

Θέματα μεταπτυχιακών & διπλωματικών εργασιών για το ακ. έτος 2026- 2027

Π. ΒΑΣΙΛΕΙΑΔΗΣ – 2026-06-30

Τα θέματα διπλωματικών και μεταπτυχιακών εργασιών συνήθως επεκτείνουν κάποια από τα υπάρχοντα εργαλεία που αναπτύσσουμε στην ομάδα μου – λίγο πιο σπάνια, ξεκινούν νέα εργαλεία.

<https://github.com/DAINTINESS-Group>

https://www.youtube.com/playlist?list=PL3G-N7ZzyDfpsMLCQcm_L9KEVLDkC454

Το παρόν κείμενο περιγράφει τα προτεινόμενα θέματα και, για να διευκολύνει την κατανόηση, δίνει και μια γενική εικόνα για τα εργαλεία που αναπτύσσουμε.

Αναγκαστικά μια εργασία πρέπει να ολοκληρωθεί αυστηρά εντός ενός έτους από την ανάληψή της. Απαιτούμενα προσόντα είναι η πολύ καλή γνώση Java & σχεδίασης ΟΟ λογισμικού. Στις περισσότερες διπλωματικές χρειάζεται να μπορείτε να συνδυάσετε frameworks — e.g., Apache Spark, και ενίοτε να πρέπει να δουλέψετε με το συνδυασμό κώδικα και βάσεων δεδομένων. **Ειδικά για το 2026-27 είμαστε στη ζώνη του front-end: τα εργαλεία θέλουμε να αποκτήσουν εύχρηστη διαπροσωπεία.**

Γενικά προχωράμε ως ομάδα με agile working methods, ήτοι **weekly runs & group meetings**. Κείμενο και κώδικας γράφονται εναλλάξ. Παντού δουλεύουμε github && εκτός ειδικών περιπτώσεων, Eclipse.

Επειδή πλέον οι καιροί είναι πονηροί, **δεν θα δεχτώ να εξετάσω διπλωματικές από φοιτητές, που δεν βλέπω στο εξάμηνο και μου παρουσιάζουν ξαφνικά ένα κείμενο και ένα κώδικα προς εξέταση στο τέλος**. Θα πρέπει να ασχολείστε συστηματικά με την εργασία σας στη διάρκεια του εξαμήνου.

Θα λείψω για 3-4 μέρες στο ελληνικό συμπόσιο διαχείρισης δεδομένων. Άρα με υπολογίζετε να είμαι πίσω στο γραφείο μετά τις 5 Ιουλίου. Είναι πιθανόν στη διάρκεια του συνεδρίου να προκύψουν και νέες ιδέες. Αν υπάρξουν προσθήκες στα θέματα, θα τις δείτε αμέσως μετά την επιστροφή μου.

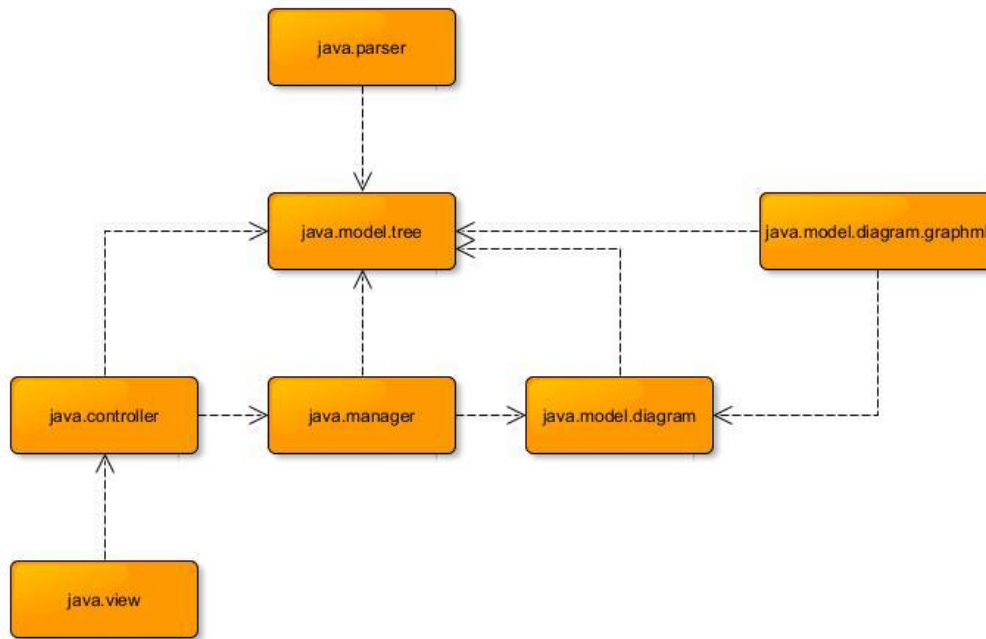
Contents

1	UML diagramming.....	2
2	Making Data Available.....	4
2.1	Data Cleaners.....	4
2.2	Πυθία: πες μου όλα όσα μπορείς για ένα data set.....	5
2.2.1	Βελτίωση της γραφικής διεπαφής.....	5
2.2.2	Προσθήκη highlights in detail.....	5

1 UML diagramming

Το ObjectAid έπαψε να υποστηρίζεται. Χρειαζόμαστε ένα εργαλείο που να κάνει reverse engineer ένα Java project και να μπορούμε να βγάλουμε UML Diagrams. Το Object-Oriented Architecture Diagrammer είναι ένα δικό μας ObjectAid για UML διαγράμματα

<https://github.com/DAINTINESS-Group/ObjectOrientedArchitectureDiagrammer>

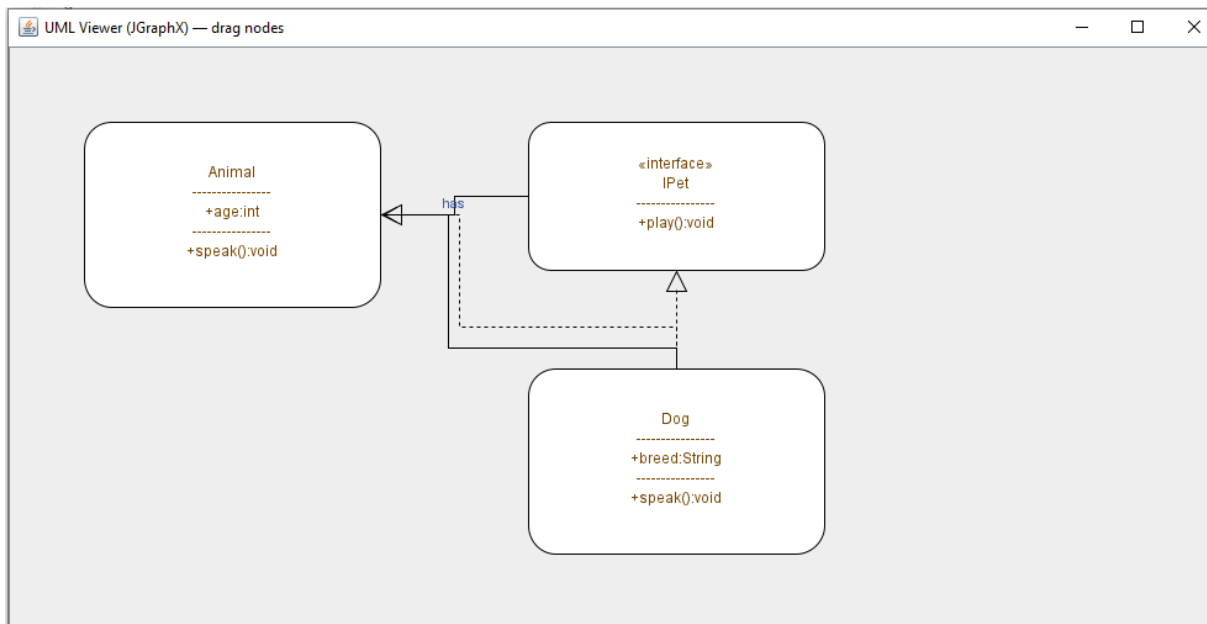


Ένα βασικό χαρακτηριστικό είναι ότι καταφέραμε να αξιοποιήσουμε parsers που επεξεργάζονται ένα java project, βγάζουν ένα AST δέντρο και από κει προκύπτει το γράφημα πακέτων, κλάσεων, και λοιπών δομικών συστατικών του source code.

Ζητούμενο: να φτιάξουμε ένα πιο lightweight εργαλείο, που να χρησιμοποιήσει αυτούσιο του parsing & domain model του OOAD και να επιτρέπει την αλληλεπίδραση με το χρήστη με (α) απλότητα, και (β) υψηλή αισθητική, σαν το ObjectAid.

Λύση 1: JGraphX μια απλή βιβλιοθήκη για κατασκευή boxes & lines

```
<dependency>
  <groupId>com.github.vlsi.mxgraph</groupId>
  <artifactId>jgraphx</artifactId>
  <version>4.2.2</version>
</dependency>
```



A simple diagram with JgraphX (100 lines of java code)

Λύση 2: plain JavaFX (the same code takes 180 lines)

Μια πρώτη προσπάθεια ξεκίνησε από τον κ. Βασίλειο Ζαφείρη (νέο μέλος ΕΔΙΠ του Τμήματος) που συμμετέχει ενεργά στην προσπάθεια: δείτε το branch draw-packages <bzafiris@uoi.gr>

<https://github.com/DAINTINESS-Group/ObjectOrientedArchitectureDiagrammer/tree/features/draw-packages>

Ενδεικτικά ανοιχτά ζητήματα:

- εμφάνιση και σωστή απεικόνιση όλων των σχέσεων μεταξύ κλάσεων
- drag n' drop για προσθήκη στοιχείων στα διαγράμματα
- φιλτράρισμα αρχείων στην πλευρική μπάρα αριστερά και εμφάνιση μόνο όσων σχετίζονται με java κώδικα
- καλύτερη εμφάνιση κλάσεων,
- υποστήριξη πολλών ανοικτών διαγραμμάτων

Δεν σημαίνει ότι θα κάνουμε τα πάντα, αλλά πάντως θα πρέπει η ευχρηστία και η αισθητική να βελτιωθούν.

Η έμφαση θα δοθεί στην ορθότητα του εργαλείου, την ευχρηστία και την (πολύ καλύτερη απ' ό,τι βλέπετε εδώ) αισθητική

2 Making Data Available

2.1 Data Cleaners

Για το υπάρχον εργαλείο **Data Cleaners** [<https://github.com/DAINTINESS-Group/DataCleaners>] θέλουμε να επεκτείνουμε τις δυνατότητές του ως εξής:

1. Βελτιωμένη αλληλεπίδραση με τον χρήστη μέσω ενός **απλού αλλά ολοκληρωμένου GUI (ideally plain swing, or in extremis, JavaFX)**

- Διαδραστική καταχώριση συνόλων δεδομένων, και διαδραστικός προσδιορισμός των εργασιών καθαρισμού δεδομένων που πρέπει να εκτελεστούν (ενδεχομένως και καθοδηγούμενος σε βήματα αν προσδιορίσουμε μια μέθοδο)
- Βελτιωμένη διαδραστική, οπτικοποιημένη αναφορά σφαλμάτων ανά είδος ελέγχου / σφαλμάτων ανά γραμμή του αρχείου / ...
- Ιδεατά, κάποιου είδους tracing + visualization της αλληλουχίας των παραγόμενων states εντός της ροής καθαρισμού δεδομένων.

Η επέκταση του λογισμικού μπορεί να γίνει είτε με νέο πακέτο για το front end, είτε με χωριστή μηχανή που χρησιμοποιεί το data cleaners ως μαύρο κουτί αν αυτό είναι εφικτό.

2. Παραγωγή offline reports προσθέτοντας τις αναφορές τύπου md (που λείπουν, τώρα έχει μόνο txt).

Για το μέλλον, επόμενη διπλωματική, αφού έχει ολοκληρωθεί το front end:

3. Πρόσθεση μετασχηματισμών, είτε στο υπάρχον εργαλείο, είτε χωριστά, π.χ.,

- Απάλειψη / ανωνυμοποίηση ευαίσθητων στηλών
- Υπολογισμός νέων στηλών
- Μετατροπή των ημερομηνιών σε μια συγκεκριμένη μορφή
- Μετατροπή των κενών τιμών, σε μια προεπιλεγμένη τιμή
- Γενικά: δίνοντας μια συνθήκη, π.χ., μια κανονική έκφραση, σε μια στήλη, αντικαταστήστε τις τιμές του αναλόγως (π.χ. εάν αριθμοί float προέρχονται από ένα csv της ΕΕ και θέλουμε ένα αρχείο csv ΗΠΑ, αντικαταστήστε όλα τα "," με ".")
- Καταχώριση δεδομένων (α) χειροκίνητη, (β) [αν είναι δυνατόν] ημιαυτόματη

Η εκτέλεση της ακολουθίας των μετασχηματισμών καταγράφεται κατάλληλα σε ένα δομημένο log file. Επίσης κρατιούνται και οι ενδιάμεσες εκδόσεις των αρχείων, όπως αυτά μετασχηματίζονται.

Η επέκταση του λογισμικού με την προσθήκη μετασχηματισμών μπορεί να είναι

- i) επέκταση της τρέχουσας DataCleanerFacade κεντρικής μηχανής, ή,
- ii) μια δεύτερη μηχανή (σε αυτή την αρχιτεκτονική, η υπάρχουσα engine είναι ο ελεγκτής ποιότητας και η νέα είναι ο μετασχηματιστής/καθαριστής)

Μπορείτε να στήσετε το Data Cleaners και να παίξετε με αυτό να δείτε πώς δουλεύει. Μπορείτε επίσης να παίξετε λίγο με ένα απλό εργαλείο data analytics, όπως π.χ., το Orange για να δείτε μια ιδέα για το πώς ο χρήστης διαδραστικά μπορεί να κάνει κάποια από αυτά τα tasks. Στη σχετική βιβλιογραφία στο φάκελο **extra material**/ έχει και βιβλιογραφία για το πώς θα έπρεπε να είναι το user experience σε ένα τέτοιο εργαλείο, καθώς και για το τι είδους ελέγχους κάνουμε στα δεδομένα.

ΕΠΙΠΕΔΟ: **Διπλωματική για Μηχανικούς**

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, Apache Spark

ΠΡΟΚΛΗΣΕΙΣ και ΟΦΕΛΗ: Η δυσκολία βρίσκεται κυρίως στη σωστή και επεκτάσιμη σχεδίαση του λογισμικού (ως συνήθως: back-end façade, strategy for connectors/algorithms, front-end controller). Τα οφέλη για ένα φοιτητή είναι: (α) τεχνογνωσία σε θέματα ποιότητας δεδομένων και (β) hands-on σε ένα ευμέγεθες κομμάτι λογισμικού, with emphasis to the front end this year. Πρέπει να μπορείτε να ανταπεξέλθετε και στη σχεδίαση λογισμικού και στην ανάπτυξη του σχετικού συστήματος.

2.2 Πυθία: πες μου όλα όσα μπορείς για ένα data set

ΠΕΡΙΛΗΨΗ: Το σύστημα Pythia κάνει ένα αυτόματο profiling ενός συνόλου δεδομένων και παραγωγή του σχετικού report <https://github.com/DAINTINESS-Group/Pythia>

ΣΥΝΤΟΜΗ ΠΕΡΙΓΡΑΦΗ: Ένα τυπικό data set δεν είναι παρά ένα αρχείο κειμένου, με γραμμές, όπου κάθε γραμμή είναι μια εγγραφή, τα πεδία της οποίας χωρίζονται με κάποιο διαχωριστικό (p.x., tabs, comma, pipe, ...). Θέλουμε, αφού εγγράψουμε ένα data set, την 100% αυτόματη παραγωγή ενός στατιστικού προφίλ για το data set, μέσω ενός οργανωμένου μοντέλου για τα highlights που κρύβονται μέσα.

Το σύστημα Pythia, που κάνει αυτόματα ένα data profile για ένα αρχείο με δεδομένα, αναπτύσσεται από το εργαστήριο εδώ και λίγο καιρό. Η διπλωματική που προτείνεται εδώ συνεχίζει δουλειά που φτιάχτηκε σε προηγούμενες διπλωματικές. Για να παραχθεί ένα dataset profile η ΠΥΘΙΑ είναι εις θέση να εκτελέσει ένα μεγάλο αριθμό από data profiling tasks: histograms, descriptive statistics, correlations, clusterings, outlier detections, regressions, ...

Μπορείτε να στήσετε την πυθία και να παίξετε με αυτή να δείτε πώς δουλεύει. Μπορείτε επίσης να παίξετε λίγο με ένα απλό εργαλείο data analytics, όπως π.χ., το Orange για να δείτε μια ιδέα για το πώς ο χρήστης διαδραστικά μπορεί να κάνει κάποια από αυτά τα tasks. Στη σχετική βιβλιογραφία στο φάκελο `extra material/` έχει και βιβλιογραφία για το πώς θα έπρεπε να είναι το user experience σε ένα τέτοιο εργαλείο.

Στη συνέχεια παραθέτω **δύο, διαφορετικές μεταξύ τους, διπλωματικές.**

2.2.1 Βελτίωση της γραφικής διεπαφής

ΚΕΝΤΡΙΚΗ ΙΔΕΑ: Προς το παρόν, η χρήση της ΠΥΘΙΑΣ γίνεται (α) με cli calls, δλδ για κάθε dataset γράφουμε κι ένα απλό client που κάνει όποιους ελέγχους θέλουμε να γίνουν και (β) με ένα πρωτόλειο swing GUI το οποίο επιδέχεται απλοποίησης και βελτίωσης στο user experience.

Η αναμόρφωση / αναδόμηση της διαπροσωπείας πρέπει να επιτρέπει στον αναλυτή εύκολα να διαχειρίζεται τους ελέγχους και να μπορεί να έχει preview των αποτελεσμάτων. Ο στόχος είναι να γίνει το εργαλείο άμεσα εύχρηστο, όπως το Orange για παράδειγμα και να μπορεί να δίνει στο χρήστη την ευκολία να ορίσει τι θα κάνει και προεπισκοπήσεις του τι παράγεται. Αυτό αφορά και την αναθεώρηση των use cases και την αναδόμηση του gui package εσωτερικά. Κατά προτίμηση δε θέλουμε να φύγουμε από το απλό swing, αλλά αν χρειαστεί, μπορούμε να εμπλέξουμε και κάτι πιο σύνθετο, π.χ. javafx.

2.2.2 Προσθήκη highlights in detail

ΚΕΝΤΡΙΚΗ ΙΔΕΑ: Θέλουμε να ενσωματώσουμε στο σύστημα τη δυνατότητα αναπαράστασης σημαντικών highlights από τα ευρήματα του εκάστοτε data profiling task. Υπάρχει ήδη μια πρωτόλεια υλοποίηση του άρθρου

https://www.cs.uoi.gr/~pvassil/publications/2025_ERForum/

την οποία και θέλουμε να υλοποιήσουμε στο 100% και να τη συνδέσουμε με όλα τα tasks data profiling που κάνει η ΠΥΘΙΑ. Για όλα τα αυτομάτως παραχθέντα tasks το σύστημα θα πρέπει να κάνει:

- Αναπαράσταση των ενδιαμέσων αποτελεσμάτων μέσω των αντίστοιχων κλάσεων && Καταγραφή ενός αναλυτικού report με όλα τα ευρήματα

- Rank and prune of the extracted highlights
- Καταγραφή ενός συνοπτικού report / επέκταση υπάρχοντων reports με τα όλα ευρήματα που έχουν τουλάχιστον ένα ελάχιστο threshold σημαντικότητας (τώρα έχουμε ένα report ανά κατηγορία ευρημάτων)

ΕΠΙΠΕΔΟ: *Διπλωματική για Μηχανικούς*

ΠΛΑΤΦΟΡΜΑ ΕΡΓΑΣΙΑΣ: Java, Apache Spark

ΠΡΟΚΛΗΣΕΙΣ και ΟΦΕΛΗ: Η δυσκολία βρίσκεται κυρίως στη σωστή και επεκτάσιμη σχεδίαση του λογισμικού. Τα οφέλη για ένα φοιτητή είναι: (α) τεχνογνωσία σε θέματα αναλυτικής δεδομένων και (β) hands-on σε ένα ευμέγεθες κομμάτι λογισμικού. Η εργασία είναι πλέον κατάλληλη για φοιτητές με ταλέντο σε προγραμματιστικά θέματα και σε θέματα αναλυτικής δεδομένων. Πρέπει να μπορείτε να ανταπεξέλθετε και στη σχεδίαση λογισμικού και στην ανάπτυξη του σχετικού συστήματος.