

Θέμα Διπλωματικής	Περιγραφή	Επιβλέπων	Τόπος	Προαπαιτούμενες γνώσεις/ικανότητες, ή ανάγκη εκπαίδευσης	Πιθανή ημερομηνία έναρξης	Παρατηρήσεις
Exploration of Neologism Formation in Sign Languages	Neologism formation in Sign Languages (SLs) presents a special interest due to the use of various SL articulators to represent new concepts. The analysis of the mechanisms exploited in composition and articulation of neologisms provides significant support to the development of SL processing tools aiming to spot, translate or otherwise process SL linguistic content from various domains. The task involves annotation of pieces of SL video, characterised as neologism intensive, as well as the semantic and morpho-phonological analysis of neologisms lists from the domain of computer use or other neologism intensive domains. The aim is to provide a description of SL neologism formation on semantic and morpho-phonological level which could support identification of this type of constructs in the framework of applications such as information retrieval from SL video and SL utterance parsing.	Ε. Ευθυμίου	εξ αποστάσεως	Γνώση μιας νοηματικής γλώσσας με προτίμηση στην Ελληνική Νοηματική (ΕΝΓ), χρήση εργαλείου επισημείωσης iLex ή ELAN	Ανοιξη 2023	
Οργάνωση γλωσσικών δεδομένων με βάση το δίκτυο WordNet	Στο πλαίσιο της εργασίας, θα γίνει αντιστοίχιση των γλωσσικών δεδομένων της δίγλωσσας (Νέα Ελληνικά – Ελληνική Νοηματική Γλώσσα) βάσης Πολύτροπον με τη βάση λεξικών δεδομένων WordNet. Συγκεκριμένα, κατά το πρώτο στάδιο της εργασίας, θα γίνει έλεγχος των συνδέσεων των εννοιών που περιέχονται στη βάση Πολύτροπον με τα αντίστοιχα συνωνυμικά σύνολα (synsets) που συνθέτουν το WordNet. Το βασικό εργαλείο που θα χρησιμοποιηθεί θα είναι το Open Multilingual WordNet. Στη συνέχεια, θα γίνει εμπλουτισμός της βάσης Πολύτροπον σύμφωνα με το δίκτυο σημασιολογικών και λεξιλογικών σχέσεων: συνωνυμία, υπωνυμία/υπερωνυμία..., με τις οποίες συνδέονται αυτά τα συνωνυμικά σύνολα μεταξύ τους. Στόχος της εργασίας είναι η βιωματική εμπλοκή στην κανονικοποίηση γλωσσικών δεδομένων με τρόπο ώστε να γίνει δυνατή η σύνδεσή τους με διάφορα είδη λεξικών και άλλων γλωσσικών πόρων ανεξαρτήτως της γλώσσας πρόσβασης. Η εμπλουτισμένη βάση που θα προκύψει θα αξιοποιηθεί για την κατασκευή πόρων και/ή εργαλείων γλωσσικής εκπαίδευσης και για την υποστήριξη της υπολογιστική επεξεργασία νοηματικών γλωσσών, τη διάκριση σημασιών και τη μηχανική μετάφραση με απώτερο στόχο την επέκταση του Multilingual Sign Language Wordnet.	Ε. Ευθυμίου	εξ αποστάσεως	Γνώση Ελληνικής Νοηματικής (ΕΝΓ)	Ανοιξη 2023	
Music-Language fusion models for concept discovery in complex compositions	Musical concepts are intertwined in various combinations that remain sometimes hidden even to expert musicians, who are able to discover them after meticulous study on musical works. Aim of this research is to examine whether and to what extent machine learning systems are able to identify such concepts in complex compositions. Some simple musical concepts will be identified beforehand (e.g., ascending/descending pitches, perfect cadence, parallel voice motion, etc) and a synthetic dataset will be created that musically implements those concepts (encoder), along with linguistic descriptions (decoder). The ability of the system to discover those concepts will be tested in real, complicated compositions.	Μ. Καλιακάτσος-Παπακώστας	εξ αποστάσεως	Γνώση προγραμματισμού	Ανοιξη 2023	

Music Q&A: Minimal adaptation of a Q&A in the domain of Music	There is an abundance of data from music shows on youtube or as podcasts that are already transcribed to text, or can be transcribed with some accuracy. Aim of this reseach is to examine aspects of domain adaptation in music with data that can become available from the web. A dataset will be created incrementally, focussing on different aspects of musical text obtained by youtube channels or podcasts with content that specializes in specific styles (e.g., jazz music, retro games music, etc.). The results will examine the capabilities of the system as different parts of the dataset are introduced, in specific Q & A tasks that will be developed.	M. Καλιακάτσος-Παπακώστας	εξ αποστάσεως	Γνώση προγραμματισμού	Άνοιξη 2023	
Study and evaluation of voice synthesis systems for audiobooks	How good are synthetic voices for reading an entire book? What makes a voice acceptable or not when we talk about entire books of literature or technical/scientific? The answer to these questions requires a deeper study of both the approaches for evaluating synthetic voice systems and the utilization and collection of appropriate data. The aim is to evaluate various modern synthesis systems on real data and identify points of improvement.	A. Κατσαμάνης	εξ αποστάσεως	Γνώση προγραμματισμού	Άνοιξη 2023	
Development of an interactive assistant to support applications in medicine	Study of the "accountability" process of nurses in clinics: At the end of each shift nurses must record their observations about each patient so that the next ones can pick up. For example: "The patient today was very good-natured, walked a lot, took the morning medications but not the afternoon ones because he complained of a headache." The language used will be studied, whether it can be recognized and organized systematically by an automatic system. Also, an interactive system will be designed to support the process.	A. Κατσαμάνης	εξ αποστάσεως	Γνώση προγραμματισμού/Βασικές γνώσεις Αναγνώρισης Φωνής/Διαλογικά Συστήματα	Άνοιξη 2023	
Personalized conversational agents	Διαλογικά συστήματα με μνήμη, συστήματα που αναγνωρίζουν τον χρήστη και του απαντούν κατάλληλα. Για παράδειγμα, του κάνουν κατάλληλες προτάσεις για μουσική ή για παιχνίδια.	A. Κατσαμάνης	εξ αποστάσεως	Γνώση Προγραμματισμού/Διαλογικά Συστήματα	Άνοιξη 2023	
Voice recognition with utilization of image data	Employees photograph folder labels, read them, and a system passes the data into special forms: this is the future of a archival material digitization company. In addition to voice data, how can the automatic system also utilize the corresponding photos?	A. Κατσαμάνης	εξ αποστάσεως	Γνώση Προγραμματισμού/Αναγνώριση φωνής	Άνοιξη 2023	
Μέθοδοι deep learning για επεξεργασία φυσικής γλώσσας		E. Κουμπάρκης	εξ αποστάσεως	Μεταπτυχιακό μάθημα "Βασικά Μηχανική Μάθηση για Επεξεργασία Φυσικής Γλώσσας" ή άλλα αντίστοιχα μαθήματα του προγράμματος	Άνοιξη 2023	
Μεγάλα γλωσσικά μοντέλα		E. Κουμπάρκης	εξ αποστάσεως	Μεταπτυχιακό μάθημα "Βασικά Μηχανική Μάθηση για Επεξεργασία Φυσικής Γλώσσας" ή άλλα αντίστοιχα μαθήματα του προγράμματος	Άνοιξη 2023	

Exploring the effect of punctuation in Natural Language Understanding tasks	Exploring the effect of punctuation in NLU tasks. Punctuation has been shown to play an unexpected role in the way models of NLI interpret it (Ek et al. 2020). In specific, it has been shown that DL models, as well as pre-trained models either interpret punctuation in ways that affect reasoning when it should play no role, or ignore punctuation across the board. The task would be to continue this preliminary work for NLU in general, target a number of other core NLU tasks in addition to NLI and check the performance of different models when systems are trained, tested or both in datasets that have been corrupted across some parameter that has to do with punctuation.	Γ. Παναγάκης	εξ αποστάσεως	Γνώση προγραμματισμού	Ανοιξη 2023	
Author Identification of Modern Greek Novelists	Author Identification of Modern Greek Novelists. Develop a model that will accurately classify text to Authors. This is a task that has some literature in NLP, but not for Greek. The tricky bit would be to find enough data. The Gutenberg project has a number of texts in Greek in .txt format that could be used as the starting point along with some other resources: https://www.ebooks4greeks.gr/category/free-ebooks/ . Eventually, one might be able to create Literature embeddings, which will be interesting to attempt.	Γ. Παναγάκης	εξ αποστάσεως	Γνώση προγραμματισμού	Ανοιξη 2023	
Υπολογιστική υφολογία: υφομετρική ανάλυση της ηλικίας του ομιλητή σε Σώματα Κειμένων από Μέσα Κοινωνικής Δικτύωσης	Η υπολογιστική υφομετρία έχει χρησιμοποιηθεί σε υφομετρικές αναλύσεις κειμένων εγκληματολογικού ενδιαφέροντος. Η συγκεκριμένη εργασία εστιάζει στην ανάλυση υφομετρικών χαρακτηριστικών που συμβάλλουν στην αναγνώριση της ηλικίας του ομιλητή με στόχο τη διαδικτυακή προστασία των ανηλίκων. Θα μελετηθούν σχετικοί δείκτες του συγγραφικού ύφους και θα εφαρμοστούν τεχνικές αναγνώρισης προτύπων. Τα απαιτούμενα δεδομένα ενηλίκων θα συλλεχθούν από Μέσα Κοινωνικής Δικτύωσης (διαθέσιμους λογαριασμούς ενηλίκων -μήπως πρέπει να πεις σε ποια γλώσσα;) ενώ ως δεδομένα ανηλίκων θα αξιοποιηθούν τα δημόσια ΗΣΚ του U.S. CyberWatch.	Β. Παπαβασιλεί	εξ αποστάσεως	Ποσοτική ανάλυση γλωσσικών δεδομένων	Ανοιξη 2023	
Υφομετρική ανάλυση των μελών διάφορων κοινοβουλευτικών ομάδων (συμπολίτευση, αντιπολίτευση) σε Σώμα Κειμένων από τα Πρακτικά του Ελληνικού Κοινοβουλίου	Η πλέον γνωστή εφαρμογή της υπολογιστικής υφολογίας είναι η αναγνώριση συγγραφέα. Η συγκεκριμένη εργασία εστιάζει στην ανάλυση υφομετρικών χαρακτηριστικών του ομιλητή, που σχετίζονται με την ένταξή του σε μια ομάδα ή/και με το ρόλο του σε αυτή.	Β. Παπαβασιλεί	εξ αποστάσεως	Ποσοτική ανάλυση γλωσσικών δεδομένων	Ανοιξη 2023	
Exploring the use of ChatGPT as a Search Engine	We will explore in depth the GPT3 language model and its use in the ChatGPT application of OpenAI. Strengths and weaknesses of GPT3 and related neural language models will be analysed and will be correlated to the performance/output of ChatGPT. Of particular interest is the potential role of such models in substituting traditional search engines. The related thesis will perform a systematic review of the related -rapidly increasing- literature on the topic and provide examples of the arguments against/for such use of generative language models.	Κ. Πάστρα	εξ αποστάσεως	Βασικές γνώσεις σχετικά με γλωσσικά μοντέλα βαθείας μάθησης	Ανοιξη 2023	

Automatic Social Story Generation	The topic focusses on social stories, i.e., on simple, usually illustrated stories (Gray 1994) that are used to explain not only how to do everyday actions (e.g., how to wash one's hands), but also complex events (COVID-19 protection or vaccination procedures) and social concepts such as bullying and online safety. The target audience ranges from children of different age groups to people with special mental needs, such as those diagnosed in the spectrum. Though a number of software platforms exists for facilitating the creation of such stories, automating such creation is still understudied. We will explore methods for generating social stories, such as combining deep learning algorithms for image caption generation and text simplification ones.	Κ. Πάστρα	εξ αποστάσεως	Βασικές γνώσεις προγραμματισμού και αλγορίθμων βαθείας μάθησης	Ανοιξη 2023	
Sentence/Document Domain Classification for Greek	The project aims at developing a model for the automatic classification of documents into one of several predefined domains based on their content. The first task of the project will be the collection and preprocessing of documents from various domains. Next, various features will be extracted, to be can be used for applying various machine learning algorithms. Following the feature extraction, is the training and evaluating of several machine learning models, such as Naive Bayes, Support Vector Machines, and Neural Networks, to determine the best-performing model.	Σ. Σοφιανόπουλος	εξ αποστάσεως	Γνώση προγραμματισμού	Ανοιξη 2023	
Implementing a Spellchecker for Greek	The aim of this project is to develop a spellchecker service for Greek. Spellcheckers are really important in the modern interactions that we have with automated systems, either for chatbots understanding a misspelled question, or for spellchecking text that will be used as a training resource. The student will have to explore the existing methods (rule-based, statistical, neural) and select one in order to build a model that identifies and fixes spelling errors.	Σ. Σοφιανόπουλος	εξ αποστάσεως	Γνώση προγραμματισμού	Ανοιξη 2023	
Development of NMT models for low-resource languages using multilingual	Summary: Zero-shot translation, translating between language pairs on which a Neural Machine Translation (NMT) system has never been trained, is an emergent property when training the system in multilingual settings. The aim here is to study the use of a multilingual setup that will allow the development of a translation system that can translate towards languages for which resources are limited.	Γ. Ταμπουρατζή	εξ αποστάσεως	Γνώση προγραμματισμού	Ανοιξη 2023	προτείνεται μετά από συζητήσεις με φοιτητή
Studying the development of an open-source off-line machine translation a	Summary: The objective is to create an offline machine translation application that will be open source. It is planned not to rely on proprietary machine translation technologies or services and run client-side for its main purpose, while offering improvements compared to similar predecessors. Collaboration with relevant projects will be pursued, to avoid reinventing the wheel.	Γ. Ταμπουρατζή	εξ αποστάσεως	Γνώση προγραμματισμού	Ανοιξη 2023	προτείνεται μετά από συζητήσεις με φοιτητή
Πρόβλεψη αριθμού σχολίων, views και likes σε online media (YouTube, TikTok)	Πρόβλεψη αριθμού σχολίων, views και likes σε online media (YouTube, TikTok)	Α. Ντούλας		Γνώση προγραμματισμού	Φθινόπωρο 2023	
Υλοποίηση πλατφόρμας για real-time A/B Testing	Υλοποίηση πλατφόρμας για real-time A/B Testing	Α. Ντούλας		Γνώση προγραμματισμού, στατιστικής	Φθινόπωρο 2023	
Product outpost: συλλογή δεδομένων και ενοποίησή τους από διάφορα online markets	Product outpost: συλλογή δεδομένων και ενοποίησή τους από διάφορα online markets	Α. Ντούλας		Γνώση προγραμματισμού	Φθινόπωρο 2023	
Υλοποίηση εφαρμογής μεγάλων δεδομένων πάνω από το Volvo API https://developer.volvocars.com/apis/	Υλοποίηση εφαρμογής μεγάλων δεδομένων πάνω από το Volvo API https://developer.volvocars.com/apis/	Α. Ντούλας		Γνώση προγραμματισμού, συστήματα μεγάλων δεδομένων	Φθινόπωρο 2023	