

## Προτεινόμενα θέματα πτυχιακών & διπλωματικών εργασιών

Ακαδ. έτος 2024-25

Data Science Lab ([www.datastories.org](http://www.datastories.org))

[μετά τις προτεινόμενες εργασίες παρατίθενται σχετικές οδηγίες εκπόνησης]

### (A) Mobility data management and analytics in the maritime domain

**Contact:** Andreas Tritsarolis ([andrewt@unipi.gr](mailto:andrewt@unipi.gr)) (2024.A1-2-3); Prof. Nikos Pelekis ([npelekis@unipi.gr](mailto:npelekis@unipi.gr)) (2024.A4)

**2024.A1-2: 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:**<sup>1</sup> .

**2024.A3: 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

<sup>1</sup> <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/>

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:**<sup>2</sup>.

**2024.A4: On the evaluation of federated-learning-based vessel location forecasting methods.** This thesis will build on the work of Vessel Location Forecasting (VLF) using LSTM neural networks, focusing on a comprehensive evaluation of some key aspects. The research will assess the performance of both the FedNautilus (federated) learning approaches, examining factors such as model accuracy, training efficiency, scalability, hyperparameter tuning and privacy trade-offs. Additionally, the study will explore real-world applicability by analyzing the impact of fragmented data sources and the challenges of federated learning in maritime environments. The findings will provide deeper insights into optimizing VLF solutions for enhanced maritime awareness. **Bib:**<sup>3</sup>.

## (B) Machine learning over spatial-temporal datasets

**Contact:** Prof. Nikos Pelekis ([npelekis@unipi.gr](mailto:npelekis@unipi.gr))

**2024.B1: Simulating trajectory datasets with machine learning methods.** The goal of this 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:**<sup>4</sup>.

**2024.B2: Machine Learning techniques meets big spatial data management.** The goal of this thesis is to study and experimentally validate machine learning techniques for big spatial data management topics, e.g., learned indexes / histograms, etc.. **Bib:**<sup>5</sup>.

**2024.B3: Machine Learning on Large-Scale Spatio-Temporal-Textual Graph Data Using Neo4j DBMS.** The proliferation of spatio-temporal-textual trajectory data from devices and applications has resulted in massive graph datasets that capture the movements and interactions of entities over time and space. Modeling this data within Neo4j DBMS allows for efficient storage and complex querying but poses challenges when applying machine learning due to scale and complexity. The goal of this thesis is to develop and implement GNN models that can interface with Neo4j to learn from the graph data and explore Neo4j's Graph Data Science library to perform tasks like node classification, link prediction, and community detection on spatio-temporal-textual data. **Bib:**<sup>6</sup>.

**2024.B4: Evaluation of the energy consumption aspects of trajectory-related AI algorithms.** This thesis will explore the energy efficiency of trajectory-related AI algorithms, a critical factor as AI systems are increasingly deployed in energy-constrained environments such as mobile devices, autonomous vehicles, and edge computing platforms. The primary objective is to

---

<sup>2</sup> <http://dx.doi.org/10.5311/JOSIS.2020.20.661>; <https://doi.org/10.1109/MDM52706.2021.00048>

<sup>3</sup> <https://www.computer.org/csdl/proceedings-article/mdm/2024/745500a083/1YEw7B5XSzm>

<sup>4</sup> <https://github.com/XingruiWang/Two-Stage-Gan-in-trajectory-generation>

<sup>5</sup> [http://people.csail.mit.edu/ibrahimsabek/pdf/21\\_tutorial\\_mdm.pdf](http://people.csail.mit.edu/ibrahimsabek/pdf/21_tutorial_mdm.pdf)

<sup>6</sup> [https://papers.nips.cc/paper\\_files/paper/2017/file/5dd9db5e033da9c6fb5ba83c7a7e9-9-Paper.pdf](https://papers.nips.cc/paper_files/paper/2017/file/5dd9db5e033da9c6fb5ba83c7a7e9-9-Paper.pdf)

evaluate the energy consumption of these algorithms when applied in different settings, such as real-time processing, large-scale datasets, and complex environments. **Bib:**<sup>7</sup> .

**2024.B5: Advanced Visualization of Spatio-Temporal Graph Data Using Neo4j DBMS.** The goal of this thesis is to develop vectorization techniques of social data and design an appropriate vector-based clustering algorithm as the mean for topic detection. Visualizing large-scale spatio-temporal graph data stored in Neo4j DBMS is crucial for data analysis and knowledge discovery but poses challenges due to data complexity and volume. This thesis aims to develop advanced visualization techniques integrated with Neo4j DBMS to enable effective exploration of big graph data. **Bib:**<sup>8</sup> .

---

## Οδηγίες εκπόνησης πτυχιακής / διπλωματικής εργασίας

### Διαδικασία – Γλώσσα συγγραφής –Χρονοδιάγραμμα

Η εργασία ξεκινά με την ανάθεσή της από τον επιβλέποντα καθηγητή (είτε τον καθ. Ι. Θεοδωρίδη είτε τον καθ. Ν. Πελέκη). Ακολουθούν αναλυτικές συζητήσεις με τον επιβλέποντα καθηγητή ή/και τον συμβουλευόντα 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 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<sup>η</sup> 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

---

<sup>7</sup> <https://asu.elsevierpure.com/en/publications/edge-intelligence-paving-the-last-mile-of-artificial-intelligence>

<sup>8</sup> <https://ieeexplore.ieee.org/document/https://ieeexplore.ieee.org/document/75269087526908>

(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 περιγράψει 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 προδιαγράψει 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 Αναφορές<sup>9</sup>

(στ) Γλωσσάρι

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

---

<sup>9</sup>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 διαβάζει 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 / μεταφράσουμε ολόκληρα κείμενα, είτε από άλλες εργασίες είτε από εργαλεία τύπου ChatGPT, και να τα ενσωματώνουμε στο δικό μας.

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<sup>ης</sup> 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: (α) τελικό κείμενο, σε πηγαία μορφή (Word ή LaTeX) και pdf, (β) πηγαίο κώδικα και σύνολα δεδομένων που χρησιμοποιήθηκαν και (γ) παρουσίαση, σε πηγαία μορφή (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.