στον αποστολέα ως προς την παράδοση των δεδομένων που διαβιβάζονται.¹⁵ Η αξιοπιστία της εφαρμογής υφίσταται καθώς δεν προκαλεί κάποιου είδους ζημία σε περίπτωση λάθους, δηλαδή αποτρέπει διπλότυπες εγγραφές, λειτουργεί χωρίς διακοπή και δεν αφήνει κατάλοιπα στον σκληρό δίσκο.

Η αποδοτικότητα μιας εφαρμογής χαρακτηρίζεται από την παροχή χρήσιμου έργου προς τον χρήστη σε σχέση με τους πόρους του συστήματος.¹⁶ Η αποδοτικότητα της εφαρμογής κρίνεται από την μη πρόκληση υπερχειλίσης στην μνήμη RAM.

Η ευελιξία-συντηρησιμότητα αφορά την τμηματική ανάπτυξη της εφαρμογής σε οντότητες (modules) ώστε εκείνη να μπορεί να εξελιχθεί καθώς αλλάζουν οι λειτουργικές απαιτήσεις.

Η ελεγχιμότητα, που βρίσκεται σε συνδυασμό με την ευελιξία, αφορά την δυνατότητα να ελεγχούν οι κώδικες php, mysql, javascript εύκολα. Αυτό υφίσταται, γιατί πέρα από την ανάπτυξη της εφαρμογής σε οντότητες, υπάρχουν ξεχωριστά αρχεία που το καθένα έχει διαφορετική λειτουργία και η ένωσή τους αποτελεί την εφαρμογή. Η συγκεκριμένη ιδιότητα αφορά την προγραμματιστική αρχιτεκτονική της εφαρμογής και είναι εκτός του παρόντος. Αντίστοιχα για την δυνατότητα επαναχρησιμοποίησης της εφαρμογής.

Η φορητότητα της εφαρμογής υφίσταται καθώς είναι προσαρμόσιμη-εγκαταστήσιμη σε περιβάλλον server και pc.

Η ευχρηστία (usability) της εφαρμογής ίσως είναι το σημαντικότερο πλεονέκτημά της. Ιδιότητες της αναλύονται παρακάτω.¹⁷

- Ορατότητα της κατάστασης του συστήματος: Το σύστημα θα πρέπει πάντα να κρατάει τους χρήστες ενήμερους για το τι συμβαίνει, μέσα από κατάλληλη ανατροφοδότηση ειδοποιήσεων εντός εύλογου χρονικού διαστήματος.
- Συμβατότητα μεταξύ του συστήματος και του πραγματικού κόσμου: Το σύστημα θα πρέπει να μιλήσει τη γλώσσα του χρήστη, με φράσεις λέξεις και έννοιες γνωστές στον χρήστη, αντί με όρους συστήματος.
- Έλεγχος και ελευθερία του χρήστη: Οι χρήστες επιλέγουν συχνά λειτουργίες από λάθος και θα χρειαστεί μια σαφή ένδειξη "έξοδος κινδύνου" για να αφήσει την ανεπιθύμητη κατάσταση, χωρίς να χρειάζεται να περάσει μέσα από ένα εκτεταμένο διάλογο.
- Συνέπεια και σαφήνεια: Οι χρήστες δεν θα πρέπει να αναρωτιούνται αν διαφορετικές λέξεις, καταστάσεις, ή ενέργειες σημαίνουν το ίδιο πράγμα.
- Πρόληψη σφαλμάτων: Ακόμα καλύτερα να υπάρχει ένας προσεκτικός σχεδιασμός από μηνύματα λάθους.

Αναγνώριση αντί ανάκλησης από την μνήμη: Ελαχιστοποίηση του φορτίου μνήμη του χρήστη, κάνοντας τα αντικείμενα, τις ενέργειες και επιλογές ορατά. Οδηγίες για τη

χρήση του συστήματος θα πρέπει να είναι ευδιάκριτη ή εύκολα να ανακτηθούν όποτε κρίνεται απαραίτητο.

- Η ευελιξία και η αποδοτικότητα της χρήσης: Επιταχυντές μπορεί συχνά να επιταχύνουν την αλληλεπίδραση με τον χρήστη, έτσι ώστε το σύστημα μπορεί να ανταποκριθεί σε δύο άπειρους και έμπειρους χρήστες.
- Αισθητική και μινιμαλιστικός σχεδιασμός: Οι ειδοποιήσεις και η αλληλεπίδραση με τον χρήστη, δεν πρέπει να περιέχουν πληροφορίες που είναι άσχετες ή σπάνια χρειάζονται.
- Υποβοήθηση των χρηστών να αναγνωρίζουν, τη διάγνωση, και να ανακάμπτουν από τα λάθη: Μηνύματα λάθους πρέπει να εκφράζονται σε απλή γλώσσα (χωρίς κωδικούς), να αναφέρουν με ακρίβεια το πρόβλημα, και επικοινωνητικά να προτείνουν μια λύση.
- Βοήθεια και εγχειρήδια χρήσης: Ακόμα κι αν είναι καλύτερα αν το σύστημα μπορεί να χρησιμοποιηθεί χωρίς εγχειρήδια χρήσης, μπορεί να χρειαστεί να παράσχει βοήθεια.





Παρακάτω αναλύονται οι διαφορετικές πτυχές της ευχρηστίας της εφαρμογής.

- Έχει γίνει αρκετά μεγάλη προσπάθεια ώστε η Γραφική Διεπιφάνεια Χρήσης (GUI) να είναι απλή και κατανοητή, σε απαλά χρώματα με αισθητική και μινιμαλισμό, χωρίς να κουράζει τον χρήστη.
- Υπάρχει σαφήνεια και συνέπεια στο GUI, ώστε αντίστοιχες λειτουργίες να γίνονται με τον ίδιο τρόπο.
- Υπάρχει διαφάνεια και έλεγχος από τον χρήστη, ώστε να φαίνονται οι υπολογισμοί που γίνονται ανά πάσα στιγμή και ο χρήστης να ελέγχει τις κινήσεις που θέλει να κάνει.
- Στην αναζήτηση υπάρχει περιοχή όπου φαίνονται οι βασικές πληροφορίες για την αναζητήτη μας και το όνομα της εργασίας, όπως προκύπτει από το σύστημα με βάση το όνομα του χρήστη, την ημερομηνία και τον αύξοντα αριθμό της εργασίας.

Signal keyword: p38
Response keyword: nfkb
Job: demou110801-001
Db: p38-nfkb,test

- Υπάρχουν διαδραστικά εγχειρήδια χρήσεως και ένα «how to get started».
- Υπάρχουν μικρές κοινοποιήσεις βοήθειας με το σύμβολο 🗨️. Τέτοια σύμβολα υπάρχουν σε όλες τις διαδικασίες, διαμέσου του GUI.
- Για την ανάδραση/πληροφόρηση κατάστασης του συστήματος, υπάρχουν ανακοινώσεις και αναφορές μετά το πέρας κάθε διαδικασίας, όπως έχουμε δει.
- Υπάρχει ειδοποίηση στις αναφορές για μη ολοκληρωμένες αναζητήσεις όπου δεν έχει αποδοθεί η θετική, αρνητική ή καμία συσχέτιση στα αποτελέσματα και ολοκληρωμένες αναζητήσεις όπου δεν έχει αποδοθεί ο συντελεστής συσχέτισης.

- Υπάρχουν συνδυαστικές αναζητήσεις στις βάσεις δεδομένων περιλήψεων, λέξεων-κλειδιών και λέξεων που δηλώνουν συσχέτιση.
- Η διαδικασία αναζήτησης στις βάσεις δεδομένων περιλήψεων και στην αναζήτηση (φαίνεται παρακάτω) γίνεται με χρήση highlight χρωμάτων, για την πιο γρήγορη εύρεση των όρων. Συγκεκριμένα στα αποτελέσματα αναζήτησης υπάρχουν 3 διαφορετικά χρώματα για τις πρωτεΐνες (στοιχεία πινάκων 1 και 2), λέξεις συσχέτισης (στοιχεία πίνακα 3) και λέξεις που δηλώνουν άρνηση (στοιχεία πίνακα 4). Επίσης τα κουμπιά για την απόδοση θετικής, αρνητικής ή καμία συσχέτισης των προτάσεων στα αποτελέσματα είναι αρκετά μεγάλα.

Sentence	Journal	Year	Cited	
A gel shift assay further showed that SalB inhibited the TNF-alpha-activated NF-kappaB and AP-1 DNA binding activities in a dose-dependent manner	J Cell Biochem	2005	0	  
We also found that SM-7368 strongly inhibits TNF-alpha-induced NF-kappaB activity but not AP-1 activity	Biochem Biophys Res Commun	2005	0	  
This cytokine induced down-regulation of the subtype 2 of TNF-alpha receptors (TNF-R2), enhanced TNF-alpha-induced NF-kappaB translocation to the nucleus and binding to DNA and increased the immunoreactivity of the protein c-IAP1, a member of the inhibitor of apoptosis (IAP) gene family whose synthesis is stimulated by NF-kappaB at the transcriptional level	Cell Signal	2005	0	  

- Υπάρχει προκαθορισμένη τιμή για την εγγύτητα των πρωτεϊνών.
- Υπάρχουν ασφαλιστικές δικλείδες πρόληψης λαθών στην εφαρμογή ώστε να γίνονται εισαγωγές, τροποποιήσεις και διαγραφές στην βάση κάθε χρήση, χωρίς να επηρεάζεται η υπόλοιπη βάση ή άλλες προτιμήσεις/αναφορές/βάσεις του χρήστη (πχ δεν μπορούν να διαγραφούν περιλήψεις που έχουν χρησιμοποιηθεί στις

αναζητήσεις/αναφορές). Επίσης, μόλις ο χρήστης αποδίδει ένα συντελεστή απόδοσης δεν μπορεί να κάνει back ώστε να τον διπλοεισάγει.

- Το ιστορικό με τις τελευταίες κινήσεις του χρήστη (logs) στο κάτω μέρος του GUI και στο Πίνακα Ελέγχου, αποτελεί μέρος της διαφάνειας του συστήματος.
- Υπάρχει σύστημα αρχείων (filesystem)

4.2 Μειονεκτήματα

Πέρα από τα πλεονεκτήματά της, η εφαρμογή έχει κάποια μειονεκτήματα.

Οι λέξεις συσχέτισης θα μπορούσαν να υποδηλώνουν θετική σχέση ή αρνητική σχέση και όταν αναγνωρίζονται από το σύστημα να φαίνονται με διαφορετικό χρώμα. Αυτό θα βοηθούσε περισσότερο στην απόδοση της σχέσης στα αποτελέσματα της αναζήτησης.

Ενα άλλο μειονέκτημα είναι ότι δεν μπορεί να γίνει αλλαγή των αναφορών και του συντελεστή συσχέτισης μετά την ολοκλήρωση της διαδικασίας απόδοσής του. Αυτό έχει ως αποτέλεσμα να πρέπει να ξαναγίνεται η διαδικασία. Εδώ θα μπορούσε να βοηθήσει ένα κουμπί ανάκλησης (undo).

Στην διεπαφή των αναφορών, ο χρήστης θα μπορούσε να έχει την αρμοδιότητα να διαγράφει προηγούμενους συντελεστές συσχέτισης για κοινά ζεύγη πρωτεϊνών.

Αναφέρουμε ότι γίνεται σύνδεση του υποκεφαλαίου των μειονεκτημάτων με το κεφάλαιο των μελλοντικών βελτιώσεων, μιας και κατ'αρχήν αφορούν την ίδια έννοια.

4.3 Ελάχιστες απαιτήσεις

Η εφαρμογή έχει αναπτυχθεί με php, mysql, xml και javascript. Αυτό την κάνει αρκετά προσαρμόσιμη σε περιβάλλον server και pc.

Σε περιβάλλον apache server απαιτείται έκδοση 1.3.33 ή νεώτερη. Οι ελάχιστες απαιτήσεις καθορίζονται από την εγκατάσταση του apache.

Σε περιβάλλον pc, εκείνο πρέπει να έχει εγκατεστημένο το wampserver και το endnote. Προφανώς οι ελάχιστες απαιτήσεις αυτού του συστήματος είναι ο κοινός τύπος των συνδυασμών των εγκαταστάσεων των δυο προγραμμάτων. Αναφέρουμε ότι είναι 1GB σκληρού δίσκου, 600MB RAM και μια απλή σύνδεση internet.

Κεφάλαιο 5^ο

ΜΕΛΛΟΝΤΙΚΕΣ ΒΕΛΤΙΩΣΕΙΣ

Η ευέλικτη και τμηματοποιημένη (modular) ανάπτυξη μιας εφαρμογής την καθιστά ευκολότερη σε έλεγχο, προσθήκες, αλλαγές και βελτιώσεις. Εδώ γίνεται η σύνδεση με το κεφάλαιο 4.2 όπου αναφέρονται τα μειονεκτήματα της εφαρμογής.

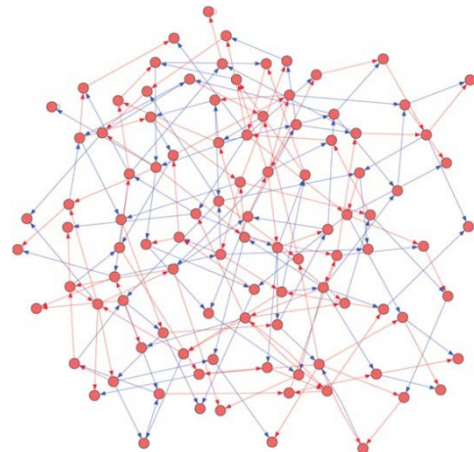
Η εφαρμογή «Signal to Response Correlator» αν και μπορεί να χρησιμοποιηθεί σε διακομιστή (server) αλλά και σε υπολογιστή (pc), θα μπορούσε να υποστεί μια βελτίωση ως προς την εγκατάστασή της. Δηλαδή ένας οδηγός εγκατάστασης, ο οποίος θα καθοδηγούσε τον χρήστη διαμέσου της διαδικασίας εγκατάστασης, θα βελτίωνε σημαντικά τη χρήση και διασπορά της εφαρμογής. Σημαντικά στοιχεία αυτού του οδηγού εγκατάστασης θα ήταν η εγκατάσταση και διαμόρφωση apache server στο τοπικό τερματικό pc.

Σε ζητήματα βάσεων δεδομένων, η εφαρμογή θα μπορούσε να έχει ομάδες προεγκατεστημένων πρωτεϊνών. Αυτό θα έκανε πιο εύκολη την αρχική χρήση της εφαρμογής.

Η αλληλεπίδραση με βάσεις δεδομένων για άρθρα και περιλήψεις θα μπορούσε επίσης να βελτιωθεί. Η χρήση μιας εσωτερικής αναζήτησης στην εφαρμογή, θα μπορούσε να αναζητήσει τα άρθρα και να τα εισάγει στην τοπική βάση, χωρίς την χρήση xml, εξάγοντας κάθε φορά τον αριθμό των citations από το Google Scholar. Αυτό θα έκανε την εφαρμογή να μην εξαρτιέται από τον Endnote.

Στο τομέα της συμβατότητας (compatibility), η εφαρμογή θα μπορούσε να συνεργαστεί με άλλες εφαρμογές βιοπληροφορικής για σύγκριση αποτελεσμάτων.

Τέλος, μετά από την πάροδο χρήσης όπου ο χρήστης θα έχει αποφανθεί για τους συντελεστές συσχέτισης ζευγαριών πρωτεϊνών, το σύστημα θα μπορούσε να αναπαραστήσει γραφικά τις σχέσεις με τους συντελεστές όλων των αλληλοεξαρτώμενων πρωτεϊνών. Αυτό θα μπορούσε να γίνει με ένα γραφικό δύο διαστάσεων διάγραμμα τοπολογίας, όπως φαίνεται δίπλα.



Παράρτημα

ΒΙΒΛΙΟΓΡΑΦΙΑ

¹ <http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm>

² http://en.wikipedia.org/wiki/Text_mining

³ Fayyad, Usama; Gregory Piatetsky-Shapiro, and Padhraic Smyth (1996). "From Data Mining to Knowledge Discovery in Databases".

⁴ Margaret Dunham, Data Mining Introductory and Advanced Topics, 2003, Pearson Education

⁵ Jiawei Han, Micheline Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2nd Edition, ISBN 1-55860-901-6, 2006.

⁶ XindongWu · Vipin Kumar · J. Ross Quinlan · Joydeep Ghosh · Qiang Yang · Hiroshi Motoda · Geoffrey J. McLachlan · Angus Ng · Bing Liu · Philip S. Yu · Zhi-Hua Zhou · Michael Steinbach · David J. Hand · Dan Steinberg "Top 10 algorithms in data mining"

⁷ <http://en.wikipedia.org/wiki/Bioinformatic>

⁸ http://en.wikipedia.org/wiki/Biomedical_text_mining

⁹ Papcun, G., K. Sentz, A. Fulmer, J. Xu, O. Lubeck, M. Wolinsky. 2003. A Construction Grammar Approach to Extracting Regulatory Relationships from Biological Literature. Pacific Symposium on Biocomputing 2003 Kauai, Hawaii.

¹⁰ Verspoor, C., G. Papcun, and K. Sentz. 2003. A Theoretical Motivation for Patterns in Information Extraction. Los Alamos Unclassified Report 03-1504.

¹¹ Τατσιόπουλος, Πρωτοσύγγελος, Πόνης, "Πληροφοριακά Συστήματα Διοίκησης στην Παραγωγή"

¹² http://en.wikipedia.org/wiki/Semantic_Web

¹³ http://en.wikipedia.org/wiki/List_of_system_quality_attributes

¹⁴ http://en.wikipedia.org/wiki/Correctness_%28computer_science%29

¹⁵ http://en.wikipedia.org/wiki/Reliability_%28computer_networking%29

¹⁶ http://en.wikipedia.org/wiki/Computer_performance

¹⁷ Ν.Μαρμαράς, Εισαγωγή στην Εργονομία