

Εργασία Εαρινού Εξαμήνου 2022-2023:

- Ημερομηνία παράδοσης: Ημερομηνία εξέτασης του μαθήματος, ώρα 23:59μμ. Ομάδες **1-4** φοιτητών. Η παράδοση της εργασίας γίνεται μέσω της πλατφόρμας e-class χωρίς παραπομπές σε εξωτερικούς συνδέσμους. Παραδίδονται:

α) η τεκμηρίωση της εργασίας σε ένα αρχείο pdf, στην πρώτη σελίδα της οποίας αναγράφονται τα ονοματεπώνυμα των φοιτητών και οι ΑΜ. Δεν θα βαθμολογηθούν εργασίες που δεν περιέχουν τεκμηρίωση ή που δεν αναφέρουν τα ονόματα των μελών της ομάδας στην τεκμηρίωση. **β)** τα αρχεία source code σε ένα συμπιεσμένο αρχείο με όνομα source2023.zip (ή .rar ή άλλη σχετική κατάληξη). **γ)** οποιαδήποτε άλλα συνοδευτικά αρχεία η ομάδα κρίνει απαραίτητα σε ένα συμπιεσμένο αρχείο με το όνομα auxiliary2023.zip (ή .rar ή άλλη σχετική κατάληξη).

- Η εργασία ισχύει και τον Σεπτέμβριο.

- Αποδεκτές γλώσσες είναι οι Octave/Matlab και Python.

- Η αντιγραφή ή η χρήση generative bots οδηγεί σε μηδενισμό στον τελικό βαθμό.

Προσοχή!!!: Δεν μπορείτε να χρησιμοποιήσετε **συνελικτικά** νευρωνικά δίκτυα ή **Recurrent Neural Networks** οποιουδήποτε τύπου. Δεν είναι αποδεκτή η χρήση έτοιμων web services ή APIs για speech recognition. Δεν μπορείτε να χρησιμοποιήσετε **transfer learning** από ήδη εκπαιδευμένα δίκτυα. Οι αντίστοιχες λύσεις μηδενίζονται.

Θέμα (4 βαθμοί): Καλείστε να υλοποιήσετε ένα ASR σύστημα, που δέχεται είσοδο μία ηχογράφιση κάθε φορά, η οποία συνιστά πρόταση αποτελούμενη από 5-10 ψηφία της Αγγλικής γλώσσας που έχουν ειπωθεί με αρκούντως μεγάλα διαστήματα παύσης.

1) Το σύστημα προχωρά στην κατάτμηση της πρότασης σε λέξεις χρησιμοποιώντας **υποχρεωτικά** έναν ταξινομητή background vs foreground της επιλογής σας. Από τις λέξεις που προκύπτουν, υπολογίστε τη θεμελιώδη συχνότητα του ομιλητή.

2) Στη συνέχεια, το σύστημα αναγνωρίζει κάθε λέξη χρησιμοποιώντας ως φασματική αναπαράσταση μόνο το απλό φασματογράφημα (επιλέξτε φασματική περιοχή αν θέλετε). Αν χρειαστείτε δεδομένα εκπαίδευσης, χρησιμοποιήστε **μόνο** δημόσια σύνολο(α) δεδομένων από το Internet. Δεν χρειάζεται να τα συμπεριλάβετε στα παραδοτέα αλλά μόνο να περιγράψετε πώς ελήφθησαν.

3) Στην έξοδο παράγεται κείμενο με τα ψηφία που αναγνωρίστηκαν.

- Δώστε έμφαση στην επεξεργασία του σήματος, προτού αρχίσουν τα στάδια κατάτμησης/αναγνώρισης (π.χ., με κατάλληλα φίλτρα, αλλαγή ρυθμού δειγματοληψίας, κ.λ.π).
- Είναι σημαντικό να **περιγράψετε** το σύστημα αλγοριθμικά (εξαγωγή χαρακτηριστικών, αλγόριθμος αναγνώρισης) και να **εξηγήσετε** τις επιδόσεις του χρησιμοποιώντας τις κατάλληλες μετρικές.
- Πρέπει να εξηγήσετε ποια δεδομένα χρησιμοποιήσατε κατά τον έλεγχο και την εκπαίδευση του συστήματος. Αν είναι δικά σας, πώς τα δημιουργήσατε.
- Προσπαθήστε να μην εξαρτάται το σύστημα από τα χαρακτηριστικά της φωνής του ομιλητή, αλλά να είναι όσο το δυνατόν ανεξάρτητο ομιλητή.

Καλή επιτυχία!