



Data Science Lab. @ University of Piraeus

Προτεινόμενες πτυχιακές & διπλωματικές εργασίες, 2022-23

Data Science Lab (www.datastories.org)

Οκτώβριος 2022

[μετά τις προτεινόμενες εργασίες παρατίθενται σχετικές οδηγίες εκπόνησης]

2022.01: TraJSuite – a suite of distributed trajectory data processing algorithms in Java. During the past years, the Data Science Lab has developed a few open-source projects related to distributed (either offline or online) streaming trajectory data processing algorithms in Java, including join and clustering operations. In this thesis, we aim to integrate this code under a uniform tool, called Trajectory Java Suite (TrajJSuite). **Bib:**¹. **Contact:** Prof. Yannis Theodoridis (ytheod@unipi.gr)

2022.02: Online network-aware trajectory clustering. Nowadays, large amounts of tracking data are generated via location enabled devices and other advanced tracking technologies. In this thesis, we will design and implement an online and distributed clustering algorithm for such streaming trajectory data moving on urban networks. **Bib:**². **Contact:** Prof. Nikos Pelekis (npelekis@unipi.gr)

2022.03: Application of machine learning techniques in big spatial data. The goal of the thesis is to study and experimentally validate machine learning techniques in big spatial data. **Bib:**³. **Contact:** Prof. Nikos Pelekis (npelekis@unipi.gr)

2022.04: 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 adapt such techniques for trajectory datasets or to develop a novel specialized algorithm for trajectory outlier detection. **Bib:**⁴. **Contact:** Prof. Nikos Pelekis (npelekis@unipi.gr)

2022.05: ETA prediction on road networks. The goal of the thesis is to design and implement a big-data-ready algorithm that will be able to predict the Estimated Time of Arrival (ETA) of objects moving on road networks. **Bib:**⁵. **Contact:** Prof. Nikos Pelekis (npelekis@unipi.gr)

2022.06: Comparative study of off-the-shelf big data clustering algorithms for trajectory clustering. The goal of the thesis is to apply an extensive comparative study of off-the-shelf clustering algorithms of the Spark MLlib platform in big trajectory datasets, investigating various transformation methods of the trajectory data to a vector representation. **Bib:**⁶. **Contact:** Prof. Nikos Pelekis (npelekis@unipi.gr)

2022.07: 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:**⁷. **Contact:** Prof. Nikos Pelekis (npelekis@unipi.gr)

¹ <https://github.com/orgs/DataStories-UniPi/repositories?q=&type=&language=java>

² <https://doi.ieeeecomputersociety.org/10.1109/BigData47090.2019.9005563>

³ http://people.csail.mit.edu/ibrahimsabek/pdf/21_tutorial_mdm.pdf

⁴ <https://arxiv.org/pdf/2209.04635.pdf>

⁵ <https://arxiv.org/abs/2108.11482>

⁶ <https://spark.apache.org/mllib/>

⁷ <https://github.com/XingruiWang/Two-Stage-Gan-in-trajectory-generation>

2022.08: 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 analyzed, 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:**⁸. **Contact:** Andreas Tritsarolis (andrewt@unipi.gr); Prof. Yannis Theodoridis (ytheod@unipi.gr)

2022.09(a-d): Large-Scale Edge Operating System (LaSEd OS). With edge devices being everywhere around us, from smart traffic lights, to pollution measuring boxes and, of course, in the pockets of each and every one of us, a fully connected future where information about any particular thing is ever flowing and instantly available in real time looks more real than it has ever looked before. The hardware is there and readily available, making the process of “smartifying” anything the simplest it has ever been. In this set of four (4) theses, we aim to develop modules towards a Large-scale edge OS, namely ‘Maestro’ (a networking framework that can enable efficient data transfer in both edge-to-edge and edge-to-cloud scenarios), ‘Skipper’ (a module purpose-built for the edge, that allows the creation of processing pipelines that are ready-to-deploy for a prespecified edge scenario), ‘Kowalski’ (an analytics framework aiming at digesting, processing and analysing the large amounts of data that can be produced by an edge system), and ‘Matrix’ (a use-case simulation framework enabling the user to accurately and reliably test multi-level architectures and software implementations based on a set of predefined variables like compute resources and network schemes), respectively. **Bib:**⁹. **Contact:** George S. Theodoropoulos (gstheo@unipi.gr); Prof. Yannis Theodoridis (ytheod@unipi.gr)

2022.10: Public transit data visualization. Nowadays thousands of urban public transport agencies publish their public transportation route and timetable information with the General Transit Feed Specification (GTFS) as the standard open format. In this thesis, we will design and implement a GTFS data visualization tool that displays spatial and temporal patterns of transit services from which qualitative information and insights can be gained. **Bib:**¹⁰. **Contact:** Dr. Eva Chondrodima (evachon@unipi.gr)

2022.11: 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:**¹¹. **Contact:** Dr. Eva Chondrodima (evachon@unipi.gr)

2022.12: UniPi-AIS platform interface. 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. **Bib:**¹². **Contact:** Andreas Tritsarolis (andrewt@unipi.gr); Prof. Yannis Theodoridis (ytheod@unipi.gr)

2022.13: 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

⁸ <http://dx.doi.org/10.5311/JOSIS.2020.20.661>; <https://doi.org/10.1109/MDM52706.2021.00048>

⁹ <https://doi.org/10.1109/MCOM.2016.7432173>; <https://doi.org/10.1109/MPRV.2015.32>

¹⁰ <https://doi.org/10.1016/j.heliyon.2020.e03729>

¹¹ <https://doi.org/10.1109/TITS.2019.2899179>

¹² <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>

aircraft trajectories of streaming data. **Bib:**¹³. **Contact:** Dr. Eva Chondrodima (evachon@unipi.gr); Yannis Kontoulis (ikontoulis@unipi.gr)

2022.14: 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 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. Towards this direction, we will utilize the pattern-aware synthetic trajectory generator ‘Hermoupolis’, with the necessary modifications and additions, to develop the generated trajectories in three dimensions. In particular, aircraft flight plans that refer to a given time frame for a specific set of airports will be used as a data source. **Bib:**¹⁴. **Contact:** Yannis Kontoulis (ikontoulis@unipi.gr); Prof. Nikos Pelekis (npelekis@unipi.gr); Prof. Yannis Theodoridis (ytheod@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 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:

- **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^η έκθεση προόδου δεν απαιτείται να ακολουθεί κάποιο συγκεκριμένο πρότυπο μορφοποίησης.

- **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

¹³ <https://doi.org/10.1007/s41060-019-00182-4>; <https://doi.org/10.1016/j.jss.2016.06.016>

¹⁴ <https://doi.org/10.1145/2937753>

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 Αναφορές¹⁵

(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 κι αυτό στη συγγραφή του τελικού κειμένου της ΠΕ/ΜΔ αφού θα έχει καθορίσει σαφώς τον «πίνακα περιεχομένων» της. Η 2^η έκθεση προόδου θα ακολουθεί το πρότυπο μορφοποίησης της ACM για τη συγγραφή ερευνητικών εργασιών στα περιοδικά της¹⁶.

- **Τελικός τόμος:** Ο τελικός τόμος θα ακολουθεί λίγο έως πολύ τη δομή της 2^{ης} έκθεσης, με τα τελικά αποτελέσματα. Καθώς η πτυχιακή / διπλωματική εργασία αποτελεί έναν τύπο μονογραφίας, ο τελικός τόμος θα ακολουθεί το πρότυπο μορφοποίησης της ACM για τη συγγραφή μονογραφιών – βιβλίων¹⁷, εκτός αν προβλέπεται κάτι διαφορετικό από τον κανονισμό του προγράμματος σπουδών. Το συνοδευτικό υλικό του τελικού τόμου περιλαμβάνει:

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

¹⁵ 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 (ή να μεταφράσουμε – το ίδιο κάνει) ολόκληρα κείμενα και να τα ενσωματώσουμε αυτούσια στο δικό μας.

¹⁶ Για σχετικές οδηγίες βλ. <https://www.acm.org/publications/taps/word-template-workflow#h-2.-the-workflow-and-templates> > STEP 1 (MS Word ή LaTeX, αναλόγως το πρόγραμμα που χρησιμοποιείτε). Ειδικά για τη μορφοποίηση των βιβλιογραφικών αναφορών συστήνεται το στυλ μορφοποίησης APA, για σχετικές οδηγίες βλ. <https://www.mendeley.com/guides/apa-citation-guide/>, https://www.lib.auth.gr/sites/default/files/docs_files/APA-examples-gre.pdf.

¹⁷ Για σχετικές οδηγίες βλ. <http://books.acm.org/authors/author-instruction>.

Η **γλώσσα συγγραφής** μπορεί να είναι η ελληνική ή η αγγλική. Η δεύτερη περίπτωση προϋποθέτει έγκριση της ΓΣ του Τμήματος.

Τυπική **διάρκεια εκπόνησης** μιας πτυχιακής (διπλωματικής) εργασίας είναι οι 12 (6, αντίστοιχα) μήνες, συνήθως με αποκλίσεις προς τα πάνω. 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.