



Data Science Lab. @ University of Piraeus

Προτεινόμενες πτυχιακές & διπλωματικές εργασίες

Data Science Lab (www.datastories.org)

Ακαδημαϊκό έτος 2023-24

[μετά τις προτεινόμενες εργασίες παρατίθενται σχετικές οδηγίες εκπόνησης]

(A) Mobility data management and analytics in the urban / maritime / aviation domain

2023.A1: Vehicle trajectory forecasting. Forecasting vehicle locations is valuable for improving road safety and designing intelligent driving systems. In this thesis, we will design and implement a neural network model in order to predict the next instance position of a vehicle based on its present and previous motion history. **Bib:**¹. **Contact:** Dr. Eva Chondrodima (evachon@unipi.gr)

2023.A2-3 (two theses jointly supervised): UniPi-AIS platform (frontend/backend). (1) The web application of the AIS antenna hosted by the Univ. Piraeus performs real-time visualization of vessels' movement within Saronic Gulf, Piraeus. In this thesis, we aim to extend the Visual Analytics (VA) functionality of the application including advanced analytics operations, such as co-movement pattern mining, route forecasting, and ETA Prediction. (2) The backend architecture of the AIS antenna consists of a database, in which the decoded AIS locations are saved and queried at periodic intervals to visualize the vessels' latest known position within the past hour. In this thesis, we aim to modernize the aforementioned workflow by using publish-subscribe frameworks (e.g., Apache Kafka) in order to facilitate data management and processing and potentially allow for higher throughput on the front-end side. **Bib:**². **Contact:** Andreas Tritsarolis (andrewt@unipi.gr); Prof. Yannis Theodoridis (ytheod@unipi.gr)

2023.A4: A realistic simulator for effectively generating synthetic trajectories of moving objects in 3D space. At the core of the mobility data mining field lies the discovery of intuitive patterns representing collective behaviors of trajectories of moving objects, information which is potentially useful for a variety of mobility data analytics tasks (e.g., semantic trajectory

¹ <https://doi.org/10.1109/TITS.2019.2899179>

² <https://www.datastories.org/unipi-ais>; <https://doi.org/10.1109/MDM52706.2021.00048>;
<https://doi.org/10.1080/13658816.2020.1834562>; <https://doi.org/10.1145/3469830.3470909>;
<http://doi.org/10.1109/MDM55031.2022.00056>; <https://arxiv.org/abs/1810.05567>;
<https://www.oreilly.com/library/view/kafka-the-definitive/9781492043072/>

prediction). In this context, an essential requirement is the exploitation of real-world datasets to support scalability experiments and evaluate the various proposals, accordingly. Data availability is not taken for granted in many cases, such as the case of the aviation domain. Real-world aircraft trajectory datasets are often unavailable due to privacy issues, among other reasons. In addition, even when datasets are available, these are usually small, while the ground truth for the underlying mobility patterns is absent. This thesis will aim to develop a realistic simulator to produce mobility pattern-driven trajectories of aircrafts, capable of supporting various types of mobility patterns, given the available ground truth information (e.g., flight plans). **Bib**³. **Contact**: Yannis Kontoulis (ikontoulis@unipi.gr); Prof. Nikos Pelekis (npelekis@unipi.gr)

2023.A5: Aircraft trajectory forecasting. Nowadays, large amounts of aircraft data are widely available due to recent advances in position broadcasting technology. In this thesis, we will design and implement a neural network model that learns from historical position data in order to predict future aircraft trajectories of streaming data. **Bib**:⁴. **Contact**: Yannis Kontoulis (ikontoulis@unipi.gr); Dr. Eva Chondrodima (evachon@unipi.gr)

(group B) Mobility / Big spatial data management and analytics [domain agnostic]

2023.B1: ST_Visions - A Python library for interactive spatio-temporal data visualization. Modern applications generate massive volumes of spatio-temporal trajectory data on daily basis that need to be collected, processed, and analysed, in order to extract useful knowledge in terms of mobility patterns. The spectrum of applications includes fleet monitoring systems, vessel and aircraft tracking services, ride-sharing apps, traffic control management, and so on. In all these domains, a common need is to provide easy support for visualizations to quickly perform Visual Analytics (VA), so that data analysts and domain experts can easily obtain an overview. In this context, we have developed ST_Visions, a Python Library that offers advanced interactive visualizations with only a few lines of code. In this thesis, we aim to extend ST_Visions with more advanced functionality, such as support for streaming data and online analytics. **Bib**:⁵. **Contact**: Andreas Tritsarolis (andrewt@unipi.gr); Prof. Yannis Theodoridis (ytheod@unipi.gr)

2023.B2: Simulating trajectory datasets with machine learning methods. The goal of the thesis is to use various machine learning methods in order to develop a trajectory data generation, simulating the digital twin of a real dataset. **Bib**:⁶. **Contact**: Prof. Nikos Pelekis (npelekis@unipi.gr)

2023.B3: Machine Learning techniques meets big spatial data management. The goal of the thesis is to study and experimentally validate machine learning techniques for big spatial data management topics, e.g., learned indexes / histograms, etc.. **Bib**:⁷. **Contact**: Prof. Nikos Pelekis (npelekis@unipi.gr)

2023.B4: Outlier detection for big trajectory datasets. Lately, a number of robust outlier detection methods have been introduced for timeseries datasets. The goal of the thesis is either to

³ <https://doi.org/10.1145/2937753>

⁴ <https://doi.org/10.1007/s41060-019-00182-4>; <https://doi.org/10.1016/j.jss.2016.06.016>

⁵ <http://dx.doi.org/10.5311/JOSIS.2020.20.661>; <https://doi.org/10.1109/MDM52706.2021.00048>

⁶ <https://github.com/XingruiWang/Two-Stage-Gan-in-trajectory-generation>

⁷ http://people.csail.mit.edu/ibrahimsabek/pdf/21_tutorial_mdm.pdf

adapt such techniques for trajectory datasets or to develop a novel specialized algorithm for trajectory outlier detection. **Bib:**⁸. **Contact:** Prof. Nikos Pelekis (npelekis@unipi.gr)

(group C) Data series / Social analytics

2023.C1: Forecasting techniques in financial data. The goal of the thesis is to apply machine learning techniques in financial data and develop an application with real-time features that is able to forecast the ‘sell/buy’ orders in stock markets. **Bib:**⁹. **Contact:** Prof. Nikos Pelekis (npelekis@unipi.gr)

2023.C2: Clustering and topic detection from mentions. The goal of the thesis is to develop vectorization techniques of social data and design an appropriate vector-based clustering algorithm as the mean for topic detection. **Bib:**¹⁰. **Contact:** Prof. Nikos Pelekis (npelekis@unipi.gr)

2023.C3-4 (two theses jointly supervised): Bot/spam/fake news detection & virality prediction on social media. The goal of the thesis is to develop various analytics techniques regarding (1) bot/spam/fake news detection and (2) virality prediction on social media data. **Bib:**¹¹. **Contact:** Prof. Nikos Pelekis (npelekis@unipi.gr)

Οδηγίες εκπόνησης πτυχιακής / διπλωματικής εργασίας

Διαδικασία – Γλώσσα συγγραφής –Χρονοδιάγραμμα

Η εργασία ξεκινά με την ανάθεσή της από τον επιβλέποντα καθηγητή. Ακολουθούν αναλυτικές συζητήσεις με τον επιβλέποντα καθηγητή ή/και τον συμβουλευοντα μεταδιδακτορικό ερευνητή ή υποψήφιο διδάκτορα για να οριστικοποιηθεί το πλαίσιο και να λυθούν απορίες. Μετά μεσολαβεί το κύριο έργο της εκπόνησης της εργασίας, κατά τη διάρκεια του οποίου ετοιμάζονται και παραδίδονται 2 ενδιάμεσες εκθέσεις προόδου. Στο τέλος κατατίθεται ένα σχέδιο (draft) του τελικού τόμου και γίνεται μια πρώτη παρουσίαση (υπό μορφή «πρόβας»). Ακολουθεί η κατάθεση του τελικού τόμου και του συνοδευτικού υλικού και η τελική παρουσίαση-εξέταση. Πιο αναλυτικά:

Βήμα 1) 1η ενδιάμεση έκθεση προόδου: Το πρώτο βήμα είναι να συγγράψετε ένα κείμενο που αφορά στον προκαταρκτικό ορισμό του θέματος που έχετε αναλάβει, τους λόγους για τους οποίους θεωρείτε ότι έχει κάποια αξία το συγκεκριμένο θέμα (πρωτοτυπία, εφαρμογή κλπ.), και μια πρώτη επισκόπηση σχετικών εργασιών με τις οποίες θα μπορούσε να «συγκριθεί» ή στις οποίες θα «επενδύσει» η δική σας. Μέσα σε λίγες (~5) σελίδες, πρέπει να περιγράψετε (α) το κίνητρο (ποιο είναι το «πρόβλημα» που καλείστε να αντιμετωπίσετε και γιατί το συγκεκριμένο αξίζει να του αφιερώσετε κάποιο από το χρόνο

⁸ <https://arxiv.org/pdf/2209.04635.pdf>

⁹ <https://www.sciencedirect.com/science/article/abs/pii/S0925231220306718>

¹⁰ <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00585-4>

¹¹ https://kdd.org/exploration_files/19-1-Article2.pdf; <https://doi.org/10.1038/srep02522>

σας;), (β) τις σχετικές εργασίες (τι έχει γίνει στο χώρο μέχρι στιγμής;) και (γ) το αντικείμενο εργασίας (ποιο αναμένεται να είναι το περιεχόμενο της εργασίας σας; ποια εργαλεία (συστήματα / μεθόδους / μοντέλα) σκοπεύετε να χρησιμοποιήσετε; τι δεδομένα θα διαχειριστείτε; κλπ.). Η 1^η έκθεση προόδου δεν απαιτείται να ακολουθεί κάποιο συγκεκριμένο πρότυπο μορφοποίησης.

Βήμα 2) 2η ενδιάμεση έκθεση προόδου: Στην πορεία ανάπτυξης της εργασίας, θα πρέπει να καταγραφούν και να συμφωνηθούν από κοινού με τον επιβλέποντα οι λεπτομέρειες (σχεδιαστικές, υλοποίησης κλπ.) της εργασίας. Η 2η έκθεση προόδου θα περιλαμβάνει σε αρκετή λεπτομέρεια τα εξής:

- (α) Ανάλυση απαιτήσεων: Αν πρόκειται για θεωρητικό θέμα, τι περιμένουμε από την ανάπτυξη ή επέκταση μιας νέας θεωρίας ή μοντέλου. Ποιες δηλαδή είναι οι προδιαγραφές του τελικού μοντέλου μαζί με μια σκιαγράφιση της μεθόδου που θα χρησιμοποιηθεί για την ανάπτυξή του. Αν πρόκειται για ανάπτυξη συστήματος / εφαρμογής, ποιες είναι οι ανάγκες των χρηστών, τι λειτουργικότητα περιμένουμε να προσφέρει το σύστημα / εφαρμογή. Αν πρόκειται για εργασία εκτενούς ανασκόπησης, ποια είναι τα συστήματα / θεωρίες / μοντέλα που θα συγκριθούν και πάνω σε ποιος άξονες.
- (β) Σχεδιασμός: Αν πρόκειται για θεωρητικό θέμα, τα βήματα που πρέπει να πραγματοποιηθούν για την ανάπτυξη ή επέκταση μιας νέας θεωρίας ή μοντέλου. Τέτοια βήματα περιλαμβάνουν τον σχεδιασμό της γλώσσας, της σημασιολογίας, του συστήματος απόδειξης, των ιδιοτήτων, κλπ. Αν πρόκειται για ανάπτυξη συστήματος ή εφαρμογής, τις λειτουργικές ενότητες (modules) του συστήματος ή της εφαρμογής και τη ροή δεδομένων μεταξύ τους, σε τυποποιημένη μορφή (π.χ. UML). Αν πρόκειται για εργασία εκτενούς ανασκόπησης, προκαταρκτική περιγραφή του «περιβάλλοντος σύγκρισης» – benchmark – και των σχετικών πειραμάτων που θα γίνουν.
- (γ) Εφαρμογές – Αποτίμηση: Αν πρόκειται για θεωρητικό θέμα, θα περιγράφει εφαρμογές της θεωρίας/μοντέλου που πρόκειται να αναπτυχθεί σε συγκεκριμένα παραδείγματα. Θα πρέπει να προσδιοριστούν οι μετρήσεις απόδοσης που θα γίνουν, στατιστικά που θα μετρηθούν (ανάλογα με την εφαρμογή), κλπ. Αν πρόκειται για σύστημα / εφαρμογή, θα περιλαμβάνει πλήρη περιγραφή της εφαρμογής που θα αναπτυχθεί (case study). Αν πρόκειται για εργασία εκτενούς ανασκόπησης, θα προδιαγράφει τα πειράματα που πρόκειται να διεξαχθούν και σε τι αποσκοπούν.
- (δ) Συμπεράσματα: 1-2 παράγραφοι ως επίλογος της έκθεσης (συνοπτικά το περιεχόμενο που είχε η εργασία σας)
- (ε) Βιβλιογραφικές Αναφορές¹²
- (στ) Γλωσσάρι

¹² Όσον αφορά στις βιβλιογραφικές αναφορές, επισημαίνονται τα εξής: (1) Όταν περιγράφουμε μια τάση, μια άποψη, μια λύση κοκ. που δεν μας ανήκει (πνευματικά) πρέπει να το στηρίζουμε με τουλάχιστον μια αναφορά. Πιο γλαφυρά, θα λέγαμε ότι πάντα σκεφτόμαστε ότι υπάρχει ένας κριτής που διαβάσει το κείμενο και συνεχώς ρωτάει "και ποιος το λέει αυτό;". (2) Όταν θέλουμε να υποστηρίξουμε το κείμενό μας με κομμάτι εργασίας κάποιου τρίτου μπορούμε είτε να κάνουμε παράφραση του πρωτότυπου κειμένου ώστε να το περιγράψουμε με το δικό μας τρόπο, πάντοτε δηλώνοντας την πηγή, είτε να βάλουμε μέσα σε εισαγωγικά το πρωτότυπο κομμάτι κειμένου που μας ενδιαφέρει (αυτό καλό είναι να περιορίζεται σε λίγες γραμμές) δηλώνοντας πάλι την πηγή. Τέλος, (3) δεν επιτρέπεται για κανένα λόγο να αντιγράψουμε (ή να μεταφράσουμε – το ίδιο κάνει) ολόκληρα κείμενα και να τα ενσωματώνουμε αυτούσια στο δικό μας.

Το κείμενο αυτό αποτελεί ουσιαστικά τον «οδηγό» με βάση τον οποίο θα προχωρήσει η εργασία στους επιμέρους άξονες (ανάλυση, σχεδίαση, υλοποίηση, αποτίμηση) και προφανώς θα βοηθήσει κι αυτό στη συγγραφή του τελικού κειμένου της ΠΕ/ΜΔ αφού θα έχει καθορίσει σαφώς τον «πίνακα περιεχομένων» της.

Βήμα 3) Τελικός τόμος: Ο τελικός τόμος θα ακολουθεί σε γενικές γραμμές τη δομή της 2^{ης} έκθεσης, με τα τελικά πια αποτελέσματα. Η γλώσσα συγγραφής μπορεί να είναι η ελληνική ή η αγγλική ενώ η μορφοποίηση του τελικού τόμου θα ακολουθεί τις οδηγίες που περιγράφονται στα παρακάτω links:

- ΠΕ: <https://www.cs.unipi.gr/> > ΣΠΟΥΔΕΣ > Προπτυχιακές Σπουδές > Κανονισμός Εκπόνησης Πτυχιακής Εργασίας
- ΜΔ: <https://www.cs.unipi.gr/> > ΣΠΟΥΔΕΣ > Μεταπτυχιακές Σπουδές > Υποβολή Μεταπτυχιακής Διατριβής στη «ΔΙΩΝΗ»

Τέλος, το συνοδευτικό υλικό του τελικού τόμου περιλαμβάνει:

- τελικό κείμενο, σε πηγαία μορφή (Word ή LaTeX) και pdf,
- πηγαίος κώδικας και σύνολα δεδομένων που χρησιμοποιήθηκαν
- παρουσίαση, σε πηγαία μορφή (π.χ. Powerpoint) και pdf.

Τυπική **διάρκεια εκπόνησης** μιας ΠΕ (ΜΔ) είναι οι 12 (6, αντίστοιχα) μήνες. Σε κάθε περίπτωση, **ανώτατο όριο θεωρούνται οι 18 (12, αντίστοιχα) μήνες**. Σε περίπτωση που ξεπεραστεί αυτό το όριο, η εκπόνηση της εργασίας θεωρείται ότι έχει διακοπεί, με ευθύνη του φοιτητή.

Το παρόν κείμενο είναι επίσης διαθέσιμο στην ιστοσελίδα του Εργαστηρίου Επιστήμης Δεδομένων: <https://www.datastories.org> > Education > Diploma Theses.