



**ΕΘΝΙΚΟ
ΜΕΤΣΟΒΙΟ
ΠΟΛΥΤΕΧΝΕΙΟ**

ΣΧΟΛΗ ΜΗΧΑΝΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ

Α.Π. : 19447
Αθήνα, 13/5/21

ΚΟΣΜΗΤΟΡΑΣ

**Προς τα Μέλη ΔΕΠ της
Σχολής Μηχ/γων
Μηχ/κών**

ΠΡΟΣΚΛΗΣΗ

Σας προσκαλούμε στην παρουσίαση της Διδακτορικής Διατριβής του **Υ.Δ. κ. ΚΟΥΤΣΑΝΔΡΕΑ Θεόδωρου**, κατόχου Διπλώματος Χημικού Μηχανικού, την οποία εκπόνησε στον Τομέα Μηχανολογικών Κατασκευών & Αυτομάτου Επιτροπής. Η παρουσίαση θα πραγματοποιηθεί την Πέμπτη 27 Μαΐου 2021, ώρα 9.30π.μ.

διαδικτυακά*. Ο ελληνικός τίτλος της Διδακτορικής Διατριβής είναι ο εξής:

**«ΑΝΑΠΤΥΞΗ ΜΕΘΟΔΩΝ ΥΠΟΛΟΓΙΣΤΙΚΗΣ ΝΟΗΜΟΣΥΝΗΣ ΓΙΑ ΤΗΝ
ΣΥΝΘΕΤΙΚΗ, ΠΟΣΟΤΙΚΗ ΚΑΙ ΣΗΜΑΣΙΟΛΟΓΙΚΗ ΑΝΑΛΥΣΗ ΚΑΙ ΕΞΑΓΩΓΗ
ΠΛΗΡΟΦΟΡΙΑΣ, ΑΠΟ ΒΙΟΛΟΓΙΚΑ ΜΕΤΑ-ΔΕΔΟΜΕΝΑ ΥΨΗΛΗΣ
ΔΙΑΣΤΑΣΙΜΟΤΗΤΑΣ»**

Και ο Αγγλικός ως εξής:

**« DEVELOPMENT OF COMPUTATIONAL METHODS FOR THE SEMANTIC
INTERPRETATION, COMPARATIVE ANALYSIS AND VISUALIZATION OF OMICS
DATA »**

Ο Κοσμητορας της Σχολής



**Ν. Μαρμαράς
Καθηγητής Ε.Μ.Π**

- Για οδηγίες για την πρόσβαση σας διαδικτυακά απευθυνθείτε στον Επιβλέποντα του Υ.Δ. Αναπλ. Καθ. Λ. Αλεξόπουλο (leo@mail.ntua.gr)

EVELOPMENT OF COMPUTATIONAL METHODS FOR THE SEMANTIC INTERPRETATION, COMPARATIVE ANALYSIS AND VISUALIZATION OF OMICS DATA

DOCTORAL THESIS

THEODOROS G. KOUTSANDREAS
GRADUATE CHEMICAL ENGINEER NTUA

Supervisor: Ass. Prof. Leonidas Alexopoulos

Abstract

In this doctoral study, a novel computational approach was devised for the interpretation of various omics data, based on the semantic network analysis, exploiting the biomedical ontologies. Biomedical ontologies constitute controlled vocabularies, which are used to organise the existing knowledge in a specific biological domain and annotate the biomolecules (mainly genes and proteins), providing a multi-faceted semantic description for each one. A novel software library was constructed in order to implement the above computational approach. The library contains versatile classes and configured workflows for the integration, processing and manipulation of biomedical ontologies, as well as the topological analysis on their graphs. Within the frames of that library, an unsupervised, automated, analytical workflow was developed for the semantic interpretation of omics data, which combines the execution of pathway analysis and gene prioritization, using the annotation and structure of biomedical ontologies. Its final goal is to transform the distribution of signals of a high-throughput experiment into a semantic network and detect the pivotal regulatory biomolecules into it. The workflow, named BioInfoMiner, integrates algorithmic modules that have been developed as separate tools in the past. These tools were redesigned so as to be optimally integrated in a unified analytical pipeline, operating in virtual machines of cloud technologies. Besides, a novel workflow was constructed for the interactive visualization of the results in front-end interfaces. An instantiation of BioInfoMiner was integrated in a Galaxy-based platform for the analysis of metagenomic data. Additional modules were developed for the comparative analysis of semantic networks, while a qualitative evaluation of semantic similarity measures was performed. Within the frames of this thesis, BioInfoMiner and modules of the software library were used in three different projects. The first project was related to the semantic interpretation of transcriptomic data, derived from the study of the role of TRAIL in the activation of NK cells, during viral infection. In the second, the semantic profile of proteostasis machinery for thousands of species was calculated, targeting to evaluate its evolutionary imprinting. Finally, a study of the SARS-CoV-2 host interactome was performed, in order to pinpoint basic motifs of its pathogenic course and elucidate its correlation with other interactomes of pathogenic viruses. All these studies led to novel findings of high merit, supporting the efficiency of BioInfoMiner, regarding the reduction of the dimensionality and complexity of various omics datasets. They also confirmed that such a system-level semantic approach may reveal the nodal cellular events and regulatory biomolecules in a phenotypic condition under study and extract critical, novel biological information.

ΑΝΑΠΤΥΞΗ ΥΠΟΛΟΓΙΣΤΙΚΩΝ ΜΕΘΟΔΩΝ ΓΙΑ ΤΗ ΣΗΜΑΣΙΟΛΟΓΙΚΗ ΕΡΜΗΝΕΙΑ, ΣΥΓΚΡΙΤΙΚΗ ΑΝΑΛΥΣΗ ΚΑΙ ΑΠΕΙΚΟΝΙΣΗ ΟΜΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

ΘΕΟΔΩΡΟΣ Γ. ΚΟΥΤΣΑΝΔΡΕΑΣ
ΔΙΠΛΩΜΑΤΟΥΧΟΣ ΧΗΜΙΚΟΣ ΜΗΧΑΝΙΚΟΣ ΕΜΠ

Επιβλέπων: Αν. Καθ. Λεωνίδας Αλεξόπουλος

Περίληψη

Στην παρούσα διατριβή, αναπτύχθηκε μια νέα υπολογιστική προσέγγιση, βασισμένη στην ιδέα της ανάλυσης σημασιολογικών δικτύων, με σκοπό την ερμηνεία ομικών δεδομένων, χρησιμοποιώντας τις βιοϊατρικές οντολογίες. Οι βιοϊατρικές οντολογίες αποτελούν ελεγχόμενα λεξικά οργάνωσης της υπάρχουσας γνώσης σε κλάδους της Βιολογίας (π.χ. ασθένειες, μοριακές λειτουργίες, μεταβολικά μονοπάτια) και χρησιμοποιούνται για το σχολιασμό των βιομορίων (κυρίως γονιδίων και πρωτεϊνών), παρέχοντας μια πολύπλευρη σημασιολογική περιγραφή για το κάθε ένα. Για την υλοποίησή της συγκεκριμένης προσέγγισης, κατασκευάστηκε μια νέα βιβλιοθήκη λογισμικού. Η βιβλιοθήκη περιέχει συναρτήσεις, κλάσεις και ροές εργασιών για την ενσωμάτωση, επεξεργασία και χρήση των οντολογιών, καθώς και την τοπολογική ανάλυση στους γράφους αυτών. Για την ερμηνεία των ομικών δεδομένων κατασκευάστηκε μια μη εποπτευόμενη, αυτοματοποιημένη, αναλυτική ροή εργασιών, η οποία συνδυάζει τις μεθόδους της ανάλυσης μονοπατιών και της ιεράρχησης γονιδίων, αξιοποιώντας το σχολιασμό και τη δομή των βιοϊατρικών οντολογιών. Σκοπός της είναι η μετατροπή της κατανομής των σημάτων ενός πειράματος υψηλής απόδοσης σε δίκτυο σημασιολογικών όρων και ο εντοπισμός των κεντρικών ρυθμιστικών βιομορίων πάνω σε αυτό. Αυτή η ροή εργασιών ονομάζεται BioInfoMiner και ενσωματώνει αλγοριθμικές ενότητες που έχουν αναπτυχθεί ως ξεχωριστά εργαλεία στο παρελθόν. Τα εργαλεία αυτά επανασχεδιάστηκαν με σκοπό να ενσωματωθούν σε ένα ενοποιημένο υπολογιστικό εργαλείο, ικανό να λειτουργήσει σε συστήματα τεχνολογίας νέφους. Επιπλέον, δημιουργήθηκε μια ροή εργασιών διαδραστικής απεικόνισης των αποτελεσμάτων σε περιβάλλον διεπαφής. Μια έκδοση του BioInfoMiner ενσωματώθηκε σε διαδικτυακή εφαρμογή, βασισμένη στο υπολογιστικό σύστημα Galaxy, για την ανάλυση μεταγενεωμικών δεδομένων. Για τη συγκριτική ανάλυση σημασιολογικών δικτύων αναπτύχθηκαν κατάλληλες παραμετροποιήσιμες συναρτήσεις, ενώ πραγματοποιήθηκε ποιοτική σύγκριση των υπαρχόντων σημασιολογικών μέτρων. Στα πλαίσια της παρούσας διατριβής, ο BioInfoMiner και η εν λόγω βιβλιοθήκη χρησιμοποιήθηκαν σε τρεις μελέτες. Η πρώτη αφορούσε στην ερμηνεία μεταγραφωμικών δεδομένων, σχετικά με τη μελέτη του ρόλου της πρωτεΐνης TRAIL στη λειτουργία των NK κυττάρων σε συνθήκες ιογενούς λοίμωξης. Στη δεύτερη μελέτη, κατασκευάστηκε το σημασιολογικό προφίλ του μηχανισμού της πρωτεόστασης για εκατοντάδες οργανισμούς, με στόχο την αξιολόγηση της εξελικτικής του αποτύπωσης. Τέλος μελετήθηκε το πρωτεϊνικό δίκτυο που αλληλεπιδρά με τις πρωτεΐνες του ιού SARS-CoV-2, ώστε να εντοπιστούν τα βασικά μοτίβα της παθογόνου δράσης του και η συσχέτιση του με αντίστοιχα

δίκτυα άλλων παθογόνων ιών. Οι παραπάνω μελέτες οδήγησαν σε νέα ευρήματα αξίας, υποστηρίζοντας την αποτελεσματικότητα του BioInfoMiner, σχετικά με τη μείωση της διαστασιμότητας και πολυπλοκότητας των συνόλων ομικών δεδομένων. Επίσης, επιβεβαίωσαν ότι η εν λόγω συστημική σημασιολογική προσέγγιση μπορεί να αναδείξει τα κομβικά κυτταρικά γεγονότα και ρυθμιστικά βιομόρια σε μια υπό μελέτη φαινοτυπική συνθήκη και να εξαγάγει κρίσιμη, νέα βιολογική πληροφορία.