



Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης
Πολυτεχνική Σχολή
Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Διπλωματική Εργασία

με θέμα

**Ανάπτυξη διαδικτυακού περιβάλλοντος διαχείρισης,
σημασιολογικής επεξεργασίας και αναγνώρισης
οπτικοακουστικού περιεχομένου**

Δημήτρης Νένος 9012

Επιβλέπων καθηγητής: Χαράλαμπος Α. Δημούλας

Θεσσαλονίκη, Ιανουάριος 2025

Ευχαριστίες

Κατά πρώτο λόγο θα ήθελα να ευχαριστήσω θερμά τον Καθηγητή και επιβλέποντα της διατριβής μου, Χαράλαμπο Δημούλα. Η επιστημονική καθοδήγησή του, οι χρήσιμες συμβουλές του και η άριστη συνεργασία μας ήταν καθοριστικές σε όλα τα στάδια εκπόνησης της εργασίας μου. Επίσης, ευχαριστώ θερμά τους διδάκτορες Λάζαρο Βρύση και Νίκο Βρύζα για τις πολύτιμες επιστημονικές συμβουλές τους. Τέλος, οφείλω ένα ευχαριστώ σε όλους τους δικούς μου ανθρώπους, την οικογένεια και τους φίλους μου για την ψυχολογική υποστήριξη και την ενθάρρυνσή τους καθ' όλη τη διάρκεια της ακαδημαϊκής μου διαδρομής.

Περίληψη

Το σύγχρονο ψηφιακό τοπίο παρουσιάζει εκθετική αύξηση περιεχομένου πολυμέσων δημιουργώντας σημαντικές προκλήσεις στην αποτελεσματική διαχείρισή του. Καθώς οι παραδοσιακές μέθοδοι οργάνωσης και ανάκτησης οπτικοακουστικού υλικού καθίστανται ολοένα και πιο ανεπαρκείς, κρίνεται επιτακτική η ανάγκη για την ύπαρξη μιας διαδικτυακής εφαρμογής που ενσωματώνει προηγμένες δυνατότητες επεξεργασίας και αναζήτησης. Στόχος της παρούσας διατριβής είναι η ανάπτυξη διαδικτυακού περιβάλλοντος για την βελτίωση της διαχείρισης, της ανάκτησης, της σημασιολογικής επεξεργασίας και αναγνώρισης οπτικοακουστικού υλικού. Το προτεινόμενο σύστημα ενσωματώνει αλγόριθμους τεχνητής νοημοσύνης για την μετατροπή ομιλίας σε κείμενο, σε συνδυασμό με την αναγνώριση ονομαστικών οντοτήτων για την εξαγωγή θεμάτων, βελτιώνοντας την αναζήτηση και ανάκτηση πολυμεσικού υλικού. Επιπλέον, χρησιμοποιείται ένας αλγόριθμος αναγνώρισης ενεργειών και αναπτύσσεται και χρησιμοποιείται ένας αλγόριθμος ταξινόμησης ήχου που αξιοποιεί φασματογραφήματα που βασίζονται σε κυματίδια για την αναπαράσταση του ηχητικού σήματος με μεγάλη ακρίβεια. Συνδυάζοντας αυτές τις τεχνολογίες, το σύστημα υπερβαίνει τις παραδοσιακές προσεγγίσεις που εξαρτώνται από μεταδεδομένα, παρέχοντας ένα αξιόπιστο και σημασιολογικά εμπλουτισμένο περιβάλλον που μπορεί να χρησιμοποιηθεί τόσο από έμπειρους όσο και από αρχάριους χρήστες. Το σύστημα έχει κατασκευαστεί με τέτοιο τρόπο ώστε να είναι εύκολα επεκτάσιμο και να μπορεί να διαχειριστεί μεγάλο όγκο δεδομένων, με τη χρήση μιας αρχιτεκτονικής για ασύγχρονη επικοινωνία και επεξεργασία που υποστηρίζεται από τα εργαλεία RabbitMQ και Celery. Απευθυνόμενη πρωτίστως στην πανεπιστημιακή κοινότητα, η εφαρμογή όχι μόνο βελτιώνει την προσβασιμότητα και την ανάκτηση οπτικοακουστικού υλικού, αλλά προάγει την καινοτομία και την έρευνα προσφέροντας μια ενιαία πλατφόρμα που μπορεί να επεκταθεί με ποικίλους τρόπους.

Abstract

The modern digital landscape is witnessing an exponential increase in multimedia content, creating significant challenges in its effective management. As traditional methods for organizing and retrieving audiovisual content become increasingly inadequate, there is an urgent need for an online application that incorporates advanced processing and search capabilities. The aim of this thesis is to develop an online environment for improving the management, retrieval, semantic processing, and recognition of audiovisual content. The proposed system incorporates advanced artificial intelligence algorithms for speech-to-text conversion, combined with named entity recognition for topic extraction, thereby enhancing the search and retrieval of multimedia content. Additionally, an action recognition algorithm and an audio classification algorithm that utilizes wavelet-based spectrograms to represent the audio signal with high accuracy are employed. By combining these technologies, the system surpasses traditional metadata dependent approaches, providing a robust and semantically enriched environment that can be utilized by both experienced and novice users. The system has been designed to be easily expandable and capable of managing large volumes of data, employing an architecture for asynchronous communication and processing supported by the tools RabbitMQ and Celery. Primarily targeting the university community, the application not only enhances the accessibility and retrieval of audiovisual content but also promotes innovation and research by offering a unified platform that can be extended in various ways.

Github repository:

https://github.com/Diminenos/audiovisual_content_management_system

Πίνακας περιεχομένων

Ευχαριστίες	1
Περίληψη.....	2
Abstract	3
Πίνακας περιεχομένων	4
Λίστα Εικόνων	7
Λίστα Πινάκων	9
1. Εισαγωγή.....	10
1.1. Θεώρηση του Προβλήματος	10
1.2. Στόχος και Καινοτομία της Εργασίας.....	11
1.3. Σύντομη παρουσίαση των κεφαλαίων της εργασίας.....	11
2. Θεωρητικό και Μεθοδολογικό Υπόβαθρο	12
2.1. Διαχείριση Οπτικοακουστικών Πόρων	12
2.1.1. Η Ανάγκη για Ευφυή Συστήματα	12
2.1.2. Ολοκληρωμένη Αρχιτεκτονική και Επικοινωνία Συστήματος	13
2.2. Μεθοδολογία ανάπτυξης λογισμικού	14
2.2.1. Συστήματα Διαχείρισης Περιεχομένου	15
2.3. Αρχιτεκτονική Επικοινωνίας του Συστήματος.....	16
2.3.1. Βασικές έννοιες και Πρωτόκολλο Ουράς Μηνυμάτων	17
2.3.2. Δομή και Ροή Μηνυμάτων.....	17
2.3.3. Διαχείριση Μηνυμάτων	19
2.4. Διαχείριση Ασύγχρονων Εργασιών.....	19
2.4.1. Κατανεμημένη Ουρά Εργασιών	20
2.4.2. Εκτέλεση Εργασιών Μέσω του Εργαλείου Celery	20
2.5. Εφαρμογές Βαθιάς Μάθησης για Ανάλυση Πολυμέσων	21
2.5.1. Μετατροπή Ομιλίας σε Κείμενο	21
2.5.2. Αναγνώριση Ονομαστικών οντοτήτων	22
2.5.3. Αναγνώριση Δράσης	22
2.6. Ταξινόμηση Ήχου	23
2.6.1. Παραδοσιακές Προσεγγίσεις	23
2.6.2. Μετασχηματισμός Κυματιδίων	24
2.6.3. Διακριτός Μετασχηματισμός Κυματιδίων.....	24
2.6.4. Δημιουργία Φασματογραφήματος με βάση τα Κυματίδια	25
2.7. Μηχανική μάθηση	26
2.8. Βαθιά Μάθηση.....	27
2.8.1. Νευρωνικά Δίκτυα Τροφοδοσίας προς τα εμπρός.....	27
2.8.2. Συνελκτικά Νευρωνικά Δίκτυα	28
2.9. Σύνοψη Δευτέρου Κεφαλαίου.....	30

3. Ανάλυση Εφαρμογής	31
3.1. Αρχική ιδέα	31
3.2. Στόχευση	31
3.3. Παρόμοιες Εφαρμογές	31
3.3.1. Αλγόριθμοι για Διαχείριση Οπτικοακουστικού Υλικού	32
3.3.2. Διαφοροποίηση της Προτεινόμενης Εφαρμογής	32
3.4. Ανάλυση Χρηστών	33
3.5. Απαιτήσεις Συστήματος	34
3.5.1. Λειτουργικές Απαιτήσεις.....	34
3.5.2. Μη λειτουργικές απαιτήσεις	34
3.6. Εργαλεία Ανάπτυξης και Τεχνολογίες	35
3.6.1. Συστήματα Διαχείρισης Περιεχομένου	35
3.6.2. Διαμεσολαβητές Μηνυμάτων	36
3.6.3. Κατανεμημένα συστήματα ουράς εργασιών	36
3.6.4. Λοιπά εργαλεία και βιβλιοθήκες	37
3.7. Οργάνωση και χρονοπρογραμματισμός εργασιών	38
3.8. Σύνοψη Τρίτου Κεφαλαίου	38
4. Σχεδίαση Εφαρμογής	39
4.1. Αρχιτεκτονική Συστήματος και Πρωτοκόλλα Επικοινωνίας	39
4.2. Αλγόριθμοι Εφαρμογής.....	40
4.3. Σενάρια Χρήσης και Ροή Εργασιών του Συστήματος	41
4.4. Σχεδιασμός Διεπαφής Χρήστη: Από το Πρωτότυπο στη Λειτουργική Εμφάνιση	45
4.4.1. Διεπαφή χρήστη για το διάγραμμα ροής δραστηριοτήτων	45
4.4.2. Σελίδα αναπαραγωγής Περιεχομένου	48
4.4.3. Αρχική Σελίδα	49
4.5. Σύνοψη Τετάρτου Κεφαλαίου	50
5. Ανάπτυξη Εφαρμογής	51
5.1. Διαμόρφωση Βασικής Αρχιτεκτονικής Συστήματος	51
5.1.1. Περιβάλλον Εγκατάστασης Συστήματος	51
5.1.2. Εγκατάσταση Περιβάλλοντος του Συστήματος Διαχείρισης Περιεχομένου... ..	51
5.1.3. Βασική Δομή Ιστοτόπου.....	52
5.1.4. Προσαρμοσμένη αναζήτηση περιεχομένου.....	52
5.1.5. Ασφάλεια Ιστοτόπου.....	53
5.2. Αρχιτεκτονική Μηχανισμού Επικοινωνίας και Διαχείρισης Εργασιών	53
5.2.1. Εγκατάσταση του Διαμεσολαβητή Μηνυμάτων	53
5.2.2. Δημιουργία Παραγωγού Μηνυμάτων	53
5.2.3. Δημιουργία Καταναλωτή και Ουράς Μηνυμάτων	54
5.2.4. Ασύγχρονη Επεξεργασία Εργασιών	55

5.2.5. Εκτέλεση Αλγορίθμων μέσω Κατανεμημένης Ουράς Εργασιών	55
5.3. Ενσωμάτωση Αλγορίθμων	56
5.4. Επεκτασιμότητα και Ενσωμάτωση Νέων Αλγορίθμων	58
5.5. Ανάπτυξη Αλγορίθμου Ταξινόμησης Ήχου	59
5.5.1. Σετ Δεδομένων και Εξαγωγή Χαρακτηριστικών	60
5.5.2. Μοντέλο Ταξινόμησης Ήχου και Εκπαίδευση	62
5.6. Γραφική Προβολή Διεπαφής Χρήστη	65
5.6.1. Αρχική Σελίδα	65
5.6.2. Σελίδα Σύνδεσης	65
5.6.3. Σελίδα Δημιουργίας Λογαριασμού	66
5.6.4. Σελίδα Προβολής Οπτικοακουστικού Περιεχομένου	67
5.6.5. Σελίδα Αναπαραγωγής περιεχομένου	68
5.6.6. Σελίδα Περιεχομένου εκάστοτε χρήστη	70
5.6.7. Προφίλ Χρηστών	70
5.7. Γραφική Προβολή Περιβάλλοντος Διαχειριστή.....	72
5.7.1. Πλατφόρμα Διαχείρισης Εικονικής Μηχανής.....	72
5.7.2. Πίνακας Διαχείρισης του Συστήματος Διαχείρισης Περιεχομένου	72
5.8. Σύνοψη Πέμπτου Κεφαλαίου	73
6. Αξιολόγηση	74
6.1. Αξιολόγηση Απόδοσης του Συστήματος	74
6.2. Αξιολόγηση Περιβάλλοντος Εφαρμογής.....	78
6.2.1. Ποιοτική Αξιολόγηση	78
6.2.2. Ποσοτική Αξιολόγηση	79
6.3. Αξιολόγηση των Μοντέλων του Συστήματος	81
6.3.1. Αξιολόγηση του Αλγορίθμου Ταξινόμησης Ήχου	82
6.4. Σύνοψη Έκτου Κεφαλαίου	92
7. Συμπεράσματα.....	93
7.1. Καινοτομία και Συνεισφορά της Εργασίας	93
7.2. Μελλοντικές Επεκτάσεις.....	94
Βιβλιογραφικές αναφορές	95

Λίστα Εικόνων

Εικόνα 2-1 Το μοντέλο του καταρράχτη Πηγή https://www.techtarget.com/searchsoftwarequality/definition/waterfall-model	14
Εικόνα 2-2 Το μοντέλο Addie (Πηγή https://educationaltechnology.net/the-addie-model-instructional-design/)	15
Εικόνα 2-3 Παράδειγμα δρομολόγησης μηνυμάτων στο πρωτόκολλο AMQP (Πηγή https://www.rabbitmq.com/tutorials/amqp-concepts.html)	17
Εικόνα 2-4 Βασική αρχιτεκτονική του εργαλείου Celery-RabbitMQ (Πηγή https://tests4geeks.com/blog/python-celery-rabbitmq-tutorial/)	20
Εικόνα 2-5 Διάγραμμα ανάλυσης σήματος πολλαπλής ανάλυσης Πηγή: https://www.researchgate.net/publication/263929587_Comparison_of_Wavelet-Based_ANN_and_Regression_Models_for_Reservoir_Inflow_Forecasting/figures	25
Εικόνα 2-6 Παράδειγμα απλού δικτύου τροφοδοσίας προς τα εμπρός (Πηγή https://www.simplilearn.com/tutorials/deep-learning-tutorial/multilayer-perceptron)	27
Εικόνα 2-7 Παράδειγμα δομής επιπέδων ενός νευρωνικού δικτύου (Πηγή https://www.analyticsvidhya.com/blog/2022/03/basic-introduction-to-feed-forward-network-in-deep-learning/)	28
Εικόνα 2-8 Αρχιτεκτονική ενός συνελκτικού νευρωνικού δικτύου για την ταξινόμηση μιας εικόνας Πηγή: https://developersbreach.com/convolution-neural-network-deep-learning/	30
Εικόνα 3-1 Πλάνο ανάπτυξης εφαρμογής (Διάγραμμα Gantt)	38
Εικόνα 4-1 Αρχιτεκτονική του συστήματος της εφαρμογής	40
Εικόνα 4-2 Διάγραμμα σεναρίων χρήσης της εφαρμογής	42
Εικόνα 4-3 Διάγραμμα δραστηριοτήτων βασικού σεναρίου της εφαρμογής	43
Εικόνα 4-4 Διάγραμμα ακολουθιών εργασιών στο σύστημα μετά την υποβολή αρχείων από συνδεδεμένους χρήστες	44
Εικόνα 4-5 Διάγραμμα ακολουθιών για την εκτέλεση αλγορίθμων ενός αρχείου που έχει φορτωθεί στο σύστημα	45
Εικόνα 4-6 Πρωτότυπο σελίδας εγγραφής	46
Εικόνα 4-7 Πρότυπο σελίδας σύνδεσης του χρήστη	46
Εικόνα 4-8 Πρότυπη σελίδα ανάρτησης αρχείου πολυμέσων	47
Εικόνα 4-9 Πρότυπη σελίδα προβολής οπτικοακουστικού υλικού	48
Εικόνα 4-10 Πρότυπη σελίδα αναπαραγωγής βίντεο	49
Εικόνα 4-11 Πρότυπη σελίδα αναπαραγωγής αρχείου ήχου	49
Εικόνα 4-12 Πρότυπη αρχική σελίδα	50
Εικόνα 5-1 Διάγραμμα Ροής αυτοματοποιημένων διαδικασιών-αλγορίθμων επεξεργασίας για ένα πολυμεσικό αρχείο	58
Εικόνα 5-2 Σχήμα ταξινόμησης των κλάσεων του αλγόριθμου ταξινόμησης ήχου	59
Εικόνα 5-3 Αναπαράσταση των βημάτων εξαγωγής χαρακτηριστικών από δεδομένα ήχου για τη δημιουργία χρονικής-συχνοτικής αναπαράστασης	61
Εικόνα 5-4 Φασματογραφήματα με βάση το κυματίδιο Daubechies-4 και 6 επίπεδα αποσύνθεσης για καθεμία από τις 3 κλάσεις, (α) Μουσική, (β) Ομιλία, (γ) Άλλα (περιβαλλοντικοί ήχοι)	61
Εικόνα 5-5 Σχηματική αναπαράσταση της βασικής αρχιτεκτονικής του μοντέλου εκπαίδευσης. Δημιουργήθηκε με τη βοήθεια του εργαλείου Netron (https://netron.app/)	63
Εικόνα 5-6 Στιγμιότυπο αρχικής σελίδας του ιστοτόπου	65
Εικόνα 5-7 Στιγμιότυπο σελίδας σύνδεσης του χρήστη	66
Εικόνα 5-8 Στιγμιότυπο σελίδας δημιουργίας λογαριασμού χρήστη	66
Εικόνα 5-9 Στιγμιότυπο σελίδας μεταφόρτωσης οπτικοακουστικού υλικού	67
Εικόνα 5-10 Στιγμιότυπο σελίδας προβολής οπτικοακουστικού υλικού	68

Εικόνα 5-11 Στιγμιότυπο σελίδας αναπαραγωγής βίντεο.....	69
Εικόνα 5-12 Στιγμιότυπο σελίδας αναπαραγωγής ήχου	69
Εικόνα 5-13 Σελίδα περιεχομένου του εκάστοτε χρήστη	70
Εικόνα 5-14 Στιγμιότυπο σελίδας λογαριασμού του χρήστη	71
Εικόνα 5-15 Στιγμιότυπο δημοσίου προφίλ χρηστών	71
Εικόνα 5-16 Εργαλείο για την παρακολούθηση και διαχείριση εικονικών μηχανών και του εικονικού περιβάλλοντος σε πραγματικό χρόνο.	72
Εικόνα 5-17 Το περιβάλλον της εικονικής μηχανής του συστήματος	72
Εικόνα 5-18 Ο Πίνακας Ελέγχου (Dashboard) του WordPress	73
Εικόνα 6-1 Διάγραμμα χρήσης της CPU του συστήματος κατά την εκτέλεση αλγορίθμων (vSphere)	75
Εικόνα 6-2 Διάγραμμα λανθάνουσας κατάστασης δίσκων (Disk Latency) του συστήματος κατά την εκτέλεση αλγορίθμων (vSphere).	75
Εικόνα 6-3 Διάγραμμα δικτυακής κυκλοφορίας του συστήματος κατά την εκτέλεση αλγορίθμων (vSphere).	76
Εικόνα 6-4 Διάγραμμα απόδοσης εικονικού δίσκου (Virtual Disk Performance) του συστήματος κατά την εκτέλεση αλγορίθμων (vSphere).	76
Εικόνα 6-5 Αναπαράσταση χρήσης μνήμης και CPU του συστήματος κατά την εκτέλεση αλγορίθμων).	77
Εικόνα 6-6 Αναπαράσταση χρήσης GPU του συστήματος κατά την εκτέλεση αλγορίθμων.	77
Εικόνα 6-7 Αναπαράσταση ραβδογράμματος που δείχνει τις μέσες βαθμολογίες για καθεμία από τις δέκα ερωτήσεις του SUS (System Usability Scale)	80
Εικόνα 6-8 Καμπύλη ROC-AUC για το κυματίδιο Daubechies-4 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 0,75 δευτερόλεπτα.	83
Εικόνα 6-9 Καμπύλη ROC-AUC για το κυματίδιο Symlet-4 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 0,75 δευτερόλεπτα.	84
Εικόνα 6-10 Καμπύλη ROC-AUC για το κυματίδιο Coiflet-5 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 0,75 δευτερόλεπτα.	85
Εικόνα 6-11 Καμπύλη ROC-AUC για το κυματίδιο Biorthogonal-4.4 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 0,75 δευτερόλεπτα.	85
Εικόνα 6-12 Σύγκριση της μέσης ακρίβειας για τις διαφορετικές οικογένειες κυματιδίων σε επίπεδα αποσύνθεσης 4,5 και 6.....	86
Εικόνα 6-13 Καμπύλη ROC-AUC για το κυματίδιο Daubechies-4 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 1 δευτερολέπτου.	87
Εικόνα 6-14 Καμπύλη ROC-AUC για το κυματίδιο Daubechies-4 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 2 δευτερολέπτων.	88
Εικόνα 6-15 Καμπύλη ROC-AUC για το κυματίδιο Daubechies-4 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 3 δευτερολέπτων.	89
Εικόνα 6-16 Σύγκριση της μέσης ακρίβειας για το κυματίδιο Daubechies-4 για διαφορετικά μήκη σήματος και επιπέδων αποσύνθεσης	89
Εικόνα 6-17 Καμπύλη ROC-AUC για το κυματίδιο Daubechies-4 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 3 δευτερολέπτων στο σύνολο των δεδομένων μετά τη χρήση τεχνικών επαύξησης.....	90
Εικόνα 6-18 Καμπύλη ROC-AUC του τελικού μοντέλου ταξινόμησης του συστήματος ..	91
Εικόνα 6-19 Οι καμπύλες απώλειας και ακρίβειας εκπαίδευσης και επικύρωσης του τελικού μοντέλου του συστήματος.	91

Λίστα Πινάκων

Πίνακας 3-1 Μη λειτουργικές απαιτήσεις του συστήματος.....	35
Πίνακας 3-2 Πίνακας Χρησιμοποιούμενων Τεχνολογιών και Εργαλείων.....	37
Πίνακας 4-1 Πίνακας παρουσίασης των επιλεγμένων μοντέλων και αλγορίθμων του συστήματος	41
Πίνακας 5-1 Χαρακτηριστικά της Εικονικής Μηχανής που παρέχει το ΚΗΔ του Α.Π.Θ....	51
Πίνακας 5-2 Λίστα Εφαρμοζόμενων Αλγορίθμων και Βοηθητικών Σεναρίων.	57
Πίνακας 5-3 Παρουσίαση αρχιτεκτονικής του μοντέλου εκπαίδευσης	64
Πίνακας 6-1 Σύνοψη Μετρικών Απόδοσης του Συστήματος κατά την Εκτέλεση Αλγορίθμων	77
Πίνακας 6-2 Χαρακτηριστικά Συμμετεχόντων στην αξιολόγηση του περιβάλλοντος της εφαρμογής.....	78
Πίνακας 6-3 Ερωτήσεις του ερωτηματολογίου της Κλίμακας Ευχρηστίας Συστήματος (System Usability Scale)	79
Πίνακας 6-4 Τα αποτελέσματα του ερωτηματολογίου της Κλίμακας Ευχρηστίας Συστήματος για κάθε χρήστη και ο συνολικός μέσος όρος.....	80
Πίνακας 6-5 Βασικές μετρικές απόδοσης των μοντέλων των αλγορίθμων του συστήματος	81
Πίνακας 6-6 Πίνακας αποτελεσμάτων για το κυματίδιο Daubechies-4 σε 3 διαφορετικά επίπεδα αποσύνθεσης με μήκος σήματος 0,75 δευτερόλεπτα.	82
Πίνακας 6-7 Πίνακας αποτελεσμάτων για το κυματίδιο Symlet-4 σε 3 διαφορετικά επίπεδα αποσύνθεσης με μήκος σήματος 0,75 δευτερόλεπτα.	83
Πίνακας 6-8 Πίνακας αποτελεσμάτων για το κυματίδιο Coiflet-5 σε 3 διαφορετικά επίπεδα αποσύνθεσης με μήκος σήματος 0,75 δευτερόλεπτα.	84
Πίνακας 6-9 Πίνακας αποτελεσμάτων για το κυματίδιο Biorthogonal-4.4 σε 3 διαφορετικά επίπεδα αποσύνθεσης με μήκος σήματος 0,75 δευτερόλεπτα.	85
Πίνακας 6-10 Πίνακας αποτελεσμάτων για το κυματίδιο Daubechies-4 σε 3 διαφορετικά επίπεδα αποσύνθεσης με μήκος σήματος 1 δευτερολέπτου.....	86
Πίνακας 6-11 Πίνακας αποτελεσμάτων για το κυματίδιο Daubechies-4 σε 3 διαφορετικά επίπεδα αποσύνθεσης με μήκος σήματος 2 δευτερολέπτων.	87
Πίνακας 6-12 Πίνακας αποτελεσμάτων για το κυματίδιο Daubechies-4 σε 3 διαφορετικά επίπεδα αποσύνθεσης με μήκος σήματος 3 δευτερολέπτων	88
Πίνακας 6-13 Πίνακας αποτελεσμάτων τελικού μοντέλου για το κυματίδιο Daubechies-4 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 3 δευτερολέπτων στο σύνολο των δεδομένων μετά τη χρήση τεχνικών επαύξησης.	90
Πίνακας 6-14 Πίνακας αποτελεσμάτων του τελικού μοντέλου ταξινόμησης του συστήματος	91

Κεφάλαιο 1^ο

1. Εισαγωγή

Στην ψηφιακή εποχή, το οπτικοακουστικό περιεχόμενο έχει αναδειχθεί σε μια από τις πιο διαδεδομένες μορφές πληροφόρησης και ψυχαγωγίας. Το YouTube και το Vimeo, μεταξύ άλλων πλατφορμών, έχουν αλλάξει τον τρόπο με τον οποίο οι άνθρωποι καταγράφουν, δημοσιεύουν και προβάλλουν πολυμεσικά αρχεία. Ο ρυθμός με τον οποίο οι άνθρωποι αναρτούν βίντεο και αρχεία ήχου έχει αυξηθεί σημαντικά, παρουσιάζοντας έτσι μια τεράστια πρόκληση στην οργάνωση, τη διαχείριση και την αναζήτηση συγκεκριμένου περιεχομένου. Τα παραδοσιακά συστήματα βασίζονται, σε μεγάλο βαθμό, στα μεταδεδομένα και στους χειροκίνητους σχολιασμούς, οι οποίοι είναι συχνά ελλιπείς, ασυνεπείς ή ανεπαρκείς για να αποτυπώσουν τον σημασιολογικό πλούτο του οπτικοακουστικού υλικού (Naphade & Huang, 2002). Αυτοί οι περιορισμοί εμποδίζουν την εμπειρία των χρηστών, ιδίως σε εφαρμογές όπου η λεπτομερής εξερεύνηση και ανάλυση οπτικοακουστικών δεδομένων είναι κρίσιμες.

Η ανάγκη για προηγμένες τεχνικές σημασιολογικής επεξεργασίας και ανάκτησης έχει αναγνωριστεί ευρέως από την ακαδημαϊκή κοινότητα. Για παράδειγμα, οι Dimoulas et al. (2015), έχουν παρουσιάσει τις προοπτικές των συστημάτων οπτικοακουστικού υλικού στον σημασιολογικό ιστό και έχουν καταδείξει το ρόλο της σημασιολογικής ανάλυσης στην αναζήτηση και χρήση πληροφοριών πολυμέσων. Ομοίως, η τρέχουσα κατάσταση της τεχνητής νοημοσύνης έχει παράσχει τα απαραίτητα εργαλεία για τη δημιουργία ιδιαίτερα εξελιγμένων συστημάτων για την ανάκτηση περιεχομένου, συμπεριλαμβανομένων μοντέλων μετατροπής ομιλίας σε κείμενο, εργαλείων εξαγωγής θεμάτων και πλαισίων αναγνώρισης ενεργειών (Jabeen et al., 2022). Παρά τις εξελίξεις αυτές, τα περισσότερα τρέχοντα συστήματα εξακολουθούν να χρησιμοποιούν απλές αναζητήσεις με βάση λέξεις-κλειδιά, περιορίζοντας τις δυνατότητες για βαθιά σημασιολογική ανάλυση και ευφυή ανάκτηση περιεχομένου. Για τον λόγο αυτό, η παρούσα διπλωματική εργασία επικεντρώνεται στην ανάπτυξη ενός διαδικτυακού περιβάλλοντος διαχείρισης, σημασιολογικής επεξεργασίας και αναγνώρισης οπτικοακουστικού περιεχομένου, το οποίο αξιοποιεί σύγχρονες τεχνικές τεχνητής νοημοσύνης.

1.1. Θεώρηση του Προβλήματος

Η εκθετική αύξηση των οπτικοακουστικών δεδομένων στις διαδικτυακές πλατφόρμες έχει καταστήσει επιτακτική την ανάγκη για αποτελεσματικότερες τεχνικές οργάνωσης, επεξεργασίας και ανάκτησης πληροφοριών. Παρά τις προόδους στις τεχνολογίες αναζήτησης, τα σύγχρονα συστήματα εξακολουθούν να βασίζονται σε μηχανισμούς μεταδεδομένων και λέξεων κλειδιά που ορίζονται από τους χρήστες, γεγονός που δημιουργεί περιορισμούς. Ένα από τα βασικότερα προβλήματα είναι η ελλιπής σημασιολογική κατανόηση του περιεχομένου. Οι παραδοσιακές μέθοδοι αναζήτησης δεν λαμβάνουν υπόψη το νοηματικό πλαίσιο των βίντεο και των ηχητικών αρχείων, με αποτέλεσμα η ανάκτηση περιεχομένου να είναι συχνά ανακριβής ή ελλιπής. Επιπλέον, η αδυναμία αξιοποίησης τεχνικών πολυτροπικής επεξεργασίας, όπως η ταυτόχρονη ανάλυση ήχου, κειμένου και εικόνας, περιορίζει σημαντικά τη δυνατότητα εντοπισμού συγκεκριμένων πληροφοριών μέσα στα πολυμέσα (Dimoulas & Symeonidis, 2015). Παράλληλα, οι σύγχρονες πλατφόρμες, όπως το YouTube, δεν ενσωματώνουν σε βάθος εργαλεία σημασιολογικής ανάλυσης. Αντίθετα, βασίζονται σε απλές τεχνικές κατηγοριοποίησης, καθιστώντας την αναζήτηση περιεχομένου λιγότερο ακριβή και στοχευμένη. Αυτή η αδυναμία είναι ιδιαίτερα εμφανής σε εφαρμογές που απαιτούν

λεπτομερή ανάλυση δεδομένων, όπως η εκπαιδευτική χρήση πολυμέσων ή η αναζήτηση συγκεκριμένων φράσεων μέσα σε μεγάλα ηχητικά αρχεία.

1.2. Στόχος και Καινοτομία της Εργασίας

Στόχος της παρούσας διατριβής είναι η αντιμετώπιση αυτών των κενών με την ανάπτυξη ενός διαδικτυακού περιβάλλοντος για τη διαχείριση, τη σημασιολογική επεξεργασία και την αναγνώριση οπτικοακουστικού περιεχομένου. Το προτεινόμενο σύστημα ενσωματώνει καινοτόμες τεχνολογίες, όπως το Whisper για τη μετατροπή ομιλίας σε κείμενο, την αναγνώριση ονομαστικών οντοτήτων (name entity recognition) για την εξαγωγή θεμάτων (topics), την αναγνώριση ενεργειών για την κατηγοριοποίηση δραστηριοτήτων και την αναγνώριση ήχων για την κατηγοριοποίησή τους. Οι χρήστες μπορούν να αναρτούν αρχεία βίντεο και ήχου, να πραγματοποιούν αναζήτηση σε προφορικό περιεχόμενο και να διερευνούν σημασιολογικές σχέσεις μεταξύ θεμάτων. Συνδυάζοντας αυτές τις λειτουργίες, το σύστημα επιτρέπει την αποτελεσματικότερη και ουσιαστικότερη αλληλεπίδραση με το οπτικοακουστικό υλικό. Στα πλαίσια της παρούσας διατριβής αναπτύχθηκε ένας αλγόριθμος ταξινόμησης ήχου με τη χρήση καινοτόμων χαρακτηριστικών, ο οποίος έχει σχεδιαστεί για την κατηγοριοποίηση του ήχου σε μουσική, ομιλία και άλλα (περιβαλλοντικοί ήχοι) με πολύ υψηλή ακρίβεια.

Η καινοτομία αυτής της εργασίας έγκειται στην ικανότητά της να ενοποιεί την απομαγνητοφώνηση (transcription), τη σημασιολογική επεξεργασία και την ταξινόμηση σε μια ενιαία κλιμακούμενη πλατφόρμα. Με την ενσωμάτωση προηγμένων εργαλείων τεχνητής νοημοσύνης, το προτεινόμενο σύστημα επιτρέπει στους χρήστες να εξερευνούν και να ανακτούν οπτικοακουστικό περιεχόμενο με πιο ουσιαστικό και αποτελεσματικό τρόπο. Επιπλέον, η ασύγχρονη αρχιτεκτονική διαχείρισης εργασιών του, που υποστηρίζεται από το RabbitMQ και το Celery, εξασφαλίζει την επεκτασιμότητα και την αξιοπιστία, καθιστώντας την πλατφόρμα κατάλληλη για την επεξεργασία μεγάλου συνόλου δεδομένων. Όσον αφορά τον αλγόριθμο ταξινόμησης χρησιμοποιήθηκαν καινοτόμα χαρακτηριστικά, τα οποία βασίζονται στη χρήση φασματογραφήματος βασισμένου σε κυματίδια, που καταγράφει λεπτομερώς τις χρονικές και τι συχνотικές ιδιότητες του σήματος, προσφέροντας μια πιο ολοκληρωμένη αναπαράσταση του ηχητικού περιεχομένου σε σχέση με τις παραδοσιακές προσεγγίσεις.

1.3. Σύντομη παρουσίαση των κεφαλαίων της εργασίας

Η παρούσα διπλωματική εργασία χωρίζεται σε επτά κεφάλαια. Αρχικά στο κεφάλαιο 1 εισάγεται το πεδίο της έρευνας και αναλύεται το πρόβλημα που επιχειρεί να αντιμετωπίσει το σύστημα. Στο κεφάλαιο 2 παρουσιάζεται το θεωρητικό και μεθοδολογικό υπόβαθρο, ενώ ταυτόχρονα περιγράφονται οι βασικές έννοιες και τεχνολογίες που απαιτούνται για την υλοποίηση της προτεινόμενης εφαρμογής. Στο κεφάλαιο 3 γίνεται αναφορά στον στόχο της εφαρμογής και πραγματοποιείται η ανάλυση των χρηστών και των βασικών απαιτήσεων του συστήματος. Επίσης, αναλύονται τα εργαλεία, οι τεχνολογίες και οι αλγόριθμοι που χρησιμοποιήθηκαν για την ανάπτυξη του συστήματος. Στο κεφάλαιο 4 πραγματοποιείται και περιγράφεται ο σχεδιασμός για την ανάπτυξη και την υλοποίηση του επιθυμητού συστήματος. Γίνεται χρήση διαγραμμάτων για την καλύτερη κατανόηση της αρχιτεκτονικής του συστήματος, της αλληλεπίδρασης των επιμέρους μονάδων αλλά και της ροής των δεδομένων. Το κεφάλαιο 5 αφορά στην τεχνική υλοποίηση του συστήματος δίνοντας έμφαση στο τρόπο ανάπτυξης των βασικών συστατικών της εφαρμογής. Επίσης αναλύεται η διαδικασία ανάπτυξης και εκπαίδευσης του αλγορίθμου ταξινόμησης του ήχου. Στο τέλος του κεφαλαίου παρουσιάζεται η γραφική προβολή της διεπαφής χρήστη. Το κεφάλαιο 6 περιγράφει την αξιολόγηση της εφαρμογής και των αλγορίθμων. Τέλος, στο κεφάλαιο 7 καταγράφονται τα συμπεράσματα που προέκυψαν από την εργασία, η καινοτομία της και παρατίθενται προτάσεις για τη μελλοντική έρευνα και την επέκταση της εφαρμογής.

Κεφάλαιο 2^ο

2. Θεωρητικό και Μεθοδολογικό Υπόβαθρο

2.1. Διαχείριση Οπτικοακουστικών Πόρων

Η παρούσα διπλωματική εργασία αφορά στη δημιουργία ενός συστήματος διαχείρισης οπτικοακουστικού υλικού (Multimedia Content Management ή Media Asset Management). Ο όρος περιεχόμενο (content) χρησιμοποιείται για τον χαρακτηρισμό κάθε είδους οπτικής, οπτικοακουστικής ή λεκτικής πληροφορίας. Λόγω της ευρείας χρήσης του, η Society of Motion Picture and Television Engineers (SMPTE) και η European Broadcasting Union (EBU) όρισε το περιεχόμενο σε δύο έννοιες. Την ουσία (essence) ως τα δεδομένα εικόνας και ήχου (media data) και τα μεταδεδομένα (metadata), τα οποία χρησιμοποιούνται για να περιγράψουν την ουσία. Περιέχουν δηλαδή πληροφορίες για τα δεδομένα ήχου και εικόνας. Ένα σύστημα διαχείρισης πολυμεσικού υλικού πρέπει να παρέχει λογικές συνδέσεις ανάμεσα στο υλικό και στα μεταδεδομένα, ενώ παράλληλα οφείλει να αναπαριστά και να διαχειρίζεται τη χρονική σχέση ανάμεσα τους (Heitmann, 1999).

Ένα ολοκληρωμένο σύστημα διαχείρισης πολυμεσικού υλικού πρέπει να εξασφαλίζει τη σύνδεση μεταξύ του περιεχομένου και των μεταδεδομένων του. Τα μεταδεδομένα περιλαμβάνουν βασικές πληροφορίες, όπως ο τίτλος, ο δημιουργός, η ημερομηνία δημιουργίας και τα τεχνικά χαρακτηριστικά (ανάλυση, μορφή αρχείου), ενώ ταυτόχρονα παρέχουν σημασιολογικά δεδομένα για το ίδιο το περιεχόμενο. Ένα αποτελεσματικό σύστημα διαχείρισης οπτικοακουστικών πόρων πρέπει να διατηρεί αυτές τις συσχετίσεις, επιτρέποντας την εύκολη αναζήτηση και ανάκτηση του πολυμεσικού υλικού. Επιπλέον, η διαχείριση πολυμέσων απαιτεί την αναπαράσταση και διαχείριση των χρονικών σχέσεων μεταξύ διαφορετικών στοιχείων περιεχομένου. Για παράδειγμα, η σωστή συγχρονισμένη αναπαραγωγή εικόνας και ήχου σε ένα βίντεο ή η αντιστοίχιση υπότιτλων με τον διάλογο αποτελούν βασικές λειτουργίες ενός τέτοιου συστήματος. Τα σύγχρονα συστήματα διαχείρισης πολυμεσικού περιεχομένου βασίζονται σε εξελιγμένες αρχιτεκτονικές βάσεων δεδομένων και υπολογιστικού νέφους (cloud), ώστε να διαχειρίζονται αποτελεσματικά μεγάλους όγκους δεδομένων. Επιπρόσθετα, οι τεχνολογίες τεχνητής νοημοσύνης επιτρέπουν την αυτοματοποιημένη κατηγοριοποίηση και αναζήτηση περιεχομένου, βελτιώνοντας την ακρίβεια και αποδοτικότητα των διαδικασιών αυτών. Τέλος, η ασφάλεια και η προστασία των πνευματικών δικαιωμάτων είναι ζητήματα ζωτικής σημασίας. Για την αποφυγή μη εξουσιοδοτημένης χρήσης και την προστασία του περιεχομένου, τα σύγχρονα συστήματα ενσωματώνουν τεχνολογίες ψηφιακού υδατοσήμου (digital watermarking) και συστήματα διαχείρισης ψηφιακών δικαιωμάτων (Digital Rights Management - DRM).

2.1.1. Η Ανάγκη για Ευφυή Συστήματα

Ζούμε σε μια εποχή, στην οποία, εξαιτίας της ραγδαίας τεχνολογικής προόδου, τα μέσα ενημέρωσης έχουν αλλάξει σημαντικά. Τα τελευταία είκοσι χρόνια, ο κλάδος των μέσων ενημέρωσης έχει υποστεί μια ολοκληρωμένη στροφή προς τις ψηφιακές πλατφόρμες, με αποτέλεσμα τη συγκρότηση ενός δυναμικού και διαρκώς μεταβαλλόμενου τοπίου (Dimoulas & Veglis, 2023). Η ανάγκη για τη δημιουργία ενός εργαλείου ψηφιακών περιουσιακών στοιχείων απορρέει από τον αυξανόμενο όγκο περιεχομένου πολυμέσων που παράγεται από τους οργανισμούς. Καθώς αυξάνεται ο όγκος των περιουσιακών στοιχείων πολυμέσων, η αποτελεσματική διαχείριση και οργάνωσή τους γίνεται όλο και πιο πολύπλοκη. Η τεχνητή νοημοσύνη μπορεί να βελτιώσει τη διαχείριση τεράστιων

ποσοτήτων περιεχομένου μέσων ενημέρωσης και να βοηθήσει τους οργανισμούς μέσων ενημέρωσης να αντλήσουν πολύτιμες πληροφορίες από τα δεδομένα τους (Schinas et al., 2023).

Τα ευφυή συστήματα αξιοποιούν τεχνικές που βασίζονται στην τεχνητή νοημοσύνη, όπως η επεξεργασία φυσικής γλώσσας, η όραση υπολογιστών και η βαθιά μάθηση για την αυτοματοποίηση της κατηγοριοποίησης των μέσων και τη βελτίωση της δυνατότητας αναζήτησης. Για παράδειγμα, τα μοντέλα βαθιάς μάθησης μπορούν να παράγουν αυτόματα μεταδεδομένα αναγνωρίζοντας αντικείμενα, ομιλία και σκηνές μέσα σε περιεχόμενο πολυμέσων. Αυτές οι δυνατότητες μειώνουν σημαντικά τις προσπάθειες χειροκίνητου σχολιασμού και επιτρέπουν την αποτελεσματικότερη διαχείριση περιεχομένου. Επιπλέον, οι αλγόριθμοι ταξινόμησης ήχου και τα μοντέλα μετατροπής ομιλίας σε κείμενο διευκολύνουν την ευρετηρίαση και ανάκτηση περιεχομένου παρέχοντας σημασιολογικές γνώσεις πέρα από τις παραδοσιακές μεθόδους που βασίζονται σε λέξεις-κλειδιά. Εκτός από την ταξινόμηση, τα ευφυή συστήματα διαδραματίζουν σημαντικό ρόλο στη σύσταση περιεχομένου, την ανίχνευση ανωμαλιών και την προσαρμοστική ροή, βελτιώνοντας την εμπειρία του χρήστη σε πλατφόρμες ψηφιακών μέσων. Καθώς η πολυπλοκότητα των συστημάτων οπτικοακουστικού υλικού αυξάνεται, η ενσωμάτωση αυτοματισμών με τη χρήση της τεχνητής νοημοσύνης καθίσταται απαραίτητη για τη διατήρηση της αποτελεσματικότητας και της επεκτασιμότητας.

2.1.2. Ολοκληρωμένη Αρχιτεκτονική και Επικοινωνία Συστήματος

Για την επιτυχή λειτουργία ενός συστήματος διαχείρισης οπτικοακουστικού υλικού που ενσωματώνει ευφυείς δυνατότητες, είναι απαραίτητη μια καλά δομημένη αρχιτεκτονική που να υποστηρίζει την ασύγχρονη και διανεμημένη επεξεργασία των δεδομένων. Σε αυτό το πλαίσιο, υιοθετούνται αρχές παράλληλων και κατανεμημένων συστημάτων (distributed computing) και της αρχιτεκτονικής μικροϋπηρεσιών, όπου κάθε λειτουργική μονάδα αναλαμβάνει ένα συγκεκριμένο κομμάτι της συνολικής λειτουργικότητας του συστήματος. Ο διαχωρισμός των λειτουργιών οδηγεί σε αυξημένη επεκτασιμότητα και αξιοπιστία, καθώς κάθε υπηρεσία μπορεί να αναπτυχθεί, να αναβαθμιστεί ή να αντικατασταθεί ανεξάρτητα χωρίς να διαταράσσεται η συνολική λειτουργία.

Η ασύγχρονη διαχείριση των εργασιών επιτυγχάνεται μέσω της ενσωμάτωσης διαμεσολαβητών μηνυμάτων, οι οποίοι διαδραματίζουν κρίσιμο ρόλο στη διαχείριση της ροής δεδομένων μεταξύ των διαφόρων μονάδων του συστήματος. Με τη χρήση ουρών μηνυμάτων, οι υπηρεσίες δεν χρειάζεται να επικοινωνούν άμεσα, πράγμα που επιτρέπει την ανεξάρτητη λειτουργία κάθε στοιχείου. Επιπλέον, η χρήση τέτοιων τεχνολογιών εξασφαλίζει την αποτελεσματική κατανομή των πόρων, καθώς οι εργασίες ανατίθενται σε διαθέσιμους επεξεργαστές βάσει του τρέχοντος φόρτου, συμβάλλοντας έτσι στη βελτίωση της απόδοσης και της απόκρισης του συστήματος. Οι ευφυείς αλγόριθμοι που χρησιμοποιούνται για την αυτόματη κατηγοριοποίηση και αναζήτηση του περιεχομένου απαιτούν σημαντική υπολογιστική ισχύ και μπορούν να επωφεληθούν από τη διανεμημένη προσέγγιση της επεξεργασίας. Η αποστολή των εργασιών μέσω των διαμεσολαβητών μηνυμάτων επιτρέπει μια εύελικτη κατανομή των πόρων και την αυτοματοποίηση της ροής δεδομένων, δημιουργώντας μια ενοποιημένη διαδικασία που συνδυάζει μεταδεδομένα και περιεχόμενο με τρόπο αποδοτικό και αξιόπιστο.

Η ενοποίηση των ευφύων συστημάτων με τις τεχνολογίες διαμεσολάβησης μηνυμάτων καθιερώνει τα θεμέλια για μια σύγχρονη και επεκτάσιμη αρχιτεκτονική. Μέσω αυτής της προσέγγισης επιτυγχάνεται η διαχείριση της αυξημένης πολυπλοκότητας των συστημάτων οπτικοακουστικού υλικού, ενώ παράλληλα ενισχύεται η αποτελεσματικότητα των αυτοματοποιημένων διαδικασιών. Στην επόμενη ενότητα θα αναλυθεί αναλυτικότερα ο ρόλος του διαμεσολαβητή μηνυμάτων, με ιδιαίτερη αναφορά

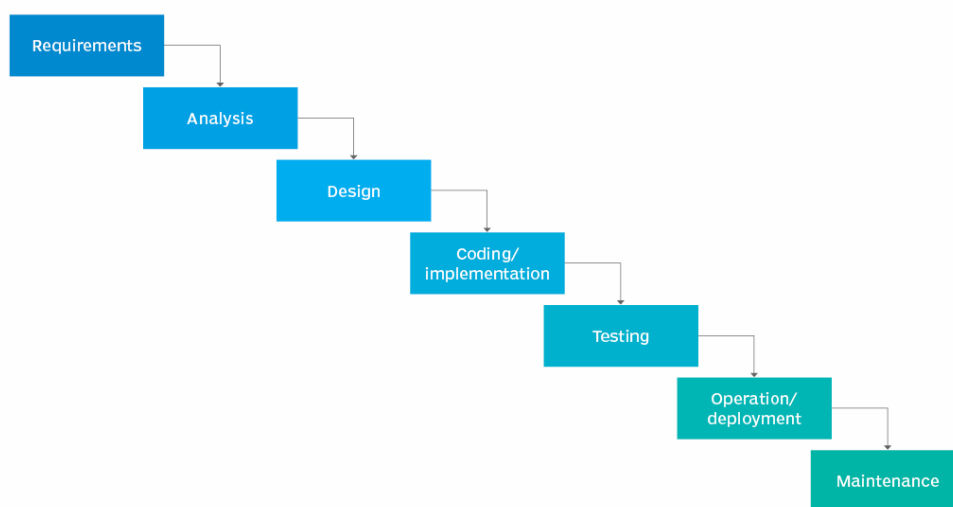
στην υλοποίηση και στις δυνατότητες που προσφέρει το RabbitMQ, ενσωματώνοντας επίσης την έννοια της διανεμημένης ουράς εργασιών.

2.2. Μεθοδολογία ανάπτυξης λογισμικού

Για την διαδικασία ανάπτυξης της παρούσας εφαρμογής συνδυάστηκαν το μοντέλο καταρράκτη και το μοντέλο ADDIE. Ο συνδυασμός αυτός επιλέχθηκε για να αντιμετωπιστούν τόσο οι τεχνικές όσο και οι χρηστο-κεντρικές (user-centric) πτυχές του προτεινόμενου συστήματος. Η πλατφόρμα, η οποία σχεδιάστηκε για τη διαχείριση, τη σημασιολογική επεξεργασία και την ανάκτηση οπτικοακουστικού περιεχομένου, απαιτούσε μια δομημένη μεθοδολογία για τον χειρισμό των πολύπλοκων τεχνικών συστατικών της, διασφαλίζοντας, παράλληλα, την ευχρηστία και την αποτελεσματικότητα της αλληλεπίδρασης της με τον χρήστη.

Το μοντέλο καταρράκτη επιλέχθηκε επειδή επικεντρώνεται στη διαδοχική ανάπτυξη, κατά την οποία, κάθε φάση βασίζεται στην προηγούμενη (Εικόνα 2-1). Αυτή η δομημένη προσέγγιση συμβαδίζει με τις απαιτήσεις του έργου, οι οποίες ήταν σαφώς καθορισμένες από την αρχή και επικεντρώθηκαν στην παροχή συγκεκριμένων λειτουργιών. Το μοντέλο χωρίζει την ανάπτυξη σε έξι βασικές φάσεις, η κάθε μία από τις οποίες ολοκληρώνεται πριν από τη μετάβαση στην επόμενη (Royce, 1970).

Waterfall model



Εικόνα 2-1 Το μοντέλο του καταρράκτη [Πηγή](https://www.techtarget.com/searchsoftwarequality/definition/waterfall-model)
<https://www.techtarget.com/searchsoftwarequality/definition/waterfall-model>

Στην παρούσα εργασία ενώ το μοντέλο του καταρράκτη δόμησε τις τεχνικές φάσεις, το μοντέλο ADDIE (Εικόνα 2-2) αφορούσε στο σχεδιασμό του συστήματος με επίκεντρο τον χρήστη. Το μοντέλο ADDIE παρείχε ένα πλαίσιο για την λειτουργικότητα του συστήματος με γνώμονα τις ανάγκες των χρηστών και την αξιολόγηση της αποτελεσματικότητάς του (Branch, 2010). Με τον τρόπο αυτό εξασφαλίστηκε ότι τα χαρακτηριστικά του συστήματος δεν ήταν μόνο άρτια τεχνικά αλλά και εύκολα προσβάσιμα.



Εικόνα 2-2 Το μοντέλο Addie (Πηγή <https://educationaltechnology.net/the-addie-model-instructional-design/>)

Το μοντέλο καταρράχτη προσέφερε μια ισχυρή δομή για την υλοποίηση του τεχνικού υποβάθρου, ενώ το ADDIE εξασφάλιζε ότι ο σχεδιασμός και η λειτουργικότητα του συστήματος αξιολογούνταν συνεχώς για τη χρηστικότητα και την αποτελεσματικότητά του. Ο συνδυασμός των παραπάνω μοντέλων ήταν κατάλληλος τόσο για την εστίαση του συστήματος στην τεχνική απόδοση όσο για την αλληλεπίδραση του συστήματος με τον χρήστη, επιτρέποντας τη δημιουργία μιας κλιμακούμενης, αποτελεσματικής και φιλικής προς τον χρήστη πλατφόρμας για τη διαχείριση οπτικοακουστικού περιεχομένου.

2.2.1. Συστήματα Διαχείρισης Περιεχομένου

Ως Σύστημα Διαχείρισης Περιεχομένου ορίζεται μια μορφή λογισμικού που περιέχει εφαρμογές, οι οποίες παρέχουν την δυνατότητα σε ένα άτομο να διαχειρίζεται το διαδικτυακό του περιεχόμενο, όπως εικόνες, κείμενα, αρχεία βίντεο κ.ο.κ. , με εύκολο τρόπο. Οι εφαρμογές διαχείρισης περιεχομένου παρέχουν την δυνατότητα της δημιουργίας μιας ιστοσελίδας χωρίς να είναι απαραίτητο να υπάρχουν ειδικές γνώσεις και χωρίς να απαιτείται η ανάπτυξη κώδικά, από έναν χρήστη, από την αρχή (Boiko, 2005). Το Σύστημα Διαχείρισης Περιεχομένου, πάνω στο οποίο βασίστηκε ο σχεδιασμός του ιστοχώρου “Ar.U.TV”, είναι το Wordpress. Είναι ένα ελεύθερο (διατίθεται υπό την GNU GPL2) και ανοιχτού κώδικα λογισμικό που χρησιμοποιείται για τη δημιουργία ιστοσελίδων και εφαρμογών. Είναι γραμμένο σε κώδικα PHP, χρησιμοποιεί την MySQL ως βάση δεδομένων και μπορεί να εγκατασταθεί σε οποιοδήποτε λειτουργικό σύστημα (Blog Tool Publishing Platform and CMS, 2024).

Το Wordpress διαθέτει φιλικό προς το χρήστη περιβάλλον και είναι εύκολο στη χρήση. Διευκολύνει την τροποποίηση της εμφάνισης, όσο και του περιεχομένου του ιστοτόπου, με εύκολο τρόπο, χωρίς την ανάγκη ανασυσγκρότησης ολόκληρων στατικών σελίδων. Αντί για αυτό, εμφανίζει τις σελίδες γρήγορα και αποτελεσματικά. Επιπλέον, οι χρήστες μπορούν εύκολα να δημιουργήσουν σελίδες μέσα από το πίνακα ελέγχου(dashboard) του Wordpress για την υλοποίηση διαφόρων ειδών ιστοτόπων. Διαθέτει πολλά θέματα (Themes), τα οποία αφορούν στον αισθητικό σχεδιασμό, στη διάταξη και τα χρώματα της ιστοσελίδας, βελτιώνοντας την ποιότητα διάδρασης του χρήστη με την ιστοσελίδα. Επιπλέον, οι προγραμματιστές μπορούν να δημιουργήσουν θέματα παιδιά (Child

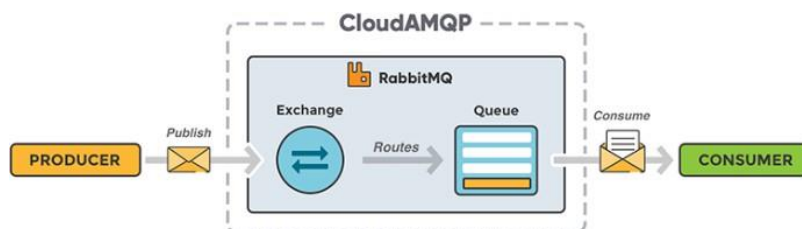
Themes), τα οποία κληρονομούν τη λειτουργικότητα και την εμφάνιση ενός κύριου θέματος, επιτρέποντας την ασφαλή προσαρμογή του ιστοτόπου χωρίς να διακυβεύεται η απώλεια των αλλαγών κατά τις ενημερώσεις του κύριου θέματος. Αξιοσημείωτο είναι το γεγονός ότι διαθέτει μία μεγάλη διαδικτυακή κοινότητα και μπορεί να συμβάλλει στην επίλυση προβλημάτων που προκύπτουν μεταξύ των προγραμματιστών. Κάθε κομμάτι κώδικα που δημιουργείται από το Wordpress συμμορφώνεται πλήρως με τα πρότυπα που ορίζει το World Wide Web Consortium (W3C). Αυτό σημαίνει ότι η ιστοσελίδα θα κατέχει την ικανότητα να λειτουργήσει σε όλους τους τωρινούς διαδικτυακούς ιστούς περιήγησης, διατηρώντας παράλληλα την συμβατότητα και με ιστούς περιήγησης επόμενης γενιάς. Διαθέτει, επίσης, ένα ευρύ φάσμα προσθέτων (Plugins), τα οποία διατίθενται στον πίνακα ελέγχου (dashboard), συνεισφέροντας, έτσι, στην ενίσχυση των λειτουργιών και των δυνατοτήτων μίας ιστοσελίδας. Τα πρόσθετα είναι τμήματα κώδικα που προστίθενται στο σύστημα και δίνουν επιπλέον δυνατότητες. Το σύστημα προσθέτων παρέχει στους χρήστες τη δυνατότητα δημιουργίας δικών τους προσθέτων, χρησιμοποιώντας τα APIs και τα Hooks του Wordpress. Τα Hooks αποτελούν ένα τρόπο αλληλεπίδρασης ή τροποποίησης μεταξύ κομματιών κώδικα σε συγκεκριμένα, προκαθορισμένα σημεία. Αποτελούν τη βάση για τον τρόπο με τον οποίο τα πρόσθετα και τα θέματα αλληλοεπιδρούν με τον πυρήνα του Wordpress, ενώ χρησιμοποιούνται και από τον ίδιο τον πυρήνα. Επιπρόσθετα, υπάρχουν οι προσαρμοσμένοι τύποι περιεχομένου (Custom Post Types - CPTs) που προσφέρουν τη δυνατότητα δημιουργίας και διαχείρισης ειδικών κατηγοριών περιεχομένου, πέρα από τις προεπιλεγμένες δυνατότητες του WordPress. Για παράδειγμα, μπορούν να χρησιμοποιηθούν για την οργάνωση αρχείων πολυμέσων, προϊόντων ή οποιασδήποτε άλλης μορφής περιεχομένου που απαιτείται από τον ιστότοπο.

Τέλος, ως πλαίσιο εφαρμογής, το Wordpress παρουσιάζει ένα εκτενές σύνολο χαρακτηριστικών για την ανάπτυξη εφαρμογών, περιλαμβάνοντας μεταφράσεις, διαχείριση χρηστών, αιτήσεις HTTP, βάσεις δεδομένων και δρομολόγηση Ενιαίου Εντοπιστή Πόρων (Uniform Resource Locator). Το REST API της πλατφόρμας επιταχύνει, επίσης, την αλληλεπίδραση. Με τους προσαρμοσμένους τύπους περιεχομένου, εκτός από τις προεπιλεγμένες δυνατότητες του Wordpress, διατίθεται η χρήση μερικών γραμμών κώδικα, ώστε να επιτρέπεται η δημιουργία προσαρμοσμένων τύπων αναρτήσεων, ταξινομιών και μεταδεδομένων.

2.3. Αρχιτεκτονική Επικοινωνίας του Συστήματος

Η τεχνολογία που χρησιμοποιήθηκε για την εκκίνηση των αλγόριθμων των αρχείων πολυμέσων, που αναρτούν οι χρήστες στην ιστοσελίδα, είναι η διαμεσολάβηση μηνυμάτων. Ένας διαμεσολαβητής μηνυμάτων μπορεί να χαρακτηριστεί ως ένα πρόγραμμα ενδιάμεσο που επιτρέπει σε διάφορες εφαρμογές (ακόμα και εάν χρησιμοποιούν διαφορετικές γλώσσες προγραμματισμού) και υπηρεσίες να επικοινωνούν μεταξύ τους και να μοιράζονται πληροφορίες. Ο κύριος ρόλος του είναι να εξασφαλίζουν ότι τα μηνύματα μεταφέρονται στον διαμεσολαβητή, όπου και αποθηκεύονται σε κατάλληλες δομές που ονομάζονται ουρές (Queues) και στη συνέχεια, δρομολογούνται στους συνδεδεμένους καταναλωτές, σύμφωνα με προκαθορισμένους κανόνες.

Ο διαμεσολαβητής μηνυμάτων, που στην προκειμένη περίπτωση είναι το RabbitMQ, αποτελεί έναν ευρέως διαδεδομένο διαμεσολαβητή μηνυμάτων ανοιχτού κώδικα ο οποίος υποστηρίζει διάφορα πρωτόκολλα. Το RabbitMQ μπορεί να αναπτυχθεί σε διανεμημένα συστήματα για να ικανοποιήσει απαιτήσεις μεγάλης κλίμακας και υψηλής διαθεσιμότητας (RabbitMQ, χ.χ).



Εικόνα 2-3 Παράδειγμα δρομολόγησης μηνυμάτων στο πρωτόκολλο AMQP ([Πηγή https://www.rabbitmq.com/tutorials/amqp-concepts.html](https://www.rabbitmq.com/tutorials/amqp-concepts.html))

2.3.1. Βασικές έννοιες και Πρωτόκολλο Ουράς Μηνυμάτων

Η λειτουργία ενός συστήματος διαμεσολάβησης μηνυμάτων βασίζεται σε ένα πρωτόκολλο επικοινωνίας που καθορίζει τους κανόνες ανταλλαγής δεδομένων. Το Προηγμένο Πρωτόκολλο Ουράς Μηνυμάτων (Advanced Message Queuing Protocol AMQP) είναι ένα πρωτόκολλο ανταλλαγής μηνυμάτων που επιτρέπει στις εφαρμογές να στέλνουν και να λαμβάνουν μηνύματα ενός διαμεσολαβητή.

Λειτουργώντας ως πρωτόκολλο στο επίπεδο εφαρμογής, χρησιμοποιεί το Transmission Control Protocol (TCP) για να διασφαλίσει την αξιόπιστη παράδοση των μηνυμάτων. Συγκεκριμένα, η έκδοση του πρωτοκόλλου που χρησιμοποιείται σηματοδοτείται ως 0-9-1. Οι λειτουργικές δυναμικές του πρωτοκόλλου ανταλλαγής μηνυμάτων αναλύονται ως εξής. Πρώτα απ' όλα τα μηνύματα διανέμονται σε πρωταρχικούς κόμβους ανταλλαγής (Exchanges) του μεσολαβητή, παρόμοιους με λειτουργικές μονάδες που θυμίζουν γραμματοκιβώτια. Στη συνέχεια, αυτοί οι πρωταρχικοί κόμβοι διανέμουν αντίγραφα των μηνυμάτων σε ουρές με βάση συγκεκριμένους κανόνες (Bindings) που καθορίζουν τη σύνδεση μεταξύ των κόμβων και των ουρών. Τελικά, ο μεσολαβητής είτε παραδίδει τα μηνύματα σε καταναλωτές που έχουν εγγραφεί σε συγκεκριμένες ουρές, είτε οι καταναλωτές λαμβάνουν μηνύματα από ουρές κατόπιν των αιτημάτων που έχουν υποβάλει.

Επιπλέον, οι εικονικοί διαμεσολαβητές (Virtual Hosts) παρέχουν τη δυνατότητα διαχείρισης πολλαπλών ανεξάρτητων περιβαλλόντων μέσα στο ίδιο σύστημα διαμεσολάβησης. Για να καταστεί δυνατό για έναν διαμεσολαβητή να φιλοξενεί πολλαπλά απομονωμένα περιβάλλοντα (ομάδες χρηστών, κόμβοι ανταλλαγής, ουρές και ούτω καθεξής), το AMQP 0-9-1 περιλαμβάνει την έννοια των εικονικών διαμεσολαβητών (vhosts). Η έννοια αυτή είναι παρόμοια με εικονικούς κεντρικούς υπολογιστές που χρησιμοποιούνται από πολλούς διαδικτυακούς διακομιστές (web servers) και παρέχουν εντελώς απομονωμένα περιβάλλοντα στα οποία ζουν οντότητες AMQP. Οι χρήστες του πρωτοκόλλου καθορίζουν ποιον vhost θέλουν να χρησιμοποιήσουν κατά την εγκαθίδρυση της σύνδεσης.

2.3.2. Δομή και Ροή Μηνυμάτων

Για να εξασφαλιστεί η σωστή δρομολόγηση των μηνυμάτων, χρησιμοποιούνται συγκεκριμένες δομές. Οι κόμβοι ανταλλαγής (Exchanges) λειτουργούν ως το αρχικό

σημείο για τη μετάδοση μηνυμάτων, αντιπροσωπεύοντας το κύριο κανάλι μέσω του οποίου κατευθύνονται τα μηνύματα. Ο ρόλος τους περιλαμβάνει την προώθηση μηνυμάτων προς καθορισμένες ουρές, σύμφωνα με προκαθορισμένους κανόνες που θεσπίζονται κατά τη δημιουργία τους. Οι κανόνες αυτοί καθορίζουν τις συσχετίσεις μεταξύ κάθε κόμβου ανταλλαγής και των ουρών, καθορίζουν τον τύπο του κόμβου ανταλλαγής, και διατυπώνουν το κλειδί σύνδεσης, το οποίο δείχνει μοναδικά τη σύνδεση μεταξύ ενός κόμβου ανταλλαγής και μιας ουράς. Στο πρωτόκολλο AMQP 0-9-1, περιγράφονται τέσσερις διαφορετικοί τύποι κόμβων ανταλλαγής. Πρώτος είναι ο άμεσος κόμβος (Direct Exchange) που προωθεί ένα μήνυμα σε μια συνδεδεμένη ουρά, μόνο εάν το κλειδί δρομολόγησης του μηνύματος ταυτίζεται με το κλειδί σύνδεσης του κόμβου ανταλλαγής με την ουρά. Έπειτα, υπάρχει ο κόμβος ευρείας μετάδοσης (Fanout Exchange), όπως υποδηλώνει και το όνομα του, αντιγράφει και δρομολογεί τα μηνύματα σε όλες τις ουρές που είναι συνδεδεμένες μαζί του, ανεξαρτήτως των κλειδιών δρομολόγησης. Αυτού του είδους οι κόμβοι είναι χρήσιμοι όταν θέλουμε να επεξεργαστούμε το ίδιο μήνυμα με διαφορετικούς τρόπους και αυτό αποτελεί το λόγο για τον οποίο χρησιμοποιήθηκε ο συγκεκριμένος κόμβος στο σύστημα που αναπτύχθηκε. Επίσης, ένα άλλος τύπος κόμβου είναι ο κόμβος θεμάτων (Topic Exchange) που είναι παρόμοιος με τον άμεσο κόμβο, αλλά, αντί για κλειδί σύνδεσης, ορίζεται ένα πρότυπο δρομολόγησης για κάθε σύνδεση του κόμβου με μια ουρά. Εάν το κλειδί δρομολόγησης των μηνυμάτων του κόμβου ταιριάζει με το πρότυπο δρομολόγησης που έχει οριστεί για τη συγκεκριμένη σύνδεση, τότε τα μηνύματα προωθούνται σε μια Ουρά. Τελευταίος τύπος κόμβου είναι ο κόμβος κεφαλίδων (Headers Exchange): Σε αυτή την περίπτωση, το κλειδί δρομολόγησης του μηνύματος αγνοείται και αντ' αυτού, ελέγχονται οι παράμετροι κεφαλίδας κάθε μηνύματος. Σε περίπτωση που ταιριάζουν με το κλειδί σύνδεσης μεταξύ ενός κόμβου ανταλλαγής κεφαλίδας και μιας ουράς, τότε προωθούνται σε αυτήν. Οι κόμβοι ανταλλαγής μπορεί να είναι είτε μόνιμοι (durable), είτε παροδικοί (transient). Οι μόνιμοι κόμβοι ανταλλαγής επιβιώνουν από μια επανεκκίνηση του διαμεσολαβητή, ενώ οι παροδικοί όχι (πρέπει να δηλωθούν ξανά όταν ο διαμεσολαβητής επαναλειτουργήσει). Επιπλέον, οι ουρές (Queues) στο RabbitMQ είναι δομές σειριακών δεδομένων που λειτουργούν με την μέθοδο FIFO- First in First Out. Αποτελούν τη βασική δομή βραχυπρόθεσμης αποθήκευσης μηνυμάτων στον διαμεσολαβητή, πριν αυτά παραδοθούν στους καταναλωτές για επεξεργασία. Μια ουρά απαιτείται να δηλωθεί για να χρησιμοποιηθεί. Η δήλωση της έχει ως αποτέλεσμα τη δημιουργία της (εάν δεν υπάρχει ήδη). Ωστόσο, εάν η ουρά υπάρχει ήδη και διαθέτει ίδια χαρακτηριστικά όπως αναφέρονται στη δήλωσή της, η δήλωση αυτή δεν έχει καμία επίπτωση. Οι ουρές μπορούν να οριστούν και αυτές ως μόνιμες (durable) ή παροδικές (transient). Στην πρώτη περίπτωση, τα μεταδεδομένα της ουράς αποθηκεύονται στον δίσκο, ενώ στη δεύτερη, αποθηκεύονται στην μνήμη, όταν αυτό είναι δυνατό. Σε περιπτώσεις που η ανθεκτικότητα είναι σημαντική, τότε εκτός από τις ουρές, θα πρέπει να ορίζονται και τα ίδια τα μηνύματα ως μόνιμα.

Οι συνδέσεις και κανάλια (Connections & Channels) επιτρέπουν την επικοινωνία μεταξύ των διαφορετικών συνιστωσών του συστήματος, διευκολύνοντας τη ροή των μηνυμάτων. Το AMQP 0-9-1 αποτελεί ένα πρωτόκολλο, οι συνδέσεις του οποίου χαρακτηρίζονται από μακρά διάρκεια. Το AMQP 0-9-1 βασίζεται στο Πρωτόκολλο Ελέγχου Μετάδοσης (TCP- Transmission Control Protocol) για την αξιόπιστη και ταχεία παράδοση μηνυμάτων. Οι συνδέσεις προστατεύονται με τη χρήση του πρωτοκόλλου Transport Layer Security (TLS). Λόγω της υψηλής κατανάλωσης πόρων και της πολυπλοκότητας από την οποία διέπεται το τείχος προστασίας (Firewall), δεν προτείνεται η διατήρηση πολλών ανοιχτών συνδέσεων TCP ταυτόχρονα, παρά το γεγονός ότι ορισμένες εφαρμογές απαιτούν πολλαπλές συνδέσεις με το διαμεσολαβητή. Στο AMQP 0-9-1 χρησιμοποιούνται κανάλια, τα οποία, επί της ουσίας, θεωρούνται «ελαφριές

συνδέσεις» που μοιράζονται μία μόνο σύνδεση TCP. Ένα κανάλι δεν μπορεί να υπάρξει αυτοτελώς. Επομένως, είναι απαραίτητη η ύπαρξη μίας σύνδεσης. Επιπρόσθετα, το κανάλι κλείνει αυτομάτως όταν κλείσει η σύνδεση. Τέλος, σε περιπτώσεις που οι εφαρμογές χρησιμοποιούν πολλαπλά νήματα/διεργασίες για επεξεργασία, είναι κοινή πρακτική να δημιουργείται ένα νέο κανάλι για κάθε νήμα/διεργασία, αποφεύγοντας, έτσι, την κοινή χρήση καναλιών μεταξύ τους.

2.3.3. Διαχείριση Μηνυμάτων

Η επεξεργασία των μηνυμάτων πραγματοποιείται από τους παραγωγούς (Producers), οι οποίοι δημιουργούν και αποστέλλουν τα μηνύματα και τους Καταναλωτές (Consumers), που τα λαμβάνουν και τα επεξεργάζονται. Για να διασφαλιστεί η ορθή παράδοση, το σύστημα χρησιμοποιεί επιβεβαιώσεις λήψης (Acknowledgements), οι οποίες ενημερώνουν ότι ένα μήνυμα παραδόθηκε επιτυχώς και μπορεί να διαγραφεί από την ουρά. Συγκεκριμένα ως παραγωγοί ονομάζονται οι εφαρμογές οι οποίες συνδέονται σε έναν διαμεσολαβητή μηνυμάτων, αυθεντικοποιούνται, εγκαθιδρύουν μια σύνδεση μέσα στην οποία ανοίγουν ένα κανάλι και τέλος, δημοσιεύουν τα παραγόμενα μηνύματά τους σε έναν κόμβο ανταλλαγής. Οι παραγωγοί έχουν μακράς διάρκειας ζωή και συνήθως, ανοίγουν τις συνδέσεις τους κατά την εκκίνηση της εφαρμογής. Συνήθως παραμένουν ενεργοί όσο καιρό παραμένει ενεργή η σύνδεση τους ή όσο η εφαρμογή εκτελείται. Έπειτα, ως καταναλωτής ορίζεται μια εφαρμογή που «καταναλώνει» μηνύματα τα οποία βρίσκονται σε μια ουρά. Στο μοντέλο AMQP 0-9-1, υπάρχουν δύο τρόποι με τους οποίους οι εφαρμογές μπορούν να επιτύχουν την κατανάλωση μηνυμάτων. Ο πρώτος είναι με push API. Με τον τρόπο αυτό, όποτε εισέρχονται νέα μηνύματα στην ουρά, αυτά ωθούνται στην εφαρμογή – καταναλωτή. Αποτελεί τη συνιστώμενη μέθοδο της εγγραφής σε ουρά. Ο δεύτερος είναι ο pull API ελέγχοντας επανειλημμένα την Ουρά για την ύπαρξη νέων μηνυμάτων σε αυτήν. Ωστόσο, θα πρέπει να αποφεύγεται στις περισσότερες περιπτώσεις διότι θεωρείται ιδιαίτερα αναποτελεσματική.

Η σύνδεση του καταναλωτή με τον διαμεσολαβητή είναι μακράς διάρκειας, δηλαδή, εγκαθιδρύεται μια φορά και μέσω αυτής της σύνδεσης, λαμβάνονται όλα τα μηνύματα τα οποία πρέπει να ληφθούν στη διάρκεια ζωής του καταναλωτή. Επίσης είναι δυνατό, περισσότεροι από έναν καταναλωτή να είναι συνδεδεμένοι σε μία ουρά.

Τα δίκτυα μπορεί να είναι αναξιόπιστα ή να συντρέξουν κάποιοι λόγοι, όπως απώλεια σύνδεσης, για τους οποίους οι καταναλωτές ενδέχεται να αποτύχουν την επεξεργασία μηνυμάτων. Για την αντιμετώπιση αυτού του προβλήματος στο AMQP πρωτόκολλο συμπεριλαμβάνονται οι Επιβεβαιώσεις Λήψης μηνυμάτων (Acknowledgements). Όταν αυτές οι λειτουργίες είναι ενεργοποιημένες, ο καταναλωτής αποστέλλει μια ειδοποίηση στον διαμεσολαβητή μηνυμάτων, σηματοδοτώντας την επιτυχή λήψη του μηνύματος και την ολοκλήρωση της σχετικής επεξεργασίας. Αυτό μπορεί να πραγματοποιηθεί είτε αυτόματα κατά την λήψη του μηνύματος από τον καταναλωτή, είτε μετά την ολοκλήρωση της επεξεργασίας του. Τα μηνύματα αφαιρούνται από τις ουρές του διαμεσολαβητή μηνυμάτων μόνο όταν αυτός λάβει την επιβεβαίωση λήψης από τον καταναλωτή. Εναλλακτικά, εάν περάσει κάποιο χρονικό όριο χωρίς να δεχθεί επιβεβαίωση λήψης για κάποιο μήνυμα ο διαμεσολαβητής, το δρομολογεί εκ νέου σε έναν καταναλωτή.

2.4. Διαχείριση Ασύγχρονων Εργασιών

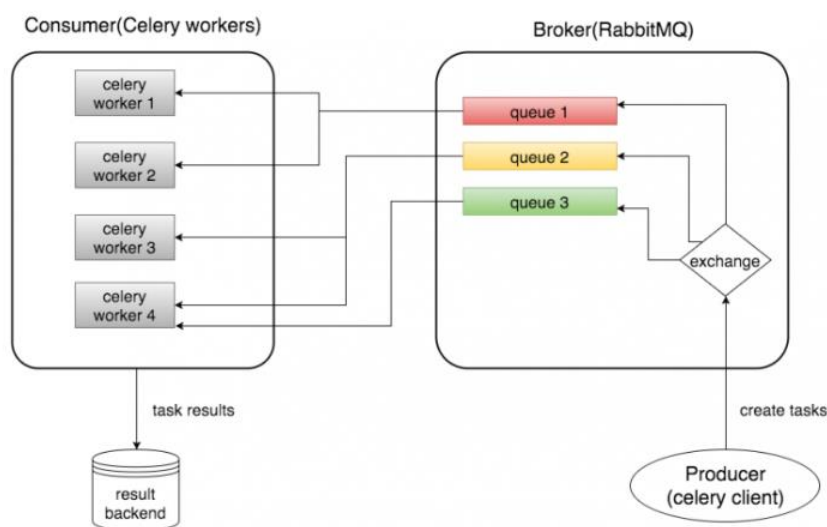
Ήδη από τα τέλη του 20^{ου} αιώνα, οι Boudreau et al. (1998) προσδιορίζουν τρόπους διάρθρωσης, επεξεργασίας και κατανομής εργασιών και επικοινωνίας με σκοπό την υπέρβαση των χρονικών και χωρικών ορίων. Η ανταλλαγή δεδομένων και μηνυμάτων μπορεί να πραγματοποιείται με τρόπο ασύγχρονο, να εκτελείται, δηλαδή, σε διαφορετικούς χρόνους (Montoya-Weiss et al., 2001). Η ασύγχρονη επικοινωνία επιτρέπει τη μετάδοση δεδομένων ή αιτημάτων χωρίς την ανάγκη διατήρησης μιας συνεχιζόμενης

σύνδεσης μεταξύ του πελάτη και του διακομιστή. Στο πλαίσιο του συστήματος, η εργαλειοθήκη Celery υλοποιεί αυτήν την αρχιτεκτονική, επιτρέποντας την εκτέλεση των αλγορίθμων, χωρίς να επηρεάζεται η ροή της κύριας εφαρμογής.

2.4.1. Κατανεμημένη Ουρά Εργασιών

Καθώς η πολυπλοκότητα των σύγχρονων εφαρμογών αυξάνεται, η ανάγκη για αποτελεσματική διαχείριση των ασύγχρονων εργασιών κρίνεται επιτακτική. Η ασύγχρονη επεξεργασία επιτρέπει την εκτέλεση των εργασιών στο παρασκήνιο, χωρίς να εμποδίζεται η ροή της κύριας εφαρμογής, γεγονός που είναι ιδιαίτερα χρήσιμο σε συστήματα που χειρίζονται υπολογιστικά εντατικές εργασίες ή απαιτούν άμεση απόκριση. Οι ουρές εργασιών (task queues) είναι εργαλεία που έχουν σχεδιαστεί για την αποτελεσματική κατανομή του φόρτου εργασίας σε πολλαπλά νήματα ή μηχανές. Κάθε ουρά εργασιών δέχεται μονάδες εργασίας γνωστές ως εργάτες (workers). Η βασική αρχιτεκτονική του Celery αποτυπώνεται στην Εικόνα 2-4.

Το Celery λειτουργεί με την ανταλλαγή μηνυμάτων, χρησιμοποιώντας, συνήθως, έναν διαμεσολαβητή μηνυμάτων για το συντονισμό της επικοινωνίας μεταξύ των πελατών και των εργατών (μονάδων εργασιών). Όταν ένας πελάτης χρειάζεται να εκτελέσει μια εργασία, στέλνει ένα μήνυμα στην ουρά. Στη συνέχεια, ο διαμεσολαβητής μηνυμάτων δρομολογεί αυτό το μήνυμα σε ένα διαθέσιμο εργάτη για επεξεργασία. Στην παρούσα εργασία ο διαμεσολαβητής μηνυμάτων είναι το RabbitMQ.



Εικόνα 2-4 Βασική αρχιτεκτονική του εργαλείου Celery-RabbitMQ
(Πηγή <https://tests4geeks.com/blog/python-celery-rabbitmq-tutorial/>)

2.4.2. Εκτέλεση Εργασιών Μέσω του Εργαλείου Celery

Το Celery είναι ένα ευρέως διαδεδομένο εργαλείο για την εκτέλεση ασύγχρονων εργασιών. Είναι απλό στην χρήση του, καθώς δεν απαιτεί αρχεία διαμόρφωσης. Οι εργάτες και οι πελάτες επαναλαμβάνουν αυτόματα την έναρξη μιας εργασίας σε περίπτωση απώλειας σύνδεσης ή αποτυχίας. Μια μεμονωμένη διεργασία Celery μπορεί να επεξεργάζεται εκατομμύρια εργασίες το λεπτό, με καθυστέρηση κυκλικής διαδρομής κάτω του χιλιοστού του δευτερολέπτου. Το "Celery" θεωρείται ευέλικτο αφού σχεδόν κάθε μέρος του μπορεί να χρησιμοποιηθεί ή να επεκταθεί. Τέλος, είναι ένα σύστημα

ανοιχτού κώδικα και είναι γραμμένο σε γλώσσα Python (*Celery - Distributed Task Queue* — *Celery 5.4.0 Documentation*, χ.χ).

Το Celery προσφέρει μια πληθώρα χαρακτηριστικών για διαχείριση εργασιών και κατανεμημένα συστήματα. Παρέχει παρακολούθηση σε πραγματικό χρόνο μέσω ροών συμβάντων, επιτρέποντας την κατανόηση της δραστηριότητας της συστάδας (cluster). Οι ροές εργασίας (work-flows) μπορούν να είναι απλές ή εξαιρετικά πολύπλοκες, κατασκευασμένες με τη χρήση πρωτογενών στοιχείων όπως η ομαδοποίηση, η αλυσιδωτή σύνδεση και η τμηματοποίηση. Οι εργασίες μπορούν να ελέγχονται με όρια χρόνου και ρυθμού, ενώ ο προγραμματισμός υποστηρίζει επαναλαμβανόμενες εργασίες με ευέλικτα διαστήματα ή εκφράσεις Crontab. Η προστασία από διαρροή πόρων είναι διαθέσιμη μέσω διαμορφώσιμων ορίων εκτέλεσης εργασιών, εξασφαλίζοντας σταθερότητα. Τέλος, το Celery επιτρέπει την εις βάθος προσαρμογή των συστατικών του εργάτη μέσω ενός γραφήματος εξάρτησης (dependency graph) που προσφέρει λεπτομερή έλεγχο της συμπεριφοράς του εργάτη.

2.5. Εφαρμογές Βαθιάς Μάθησης για Ανάλυση Πολυμέσων

Η ενσωμάτωση εφαρμογών μηχανικής μάθησης για την ανάλυση πολυμέσων επιτρέπει την καλύτερη επεξεργασία του μεγάλου σε όγκο ψηφιακού περιεχομένου. Μέσω της αξιοποίησης προηγμένων αλγορίθμων το σύστημα μπορεί να μετατρέπει ομιλία σε κείμενο, να αναγνωρίζει ονομαστικές οντότητες, να ανιχνεύει ανθρώπινες δράσεις και να ταξινομεί ήχους σε συγκεκριμένες κατηγορίες. Αυτές οι λειτουργίες συμβάλλουν στην αποτελεσματική οργάνωση, ανάλυση και ανάκτηση του περιεχομένου, ενισχύοντας τη χρηστικότητα και την ακρίβεια των πληροφοριών που εξάγονται από τα πολυμέσα.

2.5.1. Μετατροπή Ομιλίας σε Κείμενο

Ο αλγόριθμος μετατροπής ομιλίας σε κείμενο (Speech to Text) αποτελεί μία διαδικασία κατά την οποία ο προφορικός λόγος μετατρέπεται σε γραπτό κείμενο. Βασίζεται σε προηγμένους αλγορίθμους επεξεργασίας ηχητικών σημάτων, οι οποίοι εξάγουν γλωσσικά χαρακτηριστικά και τα αντιστοιχούν σε αναπαραστάσεις κειμένου. Παραδοσιακά, χρησιμοποιούνταν στατιστικά μοντέλα όπως τα Hidden Markov Models σε συνδυασμό με τα Gaussian Mixture Models, τα οποία παρείχαν αποδοτικές λύσεις αλλά παρουσίαζαν περιορισμούς στην ακρίβεια και την προσαρμοστικότητα, κυρίως σε θορυβώδη περιβάλλοντα (Nassif et al., 2019). Ωστόσο, η ραγδαία ανάπτυξη της βαθιάς μάθησης και η χρήση των νευρωνικών δικτύων βελτίωσε σημαντικά τις επιδόσεις των παλιών συστημάτων. Τα σύγχρονα συστήματα μετατροπής ομιλίας σε κείμενο χρησιμοποιούν αρχιτεκτονικές κωδικοποιητή-αποκωδικοποιητή (encoder-decoder), οι οποίες συχνά υλοποιούνται με μετασχηματιστές (transformers) ή επαναλαμβανόμενα νευρωνικά δίκτυα (Recurrent Neural Networks). Τα συστήματα αυτά εκπαιδεύονται σε μεγάλα, ποικίλα σύνολα δεδομένων για να χειρίζονται διάφορες προφορές, γλώσσες και συνθήκες θορύβου.

Σύγχρονα μοντέλα, όπως το Whisper, ενσωματώνουν μάθηση πολλαπλών εργασιών, επιτρέποντάς τους να απομαγνητοφωνούν την ομιλία σε πολλές γλώσσες και να εκτελούν μεταφράσεις (Radford et al., 2022). Οι εφαρμογές μετατροπής ομιλίας σε κείμενο κρίνονται απαραίτητες σε πολλούς τομείς, όπως σε εργαλεία δημιουργίας υποτίτλων, εικονικούς βοηθούς και υπηρεσίες απομαγνητοφώνησης. Στο υπάρχον σύστημα, ένας αλγόριθμος μετατροπής ομιλίας σε κείμενο επιτρέπει την αυτοματοποιημένη απομαγνητοφώνηση του μεταφορτωμένου οπτικοακουστικού περιεχομένου, δημιουργώντας υποτίτλους και παρέχοντας αποτελεσματικότερη αναζήτηση.

2.5.2. Αναγνώριση Ονομαστικών οντοτήτων

Η αναγνώριση ονομαστικών οντοτήτων (Named Entity Recognition) είναι ένα υποπροϊόν της επεξεργασίας της φυσικής γλώσσας (Natural Language Processing) που επικεντρώνεται στον εντοπισμό και την ταξινόμηση ονομαστικών οντοτήτων μέσα σε ένα κείμενο. Οι οντότητες αυτές περιλαμβάνουν ονόματα ανθρώπων, οργανισμών, τοποθεσιών, ημερομηνιών και άλλα. Η διαδικασία περιλαμβάνει δύο βήματα: το πρώτο είναι ο εντοπισμός των οντοτήτων, δηλαδή η εύρεση συγκεκριμένων τμημάτων κειμένου που ενδέχεται να αποτελούν ονομαστικές οντότητες, ενώ το δεύτερο βήμα είναι η ταξινόμηση των οντοτήτων σε προκαθορισμένες κατηγορίες (π.χ. άνθρωποι, οργανισμοί). Οι αλγόριθμοι αναγνώρισης οντοτήτων χρησιμοποιούν τεχνικές μηχανικής μάθησης, συχνά βασισμένες σε μοντέλα βαθιάς μάθησης, όπως τα δίκτυα BiLSTM (Bidirectional Long Short-Term Memory) που είναι επαναλαμβανόμενα νευρωνικά δίκτυα τα οποία διαβάζουν το κείμενο αμφίδρομα, δηλαδή από την αρχή προς το τέλος και αντίστροφα. Αυτό επιτρέπει τη μοντελοποίηση των συμφραζομένων σε κάθε πλευρά μιας λέξης (Jehangir et al., 2023). Επίσης, γίνεται η χρήση μετασχηματιστών, όπως το BERT (Bidirectional Encoder Representations from Transformers). Οι μετασχηματιστές παρέχουν αναπαραστάσεις λέξεων βασισμένες στο πλήρες συμφραζόμενο της πρότασης, διασφαλίζοντας ότι κάθε λέξη γίνεται κατανοητή στο πλαίσιο της πρότασης. Για παράδειγμα, η αγγλική λέξη "Apple" μπορεί να κατηγοριοποιηθεί είτε ως εταιρεία είτε ως φρούτο ανάλογα με τα συμφραζόμενα.

Η αναγνώριση ονομαστικών οντοτήτων στην ελληνική γλώσσα είναι μία πρόκληση λόγω της πλούσιας μορφολογίας της και λόγω των πολλών κλίσεων των λέξεων. Ωστόσο, εργαλεία όπως το SpaCy παρέχουν προ-εκπαιδευμένα μοντέλα για διάφορες γλώσσες, συμπεριλαμβανομένης της ελληνικής. Τα μοντέλα του SpaCy εκπαιδεύονται σε εκτεταμένα κείμενα και μπορούν να εξάγουν σημαντικές οντότητες τόσο από δομημένα όσο και από αδόμητα δεδομένα κειμένου (Honnibal et al., 2020). Στο πλαίσιο της παρούσας εφαρμογής, ο παραπάνω αλγόριθμος επεξεργάζεται τις απομαγνητοφωνήσεις που παράγονται από τον αλγόριθμο μετατροπής ομιλίας σε κείμενο. Προσδιορίζει βασικές οντότητες εντός του προφορικού λόγου, επιτρέποντας στους χρήστες να αναζητούν και να φιλτράρουν αρχεία πολυμέσων με βάση θέματα ή συγκεκριμένες λέξεις-κλειδιά. Με άλλα λόγια, δίνεται η δυνατότητα στους χρήστες να αναζητούν στο προφορικό λόγο βελτιώνοντας την ανάκτηση του οπτικοακουστικού περιεχομένου.

2.5.3. Αναγνώριση Δράσης

Η αναγνώριση ενεργειών αποτελεί έναν από τους πιο σημαντικούς τομείς της ανάλυσης βίντεο και χρησιμοποιείται για την ανίχνευση και ταξινόμηση ανθρώπινων ενεργειών ή δραστηριοτήτων μέσα σε πολυμεσικά αρχεία. Η δημιουργία ενός αποδοτικού αλγορίθμου είναι δύσκολη, καθώς οι ανθρώπινες ενέργειες μπορούν να διαφέρουν σημαντικά σε διάφορους παράγοντες, όπως οι ταχύτητες της κίνησης, η γωνία λήψης της κάμερας και οι αλλαγές στις περιβαλλοντικές συνθήκες. Επιπλέον, η δυσκολία αυξάνεται εξαιτίας της πολυπλοκότητας των ανθρώπινων ενεργειών και των πολλαπλών διαφορετικών σκηνικών και αντικειμένων που μπορεί να υπάρχουν στο βίντεο. Οι παραδοσιακές προσεγγίσεις για την αναγνώριση ενεργειών βασίζονταν στη χρήση χειροκίνητα εξαγόμενων χαρακτηριστικών και μοντέλων μηχανικής μάθησης. Ωστόσο, αυτές οι μέθοδοι συχνά αποδεικνύονταν ανεπαρκείς για τη διαχείριση σύνθετων και μεταβλητών δεδομένων. Με τη χρήση όμως των συνελκτικών νευρωνικών δικτύων δίνεται η δυνατότητα εξαγωγής χαρακτηριστικών από τα δεδομένα εισόδου και η χρήση τους για την μοντελοποίηση των χωρικών πληροφοριών των καρέ. Για την κατανόηση της χρονικής διάστασης των βίντεο γίνεται η χρήση των επαναλαμβανόμενων νευρωνικών δικτύων όπως τα Long Short-Term Memory (LSTM) δίκτυα. Τα μοντέλα αυτά είναι ικανά να καταγράφουν τη διαδοχική ροή πληροφοριών

από τα διαδοχικά καρέ και να εντοπίζουν μοτίβα δράσης που εξελίσσονται στο χρόνο. Ο συνδυασμός των συνεκτικών νευρωνικών δικτύων για τη χωρική ανάλυση και LSTM για τη χρονική ανάλυση έχει οδηγήσει σε σημαντικές βελτιώσεις στην ακρίβεια των αλγορίθμων αναγνώρισης δράσεων (Herath et al., 2017).

Οι σύγχρονες προσεγγίσεις ενσωματώνουν μοντέλα που βασίζονται σε μετασχηματιστές, οι οποίοι έχουν αποδειχθεί εξαιρετικά ισχυροί τόσο για την κατανόηση των χωρικών όσο και των χρονικών μοτίβων. Ειδικότερα, εργαλεία όπως το MMAction2, που αναπτύχθηκε από την OpenMMLab, παρέχει ευέλικτες εργαλειοθήκες για την υλοποίηση της αναγνώρισης ενεργειών. Το εργαλείο αυτό χρησιμοποιεί προ-εκπαιδευμένα μοντέλα και υποστηρίζει πολλαπλά σύνολα δεδομένων και αρχιτεκτονικές, παρέχοντας τη δυνατότητα προσαρμογής σε ποικίλες περιπτώσεις χρήσης (MMAction2 Contributors, 2020). Στο σύστημα της παρούσας διπλωματικής εργασίας, οι αλγόριθμοι αναγνώρισης ενεργειών προσθέτουν ετικέτες στα βίντεο ενισχύοντας τα μεταδεδομένα τους. Η λειτουργία αυτή καθιστά την πλατφόρμα εύχρηστη, επιτρέποντας στους χρήστες να εξερευνούν τα αρχεία πολυμέσων με βάση το περιεχόμενό τους.

2.6. Ταξινόμηση Ήχου

Η ταξινόμηση ήχου είναι η διαδικασία ανάλυσης ηχητικών σημάτων για την κατάταξή τους σε προκαθορισμένες κατηγορίες, όπως ομιλία, μουσική, περιβαλλοντικοί ήχοι ή άλλοι συγκεκριμένοι τύποι ήχου. Οι συνεχείς εξελίξεις σε αυτόν τον τομέα έχουν οδηγήσει στην ανάπτυξη βελτιωμένων αλγορίθμων ικανών να αντιμετωπίσουν την πολυπλοκότητα των δεδομένων του ήχου. Η σημασιολογική ανάλυση ήχου διαδραματίζει κρίσιμο ρόλο στις σύγχρονες εφαρμογές ήχου, οδηγώντας στην ανάγκη βελτίωσης και τελειοποίησης των αλγορίθμων ταξινόμησης (Vrysis et al., 2020). Επιπλέον, η ταξινόμηση ήχου είναι ιδιαίτερα σημαντική στα Συστήματα Διαχείρισης Πολυμέσων, καθώς επιτρέπει την αυτόματη οργάνωση, ανάκτηση και αρχειοθέτηση ηχητικών δεδομένων. Μέσω της βελτιωμένης κατηγοριοποίησης, οι χρήστες μπορούν να εντοπίζουν εύκολα συγκεκριμένα ηχητικά περιεχόμενα, βελτιώνοντας έτσι την απόδοση και την ακρίβεια τέτοιων συστημάτων.

2.6.1. Παραδοσιακές Προσεγγίσεις

Οι πρώτες προσεγγίσεις για την ταξινόμηση ήχου βασίζονταν κυρίως σε μεθόδους εξαγωγής χαρακτηριστικών, που προέρχονταν από το χρονικό πεδίο, το πεδίο συχνότητας, το φασματικό πεδίο και το πεδίο χρόνου-συχνότητας, με μεθόδους όπως οι συντελεστές φασματικού πεδίου συχνότητας (Mel-Frequency Cepstral Coefficients) και τα χαρακτηριστικά χρώματος (chroma) να διαδραματίζουν βασικό ρόλο (Sharma et al., 2020). Τα χαρακτηριστικά αυτά χρησιμοποιούνταν με τους παραδοσιακούς αλγορίθμους μηχανικής μάθησης, όπως οι Μηχανές Διανυσμάτων Στήριξης (Support Vector Machines) και τα Γκαουσιανά Μοντέλα Μίξης (Gaussian Mixture Models). Ωστόσο, αυτές οι μέθοδοι συχνά είναι δύσκολο να συλλάβουν την πολυπλοκή και ιεραρχική φύση των δεδομένων του ήχου. Με την χρήση των συνεκτικών νευρωνικών δικτύων, των επαναλαμβανόμενων νευρωνικών δικτύων και των μετασχηματιστών δίνεται η δυνατότητα στα μοντέλα να μαθαίνουν χαρακτηριστικά απευθείας από ακατέργαστα ή από ελάχιστα επεξεργασμένα δεδομένα (Zaman et al., 2023). Τα συνεκτικά νευρωνικά δίκτυα, για παράδειγμα, υπερέχουν στην επεξεργασία των φασματογραφημάτων (οπτικές αναπαραστάσεις ηχητικών σημάτων), εξάγοντας χωρικά χαρακτηριστικά, ενώ τα επαναλαμβανόμενα νευρωνικά δίκτυα και οι μετασχηματιστές συλλαμβάνουν χρονικές εξαρτήσεις και μοτίβα μεγάλης εμβέλειας.

Οι Vrysis et al. (2020) επισήμαναν την αποτελεσματικότητα τόσο των αρχιτεκτονικών των συνεκτικών νευρωνικών δικτύων με είσοδο ενός μονοδιάστατου πίνακα (1D array), όσο και των νευρωνικών με είσοδο δυσδιάστατου πίνακα (2D array) για εργασίες ταξινόμησης ήχου. Η εργασία τους συνέκρινε την απόδοση των βαθιών νευρωνικών

δικτύων με τις παραδοσιακές μεθόδους ενσωμάτωσης χρονικών χαρακτηριστικών, αναδεικνύοντας τα πλεονεκτήματα της βαθιάς μάθησης στην εξαγωγή σύνθετων χωρικών και χρονικών μοτίβων. Απέδειξαν ότι ενώ τα συνελκτικά νευρωνικά δίκτυα με δυσδιάστατη είσοδο είναι ιδιαίτερα κατάλληλα για ανάλυση με βάση τα φασματογραφήματα, τα νευρωνικά με μονοδιάστατη είσοδο μπορούν να επεξεργαστούν αποτελεσματικά ακατέργαστες κυματομορφές ήχου ή σύνολα χαρακτηριστικών χαμηλότερης διάστασης. Η μελέτη υπογραμμίζει την ευελιξία των συνελκτικών νευρωνικών δικτύων στην αντιμετώπιση ποικίλων προκλήσεων ταξινόμησης του ήχου. Η εξαγωγή χαρακτηριστικών παραμένει ένα κρίσιμο στοιχείο, καθώς η απόδοση των αλγορίθμων μηχανικής μάθησης εξαρτάται σε μεγάλο βαθμό από την ποιότητα και τη συνάφεια των εξαγόμενων χαρακτηριστικών.

Στην παρούσα διπλωματική εργασία χρησιμοποιείται μια μέθοδος που αφορά στη δημιουργία φασματογραφημάτων με βάση τα κυματίδια, τα οποία αποτελούν μια ανάλυση πολλαπλής ανάλυσης των ηχητικών σημάτων και καταγράφουν τόσο τη συχνотική όσο και τη χρονική πληροφορία. Τα φασματογραφήματα αυτά υποβάλλονται, στη συνέχεια, σε επεξεργασία από ένα συνελκτικό νευρωνικό δίκτυο για την ταξινόμηση του ήχου σε κατηγορίες όπως ομιλία, μουσική ή άλλα (περιβαλλοντικοί ήχοι). Η υβριδική αυτή προσέγγιση αξιοποιεί τα πλεονεκτήματα τόσο του διακριτού μετασχηματισμού των κυματιδίων όσο και των συνελκτικών νευρωνικών δικτύων με στόχο να επιτύχει υψηλή ακρίβεια και ανθεκτικότητα.

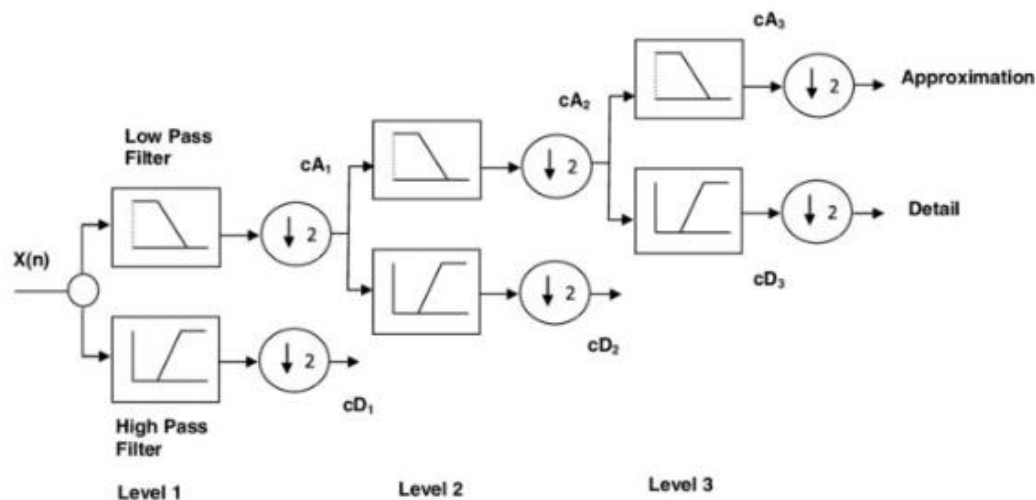
2.6.2. Μετασχηματισμός Κυματιδίων

Ο μετασχηματισμός κυματιδίων (Wavelet Transform) αναπτύχθηκε ως εναλλακτική λύση του γρήγορου μετασχηματισμού "Fourier" (Short Time Fourier Transform), για την αντιμετώπιση προβλημάτων που σχετίζονται με την ανάλυση ενός σήματος στο πεδίο του χρόνου και της συχνότητας. Συγκεκριμένα, ενώ ο γρήγορος μετασχηματισμός "Fourier" παρέχει ομοιόμορφη ανάλυση στο χρόνο και στη συχνότητα, ο διακριτός μετασχηματισμός κυματιδίων παρέχει υψηλή ανάλυση χρόνου και χαμηλή ανάλυση συχνότητας στις υψηλές συχνότητες και χαμηλή ανάλυση χρόνου και υψηλή ανάλυση συχνότητας στις χαμηλές συχνότητες. Παρόμοια χαρακτηριστικά χρονικής και συχνотικής ανάλυσης παρουσιάζει και το ανθρώπινο αυτί (Tzanetakis et al., 2001).

Για να πραγματοποιηθεί ένας μετασχηματισμός κυματιδίων απαιτείται η χρήση ενός κυματιδίου, δηλαδή, ενός μικρού κύματος που η ενέργεια του είναι συγκεντρωμένη στο τμήμα του χρόνου. Ένα κύμα είναι ομαλό, κανονικό και μπορεί να είναι ατελείωτο, ενώ το κυματίδιο είναι συχνά ανομοιόμορφο και έχει περιορισμένη διάρκεια. Κύματα όπως το ημίτονο και συνημίτονο είναι αιτιοκρατικά πρότυπα προς ανάλυση στον μετασχηματισμό "Fourier" για στάσιμα σήματα. Αντίθετα, τα κυματίδια λειτουργούν για αιτιοκρατικά πρότυπα και μη, για ανάλυση μη στάσιμων σημάτων, όπως τα σήματα ήχου, αποσυνθέτοντας το σήμα, τόσο στο χρονικό όσο και στο συχνотικό πεδίο. (Χατζηλεοντιάδης, 2021).

2.6.3. Διακριτός Μετασχηματισμός Κυματιδίων

Ο διακριτός μετασχηματισμός κυματιδίων, όπως καθιερώθηκε στο πλαίσιο της ανάλυσης πολλαπλής ανάλυσης (Multiresolution Analysis) (Mallat, 1989), αποσυνθέτει ένα σήμα σε διαφορετικές ζώνες συχνотήτων χρησιμοποιώντας έναν συνδυασμό φιλτραρίσματος και υποδειγματοληψίας. Η διαδικασία περιλαμβάνει, όπως φαίνεται και στην Εικόνα 2.5, την εφαρμογή ενός χαμηλοπερατού φίλτρου για την εξαγωγή συντελεστών προσέγγισης, οι οποίοι αντιπροσωπεύουν το περιεχόμενο χαμηλής συχνότητας του σήματος, και ενός υψηλοπερατού φίλτρου για τον υπολογισμό των συντελεστών λεπτομέρειας, οι οποίοι αποτυπώνουν πληροφορίες υψηλής συχνότητας. Αυτή η αποσύνθεση επαναλαμβάνεται επαναληπτικά στους συντελεστές προσέγγισης, δημιουργώντας μια ιεραρχική δομή γνώστη ως δέντρο αποσύνθεσης κυματιδίων.



Εικόνα 2-5 Διάγραμμα ανάλυσης σήματος πολλαπλής ανάλυσης [Πηγή:
https://www.researchgate.net/publication/263929587_Comparison_of_Wavelet-Based_ANN_and_Regression_Models_for_Reservoir_Inflow_Forecasting/figures](https://www.researchgate.net/publication/263929587_Comparison_of_Wavelet-Based_ANN_and_Regression_Models_for_Reservoir_Inflow_Forecasting/figures)

2.6.4. Δημιουργία Φασματογραφήματος με βάση τα Κυματίδια

Το φασματογράφημα είναι μια ισχυρή μέθοδος για την ανάλυση των χαρακτηριστικών χρόνου-συχνότητας ενός σήματος. Συνήθως, ένα φασματογράφημα δημιουργείται με τη χρήση του μετασχηματισμού Fourier. Στην παρούσα, όμως, διπλωματική εργασία χρησιμοποιείται ο διακριτός μετασχηματισμός των κυματιδίων, ο οποίος αποσυνθέτει ένα σήμα σε διαφορετικές συχνοτικές συνιστώσες μέσω μιας ανάλυσης πολλαπλών επιπέδων. Η τεχνική αυτή είναι ιδιαίτερα χρήσιμη για την αποτύπωση της χρονικά μεταβαλλόμενης φύσης των σημάτων, ιδίως όταν αυτά παρουσιάζουν χαρακτηριστικά τόσο χαμηλής όσο και υψηλής συχνότητας με την πάροδο του χρόνου. Οι Dimoulas et al. (2007) εισήγαγαν μια παρόμοια μεθοδολογία που χρησιμοποιεί την αποσύνθεση των κυματιδίων για την παρακολούθηση της εξέλιξης των φασματικών χαρακτηριστικών σε μακροχρόνιες καταγραφές. Στην προσέγγισή τους, εφάρμοσαν καμπύλες περιβάλλουσας ισχύος για να αναπαραστήσουν τις μεταβολές της ενέργειας με την πάροδο του χρόνου, όπου είναι ιδιαίτερα χρήσιμο για την παρακολούθηση σημάτων όπως οι βιοηλεκτρικές καταγραφές και τα μανομετρικά επίπεδα πίεσης. Η μεθοδολογία τους περιλαμβάνει την εκτίμηση τόσο του χρόνου όσο και των περιλήψεων του χρόνου-συχνότητας με τη χρήση των επεξεργασμένων συντελεστών των κυματιδίων.

Η δημιουργία ενός φασματογραφήματος βασισμένου σε κυματίδια ξεκινά με την αποσύνθεση του σήματος εισόδου με τη χρήση ενός διακριτού μετασχηματισμού κυματιδίων. Η διαδικασία αυτή διαιρεί το σήμα σε πολλαπλές ζώνες συχνοτήτων, καθεμία από τις οποίες αντιστοιχεί σε συγκεκριμένη ανάλυση, με βάση την επιλεγμένη συνάρτηση κυματιδίου και τον αριθμό των επιπέδων αποσύνθεσης. Ο μετασχηματισμός κυματιδίων εξάγει ένα σύνολο συντελεστών προσέγγισης, οι οποίοι αντιπροσωπεύουν τις συνιστώσες χαμηλότερης συχνότητας και συντελεστών λεπτομέρειας, που αποτυπώνουν τις συνιστώσες υψηλότερης συχνότητας. Το σήμα χωρίζεται σε μη επικαλυπτόμενα τμήματα σταθερού μεγέθους. Για κάθε τμήμα πραγματοποιείται η αποσύνθεση κυματιδίων και αναλύονται οι συντελεστές για κάθε ζώνη συχνοτήτων. Υπολογίζεται η ενέργεια των συντελεστών σε κάθε ζώνη, η οποία αντικατοπτρίζει την κατανομή ισχύος του σήματος σε όλες τις συχνότητες. Στη συνέχεια, αυτές οι τιμές

ενέργειας μετατρέπονται σε λογαριθμική κλίμακα για να βελτιωθεί η ικανότητα ερμηνείας και να αποφευχθούν πιθανά αριθμητικά ζητήματα κατά τον υπολογισμό.

Αυτές οι τιμές ενέργειας οργανώνονται σε έναν πίνακα, όπου κάθε γραμμή αντιπροσωπεύει μια ζώνη συχνοτήτων και κάθε στήλη αντιστοιχεί σε ένα χρονικό τμήμα. Αυτός ο πίνακας σχηματίζει το φασματογράφημα με βάση τα κυματίδια, παρέχοντας μια αναπαράσταση χρόνου-συχνότητας του σήματος. Ο άξονας χρόνου προκύπτει από την κατάτμηση του σήματος σε τμήματα, ενώ ο άξονας συχνότητας αντιστοιχεί στις κεντρικές συχνότητες των ζωνών των κυματιδίων. Η μέθοδος αυτή αποτυπώνει τόσο τη χρονική εξέλιξη όσο και τα φασματικά χαρακτηριστικά του σήματος, επιτρέποντας τη λεπτομερή ανάλυση του χρονικά μεταβαλλόμενου συχνοτικού περιεχομένου.

2.7. Μηχανική μάθηση

Η μηχανική μάθηση αντιπροσωπεύει ένα κλάδο της τεχνητής νοημοσύνης (Artificial Intelligence), ο οποίος επικεντρώνεται στην υλοποίηση αλγορίθμων με σκοπό την βελτιστοποίηση της απόδοσης ή διεξαγωγή ακριβών προβλέψεων με βάση την «εμπειρία». Με άλλα λόγια, ένα σύστημα μηχανικής μάθησης μαθαίνει από τα δεδομένα που του παρέχονται, αναγνωρίζοντας μοτίβα μέσα στα δεδομένα και λαμβάνοντας αποφάσεις με ελάχιστη ανθρώπινη παρέμβαση (Jordan & Mitchell, 2015; Hastie et al., 2009). Ο συγκεκριμένος κλάδος έχει αναπτυχθεί ραγδαία τα τελευταία χρόνια, καθώς ολοένα και περισσότεροι αλγόριθμοι μηχανικής μάθησης υλοποιούνται και έχουν εφαρμογή σε πολλά πεδία. Οι αλγόριθμοι μηχανικής μάθησης μπορούν να χωριστούν σε τέσσερις κατηγορίες. Παρακάτω θα αναφερθούν και περιγράφουν εν συντομία οι κατηγορίες αυτές. Η πρώτη κατηγορία αφορά την επιβλεπόμενη μάθηση είναι μια διαδικασία εκμάθησης σχετικά με μια συνάρτηση αντιστοίχισης μιας εισόδου με μια έξοδο, χρησιμοποιώντας ζεύγη δειγμάτων δεδομένων εισόδου-εξόδου (Han et al., 2012). Τα δεδομένα εκπαίδευσης περιλαμβάνουν ετικέτες (label), που υποδηλώνουν την τιμή εξόδου που αντιστοιχεί σε κάθε είσοδο και χρησιμοποιούνται στην διαδικασίας μάθησής για τον καθορισμό της χαρακτηριστικής συνάρτησης αντιστοίχισης. Η πιο συνηθισμένη τεχνική είναι αυτή της κατηγοριοποίησης, καθώς είναι και σημαντική η χρήση της στην επεξεργασία δεδομένων πολυμέσων. Επόμενη κατηγορία μηχανικής μάθησης είναι η μη επιβλεπόμενη μάθηση αφορά στη διαδικασία ανάλυσης δεδομένων που δεν έχουν ετικέτες, γεγονός που σημαίνει ότι δεν υπάρχει προκαθορισμένη έξοδος για μια συγκεκριμένη είσοδο. Η συγκεκριμένη διαδικασία χρησιμοποιείται ευρέως για την εξαγωγή γενετικών χαρακτηριστικών καθώς και τον εντοπισμό και την κατηγοριοποίηση τάσεων από ένα σύνολο δεδομένων (Han et al., 2012). Επίσης, η ημι-επιβλεπόμενη μάθηση αποτελεί μια σύνθετη προσέγγιση μάθησης που ενσωματώνει στοιχεία τόσο από την επιβλεπόμενη μάθηση όσο και από την μη επιβλεπόμενη μάθηση. Αυτή η μέθοδος λειτουργεί με σύνολα δεδομένων που μπορεί να είναι επισημασμένα αλλά και μη-επισημασμένα (Han et al., 2012; Sarker et al., 2020). Η ημι-επιβλεπόμενη μάθηση καθίσταται ιδιαίτερα χρήσιμη, καθώς τα σύνολα δεδομένων με ετικέτες είναι πολύ λιγότερα συγκριτικά με τα δεδομένα χωρίς ετικέτες (Jordan & Mitchell, 2015). Ο κύριος στόχος ενός μοντέλου ημι-επιβλεπόμενης μάθησης είναι να παρέχει πιο ακριβείς προβλέψεις σε τιμές εξόδου σε σύγκριση με τα μοντέλα που χρησιμοποιούν μόνο επισημασμένα δεδομένα, αξιοποιώντας ταυτόχρονα και μη επισημασμένα δεδομένα. Τελευταία κατηγορία της μηχανικής μάθησης είναι η ενισχυμένη μάθηση ξεχωρίζει από τις προηγούμενες προσεγγίσεις καθώς δεν απαιτεί σύνολα δεδομένων για την εκπαίδευση. Αντίθετα, η μάθηση επιτυγχάνεται μέσω μιας διαδικασίας «δοκιμής και ανατροφοδότησης» (François-Lavet et al., 2018). Από την ίδια αναλυτική διαδικασία, όταν η δράση συμβάλλει στη επίτευξη του στόχου, ο αλγόριθμος λαμβάνει ανατροφοδότηση με μορφή «ανταμοιβών», ή αντίστροφα, όταν η ενέργεια δεν συμβάλλει στην επίτευξη του στόχου, η ανατροφοδότηση λαμβάνεται με την μορφή

«κυρώσεων». Ο αλγόριθμος έχει ως στόχο την ελαχιστοποίηση της «κύρωσης» και την μεγιστοποίηση της «ανταμοιβής» (Sutton & Barto, 2018).

2.8. Βαθιά Μάθηση

Η βαθιά μάθηση (deep learning), αποτελεί υποσύνολο της γενικότερης έννοιας της μηχανικής μάθησης και βασίζεται στην εκμάθηση από αναπαράστασεις δεδομένων. (LeCun et al., 2015). Επί της ουσίας, αποτελεί ένα νευρωνικό δίκτυο με πολλαπλά επίπεδα έχοντας την ικανότητα εξαγωγής πληροφοριών από μεγάλες ποσότητες δεδομένων και προσπαθώντας να προσομοιώσει τη συμπεριφορά του ανθρώπινου εγκεφάλου (Goodfellow et al., 2016). Τα νευρωνικά δίκτυα τροφοδοσίας προς τα εμπρός (Feedforward Neural Networks), τα οποία είναι γνωστά στη βιβλιογραφία ως MLP, αποτελούν τη βασική κατηγορία μοντέλων βαθιάς μάθησης και παράγουν μη γραμμικά μοντέλα που επιτρέπουν την πρόβλεψη δεδομένων ανάλογα με τα δεδομένα της εισόδου (Taud & Mas, 2018).

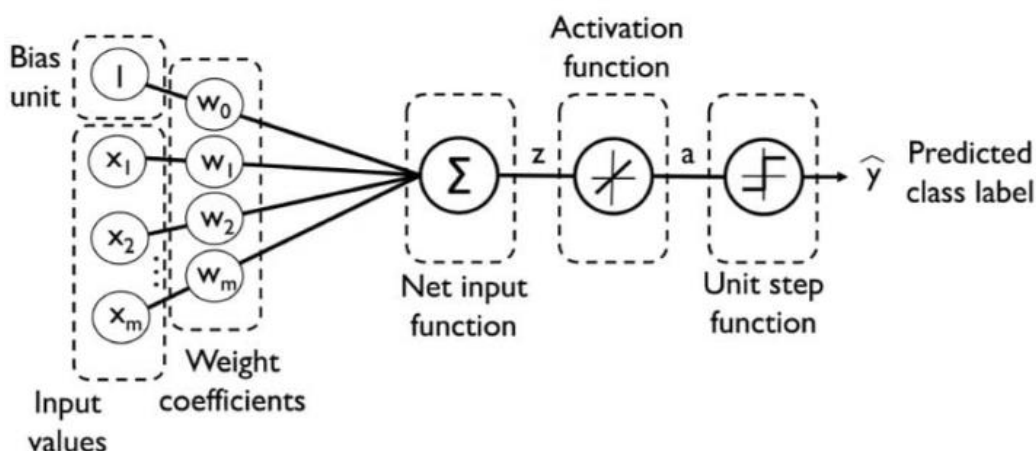
2.8.1. Νευρωνικά Δίκτυα Τροφοδοσίας προς τα εμπρός

Το πιο βασικό στοιχείο των νευρωνικών δικτύων είναι ο τεχνητός νευρώνας, γνωστός ως perceptron. Η πιο απλή μορφή ενός νευρωνικού δικτύου είναι η ύπαρξη ενός νευρώνα και ενός επιπέδου, δηλαδή, υπάρχει μια έξοδος για όλες τις εισόδους που είναι συνδεδεμένες. Εάν υποθέσουμε ότι για $i = 0, 1, \dots, n$, όπου n είναι ο αριθμός των εισόδων και τα w_i είναι βάρη του νευρώνα, οι εισοδοί x_i αντιστοιχούν σε χαρακτηριστικά ή μεταβλητές και η έξοδος y αντιστοιχεί στην προβλεπόμενη κλάση. Για τον σχηματισμό ενός μοντέλου νευρώνα απαιτούνται τρία βήματα.

Το πρώτο βήμα περιλαμβάνει τον πολλαπλασιασμό κάθε τιμής εισόδου με το αντίστοιχο βάρος, κατά το δεύτερο βήμα πραγματοποιείται το άθροισμα του προηγούμενου πολλαπλασιασμού και τέλος, στο τρίτο βήμα εφαρμόζεται μια συνάρτηση ενεργοποίησης (μεταφοράς) στο προηγούμενο άθροισμα, με αποτέλεσμα να παράγεται η έξοδος y , όπως φαίνεται και παρακάτω:

$$y = f\left(\sum_{j=0}^n w_j x_j\right) \quad (2.1)$$

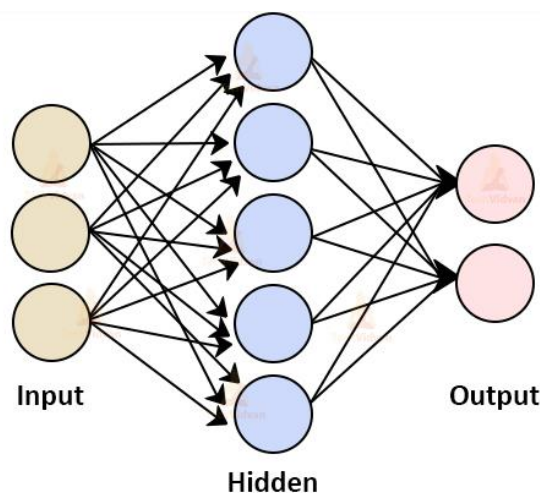
Ο όρος που προκύπτει για $i=0$ ονομάζεται όρος πόλωσης (bias term), όπου $x_0 = 1$ και $w_0 = \text{bias}$. Η Εικόνα 2-6 αποτυπώνει την απλοποιημένη αναπαράστασή του.



Εικόνα 2-6 Παράδειγμα απλού δικτύου τροφοδοσίας προς τα εμπρός (Πηγή: <https://www.simplilearn.com/tutorials/deep-learning-tutorial/multilayer-perceptron/>)

Ο στόχος ενός νευρωνικού δικτύου είναι να βρει την ακριβέστερη συνάρτηση $f(x)$ που να προσεγγίζει καλύτερα την αληθή συνάρτηση αντιστοίχισης $f^*(x)$. Για να επιτευχθεί αυτό, το δίκτυο εκπαιδεύεται με προσεγγίσεις της $f^*(x)$, για διαφορετικές τιμές εισόδου x , όπου κάθε τιμή εισόδου συνοδεύεται και με μια ετικέτα (κλάση) y . Σκοπός της εκπαίδευσης-μάθησης είναι η βελτιστοποίηση των βαρών με την ελαχιστοποίηση μιας συνάρτησης κόστους, η οποία είναι συνήθως το τετραγωνικό σφάλμα μεταξύ της τιμής της εξόδου του νευρωνικού με την επιθυμητή τιμή.

Αναλυτικές τεχνικές, όπως η κατάβαση δυναμικού (Gradient Descent), καθορίζουν το βέλτιστο διάνυσμα βάρους. Ο αλγόριθμος συγκλίνει σε μια λύση φτάνοντας σε μια λειτουργική διαμόρφωση δικτύου. Η επικύρωση του μοντέλου επιτυγχάνεται με τη χρήση νέων δεδομένων για να δείξει πως η διαμόρφωση μπορεί να γενικευτεί σε νέες καταστάσεις. Το απλό νευρωνικό δίκτυο, που αναφέρθηκε παραπάνω, δεν μπορεί να αναπαραστήσει μη γραμμικές αλληλεπιδράσεις μεταξύ εισόδου και εξόδου. Τη λύση στο πρόβλημα δίνουν τα νευρωνικά δίκτυα τροφοδοσίας προς τα εμπρός που αποτελούνται από συνδεδεμένους τεχνητούς νευρώνες που οργανώνονται σε επίπεδα. Η ροή της πληροφορίας, ξεκινά από το πρώτο επίπεδο, γνωστό ως επίπεδο εισόδου (input), προς τα ενδιάμεσα επίπεδα, γνωστά ως κρυφά (hidden) και καταλήγει στο τελευταίο, γνωστό ως επίπεδο εξόδου (output) (Bishop, 1995). Στην Εικόνα 2-7 παρουσιάζεται η δομή ενός νευρωνικού δικτύου, οπισθοδιάδοσης (backpropagation), που αποτελεί γενίκευση του κανόνα ελαχίστων τετραγώνων (Du & Swamy, 2013)



Εικόνα 2-7 Παράδειγμα δομής επιπέδων ενός νευρωνικού δικτύου ([Πηγή: https://www.analyticsvidhya.com/blog/2022/03/basic-introduction-to-feed-forward-network-in-deep-learning/](https://www.analyticsvidhya.com/blog/2022/03/basic-introduction-to-feed-forward-network-in-deep-learning/))

Τα βάρη του νευρώνα διορθώνονται ξεκινώντας από το επίπεδο εξόδου προς τα πίσω, εξηγώντας τον όρο της οπισθοδιάδοσης. Η απόδοση του τελικού μοντέλου εξαρτάται, τόσο από την επιλογή των μεταβλητών, τον αριθμό στρωμάτων, των κόμβων και των δεδομένων εκπαίδευσης όσο και από τις παραμέτρους εκπαίδευσης, όπως την ταχύτητα μάθησης και τον αριθμό των επαναλήψεων.

2.8.2. Συνελικτικά Νευρωνικά Δίκτυα

Τα συνελικτικά νευρωνικά δίκτυα (Convolutional Neural Networks) αντιπροσωπεύουν μια εξειδικευμένη κατηγορία αρχιτεκτονικών βαθιάς μάθησης που εισήχθησαν αρχικά από τους LeCun et al. (1989) στην πρωτοποριακή εργασία τους για την αναγνώριση χειρόγραφων ψηφίων και από τότε έχουν εξελιχθεί σε θεμελιώδες εργαλείο της μηχανικής μάθησης. Αυτή η προσέγγιση της μάθησης από άκρο σε άκρο έχει διερευνηθεί κυρίως

στον τομέα της όρασης υπολογιστών, ιδίως μετά τις σημαντικές εξελίξεις στην ταξινόμηση εικόνων (Krizhevsky et al., 2012). Πλέον χρησιμοποιούνται σε πολλούς τομείς όπως σηματολογική ανάλυση οπτικοακουστικού υλικού και σε άλλους τομείς που απαιτείται η αναγνώριση προτύπων.

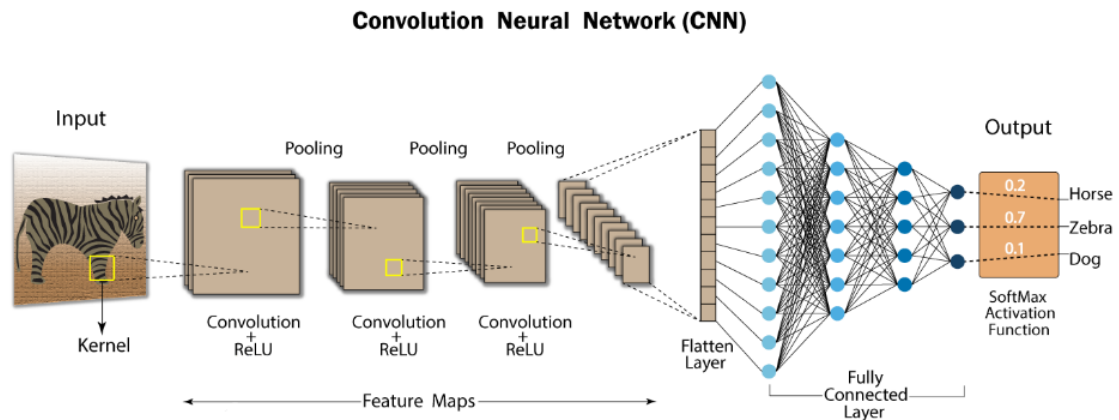
Στο πυρήνα ενός συνελικτικού νευρωνικού δικτύου υπάρχουν τα επίπεδα συνελίξεων (convolutional layers), τα οποία εξαγωγή χαρακτηριστικά εφαρμόζοντας φίλτρα (kernels) στα δεδομένα εισόδου. Αυτά τα φίλτρα ολισθαίνουν στο χώρο εισόδου, υπολογίζοντας τοπικά γινόμενα σημείων για τη δημιουργία χαρτών χαρακτηριστικών (feature maps) που αναπαριστούν χωρικά συσχετιζόμενα χαρακτηριστικά. Μαθηματικά, η λειτουργία της συνέλιξης μπορεί να περιγραφεί ως το σταθμισμένο άθροισμα των επικαλυπτόμενων περιοχών της εισόδου με ένα φίλτρο, που εκφράζεται στην εξίσωση ως εξής:

$$(f * g)(x, y) = \sum_{m=-k}^k \sum_{n=-k}^k g(m, n) \cdot f(x - m, y - n), \quad (2.6)$$

όπου $f(x, y)$ δηλώνει την είσοδο, $g(m, n)$ αντιπροσωπεύει το φίλτρο και k είναι το μέγεθος του φίλτρου. Συναρτήσεις ενεργοποίησης, όπως η διορθωμένη γραμμική μονάδα, εφαρμόζονται στη συνέχεια στους χάρτες χαρακτηριστικών που προκύπτουν, μετά την συνέλιξη, για την εισαγωγή μη γραμμικότητας, επιτρέποντας στο δίκτυο να μοντελοποιεί σύνθετα μοτίβα και σχέσεις στα δεδομένα.

Εκτός από τα επίπεδα συνελίξεων περιλαμβάνονται τα επίπεδα δειγματοληψίας (pooling layers) τα οποία μειώνουν τις διαστάσεις των χαρτών χαρακτηριστικών, διατηρώντας παράλληλα τις πληροφορίες. Η πιο γνωστή τεχνική δειγματοληψίας είναι της μέγιστης (max pooling), στην οποία γίνεται η επιλογή της μεγαλύτερης τιμής από μια περιοχή του χάρτη χαρακτηριστικών, αλλά και της μέσης δειγματοληψίας (average pooling) που υπολογίζει τον μέσο όρο. Η διαδικασία αυτή συμβάλλει στη μείωση της υπολογιστικής πολυπλοκότητας και στην βελτίωση της γενίκευσης, μειώνοντας τον κίνδυνο υπερπροσαρμογής.

Τα τελευταία επίπεδα ενός τέτοιου δικτύου είναι τα πλήρως συνδεδεμένα, όπου τα εξαγόμενα χαρακτηριστικά συγκεντρώνονται και αντιστοιχίζονται στην επιθυμητή έξοδο, όπως επικέτες ταξινόμησης ή τιμές παλινδρόμησης. Αυτά τα επίπεδα ερμηνεύουν τα πρότυπα που μαθαίνει το δίκτυο και παράγουν προβλέψεις με βάση τα δεδομένα εισόδου. Στο Εικόνα 2-8 φαίνονται τα βασικά τμήματα ενός συνελικτικού νευρωνικού δικτύου όπως αναφέρθηκαν παραπάνω.



Εικόνα 2-8 Αρχιτεκτονική ενός συνεκτικού νευρωνικού δικτύου για την ταξινόμηση μιας εικόνας
 Πηγή: <https://developersbreach.com/convolution-neural-network-deep-learning/>

2.9. Σύνοψη Δευτέρου Κεφαλαίου

Στο δεύτερο κεφάλαιο έγινε αναφορά στην ανάγκη για καλύτερη διαχείριση και ανάκτηση οπτικοακουστικών πόρων και πως αυτή μπορεί να επιτευχθεί με τη χρήση ευφυών συστημάτων που διαθέτουν αλγορίθμους μηχανικής μάθησης. Γίνεται αναφορά στην μεθοδολογία ανάπτυξης λογισμικού και το σύστημα διαχείρισης περιεχομένου που χρησιμοποιήθηκε. Για τη δημιουργία ενός τέτοιου συστήματος απαιτείται μια αρχιτεκτονική που θα μπορεί να εκτελεί αλγορίθμους στο παρασκήνιο του συστήματος. Γίνεται ανάλυση των εργαλείων της διαμεσολάβησης μηνυμάτων για ασύγχρονη επικοινωνία και επεξεργασία δεδομένων, ώστε να μπορούν οι αλγόριθμοι να τρέχουν ασύγχρονα και χωρίς να επηρεάζεται η κύρια εφαρμογή. Επιπλέον, εξετάζονται τεχνικές βαθιάς μάθησης, όπως η μετατροπή ομιλίας σε κείμενο, η αναγνώριση οντοτήτων και δράσεων και πως αυτές μπορούν να βοηθήσουν στην καλύτερη διαχείριση και ανάλυση πολυμέσων. Ιδιαίτερη έμφαση δίνεται στην ταξινόμηση ήχου και πως μπορεί να δημιουργηθεί ένας τέτοιος αλγόριθμος με την χρήση φασματογραφημάτων με βάση το διακριτό μετασχηματισμό των κυματιδίων ως είσοδο σε ένα νευρωνικό δίκτυο. Αφού παρουσιάστηκαν η θεωρία και η μεθοδολογία της παρούσας διπλωματικής στο επόμενο κεφάλαιο θα γίνει ανάλυση της εφαρμογής, των αλγορίθμων, των απαιτήσεων χρηστών καθώς και όλων των διαθέσιμων εργαλείων και τεχνολογιών που απαιτούνται για τη δημιουργία της εφαρμογής.

Κεφάλαιο 3ο

3. Ανάλυση Εφαρμογής

3.1. Αρχική ιδέα

Στα πλαίσια του μαθήματος «Εργαστήριο Πολυμεσικής Παραγωγής» ξεκίνησε η πρώτη προσπάθεια ανάπτυξης του Ar.U.TV ως τηλεοπτικός ιστοχώρος του Α.Π.Θ. (Ομάδα εργασίας Εργαστηρίου Πολυμεσικής Παραγωγής, 2009). Έπειτα, το επόμενο στάδιο ήταν ο εντοπισμός των αδυναμιών της πρώτης προσπάθειας, με σκοπό την αναδιαμόρφωση του τρόπου προβολής των πληροφοριών και του υλικού, έτσι ώστε να παρέχεται στο χρήστη καλύτερα και ευκολότερα η δυνατότητα διαχείρισης και αναζήτησης πολυμεσικού περιεχομένου με χρήση περιγραφέων MPEG-7, όπως χρώμα και κίνηση, αλλά και ετικετών και κατηγοριών που προστίθονταν από τους χρήστες (Ζαπάρτας Παναγιώτης & Σαρτσένκο Ιουλία, 2012). Αναγνωρίστηκε, επιπρόσθετα, η ανάγκη για περαιτέρω κατηγοριοποίηση του οπτικοακουστικού υλικού (Tsiras et al., 2015).

3.2. Στόχευση

Ο κύριος στόχος της παρούσας διατριβής είναι η ανακατασκευή και επαναλειτουργία του ιστοτόπου Ar.U.TV, εμπλουτίζοντας τον με καινοτόμες λειτουργίες και αλγόριθμους μηχανικής μάθησης τελευταίας τεχνολογίας. Δευτερεύον στόχος είναι να αναπτυχθεί ένας αλγόριθμος ταξινόμησης ήχου ο οποίος θα προστεθεί στο σύστημα και χρησιμοποιεί καινοτόμα χαρακτηριστικά. Συγκεκριμένα, η διατριβή εστιάζει στην δημιουργία μιας αρθρωτής αρχιτεκτονικής που θα πραγματοποιείται προσθήκη αλγόριθμων που να μπορούν να εξαγάγουν χαρακτηριστικά από πολυμεσικά αρχεία που αναρτώνται από τους χρήστες, με αποτέλεσμα την καλύτερη κατηγοριοποίηση και κατ' επέκταση, ανάκτησή τους. Για παράδειγμα, ο αλγόριθμος που μετατρέπει την ομιλία σε κείμενο (speech-to-text) σε συνδυασμό με τη αναγνώριση οντοτήτων συμβάλουν στην καλύτερη αναζήτηση περιεχομένου, βάσει αυτών που έχουν ειπωθεί. Ένας αλγόριθμος για κατηγοριοποίηση του ήχου σε ομιλία, μουσική ή άλλα (περιβαλλοντικοί ήχοι) συμβάλει στην καλύτερη κατηγοριοποίηση και ανάκτηση ανάλογα με το περιεχόμενο που θέλει να ανακτήσει ένας χρήστης. Επιπρόσθετα, υπάρχει και ο αλγόριθμος αναγνώρισης ενέργειας που επιτρέπει μια επιπλέον κατηγοριοποίηση στο αναρτημένο υλικό. Όλα τα παραπάνω χαρακτηριστικά δημιουργούνται αυτόματα με την κάθε ανάρτηση υλικού στην ιστοσελίδα, χωρίς να επηρεάζεται το κύριο μέρος της εφαρμογής, το οποίο είναι η περιήγηση στην ιστοσελίδα. Η αφθονία και η συνεχώς αυξανόμενη εξάπλωση του πολυμεσικού περιεχομένου, η οποία διέπει το σημερινό ψηφιακό τοπίο, καθιστά επιτακτική την ανάγκη για την οργάνωση και διαχείριση του. Ο γενικός στόχος, επομένως, της παρούσας διατριβής είναι να συνεισφέρει στην αναβάθμιση του Ar.U.TV και να συμβάλλει στην αντιμετώπιση των συγχρόνων προκλήσεων με τις οποίες έρχεται αντιμέτωπη η επιστημονική κοινότητα, που αφορούν στην οργάνωση, κατηγοριοποίηση και ανάκτηση του οπτικοακουστικού υλικού.

3.3. Παρόμοιες Εφαρμογές

Το YouTube και το Vimeo είναι οι δύο πιο αναγνωρισμένες πλατφόρμες διαχείρισης, προβολής και ανταλλαγής οπτικοακουστικού υλικού. Αποτελούν τα κύρια πρότυπα για τις πλατφόρμες πολυμεσικού περιεχομένου που δημιουργούνται από χρήστες, αλλά διαθέτουν αξιοσημείωτους περιορισμούς όσον αφορά την προηγμένη σημασιολογική επεξεργασία και τη λειτουργικότητα αναζήτησης. Παρακάτω, εξετάζονται τα βασικά τους χαρακτηριστικά και οι ελλείψεις τους.

Το YouTube θεωρείται δημοφιλές εξαιτίας της παγκόσμιας επεκτασιμότητάς του και της ευκολίας χρήσης του τόσο για τους δημιουργούς περιεχομένου όσο και για τους χρήστες. Ωστόσο, η εξάρτησή του από τα μεταδεδομένα τα οποία δημιουργούνται χειροκίνητα από τους χρήστες, περιορίζει τη λειτουργία της αναζήτησης. Συγκεκριμένα, δεν προσφέρεται στους χρήστες η δυνατότητα να πραγματοποιούν σημασιολογικές αναζητήσεις εντός του περιεχομένου της ομιλίας και δε διατίθενται εργαλεία για αυτόματη εξαγωγή θεμάτων ή λεπτομερή ταξινόμηση περιεχομένου.

Το Vimeo είναι, κυρίως, δημοφιλές για τα χαρακτηριστικά του σε επαγγελματικό επίπεδο, συμπεριλαμβανομένων των προσαρμοσίμων προγραμμάτων αναπαραγωγής βίντεο και της υψηλής ποιότητας ροής. Σε αντίθεση με το YouTube, το Vimeo απευθύνεται σε δημιουργούς περιεχομένου που αναζητούν περισσότερο έλεγχο στην παρουσίαση του περιεχομένου τους. Παρά τα πλεονεκτήματα αυτά, διαθέτει παρόμοιους περιορισμούς όσον αφορά την σημασιολογική επεξεργασία και την έλλειψη χαρακτηριστικών βασισμένων στην τεχνητή νοημοσύνη για την αναζήτηση και ανάκτηση περιεχομένου.

Οι παραπάνω πλατφόρμες αποτελούν πρότυπο για τη διαχείριση οπτικοακουστικού περιεχομένου. Παρόλα αυτά, αποτυγχάνουν να αντιμετωπίσουν βαθύτερες ανάγκες, όπως η απομαγνητοφώνηση, η οποία προσφέρει ένα νέο είδος αναζήτησης, αλλά και η κατηγοριοποίηση περιεχομένου με βάση σημασιολογικά χαρακτηριστικά.

3.3.1. Αλγόριθμοι για Διαχείριση Οπτικοακουστικού Υλικού

Στο πλαίσιο της ανάλυσης των τεχνολογικών λύσεων για το σύστημα διαχείρισης οπτικοακουστικού υλικού, διαπιστώνεται ότι πολλές ερευνητικές και καινοτόμες λύσεις. Οι λύσεις αυτές επικεντρώνονται σε αλγόριθμους με συγκεκριμένες εργασίες, όπως η μετατροπή ομιλίας σε κείμενο, η αναγνώριση ονομαστικών οντοτήτων, η αναγνώριση ενεργειών του ανθρώπου και η κατηγοριοποίηση του ήχου. Ο αλγόριθμος Whisper, για παράδειγμα, αποτελεί έναν ευρύτατα διαδεδομένο αλγόριθμο που έχει σημειώσει σημαντική πρόοδο στη μετατροπή της ομιλίας σε κείμενο. Επιδεικνύει υψηλή ακρίβεια στην απομαγνητοφώνηση, γεγονός που τον καθιστά δημοφιλή επιλογή για τη δημιουργία υποτίτλων, ενώ μπορεί να συνδυαστεί εύκολα και με αλγόριθμους επεξεργασίας κειμένων. Ωστόσο, τα εργαλεία αυτά είναι συνήθως αυτόνομες εφαρμογές και δεν ενσωματώνονται σε ολοκληρωμένα συστήματα διαχείρισης οπτικοακουστικού υλικού. Επιπλέον οι αλγόριθμοι αναγνώρισης ονομαστικών οντοτήτων, χρησιμοποιούνται στην επεξεργασία κειμένου για τον εντοπισμό βασικών οντοτήτων, όπως πρόσωπα, τοποθεσίες και γεγονότα. Παρόλο που συναντώνται στην ανάλυση γραπτού κειμένου, η εφαρμογή τους σε απομαγνητοφωνήσεις οπτικοακουστικού περιεχομένου είναι σχετικά σπάνια, παρουσιάζοντας μια ευκαιρία για καινοτομία σε αυτόν τον τομέα.

Τέλος, οι αλγόριθμοι αναγνώρισης δράσης έχουν χρησιμοποιηθεί εκτενώς για εφαρμογές στην επιτήρηση και την αθλητική ανάλυση, όπου ο εντοπισμός συγκεκριμένων κινήσεων ή ενεργειών είναι κρίσιμος. Ομοίως, οι τεχνικές ταξινόμησης ήχου χρησιμοποιούνται σε τομείς όπως η κατηγοριοποίηση μουσικής και η ανάλυση ήχου. Οι τεχνολογίες αυτές, ωστόσο, χρησιμοποιούνται συχνά μεμονωμένα και όχι ως μέρος ολοκληρωμένων συστημάτων διαχείρισης οπτικοακουστικού υλικού.

3.3.2. Διαφοροποίηση της Προτεινόμενης Εφαρμογής

Το προτεινόμενο σύστημα συνδυάζει τα πλεονεκτήματα των κύριων πλατφορμών και των εξειδικευμένων τεχνολογιών, ενώ παράλληλα αντιμετωπίζει τις αντίστοιχες ελλείψεις τους. Εισάγει ένα ενοποιημένο περιβάλλον για τη διαχείριση, την ανάλυση και την ανάκτηση οπτικοακουστικού περιεχομένου με χαρακτηριστικά που δεν είναι διαθέσιμα στις υφιστάμενες εφαρμογές. Το προτεινόμενο σύστημα συνδυάζει τα πλεονεκτήματα των κύριων πλατφορμών και των εξειδικευμένων τεχνολογιών, ενώ παράλληλα

αντιμετωπίζει τις αντίστοιχες ελλείψεις τους. Εισάγει ένα ενοποιημένο περιβάλλον για τη διαχείριση, την ανάλυση και την ανάκτηση οπτικοακουστικού περιεχομένου με χαρακτηριστικά που δεν είναι διαθέσιμα στις υφιστάμενες εφαρμογές.

Σε αντίθεση με το YouTube και το Vimeo, το σύστημα χρησιμοποιεί αλγόριθμους αυτόματης αναγνώρισης ομιλίας, όπως το Whisper, για τη μετατροπή της ομιλίας σε κείμενο. Τα κείμενα που δημιουργούνται επεξεργάζονται περαιτέρω με αλγόριθμους αναγνώρισης ονομαστικών οντοτήτων, επιτρέποντας την εξαγωγή θεμάτων και την αναζήτηση περιεχομένου στον προφορικό λόγο. Παράλληλα, αξιοποιεί αλγορίθμους αναγνώρισης ενεργειών και ταξινόμησης ήχου, προσφέροντας πιο λεπτομερή ανάλυση και καλύτερη οργάνωση συγκριτικά με τα συστήματα που βασίζονται αποκλειστικά σε χειροκίνητα μεταδεδομένα.

Η ενσωμάτωση εργαλείων ασύγχρονης διαχείρισης εργασιών επιτρέπει μια κλιμακούμενη και αυτοματοποιημένη ροή εργασιών. Σε αντίθεση με τις υπάρχουσες πλατφόρμες που συχνά απαιτούν χειροκίνητη παρέμβαση για την απομαγνητοφώνηση ή την κατηγοριοποίηση, το προτεινόμενο σύστημα επεξεργάζεται αυτόματα τα αρχεία πολυμέσων από τη μεταφόρτωση έως την ανάλυση, εξοικονομώντας χρόνο και βελτιώνοντας την αποδοτικότητα. Επιπλέον, παρέχει ενοποιημένα χαρακτηριστικά που υποστηρίζονται από την τεχνητή νοημοσύνη. Σε αντίθεση με τις πλατφόρμες που προσφέρουν μεμονωμένες λειτουργίες, όπως οι αυτόματοι υπότιτλοι του YouTube, το σύστημα ενσωματώνει ταξινόμηση και σημασιολογική διερεύνηση σε μια ενιαία πλατφόρμα, σχεδιασμένη ειδικά για τη διαχείριση και την ανάλυση πολυμέσων.

3.4. Ανάλυση Χρηστών

Ο ιστότοπος του Ag.U.TV. απευθύνεται κυρίως σε όλους τους φοιτητές αλλά και γενικά σε όλη την ακαδημαϊκή κοινότητα του Α.Π.Θ. Στα κύρια χαρακτηριστικά του προφίλ των χρηστών, περιλαμβάνονται μια κατανόηση τεχνολογίας πολυμέσων, ενδιαφέρον για διαδραστικές πληροφορίες και συνεχή αναζήτηση για οικονομικούς τρόπους μελέτης, επικοινωνίας και ψυχαγωγίας. Η συγκεκριμένη ιστοσελίδα απευθύνεται, παράλληλα, σε άτομα τα οποία έχουν έντονη καλλιτεχνική δημιουργικότητα και θέλουν να χρησιμοποιήσουν την ιστοσελίδα για να μοιραστούν και να αποθηκεύσουν οποιαδήποτε παραγωγή τους.

Η παρουσίαση του περιεχομένου είναι απλή και εύκολα προσβάσιμη από οποιοδήποτε πρόγραμμα περιήγησης ιστού, καθώς δεν απαιτούνται εξειδικευμένες γνώσεις. Το μόνο που χρειάζεται ένας χρήστης είναι βασικές γνώσεις χειρισμού υπολογιστή και σύνδεση στο διαδίκτυο (Ζαπάρτας Παναγιώτης & Στартσένκο Ιουλία, 2012).

Συνολικά στο σύστημα θα υπάρχουν τρία είδη χρηστών, ανάλογα με τα δικαιώματα και τις ικανότητες που θα έχουν στην ιστοσελίδα. Στην πρώτη κατηγορία ανήκουν οι απλοί επισκέπτες που θα έχουν πρόσβαση στο πολυμεσικό υλικό που διατίθεται στην ιστοσελίδα αλλά δε θα έχουν τη δυνατότητα να ανεβάσουν οι ίδιοι υλικό σε αυτήν. Στη δεύτερη κατηγορία ανήκουν οι χρήστες, οι εγγεγραμμένοι, οι οποίοι θα έχουν τη δυνατότητα να βλέπουν και να αναρτούν υλικό. Οι εγγεγραμμένοι χρήστες, έχουν επίσης την δυνατότητα να μορφοποιήσουν ή να διαγράψουν το υλικό τους από την ιστοσελίδα. Για παράδειγμα, αν ένας αλγόριθμος δημιουργήσει μια ετικέτα που περιγράφει λανθασμένα το υλικό, ο χρήστης θα μπορεί να τη διαγράψει. Στην τελευταία κατηγορία ανήκουν οι χρήστες με δικαιώματα διαχειριστών, οι οποίοι είναι αρμόδιοι είτε για τις εφαρμογές του WordPress, είτε για τον έλεγχο σωστής λειτουργίας των αλγορίθμων και συνολικά, του συστήματος.

3.5. Απαιτήσεις Συστήματος

Στη παρούσα ενότητα θα παρατεθούν οι λειτουργικές και μη λειτουργικές απαιτήσεις της διαδικτυακής εφαρμογής. Στις λειτουργικές απαιτήσεις δεν περιέχεται η ακριβής σχεδίαση του συστήματος αλλά περιγράφονται οι βασικές λειτουργίες. Οι μη λειτουργικές απαιτήσεις ορίζουν τις ιδιότητες και τους περιορισμούς του συστήματος.

3.5.1. Λειτουργικές Απαιτήσεις

- Ο χρήστης θα πρέπει να μπορεί να βλέπει όλο το υλικό (μαζί με τα μεταδεδομένα του) που έχει αναρτηθεί στην διαδικτυακή εφαρμογή (Τίτλος, Κατηγορία, Ετικέτες, Κυρίως Κείμενο, Φωτογραφία).
- Ο χρήστης θα πρέπει να μπορεί να κάνει εγγραφή στην διαδικτυακή εφαρμογή.
- Ο χρήστης θα πρέπει να μπορεί να συνδεθεί στην διαδικτυακή εφαρμογή.
- Ο εγγεγραμμένος χρήστης θα πρέπει να έχει την δυνατότητα να αναρτήσει οπτικοακουστικό υλικό στο σύστημα.
- Ο εγγεγραμμένος χρήστης θα πρέπει να έχει την δυνατότητα της επεξεργασίας ή διαγραφής του περιεχομένου που αναρτά.
- Ο εγγεγραμμένος χρήστης θα πρέπει να μπορεί να βλέπει στην κατάλληλη σελίδα το οπτικοακουστικό υλικό που έχει ο ίδιος αναρτήσει.
- Ο χρήστης θα πρέπει να μπορεί να αναζητήσει οπτικοακουστικό υλικό με βάση την κατηγορία που θέλει ή κάποια λέξη κλειδί.
- Το σύστημα θα πρέπει να μπορεί να δέχεται αρχεία μεγάλου μεγέθους.
- Το σύστημα θα πρέπει να μπορεί να εκκινήσει αυτόματα τους αλγόριθμους για κάθε αρχείο πολυμέσου που ανεβαίνει στην ιστοσελίδα.
- Το σύστημα θα πρέπει να μπορεί να αποθηκεύει αυτόματα τα αποτελέσματα των αλγόριθμων στην βάση δεδομένων στο κατάλληλο πολυμεσικό αρχείο.
- Το σύστημα θα πρέπει να μπορεί να προβάλλει τα αποτελέσματα που έχουν αποθηκευτεί στη βάση δεδομένων και να αντιστοιχίζονται στο κατάλληλο πολυμεσικό αρχείο.
- Το σύστημα θα πρέπει να μπορεί να δημιουργεί μια εικόνα για κάθε αρχείο που αναρτάται και να προβάλλεται στην ιστοσελίδα.
- Το σύστημα θα πρέπει να μπορεί να αποθηκεύει αυτόματα όλα τα πολυμεσικά αρχεία που αναρτώνται στην ιστοσελίδα μαζί και με τα μεταδεδομένα τους στη βάση δεδομένων.
- Ο διαχειριστής θα πρέπει να μπορεί να επεξεργάζεται, να διαγράφει και να πραγματοποιεί ενημέρωση στις εφαρμογές του WordPress.

3.5.2. Μη λειτουργικές απαιτήσεις

Οι μη λειτουργικές απαιτήσεις του συστήματος συνοψίζονται στον Πίνακα 3-1.

Πίνακας 3-1 Μη λειτουργικές απαιτήσεις του συστήματος

Κατηγορία	Απαιτηση
Απόδοση	Το σύστημα θα πρέπει να ανταποκρίνεται στις αλληλεπιδράσεις του χρήστη σε διάστημα εντός 5 δευτερολέπτων.
	Το σύστημα θα πρέπει να επιτρέπει αρχεία μεγέθους μέχρι 2048Mbytes.
	Το σύστημα θα πρέπει να επιτρέπει αρχεία ήχου της μορφής mp3 και wav.
	Το σύστημα θα πρέπει να επιτρέπει αρχεία βίντεο της μορφής mp4.
	Το σύστημα θα πρέπει να κατέχει "RAM" μεγέθους τουλάχιστον 16GB και "GPU" μεγέθους τουλάχιστον 8GB.
	Το σύστημα θα πρέπει να κατέχει τουλάχιστον 4 πυρήνες επεξεργαστή.
Ευχρηστία	Το σύστημα θα πρέπει να είναι προσβάσιμο από όλους τους περιηγητές ιστών και από οποιαδήποτε συσκευή.
Μεταφερσιμότητα	Το σύστημα θα πρέπει να επιτρέπει ταυτόχρονες μεταφορτώσεις από τουλάχιστον 10 χρήστες.
	Το σύστημα θα πρέπει να επιτρέπει τουλάχιστον 50 χρήστες να είναι ταυτόχρονα συνδεδεμένοι σε αυτό.
Σταθερότητα	Το σύστημα θα πρέπει να έχει χρόνο επανεκκίνησης μετά από σφάλμα 1 λεπτό.
	Το σύστημα θα πρέπει να έχει ποσοστό 0.1% ενεργειών που οδηγούν σε σφάλμα.
	Το σύστημα θα πρέπει να έχει πιθανότητα απώλειας δεδομένων 0.01%.
Νομικά	Το σύστημα θα πρέπει να προστατεύει τα προσωπικά στοιχεία του χρήστη.

3.6. Εργαλεία Ανάπτυξης και Τεχνολογίες

3.6.1. Συστήματα Διαχείρισης Περιεχομένου

Η διαχείριση του περιεχομένου αποτελεί κρίσιμο παράγοντα για την αποτελεσματική λειτουργία της ιστοσελίδας, δεδομένου ότι επηρεάζει άμεσα την ευκολία ενημέρωσης, την απόδοση και τη συνεχή ανάπτυξη του συστήματος. Στη σύγχρονη αγορά κυριαρχούν πλατφόρμες όπως το WordPress, το Joomla και το Drupal. Ο κύριος λόγος που χρησιμοποιήθηκε ένα υψηλού επιπέδου πλαίσιο εργασίας όπως το WordPress και όχι ένα χαμηλού επιπέδου όπως το Django ή Flask είναι η ικανότητά του να διαχειρίζεται και να συνεχίζει να εξελίσσεται από φοιτητές, ασχέτως αν έχουν εξειδικευμένες γνώσεις ή όχι. Έτσι, ένας φοιτητής που διαθέτει ειδικές γνώσεις μπορεί, μετέπειτα, να προσθέσει και άλλους αλγορίθμους που θα προσφέρουν επιπλέον λειτουργικότητες, ενώ ένας φοιτητής που δεν διαθέτει ειδικές γνώσεις θα μπορεί να επικεντρωθεί στη βελτίωση του δημοσίου τμήματος του Ar.U.TV. Το WordPress έχει αποκτήσει ιδιαίτερη φήμη για την ευχρηστία του και την ευρεία υποστήριξη που προσφέρει μέσω ενός εκτενούς οικοσυστήματος πρόσθετων και θεμάτων, γεγονός που επιτρέπει την ταχεία προσαρμογή και την προσθήκη νέων λειτουργιών από χρήστες με διάφορα επίπεδα εξειδίκευσης. Από την άλλη πλευρά, οι πλατφόρμες όπως το Joomla και το Drupal προσφέρουν περισσότερες δυνατότητες προσαρμογής και μεγαλύτερο βαθμό ευελιξίας σε πιο σύνθετα έργα, αλλά απαιτούν και πιο εξειδικευμένες γνώσεις για τη σωστή υλοποίησή τους. Το WordPress υπερτερεί συγκριτικά με το Joomla και το Drupal σε χαρακτηριστικά απόδοσης που σχετίζονται με το χρόνο φόρτωσης σελίδας, τις αιτήσεις στο διακομιστή, τη διαχείριση αρχείων CSS και την προσωρινή αποθήκευση δεδομένων. Αποτελεί, επιπρόσθετα, το βέλτιστο (και ιδιαίτερα ισχυρό) σύστημα

διαχείρισης περιεχομένου κυρίως για δυναμικές και πραγματικού χρόνου διαδικτυακές εφαρμογές. Τέλος, προσφέρει χρήσιμες λειτουργίες στην περιοχή διαχείρισης, παρέχοντας δωρεάν εξυπηρετητή και είναι ισχυρό για μηχανές αναζήτησης (Martínez-Caro et al., 2018; Patel et al., 2011a; Patel et al., 2011b). Όλα αυτά τα χαρακτηριστικά καθιστούν το WordPress μια χρηστική επιλογή για ποικίλες ανάγκες ανάπτυξης ιστού.

3.6.2. Διαμεσολαβητές Μηνυμάτων

Για την ενίσχυση της επικοινωνίας μεταξύ των υποσυστημάτων της εφαρμογής, η επιλογή ενός αξιόπιστου διαμεσολαβητή μηνυμάτων είναι ζωτικής σημασίας. Σε αυτό το πλαίσιο, τεχνολογίες όπως το RabbitMQ, το Kafka, το ActiveMQ και το NATS έχουν δοκιμαστεί και αξιολογηθεί διεξοδικά. Το RabbitMQ αποτελεί καλύτερη επιλογή σε θέματα επεκτασιμότητας και απόδοσης, προσφέροντας σταθερότητα σε συνθήκες αυξημένου όγκου μηνυμάτων. Οι δυνατότητες ασύγχρονης επικοινωνίας του RabbitMQ ενισχύουν περαιτέρω την αποτελεσματικότητα των συστημάτων, επιτρέποντας την εκτέλεση των αλγοριθμικών εργασιών στο παρασκήνιο και βελτιώνοντας, κατ' επέκταση, την εμπειρία του χρήστη στην ιστοσελίδα (Hong et al., 2018). Αντίθετα, ενώ το Kafka είναι εξαιρετικά αποδοτικό σε περιβάλλοντα με πολύ υψηλές ροές δεδομένων, η πολυπλοκότητα στη ρύθμιση και η απαιτούμενη διαχείριση το καθιστούν λιγότερο ελκυστικό για το συγκεκριμένο έργο. Τεχνολογίες όπως το ActiveMQ και το NATS επίσης παρουσιάζουν ενδιαφέρουσες δυνατότητες, αλλά δεν ανταποκρίνονται πλήρως στις απαιτήσεις της εφαρμογής όσον αφορά την ευελιξία και τη λειτουργικότητα με άλλα συστήματα.

Ανάλογα με την αρχιτεκτονική του κάθε συστήματος και τις απαιτήσεις που υπάρχουν, πραγματοποιείται και η κατάλληλη επιλογή του εκάστοτε συστήματος. Το RabbitMQ είναι ένας ευέλικτος διαμεσολαβητής ο οποίος προσφέρει τη δυνατότητα δημιουργίας απλών και πολύπλοκων συστημάτων δρομολόγησης μηνυμάτων. Συγκριτικά με του υπόλοιπους διαμεσολαβητές, λόγω της λειτουργίας βεβαίωσης λήψης μηνυμάτων, μπορεί να εγγραφεί ότι όλα τα μηνύματα θα σταλούν στους καταναλωτές, γεγονός το οποίο συνεπάγεται την καθολική εκτέλεση των αλγορίθμων για κάθε πολυμεσικό αρχείο που αναρτάται (Sharvari & Sowmya, 2019). Ένα ακόμα πλεονέκτημα του είναι το γεγονός ότι, λόγω του πρωτοκόλλου AMQP, μπορεί να πραγματοποιήσει συνδέσεις μεταξύ διαφορετικών λογισμικών με ενδεχόμενη διαφορετική αρχιτεκτονική, γλώσσα προγραμματισμού ή λειτουργικό σύστημα (Ionescu, 2015). Συνεπώς, πραγματοποιείται πιο εύκολα η σύνδεση του με το WordPress, το οποίο είναι γραμμένο σε γλώσσα προγραμματισμού PHP, σε αντίθεση με τους αλγορίθμους οι οποίοι είναι γραμμένοι στη γλώσσα προγραμματισμού Python. Τέλος, προσδίδει μια πιο απλή προσέγγιση σε ό,τι αφορά τα θέματα επεκτασιμότητας, ενώ μπορεί να λειτουργήσει σε συστήματα με μειωμένους πόρους (Dobbelaere & Sheykh Esmaili, 2017).

3.6.3. Κατανεμημένα συστήματα ουράς εργασιών

Η αποτελεσματική διαχείριση των ασύγχρονων εργασιών και η κατανομή τους σε πολλούς επεξεργαστές είναι κρίσιμη για την απόδοση εφαρμογών που υλοποιούν απαιτητικές υπολογιστικές διεργασίες, όπως είναι οι αλγόριθμοι μηχανικής μάθησης και η επεξεργασία πολυμέσων. Στον χώρο αυτό, το Celery, ένα κατανεμημένο σύστημα ουράς εργασιών γραμμένο σε Python, έχει αναδειχθεί ως μια από τις πιο αξιόπιστες λύσεις, λόγω της δυνατότητάς του να διαχειρίζεται παράλληλες διεργασίες και να ενσωματώνεται εύκολα με διαμεσολαβητές μηνυμάτων όπως το RabbitMQ. Έχει σχεδιαστεί για να απλοποιεί την εκτέλεση ασύγχρονων εργασιών σε διάφορες εφαρμογές. Αναθέτει την εκτέλεση των απαιτητικών χρονικά εργασιών, οι οποίες εκτελούνται από τους εργάτες (workers) στο παρασκήνιο. Αυτό έχει ως αποτέλεσμα τη βελτίωση της συνολικής απόδοσης και ανταπόκρισης των εφαρμογών (Harshjeet et al., 2024). Εναλλακτικές λύσεις, όπως το Redis, παρουσιάζουν επίσης ενδιαφέρον, αλλά συχνά δεν παρέχουν το ίδιο επίπεδο ευελιξίας και επεκτασιμότητας, ειδικά όταν οι

ανάγκες του συστήματος περιλαμβάνουν την ολοκληρωμένη διαχείριση μακροχρόνιων και απαιτητικών εργασιών. Η ανάλυση των διαφορετικών τεχνολογιών υποδεικνύει ότι το Celery, σε συνδυασμό με το RabbitMQ, προσφέρει ένα ολοκληρωμένο και αξιόπιστο περιβάλλον για την κατανομή των εργασιών, διασφαλίζοντας έτσι ότι οι βαριές υπολογιστικές διεργασίες δεν επηρεάζουν την κύρια εφαρμογή.

3.6.4. Λοιπά εργαλεία και βιβλιοθήκες

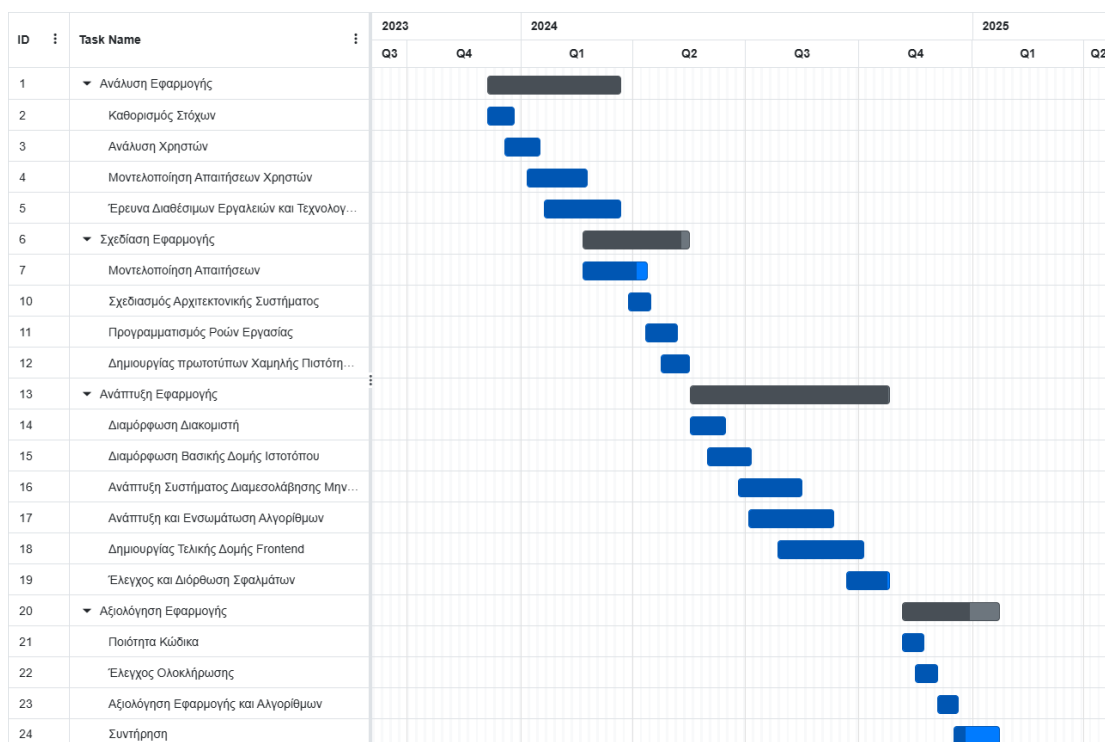
Εκτός από τα παραπάνω εργαλεία και τεχνολογίες, για την σωστή και επιτυχή ανάπτυξη του διαδικτυακού ιστοτόπου και της δημιουργίας του αλγορίθμου, χρησιμοποιήθηκε ένα εύρος εργαλείων και βιβλιοθηκών. Το σύνολο αυτών αναφέρεται στο Πίνακα 3-2.

Πίνακας 3-2 Πίνακας Χρησιμοποιούμενων Τεχνολογιών και Εργαλείων

Εργαλείο/Βιβλιοθήκη	Περιγραφή	Χρήση
MySQL	Σχεσιακή βάση δεδομένων συμβατή με το WordPress.	Αποθήκευση δεδομένων συστήματος και χρηστών.
PHP	Γλώσσα προγραμματισμού για προσθήκες στο WordPress.	Δημιουργία plugins για επιπλέον λειτουργικότητα.
Python	Γλώσσα προγραμματισμού για αλγορίθμους και διαχείριση RabbitMQ.	Εκτέλεση αλγορίθμων και αλληλεπίδραση με το RabbitMQ.
Javascript	Γλώσσα για διαδραστικά στοιχεία στο περιβάλλον χρήστη.	Δημιουργία λειτουργικότητας αναζήτησης και κουμπιών.
php-amqp	Υπο-βιβλιοθήκη PHP για AMQP και RabbitMQ.	Αλληλεπίδραση του WordPress με το RabbitMQ.
Pika	Βιβλιοθήκη Python για RabbitMQ.	Αλληλεπίδραση Python εφαρμογών με RabbitMQ.
PymySQL	Βιβλιοθήκη για σύνδεση με τη βάση δεδομένων MySQL.	Σύνδεση Python εφαρμογών με MySQL.
Apache HTTP Server	Διακομιστής ιστού για WordPress.	Παράδοση περιεχομένου και διαχείριση αρχείων.
Visual Studio Code	Προγραμματιστικό περιβάλλον.	Εγγραφή και επεξεργασία κώδικα.
NVIDIA GPU	Υποστήριξη για επιτάχυνση εκτέλεσης αλγορίθμων.	Βελτίωση ταχύτητας εκτέλεσης αλγορίθμων.
Ffmpeg	Λογισμικό ανοιχτού κώδικα για μετατροπή πολυμέσων.	Μετατροπή μορφών αρχείων για την καλύτερη χρήση τους.
PyTorch	Πλαίσιο μηχανικής μάθησης.	Εκπαίδευση μοντέλων και χρήση GPU.
OpenCV	Βιβλιοθήκη για επεξεργασία εικόνας και βίντεο.	Ανάλυση καρέ και εξαγωγή χαρακτηριστικών.
SSL Certificate	Ψηφιακό πιστοποιητικό ασφαλείας.	Ασφάλεια μετάδοσης δεδομένων.
Ubuntu Services	Υπηρεσίες με το systemd για Celery και RabbitMQ.	Αυτόματη εκκίνηση και διαχείριση ασύγχρονων εργασιών.
Anaconda Environments	Απομόνωση εξαρτήσεων.	Δημιουργία ξεχωριστών περιβαλλόντων για κάθε αλγόριθμο.
PyWavelets	Βιβλιοθήκη python για το μετασχηματισμό κυματιδίων.	Υπολογισμός του διακριτού μετασχηματισμού των κυματιδίων για τη δημιουργία χαρακτηριστικών για την ταξινόμηση ήχου.
vSphere Client	Κεντρική διεπαφή για τη διαχείριση και τον έλεγχο απόδοσης του διακομιστή.	Διαχείριση και παρακολούθηση διακομιστή.

3.7. Οργάνωση και χρονοπρογραμματισμός εργασιών

Η ανάπτυξη της εφαρμογής σχεδιάστηκε και εκτελέστηκε σχολαστικά μέσω ενός δομημένου χρονοδιαγράμματος, όπως απεικονίζεται στο διάγραμμα Gantt (Εικόνα 3-1). Αυτό το διάγραμμα παρέχει μια ολοκληρωμένη επισκόπηση των βασικών φάσεων, εξασφαλίζοντας μια οργανωμένη προσέγγιση για την επίτευξη των στόχων του έργου.



Εικόνα 3-1 Πλάνο ανάπτυξης εφαρμογής (Διάγραμμα Gantt)

3.8. Σύνοψη Τρίτου Κεφαλαίου

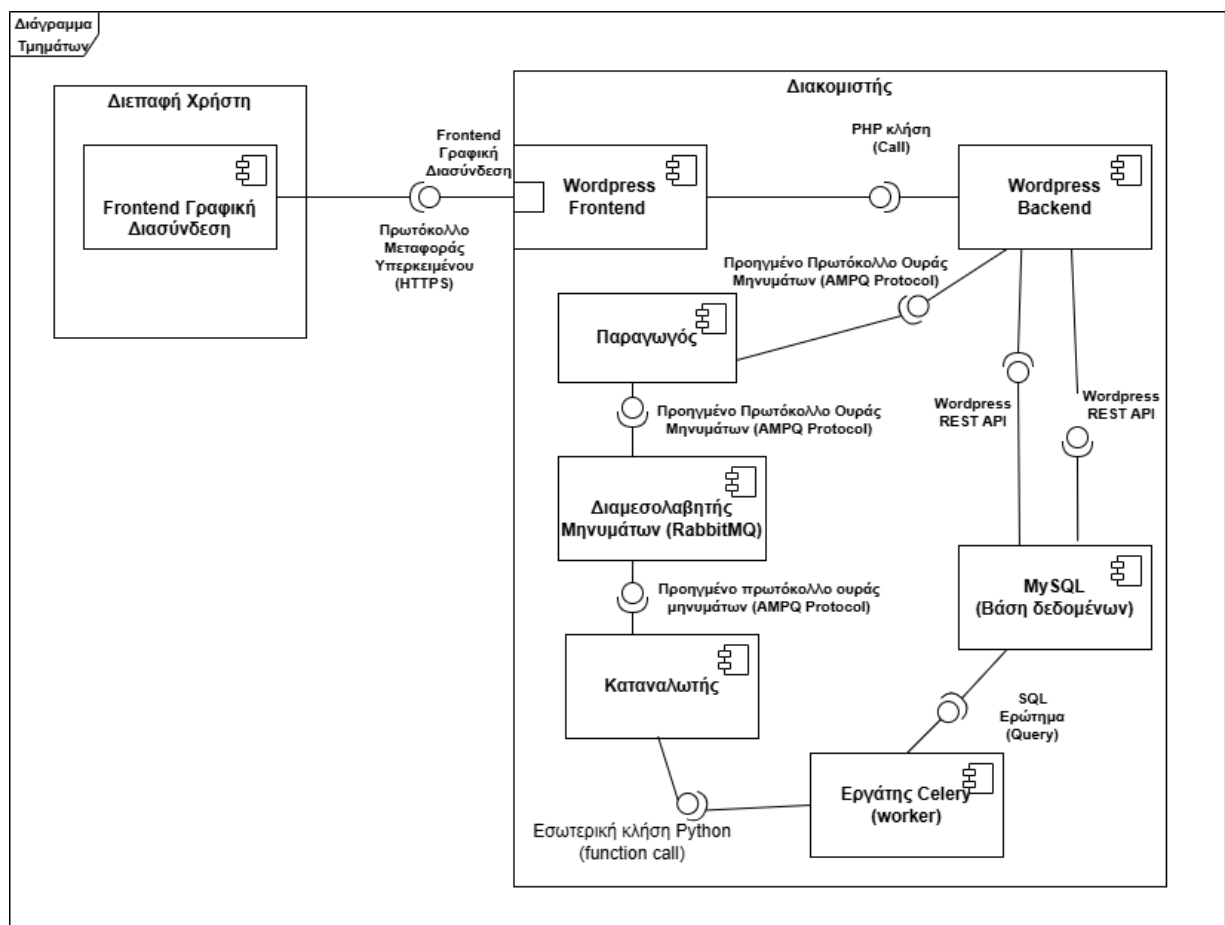
Το κεφάλαιο αυτό παρουσιάζει την αρχική ιδέα και την εξέλιξη του Ar.U.TV, ενός τηλεοπτικού ιστοτόπου του Α.Π.Θ., καθώς και τους στόχους της παρούσας διατριβής για την ανακατασκευή του με σύγχρονες τεχνολογίες μηχανικής μάθησης. Συγκρίνονται οι δυνατότητες του Ar.U.TV με τις δημοφιλείς πλατφόρμες YouTube και Vimeo, αναδεικνύοντας τη διαφοροποίησή του μέσω της ενσωμάτωσης προηγμένων τεχνολογιών τεχνητής νοημοσύνης. Γίνεται ανάλυση των αλγορίθμων που μπορούν να συνεισφέρουν στην καλύτερη διαχείριση, οργάνωση και ανάκτηση οπτικοακουστικού υλικού. Παρουσιάζονται οι απαιτήσεις του συστήματος, η ανάλυση χρηστών, τα εργαλεία ανάπτυξης και οι διαθέσιμες τεχνολογίες. Τέλος, αναλύεται ο προγραμματισμός εργασιών, διασφαλίζοντας την αποτελεσματική υλοποίηση του έργου. Στο επόμενο κεφάλαιο θα ξεκινήσει η σχεδίαση του συστήματος από την βασική αρχιτεκτονική, πώς επικοινωνούν τα διάφορα συστατικά της, ποιοι αλγόριθμοι επιλέχθηκαν και πώς κινείται η ροή της πληροφορίας μέσα στο σύστημα.

Κεφάλαιο 4ο

4. Σχεδίαση Εφαρμογής

4.1. Αρχιτεκτονική Συστήματος και Πρωτοκόλλα Επικοινωνίας

Ως σύστημα διαχείρισης περιεχομένου χρησιμοποιήθηκε το Wordpress. Η διεπαφή χρήστη (Frontend) αλληλεπιδρά με το WordPress Frontend μέσω του HTTPS (Hypertext Transfer Protocol Secure), το οποίο χρησιμεύει ως σημείο εισόδου για τις αλληλεπιδράσεις των χρηστών. Το WordPress Backend χειρίζεται την επιχειρησιακή λογική και την επεξεργασία δεδομένων, εκθέτοντας τη λειτουργικότητα μέσω του WordPress REST API για επικοινωνία με άλλα στοιχεία του συστήματος. Ένα πρόσθετο WordPress (plugin) εντός του backend ενεργεί ως παραγωγός, υπεύθυνος για τη δημοσίευση μηνυμάτων στον διαμεσολαβητή μηνυμάτων (RabbitMQ). Ο διαμεσολαβητής διευκολύνει την ασύγχρονη ανταλλαγή μηνυμάτων, εξασφαλίζοντας αξιόπιστη ροή δεδομένων μεταξύ των στοιχείων χρησιμοποιώντας το πρωτόκολλο AMQP. Τα μηνύματα καταναλώνονται από έναν Καταναλωτή, ο οποίος τα επεξεργάζεται και αναθέτει ασύγχρονα τις εργασίες (εκτέλεση αλγορίθμων) σε μια κατανεμημένη ουρά εργασιών, δηλαδή σε έναν Celery Worker, μέσω κλήσεων συναρτήσεων Python (Function Calls). Ο Celery Worker εκτελεί τους αλγορίθμους στο παρασκήνιο και αποθηκεύει τα αποτελέσματα στη βάση δεδομένων MySQL μέσω μιας σύνδεσης SQL. Αυτή η αρχιτεκτονική εξασφαλίζει αρθρωτότητα, επεκτασιμότητα και αποτελεσματικό χειρισμό εργασιών, αξιοποιώντας την επικοινωνία RESTful για σύγχρονα αιτήματα και το AMQP για ασύγχρονη επεξεργασία. Η Εικόνα 4-1 αποτελεί ένα διάγραμμα τμημάτων (UML Component diagram, όπου σκοπός της χρήσης του είναι η παρουσίαση της αρχιτεκτονικής του συστήματος σε υψηλό επίπεδο, αποτυπώνοντας τις συσχετίσεις και τις συνδέσεις μεταξύ των διαφορετικών τμημάτων του συστήματος, καθώς και πως επικοινωνούν μεταξύ τους.



Εικόνα 4-1 Αρχιτεκτονική του συστήματος της εφαρμογής

4.2. Αλγόριθμοι Εφαρμογής

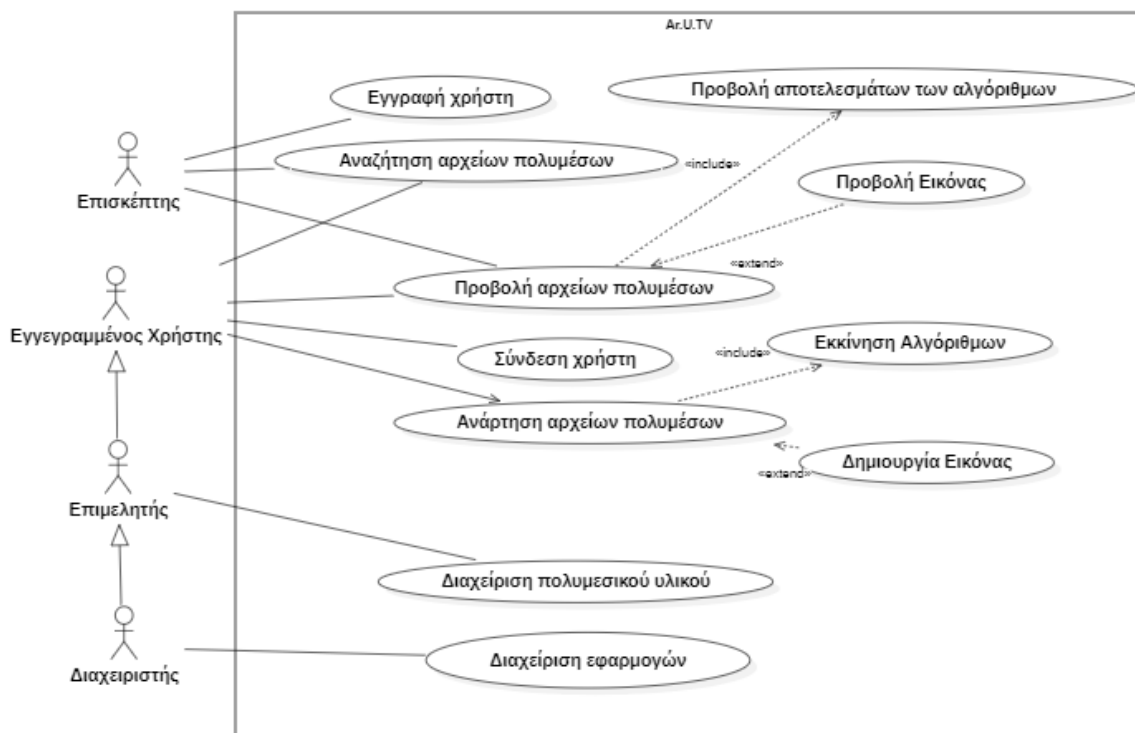
Ο Πίνακας 4-1 συνοψίζει τους αλγόριθμους που επιλέχθηκαν για την ανάπτυξη του συστήματος διαχείρισης οπτικοακουστικού υλικού, περιγράφοντας τα κύρια χαρακτηριστικά τους και τη συνεισφορά τους στη λειτουργικότητα του συστήματος.

Πίνακας 4-1 Πίνακας παρουσίασης των επιλεγμένων μοντέλων και αλγορίθμων του συστήματος

Αλγόριθμος	Περιγραφή	Κύρια Χαρακτηριστικά	Επιλογή Μοντέλου	Λόγος Επιλογής	Ρόλος στο Σύστημα
Whisper	Μοντέλο αναγνώρισης ομιλίας για απομαγνητοφώνηση της ομιλίας σε κείμενο.	Υποστήριξη πολλών γλωσσών, απομαγνητοφώνηση σε πραγματικό χρόνο, πολύ καλή ακρίβεια σε ποικίλες συνθήκες ήχου.	Μέγεθος μοντέλου: Medium (ισορροπία ταχύτητας, ακρίβειας και χρήση μνήμης).	Ισορροπία στην ταχύτητα και την ακρίβεια, κατάλληλο για συστήματα με περιορισμένους πόρους μνήμης.	Δημιουργία υποτίτλων αλλά και για περαιτέρω ανάλυση του προφορικού λόγου.
SpaCy	Βιβλιοθήκη φυσικής γλωσσικής επεξεργασίας για αναγνώριση ονομαστικών οντοτήτων (π.χ., ονόματα, τοποθεσίες, οργανισμοί).	Γρήγορη και ακριβής ανάλυση φυσικής γλώσσας. Υποστηρίζει αναγνώριση οντοτήτων, κατηγοριοποίηση και εξαγωγή σχέσεων. Διαθέτει προ-εκπαιδευμένο μοντέλο για την ελληνική γλώσσα.	el_core_news_lg (μεγάλο μοντέλο για την ελληνική γλώσσα).	Προσφέρει υψηλή ακρίβεια στην αναγνώριση οντοτήτων στο ελληνικό κείμενο και είναι κατάλληλο για μεγαλύτερα κείμενα, λόγω του μεγέθους του μοντέλου.	Προσδιορίζει και κατηγοριοποιεί οντότητες από το κείμενο της απομαγνητοφώνηση, βελτιώνοντας τα μεταδεδομένα για προηγμένες λειτουργίες αναζήτησης.
MMAction2	Εργαλειοθήκη ανοικτού κώδικα που υποστηρίζει πολυάριθμα μοντέλα κατανόησης βίντεο, συμπεριλαμβανομένης της αναγνώρισης δράσης.	Ανάλυση και αναγνώριση δράσεων σε βίντεο με χρήση προ-εκπαιδευμένων μοντέλων. Διαθέτει χωροχρονική ανάλυση με βάση τη βαθιά μάθηση.	ResNet-50, προ-εκπαιδευμένο στο Kinetics-400.	Προσφέρει ισχυρές δυνατότητες αναγνώρισης ανθρωπίνων ενεργειών σε βίντεο με γρήγορες κινήσεις.	Δημιουργεί μεταδεδομένα βασισμένα σε ενέργειες, υποστηρίζοντας την κατηγοριοποίηση και την ανάκτηση βίντεο με βάση τις ανθρώπινες δραστηριότητες.
Ταξινόμηση Ήχου	Αλγόριθμος που αναπτύχθηκε στα πλαίσια της διπλωματικής εργασίας για την ταξινόμηση του ήχου σε κατηγορίες όπως ομιλία, μουσική και άλλα (περιβαλλοντικοί ήχοι).	Χρησιμοποιεί φασματογραφήματα με βάση τα κυματίδια για καλύτερη εξαγωγή χαρακτηριστικών τόσο από το χρόνο όσο και από τη συχνότητα.	Το βέλτιστο μοντέλο που προέκυψε από την εκπαίδευση και αξιολόγηση σε δεδομένα εκπαίδευσης.	Παρέχει υψηλή ακρίβεια και αξιοπιστία στην ταξινόμηση ηχητικών δεδομένων, ενώ καταναλώνει ελάχιστους πόρους.	Δημιουργεί ετικέτες με βάση το ηχητικό περιεχόμενο, βελτιώνοντας την κατηγοριοποίηση και ανάκτηση του οπτικοακουστικού υλικού.

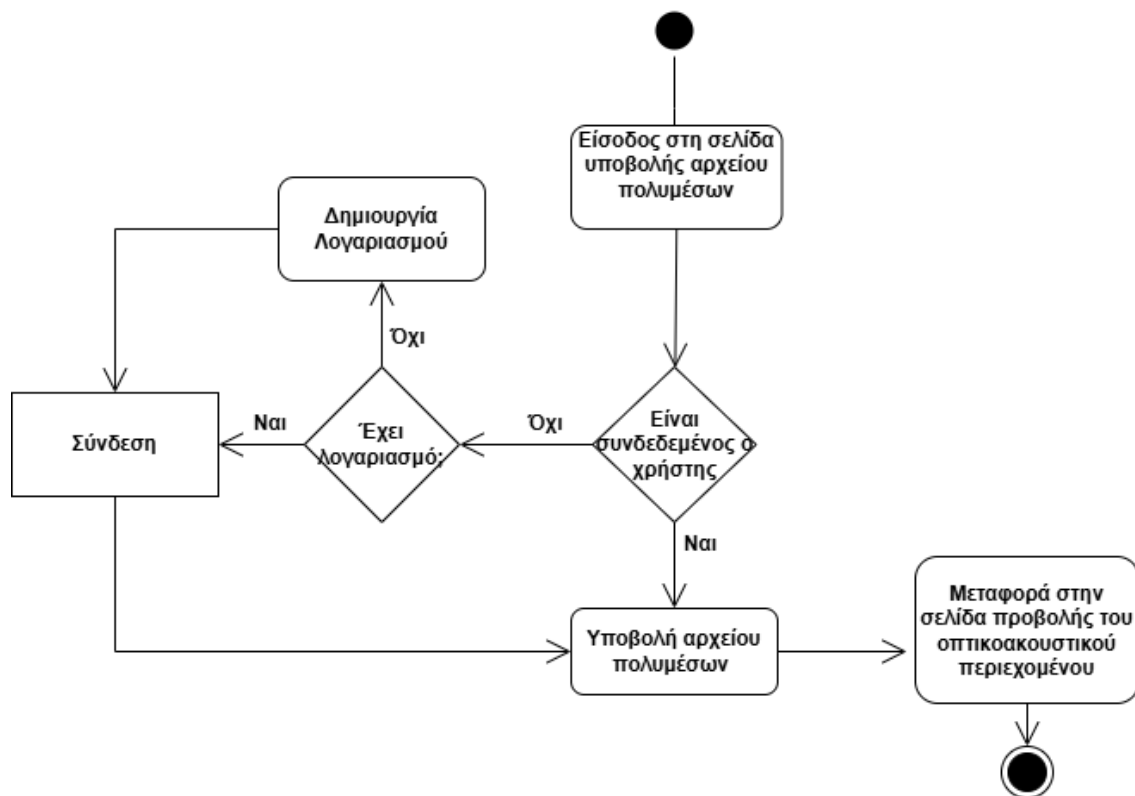
4.3. Σενάρια Χρήσης και Ροή Εργασιών του Συστήματος

Το διάγραμμα σεναρίων χρήσης (UML use-cases) λειτουργεί ως συνοπτική αναπαράσταση των λειτουργικών απαιτήσεων του συστήματος, που παρουσιάστηκαν στο προηγούμενο κεφάλαιο. Περιγράφει τις αλληλεπιδράσεις των χρηστών με το σύστημα, καταγράφοντας τα βασικά σενάρια χρήσης, χωρίς να εξηγεί τα βήματα που ακολουθούνται για να πραγματοποιηθούν τα σενάρια αυτά. Παρακάτω δίνεται το διάγραμμα σεναρίων χρήσης που περιέχει τις δυνατότητες των διαφόρων χρηστών στο σύστημα (Εικόνα 4-2).



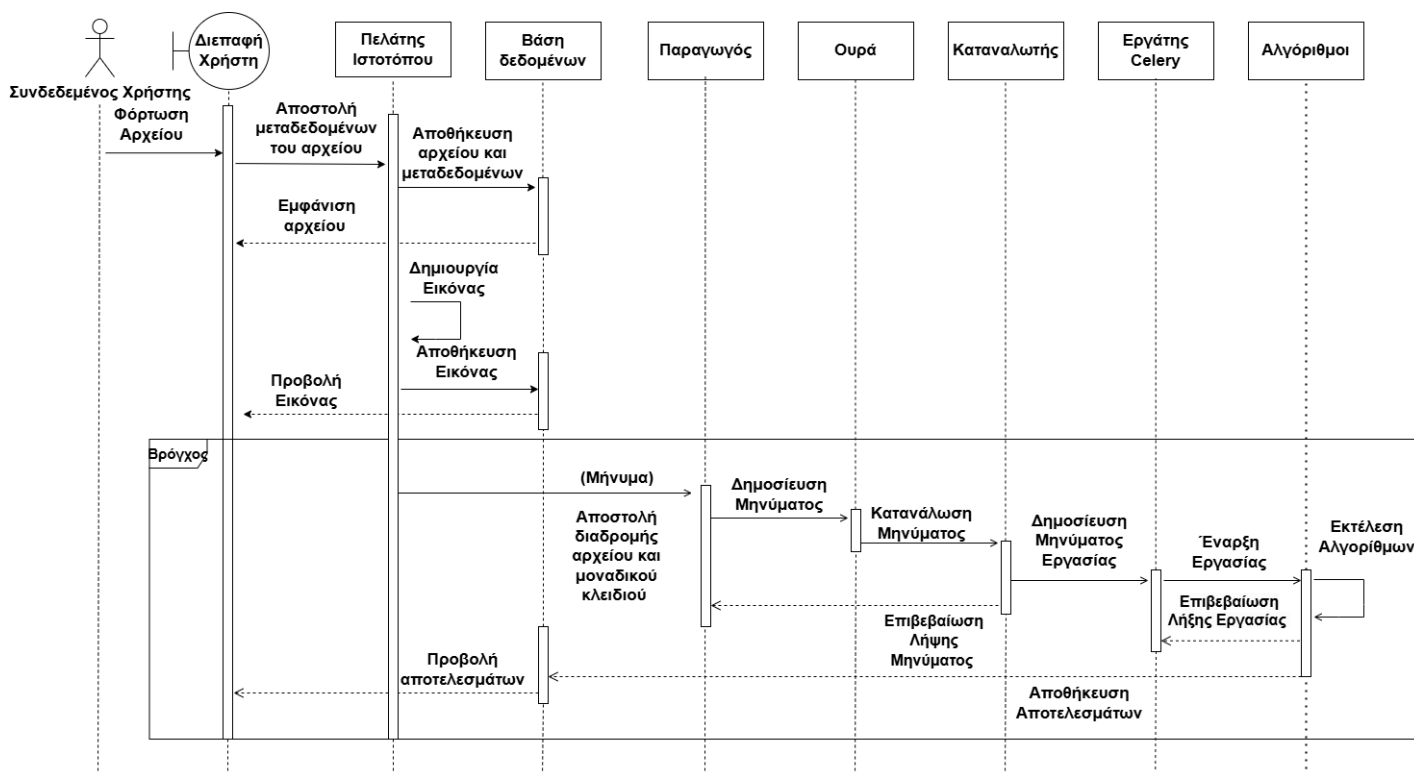
Εικόνα 4-2 Διάγραμμα σεναρίων χρήσης της εφαρμογής

Το βασικό σενάριο χρήσης του συστήματος είναι η ανάρτηση ενός πολυμεσικού αρχείου από έναν χρήστη στην ιστοσελίδα. Ένα διάγραμμα δραστηριοτήτων αποτυπώνει όλες τις δράσεις ενός χρήστη που προκύπτουν για ένα σενάριο χρήσης. Παρακάτω παρατίθεται ένα διάγραμμα δραστηριοτήτων για την ανάρτηση ενός βίντεο (Εικόνα 4.3).



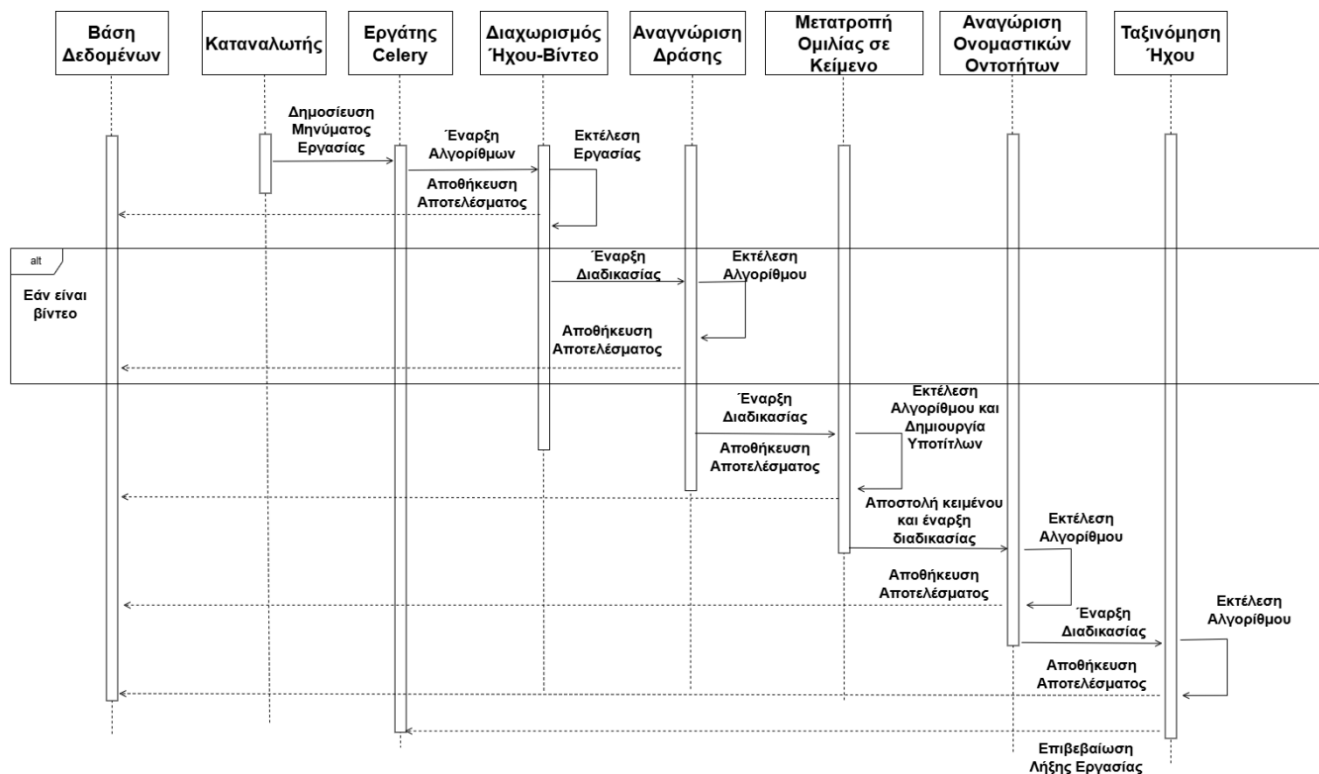
Εικόνα 4-3 Διάγραμμα δραστηριοτήτων βασικού σεναρίου της εφαρμογής

Η βασική διαφορά μεταξύ των δύο διαγραμμάτων είναι ότι στο διάγραμμα ακολουθιών αποτυπώνεται η αλληλεπίδραση των διαφόρων αντικειμένων του συστήματος ώστε να αναρτηθεί το υλικό στην ιστοσελίδα. Αντιθέτως, στο διάγραμμα δραστηριοτήτων αποτυπώνονται όλες η δράσεις που πραγματοποιεί ο χρήστης χωρίς, όμως, να φαίνεται ό,τι συμβαίνει στο σύστημα. Παρακάτω παρουσιάζεται το διάγραμμα ακολουθιών, με την προϋπόθεση ότι ο χρήστης είναι συνδεδεμένος στο σύστημα (Εικόνα 4-4). Μέσω της Εικόνας 4-4, γίνεται κατανοητό ότι αρχικά αναρτάται το αρχείο στο σύστημα, γίνεται η αποθήκευση του στη βάση δεδομένων, δημιουργείται μια εικόνα για το αρχείο αυτό και έπειτα, προβάλλεται στην ιστοσελίδα. Στη συνέχεια, με ασύγχρονη επικοινωνία, που επιτυγχάνεται μέσω τον συνδυασμό των συστατικών της αρχιτεκτονικής του RabbitMQ (καταναλωτές, ουρές και παραγωγοί) και των εργατών του Celery, αρχίζει η εκτέλεση των αλγόριθμων. Δηλαδή, μόλις αναρτηθεί το αρχείο στο σύστημα, αποστέλλεται ένα μήνυμα από τον παραγωγό σε μια ουρά, ώστε, στη συνέχεια, να ξεκινήσει ο αλγόριθμος για αυτό το αρχείο. Το μήνυμα αυτό περιέχει την ακριβή τοποθεσία του αρχείου μέσα στο σύστημα αρχείων του συστήματος και ένα μοναδικό κλειδί, με το οποίο έχει αντιστοιχηθεί η ανάρτηση του αρχείου στη βάση δεδομένων. Αφού, λοιπόν, ένα μήνυμα βρίσκεται στη ουρά, μεταφέρεται σε έναν καταναλωτή. Ο καταναλωτής με τη σειρά του στέλνει το μήνυμα αυτό στον εργάτη του Celery, ώστε να ξεκινήσουν να εκτελούνται οι αλγόριθμοι, οι οποίοι εκτελούνται στο παρασκήνιο του συστήματος. Αφού ξεκινήσει η διαδικασία εκτέλεσης των αλγορίθμων ο καταναλωτής στέλνει ένα μήνυμα επιβεβαίωσης ότι ο εργάτης έλαβε το μήνυμα και ξεκίνησε την εκτέλεση της εργασίας. Μετά το πέρας της εκτέλεσης κάθε αλγόριθμου, αποθηκεύονται τα αποτελέσματα στη βάση δεδομένων στην κατάλληλη ανάρτηση σύμφωνα με το κλειδί που δόθηκε, ώστε να γίνει και η προβολή τους στον ιστότοπο και στην κατάλληλη ανάρτηση. Η διαδικασία αυτή επαναλαμβάνεται για κάθε υλικό που αναρτάται στην ιστοσελίδα.



Εικόνα 4-4 Διάγραμμα ακολουθιών εργασιών στο σύστημα μετά την υποβολή αρχείων από συνδεδεμένους χρήστες

Στην Εικόνα 4-4 δε δίνεται έμφαση στον ακριβή τρόπο με τον οποίο εκτελούνται οι αλγόριθμοι. Για την καλύτερη κατανόηση του συστήματος γίνεται χρήση ενός ακόμα διαγράμματος ακολουθιών (Εικόνα 4-5) που περιγράφει αναλυτικά ό,τι συμβαίνει στο σύστημα όταν ο καταναλωτής στέλνει το μήνυμα στον εργάτη του Celery για την εκτέλεση των αλγορίθμων ενός αρχείου. Εξαιτίας της αυξημένης υπολογιστικής ανάγκης των αλγορίθμων σε συνδυασμό με το ελλιπές σε πόρους διαθέσιμο σύστημα, οι αλγόριθμοι για κάθε αρχείο εκτελούνται σειριακά. Όταν, λοιπόν, ο καταναλωτής στέλνει μια εργασία στον εργάτη του Celery τότε ξεκινά το πρώτο αρχείο κώδικα που αφορά στο διαχωρισμό των αρχείων ήχου με τα βίντεο. Λόγω της ύπαρξης βίντεο με στατικές εικόνες (ίδια εικόνα σε όλο το βίντεο) το αρχείο κώδικα έχει προγραμματιστεί με τέτοιο τρόπο ώστε να αναγνωρίζει τα βίντεο με τις στατικές εικόνες και να τα κατατάσσει στα αρχεία ήχου. Αφού αποθηκευτεί στο σύστημα το είδος του αρχείου, έπειτα ξεκινά η εκτέλεση του αλγόριθμου μετατροπής της ομιλίας σε κείμενο και με το πέρας του αλγόριθμου δημιουργούνται ελληνικοί και αγγλικοί υπότιτλοι και τα αποτελέσματα αποθηκεύονται στη βάση δεδομένων. Στη συνέχεια στέλνεται το μη δομημένο κείμενο του αλγόριθμου στον επόμενο αλγόριθμο αναγνώρισης οντοτήτων, ο οποίος και αυτός με τη σειρά του αποθηκεύει τα αποτελέσματά του στη βάση. Τέλος, εάν το αρχείο είναι βίντεο, εκτελείται και ο αλγόριθμος αναγνώρισης δράσης. Με το πέρας της εκτέλεσης όλων των αλγορίθμων στέλνεται αυτόματα ένα μήνυμα επιβεβαίωσης ότι εκτελέστηκαν όλοι οι αλγόριθμοι για το ένα αρχείο και μπορεί να ξεκινήσει η ίδια διαδικασία για το επόμενο μήνυμα.



Εικόνα 4-5 Διάγραμμα ακολουθιών για την εκτέλεση αλγορίθμων ενός αρχείου που έχει φορτωθεί στο σύστημα

4.4. Σχεδιασμός Διεπαφής Χρήστη: Από το Πρωτότυπο στη Λειτουργική Εμφάνιση

Η δημιουργία πρωτοτύπων χαμηλής πιστότητας είναι ένα πρώιμο βήμα στη διαδικασία σχεδιασμού που περιλαμβάνει τη δημιουργία βασικών σκίσεων, είτε χειρόγραφα είτε ψηφιακά, για την αναπαράσταση της δομής και της λειτουργικότητας ενός συστήματος. Αυτή η προσέγγιση επιτρέπει στους σχεδιαστές και τους προγραμματιστές να δοκιμάζουν ιδέες, να βελτιώνουν τις έννοιες και να αντιμετωπίζουν πιθανές προκλήσεις ευχρηστίας πριν επενδύσουν σε λεπτομερή ανάπτυξη. Όσο αφορά το συγκεκριμένο σύστημα, τα πρωτότυπα χαμηλής πιστότητας διαδραμάτισαν καθοριστικό ρόλο στη διαμόρφωση της διεπαφής του χρήστη και της εμπειρίας του συστήματος διαχείρισης οπτικοακουστικού περιεχομένου. Για την επίτευξη όλων των παραπάνω χρησιμοποιήθηκαν συρματογραφικά πλαίσια (wireframes), τα οποία επικεντρώνονται περισσότερο στα λειτουργικά χαρακτηριστικά. Για την δημιουργία των προτύπων χαμηλής πιστότητας χρησιμοποιήθηκε η εφαρμογή MockFlow Wireframepro.

4.4.1. Διεπαφή χρήστη για το διάγραμμα ροής δραστηριοτήτων

Το πρώτο βήμα είναι η δημιουργία των διεπαφών για τη βασικό σενάριο χρήσης που αφορά στην ανάρτηση ενός βίντεο, όπως αναφέρεται και στην Εικόνα 4-6. Θα υποθέσουμε ότι ο χρήστης δεν έχει λογαριασμό στην ιστοσελίδα οπότε αρχικά θα παρουσιάσουμε την σελίδα δημιουργίας λογαριασμού του χρήστη. Σε αυτή ο χρήστης συμπληρώνει το όνομα χρήστη, όπως θέλει να φαίνεται στην ιστοσελίδα, και στη συνέχεια ονοματεπώνυμο, το ηλεκτρονικό του ταχυδρομείο και τον κωδικό του.

Home AudioVisual Content Content Upload My content Login Register

Username
Your Username

First Name
Your First Name

Last Name
Your Last Name

mail@example.com

Minimum of 8 characters

Register

Εικόνα 4-6 Πρωτότυπο σελίδας εγγραφής

Στη συνέχεια παρουσιάζεται το πρότυπο της σελίδας που ο χρήστης συνδέεται στην ιστοσελίδα, όπου θα απαιτείται να συμπληρώσει το όνομα χρήστη και τον κωδικό του. Εφόσον ο χρήστης δεν διαθέτει λογαριασμό θα μπορεί να δημιουργήσει πατώντας τον αντίστοιχο σύνδεσμο ώστε να μεταβεί στην σελίδα δημιουργίας λογαριασμού (Εικόνα 4-7).

Home AudioVisual Content Content Upload My content Login Register

Username

Password

☐ Keep me signed in

Login Register

Forgot your password?

Εικόνα 4-7 Πρότυπο σελίδας σύνδεσης του χρήστη

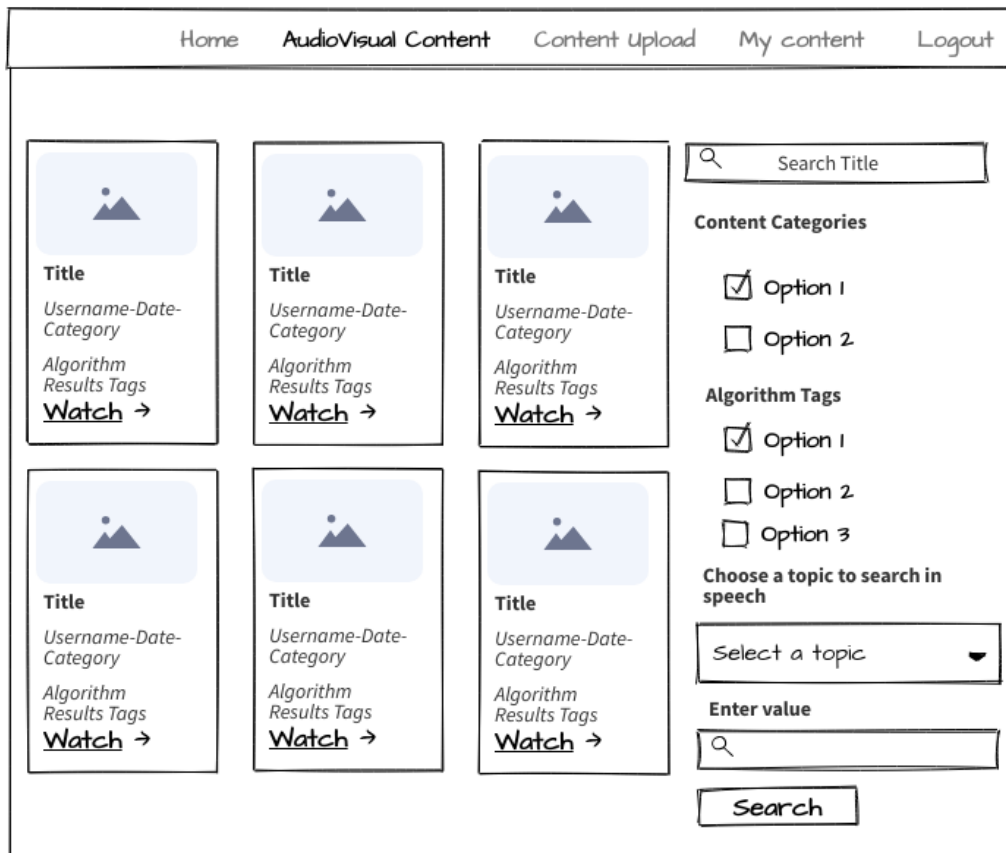
Εν συνεχεία, αφού ο χρήστης είναι συνδεδεμένος μπορεί να αναρτήσει ένα αρχείο ήχου ή βίντεο. Στην σελίδα ανάρτησης περιεχομένου θα συμπληρώνεται αυτόματα το όνομα του κάθε χρήστη, ενώ ο χρήστης θα μπορεί να ονομάσει το αρχείο του και να επιλέξει μια βασική κατηγορία στην οποία θεωρεί πως ανήκει (Εικόνα 4-8).

The image shows a web interface for uploading content. At the top, there is a navigation bar with links: Home, AudioVisual Content, Content upload, My content, and Logout. The main content area contains a form with the following elements:

- A text input field labeled "Username".
- A text input field labeled "Content Title".
- A dropdown menu labeled "Content Category" with a downward arrow. The dropdown is open, showing five options: Option 1, Option 2, Option 3, Option 4, and Option 5.
- A button labeled "Choose File To Upload".
- A button labeled "Upload".

Εικόνα 4-8 Πρότυπη σελίδα ανάρτησης αρχείου πολυμέσων

Τέλος, αφού ο χρήστης αναρτήσει το αρχείο μεταφέρεται στην σελίδα προβολής του οπτικοακουστικού υλικού (Εικόνα 4-9).



Εικόνα 4-9 Πρότυπη σελίδα προβολής οπτικοακουστικού υλικού

Σε αυτή τη σελίδα θα υπάρχουν όλα τα αρχεία που έχουν αναρτηθεί στην ιστοσελίδα. Κάθε αρχείο διαθέτει μια εικόνα και μερικά μεταδεδομένα, όπως το όνομα του χρήστη που το ανέβηκε, την ημερομηνία ανάρτησης και την γενική κατηγορία. Επίσης, θα προβάλλονται και τα αντίστοιχα αποτελέσματα των αλγορίθμων. Τέλος, ο χρήστης θα μπορεί να αναζητήσει είτε στους τίτλους των αρχείων είτε σε όσα έχουν ειπωθεί και παράλληλα θα μπορεί να επιλέξει να προβληθούν μόνο συγκριμένες κατηγορίες ή ετικέτες από τα αποτελέσματα των αλγορίθμων.

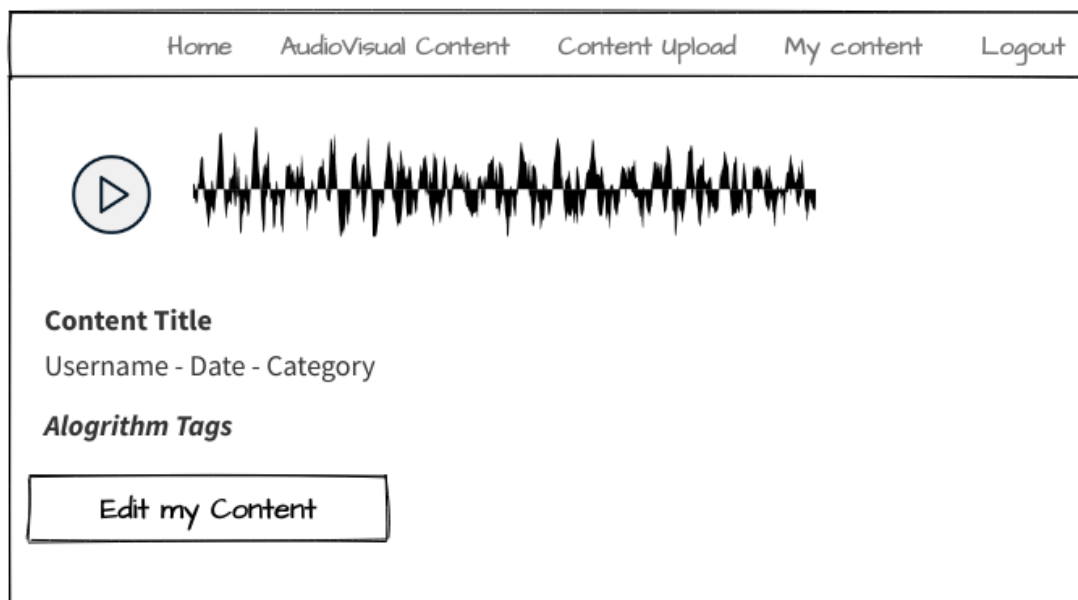
4.4.2. Σελίδα αναπαραγωγής Περιεχομένου

Το περιεχόμενο του συστήματος μπορεί να είναι είτε βίντεο είτε αρχείο ήχου. Σε κάθε περίπτωση η σελίδα αναπαραγωγής θα διαφέρει. Αν το περιεχόμενο είναι βίντεο τότε η σελίδα αναπαραγωγής θα εμφανίζεται με τον ίδιο τρόπο που αυτή απεικονίζεται στην Εικόνα 4-10. Σε αυτήν ο χρήστης θα μπορεί να παρακολουθεί το βίντεο και να προσθέσει και υπότιτλους, εάν το επιθυμεί, ενώ ταυτόχρονα, θα προβάλλονται τα μεταδεδομένα και τα αποτελέσματα των αλγορίθμων. Ανάμεσα σε αυτά θα είναι και το αποτέλεσμα του αλγορίθμου αναγνώρισης δράσης, όπου πατώντας το κουμπί ο χρήστης θα μεταφέρεται στο αντίστοιχο χρονικό σημείο του αποτελέσματος του αλγορίθμου στο βίντεο. Τέλος δίνεται η δυνατότητα επεξεργασίας του περιεχομένου, εφόσον το περιεχόμενο ανήκει στον χρήστη που το ανέβασε στην ιστοσελίδα.



Εικόνα 4-10 Πρότυπη σελίδα αναπαραγωγής βίντεο

Η διαφορά προβολής ενός αρχείου ήχου με ένα αρχείο βίντεο, έγκειται στο γεγονός ότι στο αρχείο ήχου δεν θα περιέχονται τα κουμπιά της αναγνώρισης δράσης και αντί της προβολής βίντεο θα προβάλλεται η κυματομορφή του ήχου (Εικόνα 4-11).

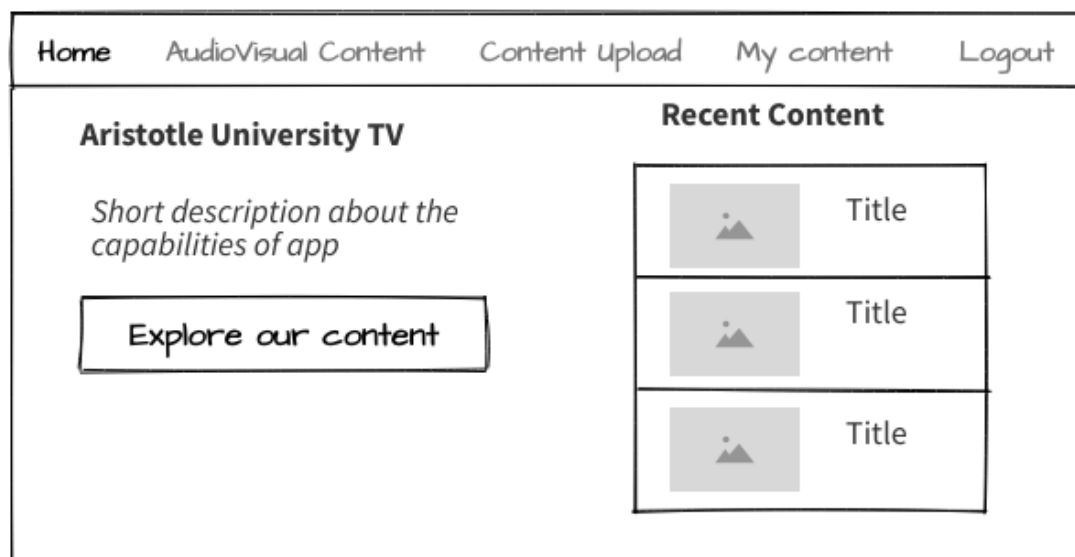


Εικόνα 4-11 Πρότυπη σελίδα αναπαραγωγής αρχείου ήχου

4.4.3. Αρχική Σελίδα

Η αρχική σελίδα του συστήματος (Εικόνα 4-12) δεν θα είναι η προβολή όλου του οπτικοακουστικού υλικού, όπως συμβαίνει στα υπόλοιπα εργαλεία του είδους, αλλά θα περιέχει μια μικρή περιγραφή σχετικά με την ιστοσελίδα και τις δυνατότητές της. Αυτό έχει

ως σκοπό να γνωρίσουν οι χρήστες τι πραγματεύεται η ιστοσελίδα και τι δυνατότητες έχει.



Εικόνα 4-12 Πρότυπη αρχική σελίδα

4.5. Σύνοψη Τετάρτου Κεφαλαίου

Σε αυτό το κεφάλαιο περιγράφεται η αρχιτεκτονική του συστήματος και τα πρωτόκολλα επικοινωνίας μεταξύ των επιμέρους μονάδων του συστήματος. Ως σύστημα διαχείρισης περιεχομένων χρησιμοποιείται το Wordpress, ενώ υπεύθυνο για την εκτέλεση των αλγορίθμων είναι ο διαμεσολαβητής μηνυμάτων RabbitMQ σε συνδυασμό με την κατανεμημένη ουρά εργασιών Celery. Στη συνέχεια παρουσιάζονται μοντέλα του συστήματος που επιλέχθηκαν, αυτά είναι το Whisper για μετατροπή ομιλίας σε κείμενο, το SraCy για αναγνώριση οντοτήτων, το MMAAction2 για ανάλυση δράσεων και ένας αλγόριθμος ταξινόμησης ήχου που θα αναπτυχθεί στα πλαίσια της διπλωματικής. Επίσης, εξηγούνται σενάρια χρήσης, όπου οι χρήστες αναρτούν, αναζητούν και αλληλοεπιδρούν με πολυμεσικό περιεχόμενο, ενώ δίνεται έμφαση στη ροή εργασιών του συστήματος κατά την ανάρτηση ενός αρχείου πολυμέσων. Τέλος, παρουσιάζεται ο σχεδιασμός της διεπαφής χρήστη, με έμφαση στην ευχρηστία, την οργάνωση περιεχομένου και τις δυνατότητες αναπαραγωγής και επεξεργασίας πολυμέσων. Στο επόμενο κεφάλαιο πραγματοποιείται η υλοποίηση τόσο του συστήματος όσο και του αλγορίθμου ταξινόμησης ήχου, ενώ θα παρουσιαστούν και οι γραφικές προβολές της τελικής εφαρμογής.

Κεφάλαιο 5ο

5. Ανάπτυξη Εφαρμογής

5.1. Διαμόρφωση Βασικής Αρχιτεκτονικής Συστήματος

5.1.1. Περιβάλλον Εγκατάστασης Συστήματος

Για την υλοποίηση ενός συστήματος σαν και αυτό υπάρχουν πολλές αρχιτεκτονικές που μπορούν να εφαρμοστούν, πιο συγκεκριμένα, η ύπαρξη ενός μόνο ή περισσοτέρων εξυπηρετητών (server) , ή η χρήση ενός εξυπηρετητή σε συνδυασμό με κάποια υπηρεσία αποθήκευσης νέφους (cloud). Στο παρόν σύστημα εφαρμόστηκε η χρήση ενός μόνο εξυπηρετητή, στον οποίο εκτελούνται η ιστοσελίδα και η βάση δεδομένων. Τα αρχεία αποθηκεύονται στο σύστημα αρχείων του εξυπηρετητή που εκτελείται η ιστοσελίδα, έχοντας, έτσι, τον πλήρη έλεγχο τους και αποθηκεύονται τα indexes και το path των αρχείων στην MySQL βάση δεδομένων του WordPress. Επομένως, μπορεί πολύ εύκολα να πραγματοποιηθεί η προβολή τους στην ιστοσελίδα. Για να έχουν, όμως, τη δυνατότητα, σε συνδυασμό με τα παραπάνω, να εκτελούνται και οι αλγόριθμοι στο παρασκήνιο του συστήματός, απαιτείται ένας εξυπηρετητής με καλούς διαθέσιμους πόρους. Η λύση δόθηκε από το Κέντρο Ηλεκτρονικής Διακυβέρνησης (ΚΗΔ) του Α.Π.Θ., το οποίο παρείχε μια εικονική μηχανή τα χαρακτηριστικά της οποίας παρουσιάζονται στο πίνακα 5-1.

Πίνακας 5-1 Χαρακτηριστικά της Εικονικής Μηχανής που παρείχε το ΚΗΔ του Α.Π.Θ.

ΧΑΡΑΚΤΗΡΙΣΤΙΚΟ	ΠΕΡΙΓΡΑΦΗ
ΕΠΕΞΕΡΓΑΣΤΗΣ (CPU)	4 Πυρήνες
ΜΝΗΜΗ RAM	16 GB
ΑΠΟΘΗΚΕΥΤΙΚΟΣ ΧΩΡΟΣ	2 TB (Σκληρός Δίσκος)
ΚΑΡΤΑ ΓΡΑΦΙΚΩΝ (GPU)	8 GB Μνήμη
ΤΑΧΥΤΗΤΑ ΔΙΑΔΙΚΤΥΟΥ	800 Mbps (Αποστολή Δεδομένων)
ΛΕΙΤΟΥΡΓΙΚΟ ΣΥΣΤΗΜΑ	Linux Ubuntu 22.02
ΠΑΡΟΧΟΣ ΦΙΛΟΞΕΝΙΑΣ	https://arutv.jour.auth.gr/
GPU DRIVERS	Nvidia 535.129.03
ΈΚΔΟΣΗ CUDA	12.2

5.1.2. Εγκατάσταση Περιβάλλοντος του Συστήματος Διαχείρισης Περιεχομένου

Για την εγκατάσταση ενός συστήματος διαχείρισης περιεχομένου, όπως το WordPress, σε λειτουργικό σύστημα Linux απαιτείται η παραμετροποίηση μιας στοίβας που ονομάζεται LAMP (Linux, Apache, MySQL, PHP). Στο τομέα της ανάπτυξης ιστού, η στοίβα LAMP αποτελεί μια ομάδα λογισμικών ανοιχτού κώδικα που, συνήθως, εγκαθίστανται όλα μαζί, έτσι ώστε να ενεργοποιήσουν ένα διακομιστή για να φιλοξενήσει εφαρμογές ιστού γραμμένες σε PHP. Έτσι, αφού υπάρχει ένα σύστημα με λειτουργικό Linux, είναι επιβεβλημένη η εγκατάσταση των Apache, MySQL και PHP επεκτάσεων. Εν συνεχεία, πραγματοποιείται η λήψη της τελευταίας έκδοσης του WordPress. Για την σωστή λειτουργία του απαιτείται η δημιουργία μιας βάσης δεδομένων και ενός χρήστη. Σε αυτόν παραχωρούνται όλα τα προνόμια, γεγονός που σημαίνει ότι μπορεί να έχει πρόσβαση στην βάση δεδομένων, να επεξεργαστεί, να δημιουργήσει και να διαγράψει δεδομένα μέσα σε αυτήν. Τα αρχεία του WordPress μεταφέρονται μέσα στο σύστημα

αρχείων του συστήματος, όπου γίνεται εκχώρηση κατάλληλων δικαιωμάτων για να διασφαλιστεί η ασφαλής πρόσβασή και ακεραιότητα των δεδομένων.

Έπειτα επιβάλλεται ο εξυπηρετητής Apache να διαμορφωθεί με τέτοιο τρόπο, ώστε να αναγνωρίσει την εγκατάσταση του WordPress. Αυτό επιτυγχάνεται με τη δημιουργία ενός αρχείου παραμετροποίησης, το οποίο δείχνει που βρίσκονται τα αρχεία του WordPress στο σύστημα αρχείων, παρέχοντας, ταυτόχρονα, και τα απαραίτητα δικαιώματα. Υπάρχει ακόμα ένα αρχείο παραμετροποίησης μέσα στο κατάλογο αρχείων του WordPress που ονομάζεται `wp-config.php`, με το οποίο πραγματοποιείται η σύνδεση του WordPress με τη βάση δεδομένων που δημιουργήθηκε.

Τέλος, η διαδικασία εγκατάστασης του WordPress ξεκινάει με την πρόσβαση του domain του διακομιστή μέσω ενός προγράμματος περιήγησης ιστού και ολοκληρώνεται με την παροχή απαραίτητων πληροφοριών, όπως ο τίτλος της ιστοσελίδας, το όνομα διαχειριστή, ο κωδικός πρόσβασης και η διεύθυνση ηλεκτρονικού ταχυδρομείου. Όλα τα παραπάνω βήματα αποτελούν τη βάση για την δημιουργία ενός συστήματος διαχείρισης περιεχομένου σε σύστημα Linux.

5.1.3. Βασική Δομή Ιστοτόπου

Αφού πραγματοποιηθεί με επιτυχία η εγκατάσταση του WordPress, το επόμενο στάδιο είναι η δημιουργία μιας βασικής δομής για τον ιστότοπο του ΑΓ.Υ.Τ.Β. Δε θα δοθεί έμφαση στον τρόπο με τον οποίο δημιουργήθηκε η βασική δομή της ιστοσελίδας. Παρόλα αυτά, αξίζει να σημειωθεί η χρήση ενός Child Theme για την προσαρμογή του θέματος, ώστε να διατηρηθεί η δυνατότητα ενημερώσεων του βασικού θέματος χωρίς να χαθούν οι προσαρμογές.

Επιπλέον, χρησιμοποιήθηκε ένας προσαρμοσμένος τύπος περιεχομένου, με την ονομασία Content, για την οργάνωση και διαχείριση του οπτικοακουστικού περιεχομένου. Η επιλογή αυτή επέτρεψε τη διαμόρφωση ενός ξεχωριστού τύπου περιεχομένου, στον οποίο αποθηκεύονται και παρουσιάζονται τα αρχεία που ανεβάζουν οι χρήστες, μαζί με τις ετικέτες που παράγονται από τους αλγόριθμους. Θα δοθούν, παρόλα αυτά, λεπτομέρειες σε βασικά σημεία που έγιναν προσθήκες κώδικα και φυσικά, θα περιγράψει λεπτομερώς ο τρόπος με τον οποίο ενσωματώθηκαν το RabbitMQ και το Celery στο σύστημα, ώστε να εκτελούνται οι αλγόριθμοι για τα αρχεία οπτικοακουστικού υλικού που ανεβαίνουν στο σύστημα.

Για να δημιουργηθεί ένα πρόσθετο και να συνδεθεί με το WordPress, δημιουργείται ένα αρχείο PHP μέσα σε ένα φάκελο, ο οποίος αποθηκεύεται μέσα στην εγκατάσταση του WordPress στο φάκελο `wp-include/plugins/`. Έπειτα, το πρόσθετο εμφανίζεται στον πίνακα ελέγχου και από εκεί ενεργοποιείται. Εν συνεχεία, παρουσιάζονται οι προσθήκες προσθέτων που δημιουργήθηκαν.

5.1.4. Προσαρμοσμένη αναζήτηση περιεχομένου

Το πρόσθετο προσαρμοσμένης αναζήτησης περιεχομένου προσφέρει μια εξειδικευμένη λύση για την αποτελεσματική αναζήτηση περιεχομένου στα αποτελέσματα του αλγορίθμου αναγνώρισης οντοτήτων. Επιτρέπει στους χρήστες να φιλτράρουν τα περιεχόμενα με βάση όσα έχουν ειπωθεί σε πεδία, όπως «Όνομα εκδήλωσης», «Γεωπολιτική τοποθεσία» ή «Όνομα προσώπου», παρέχοντας μια στοχευμένη προσέγγιση για την ανάκτηση συγκεκριμένου περιεχομένου. Επιπλέον, οι χρήστες έχουν τη δυνατότητα να πραγματοποιούν αναζητήσεις πλήρους κειμένου σε όλο το περιεχόμενο ομιλίας που σχετίζεται με τα βίντεο, επιτρέποντας πιο ολοκληρωμένες δυνατότητες αναζήτησης.

Το πρόσθετο λειτουργεί συνδυαστικά με το φίλτρο `pre_get_posts` του WordPress για την προσαρμογή του κύριου ερωτήματος βάσει της εισόδου του χρήστη. Ανάλογα με τα

κριτήρια αναζήτησης που εισάγονται, εφαρμόζει δυναμικά προσαρμοσμένα μετα-ερωτήματα για να φιλτράρει το περιεχόμενο με βάση συγκεκριμένες τιμές μεταδεδομένων ή εκτελεί μια ευρύτερη αναζήτηση για να ταιριάξει με κείμενο εντός του περιεχομένου. Αυτή η λειτουργικότητα διασφαλίζει ότι οι χρήστες μπορούν να εξερευνήσουν τα αποτελέσματα του αλγορίθμου αναγνώρισης οντοτήτων, βελτιώνοντας σημαντικά τη χρηστικότητα και την προσβασιμότητα των περιεχομένων εντός του συστήματος.

5.1.5. Ασφάλεια Ιστοτόπου

Στο σύστημα προστέθηκε, επίσης, ένα πιστοποιητικό SSL(Secure Sockets Layer), το οποίο περιλαμβάνει την ασφάλιση της επικοινωνίας μεταξύ του διακομιστή ιστού και των προγραμμάτων περιήγησης των χρηστών, μέσω της κρυπτογράφησης των δεδομένων που ανταλλάσσονται. Το πιστοποιητικό λαμβάνεται από μια αξιόπιστη πηγή και περιλαμβάνει κρυπτογραφικά κλειδιά που θεσπίζουν μια ασφαλή σύνδεση. Αφού, λοιπόν, αποκτηθεί το πιστοποιητικό γίνεται εγκατάσταση του στον διακομιστή ιστού, ενεργοποιώντας το HTTPS (HTTP Secure). Η αλλαγή αυτή δημιουργεί κρυπτογραφημένη σύνδεση, διασφαλίζοντας την ασφαλή μετάδοση ευαίσθητων πληροφοριών, όπως στοιχεία σύνδεσης ή προσωπικά δεδομένα. Επιπλέον, οι διευθύνσεις URL του ιστοτόπου μετατρέπονται από "http://" σε "https://", επιβεβαιώνοντας, με τον τρόπο αυτόν, την επιτυχή εγκατάσταση του πιστοποιητικού.

5.2. Αρχιτεκτονική Μηχανισμού Επικοινωνίας και Διαχείρισης Εργασιών

5.2.1. Εγκατάσταση του Διαμεσολαβητή Μηνυμάτων

Αρχικά, απαιτείται η εγκατάσταση του RabbitMQ στο διακομιστή του συστήματος, που διαθέτει κεντρικό ρόλο στην αρχιτεκτονική της εφαρμογής, καθώς ενεργεί ως διαμεσολαβητής μηνυμάτων για τη διευκόλυνση της ασύγχρονης επεξεργασίας των εργασιών. Αφού ολοκληρωθεί με επιτυχία η εγκατάσταση, απαιτείται η δημιουργία ενός νέου διαχειριστή-χρήστη και ενός εικονικού διακομιστή, μέσω της πρόσβασης στην κονσόλα διαχείρισης του RabbitMQ. Η δημιουργία χρήστη συμβάλλει στη διαμόρφωση αδειών πρόσβασης, σύμφωνα με τις εκάστοτε λειτουργικές ανάγκες, ενισχύοντας, ταυτόχρονα, και τη συνολική ασφάλεια του συστήματος. Παράλληλα, η ρύθμιση ενός εικονικού διακομιστή εξασφαλίζει την ύπαρξη ενός απομονωμένου περιβάλλοντος του RabbitMQ, μειώνοντας τις πιθανές συγκρούσεις και διευκολύνοντας τη διαχείριση πολλαπλών εφαρμογών.

Για να πραγματοποιηθεί η ενσωμάτωση του διαμεσολαβητή RabbitMQ στο σύστημα αρχικά απαιτείται η δημιουργία ενός Composer, ο οποίος είναι ένας διαχειριστής εξαρτήσεων για την PHP και βελτιστοποιεί την διαχείριση των εξωτερικών βιβλιοθηκών και πακέτων. Αφού γίνει η εγκατάσταση του, είναι απαραίτητος ο προσδιορισμός των εξαρτήσεων. Το composer.json λειτουργεί σαν αρχείο δήλωσης, προσδιορίζοντας τα απαραίτητα πακέτα, συμπεριλαμβανομένης και της βιβλιοθήκης php-amqp, η οποία περιέχει μια υλοποίηση "PHP" για την αλληλεπίδραση του WordPress με το RabbitMQserver.

5.2.2. Δημιουργία Παραγωγού Μηνυμάτων

Αρχικά, δημιουργείται ένα αρχείο PHP, στο οποίο προσδιορίζονται οι παραπάνω βιβλιοθήκες. Περιλαμβάνει τον καθορισμό βασικών παραμέτρων, όπως τον διακομιστή, τη θύρα, τα διαπιστευτήρια χρήστη και τις πληροφορίες του εικονικού διακομιστή, για να είναι εφικτή η σύνδεση στο RabbitMQ. Το αρχείο αυτό λειτουργεί ως γέφυρα της εφαρμογής του WordPress και του διακομιστή RabbitMQ, εξασφαλίζοντας την αποτελεσματική μεταξύ τους επικοινωνία. Λειτουργεί, όμως, και σαν παραγωγός μηνυμάτων στις ουρές του συστήματος. Αποθηκεύεται μαζί με τα παραπάνω, composer και βιβλιοθήκη php-amqp, σε ένα φάκελο που αποθηκεύεται στον κατάλογο αρχείων wp-include/plugins/, δημιουργώντας ένα πρόσθετο.

Μέσα σε αυτό το αρχείο χρησιμοποιούνται τα ίδια "hooks" με το πρόσθετο της δημιουργίας εικόνων. Συγκεκριμένα, όταν αναρτάται ένα αρχείο στο σύστημα λαμβάνονται δύο στοιχεία: το ακριβές μονοπάτι του αρχείου που αναρτήθηκε και το μοναδικό κλειδί της ανάρτησης που δημιουργήθηκε όταν αποθηκεύτηκε στη βάση δεδομένων. Τότε, ο παραγωγός, ο κώδικας του οποίου βρίσκεται μέσα στο ίδιο αρχείο, ενεργοποιείται και στέλνει τα δύο αυτά στοιχεία σε ένα .json μήνυμα στις ουρές του συστήματος. Αξίζει να σημειωθεί ότι κατά την ενεργοποίηση του παραγωγού, δημιουργείται μία σύνδεση και ένα κανάλι, με σκοπό τη δημοσίευση μηνυμάτων στο RabbitMQ. Έπειτα, δημιουργείται ένας κόμβος ανταλλαγής και συγκεκριμένα, ο κόμβος ευρείας μετάδοσης, ώστε το ίδιο μήνυμα (τα δύο στοιχεία του αρχείου) να μεταδοθεί στον καταναλωτή που ακούει στον κόμβο αυτόν. Κατά τον ορισμό του κόμβου, επιλέγεται ο κόμβος να είναι μόνιμος, να διαγράφεται αυτόματα αν δεν υπάρχει ουρά συνδεδεμένη σε αυτόν. Το μήνυμα που δημοσιεύεται δηλώνεται σαν συνεχές (persistent), δηλαδή, αποθηκεύεται στο σκληρό δίσκο του συστήματος ώστε να μην χαθεί σε περίπτωση επανεκκίνησης του διαμεσολαβητή. Αφού δημοσιευτεί το μήνυμα απενεργοποιείται η σύνδεση και το κανάλι ώστε να μην καταναλώνει το σύστημα περιττούς πόρους.

5.2.3. Δημιουργία Καταναλωτή και Ουράς Μηνυμάτων

Στο υπάρχον σύστημα υπάρχει ένας παραγωγός που στέλνει μηνύματα (στοιχεία των αρχείων) και ένας καταναλωτής που το λαμβάνει. Το αρχείο του καταναλωτή είναι γραμμένο σε γλώσσα προγραμματισμού Python και με τη χρήση της βιβλιοθήκης pika, που επιλέχθηκε για την ισχυρή υποστήριξή του για επικοινωνία βασισμένη στο AMQP. Καθορίζονται, όπως και στον παραγωγό, οι βασικές παράμετροι, όπως ο διακομιστής, η θύρα, τα διαπιστευτήρια χρήστη και οι πληροφορίες του εικονικού διακομιστή, για να είναι εφικτή η σύνδεση στο RabbitMQ. Επίσης, γίνεται η χρήση της παραμέτρου heartbeat, η οποία ορίζεται στα 600 δευτερόλεπτα και της παραμέτρου blocked_connection_timeout στα 300 δευτερόλεπτα. Το heartbeat αποτελεί ένα περιοδικό σήμα που στέλνει ο πελάτης στο διακομιστή για να βεβαιωθεί ότι η σύνδεση είναι ενεργή, ενώ το blocked_connection_time θέτει ένα χρονικό περιθώριο για μια μπλοκαρισμένη σύνδεση. Το χρονικό αυτό περιθώριο προσδιορίζει πόσο χρόνο θα αναμένει ο πελάτης, ώστε η σύνδεση να θεωρηθεί ότι δεν χρησιμοποιείται. Με την χρήση των παραπάνω παραμέτρων εξασφαλίζεται η αξιοπιστία και η ανθεκτικότητα της σύνδεσης σε προβλήματα, όπως η έλλειψη διαθέσιμων πόρων και η αδράνεια του δικτύου.

Έπειτα από την καθιέρωση της σύνδεσης και του καναλιού, πραγματοποιείται ο καθορισμός της ουράς. Ορίζεται το όνομά της (algo_queue), καθώς και το όνομα του κόμβου ανταλλαγής (algo_exchange) στον οποίο θα συνδεθεί η ουρά για να μπορεί να λαμβάνει μηνύματα. Δημιουργείται μια ανθεκτική σύνδεση κατά την οποία δηλώνεται η ουρά ως μόνιμη, εξασφαλίζοντας την αξιοπιστία των μηνυμάτων ακόμη και σε περίπτωση αποτυχίας ή επανεκκίνησης του διακομιστή. Ο κόμβος ανταλλαγής που χρησιμοποιείται για να μεταδίδει τα μηνύματα είναι ο fanout, καθώς υπάρχει ένας μόνο καταναλωτής. Ο καταναλωτής ακούει συνεχώς για νέα μηνύματα και τα επεξεργάζεται καθώς φθάνουν. Τα μηνύματα είναι αρχεία JSON, που περιέχουν, συνήθως, ένα μοναδικό κλειδί (post_id) και την αντίστοιχη διαδρομή αρχείου του βίντεο που πρόκειται να υποβληθεί σε επεξεργασία.

Ο καταναλωτής έχει ρυθμιστεί ώστε να χειρίζεται ένα μήνυμα κάθε φορά χρησιμοποιώντας έναν μηχανισμό ποιότητας υπηρεσιών. Μετά την επεξεργασία, το μήνυμα επιβεβαιώνεται, σηματοδοτώντας στο RabbitMQ ότι έχει εκτελεστεί με επιτυχία. Εάν προκύψουν σφάλματα κατά την ανάλυση του μηνύματος ή την υποβολή εργασιών, εφαρμόζεται ισχυρός χειρισμός εξαιρέσεων για την καταγραφή των σφαλμάτων και την αποφυγή διακοπών κατά τη ροή εργασίας.

5.2.4. Ασύγχρονη Επεξεργασία Εργασιών

Στο σύστημα χρησιμοποιείται το Celery για τη διαδοχική και ασύγχρονη εκτέλεση των αλγορίθμων επεξεργασίας οπτικοακουστικού υλικού. Σε συνδυασμό με το RabbitMQ ως διαμεσολαβητή μηνυμάτων, το Celery διασφαλίζει την αποτελεσματική εκτέλεση εργασιών σε ουρές, καθιστώντας το σύστημα αξιόπιστο και επεκτάσιμο. Τα μηνύματα που τοποθετούνται στην ουρά `algo_queue` από τον παραγωγό καταναλώνονται και παραδίδονται στο Celery για επεξεργασία. Οι εργάτες του Celery, που έχουν ρυθμιστεί ανεξάρτητα από το RabbitMQ, χειρίζονται την πραγματική εκτέλεση αυτών των εργασιών (εκτέλεση αλγορίθμων). Αυτός ο διαχωρισμός εξασφαλίζει ότι η εφαρμογή παραμένει ευέλικτη ακόμη και κατά την επεξεργασία υπολογιστικά εντατικών εργασιών.

Ο καταναλωτής RabbitMQ είναι υπεύθυνος για την ανάλυση των εισερχόμενων μηνυμάτων και την κλήση των εργασιών Celery. Όταν λαμβάνεται ένα μήνυμα, αποκωδικοποιείται και επικυρώνεται. Χρησιμοποιείται η μέθοδος `apply_async` για την ανάθεση του έργου της εκτέλεσης των αλγορίθμων σε εργάτες "Celery". Η ασύγχρονη αυτή επικοινωνία επιτρέπει στον καταναλωτή να συνεχίσει την επεξεργασία νέων μηνυμάτων χωρίς να μπλοκάρεται από τρέχουσες εργασίες.

Οι εργάτες Celery εκτελούν τους αλγορίθμους στο παρασκήνιο. Το παρόν σύστημα υποστηρίζει την οριζόντια κλιμάκωση, όπου μπορούν να προστεθούν επιπλέον εργάτες για να χειριστούν αυξημένο φόρτο εργασίας χωρίς να επηρεαστεί το υπόλοιπο σύστημα. Αυτή η ενσωμάτωση βελτιώνει επίσης την ανοχή σε σφάλματα. Δηλαδή, το RabbitMQ διασφαλίζει την αξιόπιστη παράδοση των μηνυμάτων, ενώ το εργαλείο του Celery παρέχει ισχυρούς μηχανισμούς χειρισμού σφαλμάτων και επανάληψης για την εκτέλεση εργασιών. Μαζί, αυτά τα στοιχεία δημιουργούν μια απρόσκοπτη, κλιμακούμενη και αποδοτική αρχιτεκτονική για ασύγχρονη επεξεργασία. Συνδυάζοντας τις δυνατότητες των παραπάνω εργαλείων, επιτυγχάνεται η δημιουργία ενός συστήματος υψηλής απόκρισης και ανοχής σε σφάλματα για τη εκτέλεση υπολογιστικά εντατικών διεργασιών, όπως η επεξεργασία των βίντεο.

5.2.5. Εκτέλεση Αλγορίθμων μέσω Κατανεμημένης Ουράς Εργασιών

Το εργαλείο Celery διαθέτει ένα μηχανισμό ορισμού του αριθμού των εργατών. Η απουσία ισχυρών πόρων στο σύστημα καθιστά την ανάγκη για σειριακή εκτέλεση των αλγορίθμων. Για τον λόγο αυτό ορίστηκε ένας εργάτης. Αυτή η ρύθμιση είναι ζωτικής σημασίας για τη διατήρηση της προκαθορισμένης ακολουθίας εκτέλεσης αλγορίθμων. Προηγμένες διαμορφώσεις, όπως οι καθυστερημένες επιβεβαιώσεις (`acks_late`) και η αποκατάσταση συντριβής του εργάτη (`task_reject_on_worker_lost`), ενισχύουν την αξιοπιστία με την επαναφορά εργασιών που δεν ολοκληρώνονται επιτυχώς. Επιπλέον, ένας μηχανισμός χρονικού ορίου ορατότητας αποτρέπει την επ' αόριστον καθυστέρηση εργασιών.

Η πρωταρχική εργασία Celery, είναι υπεύθυνη για τη διαχείριση της διαδοχικής εκτέλεσης πολλαπλών αλγορίθμων επεξεργασίας οπτικοακουστικού υλικού. Οι αλγόριθμοι που εκτελούνται περιλαμβάνουν διαχωρισμό ήχου και βίντεο, κατηγοριοποίηση ήχου, δημιουργία υπότιτλων με τον αλγόριθμο μετατροπής ομιλίας σε κείμενο, αναγνώριση ονομαστικών οντοτήτων και αναγνώριση δράσης. Για τη βελτιστοποίηση της χρήσης της κεντρική μονάδας επεξεργασίας (CPU), χρησιμοποιείται η μέθοδος `subprocess.run` για την εκτέλεση κάθε αλγορίθμου στο κατάλληλο περιβάλλον Python. Αυτή η προσέγγιση εξασφαλίζει την εκτέλεση των αλγορίθμων σε απομονωμένα περιβάλλοντα διατηρώντας την αποδοτική χρήση των πόρων του συστήματος. Το Celery περιέχει έναν επιπλέον ισχυρό μηχανισμό επανάληψης, ο οποίος επαναφέρει αυτόματα τις εργασίες σε ουρά με εκθετικό `backoff` σε περίπτωση σφάλματος. Αυτό το ολοκληρωμένο πλαίσιο παρέχει κλιμακούμενο, αξιόπιστο και αποτελεσματικό χειρισμό των εντατικών σε

υπολογισμούς ροών εργασίας εκτέλεσης των αλγορίθμων, αποτελώντας βασικό συστατικό της αρχιτεκτονικής του συστήματος.

Τέλος, για να εξασφαλιστεί η συνεχής και αξιόπιστη λειτουργία του συστήματος, έχουν ρυθμιστεί δύο κρίσιμες υπηρεσίες με τη χρήση του systemd. Η πρώτη υπηρεσία επικεντρώνεται στην εκτέλεση του σεναρίου καταναλωτή RabbitMQ, το οποίο ακούει ενεργά την καθορισμένη ουρά για εισερχόμενα μηνύματα και αποστέλλει εργασίες στους εργάτες Celery. Η δεύτερη υπηρεσία είναι υπεύθυνη για τη διαχείριση του εργάτη Celery, ο οποίος εκτελεί τους αλγορίθμους επεξεργασίας περιεχομένου με διαδοχικό και ασύγχρονο τρόπο. Αυτές οι υπηρεσίες έχουν ρυθμιστεί προσεκτικά ώστε να ξεκινούν αυτόματα κατά την εκκίνηση του συστήματος, διασφαλίζοντας ότι το σύστημα τίθεται σε λειτουργία χωρίς να απαιτείται χειροκίνητη παρέμβαση. Επιπλέον, και οι δύο υπηρεσίες έχουν ρυθμιστεί να επανεκκινούνται αυτόματα σε περίπτωση αποτυχίας, χρησιμοποιώντας τις πολιτικές επανεκκίνησης του systemd. Αυτή η προσέγγιση όχι μόνο ελαχιστοποιεί το χρόνο διακοπής λειτουργίας, αλλά παρέχει επίσης έναν ισχυρό μηχανισμό για τη διατήρηση της σταθερότητας και της αξιοπιστίας της ροής εργασιών του συστήματος, ακόμη και σε περίπτωση απροσδόκητων διακοπών ή επανεκκινήσεων του διακομιστή.

5.3. Ενσωμάτωση Αλγορίθμων

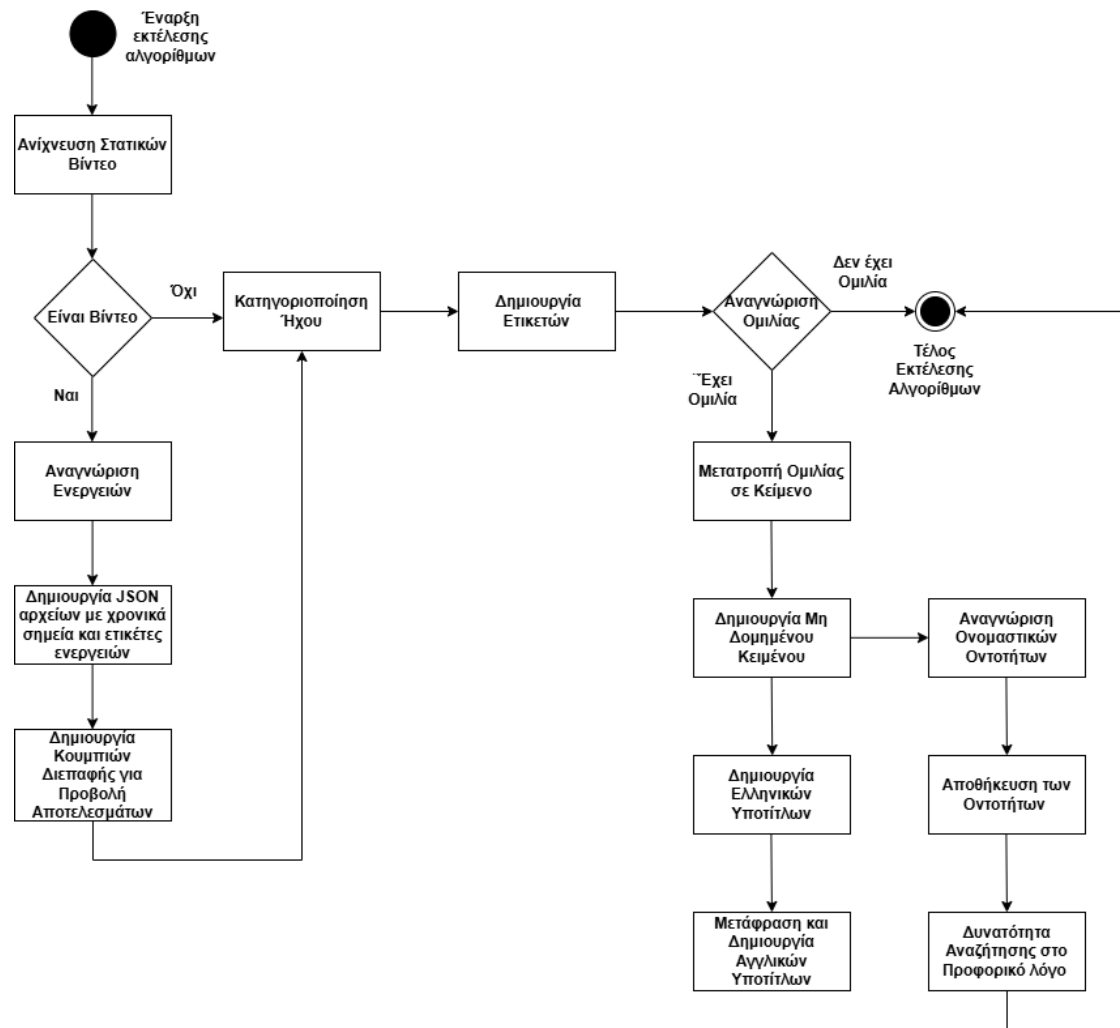
Κάθε αλγόριθμος που ενσωματώνεται στο σύστημα χρησιμοποιεί ένα προ-εκπαιδευμένο μοντέλο που λαμβάνεται από διαθέσιμες έρευνες, τα ονόματα των οποίων και ο λόγος επιλογής τους παρατίθενται στον Πίνακα 4.1. Για να διασφαλιστεί η ομαλή εκτέλεση των αλγορίθμων και να αποφευχθούν οι συγκρούσεις εξαρτήσεων, κάθε αλγόριθμος εκτελείται στο ειδικό περιβάλλον Conda. Αυτή η προσέγγιση απομονώνει τις εξαρτήσεις και αξιοποιεί τις διαμορφώσεις που αφορούν το συγκεκριμένο περιβάλλον, διευκολύνοντας την αποτελεσματική και χωρίς συγκρούσεις εκτέλεση.

Σε όλο τον κώδικα έχουν ενσωματωθεί μηχανισμοί χειρισμού εξαιρέσεων για τη διευθέτηση πιθανών σφαλμάτων και τη διασφάλιση της ανθεκτικότητας του συστήματος. Γίνεται χρήση των τμημάτων try-except σε κρίσιμα τμήματα του κώδικα, επιτρέποντας την ευκολότερη διαχείριση εξαιρέσεων που ενδέχεται να προκύψουν κατά τη διάρκεια λειτουργιών της βάσης δεδομένων, των αλληλεπιδράσεων με το RabbitMQ ή κατά τη διάρκεια εκτέλεσης του αλγόριθμου. Με τη χρήση των εξαιρέσεων, η δέσμη ενεργειών του διακομιστή παρέχει μηνύματα σφαλμάτων με τη χρήση logs, συμβάλλοντας στην ανθεκτικότητα του συστήματος. Παράλληλα βοηθάει στην ομαλή λειτουργία του κώδικα. Για παράδειγμα, εάν ένας αλγόριθμος σταματήσει τη εκτέλεσή του για οποιοδήποτε λόγο, τότε ο καταναλωτής μπορεί να καταναλώσει το επόμενο μήνυμα της ουράς.

Στη συνέχεια στο Πίνακα 5-2 παρουσιάζονται τα αρχεία κώδικα που εκτελούνται σειριακά για ένα αρχείο πολυμέσων που ανεβαίνει στο σύστημα από την εκτέλεση των αλγορίθμων μέχρι και την αποθήκευση των αποτελεσμάτων τους. Ενώ στην Εικόνα 5-1 παρουσιάζεται ένα διάγραμμα ροής που παρουσιάζονται όλες οι διαδικασίες επεξεργασίας που πραγματοποιούνται από τη στιγμή που ξεκινά η εκτέλεση των αλγορίθμων για ένα αρχείο πολυμέσων μέχρι το τέλος της εκτέλεσης τους.

Πίνακας 5-2 Λίστα Εφαρμοζόμενων Αλγορίθμων και Βοηθητικών Σεναρίων.

Συστατικό/ Αλγόριθμος	Περιγραφή	Λεπτομέρειες Υλοποίησης	Σκοπός
Ανίχνευση Στατικών Βίντεο	Αναγνωρίζει στατικά βίντεο (ίδιο καρέ σε όλο το βίντεο).	Χρησιμοποιεί το Μέτρο Δείκτη Δομικής Ομοιότητας (Structural Similarity Index Measure) με OpenCV. Μετατρέπει τα καρέ σε κλίμακα του γκρι για να μειώσει την πολυπλοκότητα χωρίς να επηρεάζεται η αποδοτικότητα. Για κάθε ζεύγος διαδοχικών καρέ, υπολογίζεται το SSIM. Εάν η βαθμολογία ομοιότητας μεταξύ δύο καρέ πέσει κάτω από ένα προκαθορισμένο όριο (π.χ. 0,99), το βίντεο ταξινομείται ως δυναμικό, υποδεικνύοντας αλλαγές κίνησης. Σε άλλη περίπτωση θεωρείται ως στατικό.	Αποτρέπει την περιττή κατανάλωση πόρων παραλείποντας την αναγνώριση ενεργειών για στατικά βίντεο, τα οποία αντιμετωπίζονται ως αρχεία ήχου.
Κατηγοριοποίηση Ήχου	Κατηγοριοποιεί τον ήχο ως ομιλία, μουσική ή άλλα (περιβαλλοντικοί ήχοι).	Αν είναι αρχείο βίντεο το μετατρέπει σε αρχείο ήχου με τη χρήση του ffmpeg και στη συνέχεια υπολογίζει τα φασματογραφήματα με βάση τα κυματίδια. Τα χαρακτηριστικά αυτά τα εισάγει στο καλύτερο μοντέλο που έχει προκύψει από την εκπαίδευση.	Δημιουργεί αυτόματες περιγραφικές ετικέτες για το περιεχόμενο του ήχου, αποθηκευόντάς τα στο μεταδεδομένα του WordPress.
Μετατροπή Ομιλίας σε Κείμενο	Μετατρέπει την ομιλία σε κείμενο στα Ελληνικά και Αγγλικά.	Χρησιμοποιεί το medium μοντέλο Whisper της OpenAI. Δημιουργεί υπότιτλους WebVTT στα Ελληνικά και τους μεταφράζει στα Αγγλικά. Αποθηκεύει και τις δύο εκδοχές δυναμικά στο σύστημα.	Παρέχει υπότιτλους σε πολλές γλώσσες για τα βίντεο, βελτιώνοντας την προσβασιμότητα.
Αναγνώριση Οντοτήτων	Εντοπίζει οντότητες (ονόματα, τοποθεσίες κλπ.) στο ελληνικό κείμενο.	Επεξεργάζεται το κείμενο που παράγεται από το Whisper χρησιμοποιώντας το SpaCy για αναγνώριση οντοτήτων. Οι κατηγορίες περιλαμβάνουν οργανισμούς, προϊόντα, άτομα, εκδηλώσεις και τοποθεσίες. Μεταφράζει τις οντότητες στα Αγγλικά με το Google Translate. Αποθηκεύει τα μεταδεδομένα στο WordPress.	Εμπλουτίζει τα μεταδεδομένα με οντότητες για καλύτερη αναζήτηση περιεχομένου.
Αναγνώριση Ενεργειών	Αναγνωρίζει ενέργειες στο περιεχόμενο των βίντεο.	Χρησιμοποιεί το προ-εκπαιδευμένο μοντέλο ResNet-50 του MMAction2. Επεξεργάζεται τα καρέ του βίντεο σε παρτίδες και δημιουργεί ετικέτες ενεργειών με χρονικά σημεία. Ενημερώνει τις ετικέτες του WordPress και δημιουργεί αρχεία JSON για ανιχνευμένες ενέργειες, επιτρέποντας την αλληλεπίδραση των χρηστών με κουμπιά διεπαφής.	Ενισχύει την αλληλεπίδραση επιτρέποντας στους χρήστες να πλοηγούνται απευθείας σε σημεία ενεργειών στα βίντεο.
Ενημέρωση Αποτελεσμάτων	Ενημερώνει υπότιτλους και αποτελέσματα αλγορίθμων στο σύστημα.	Οι υπότιτλοι αποθηκεύονται στον φάκελο πολυμέσων του διακομιστή. Τα μεταδεδομένα και τα προσαρμοσμένα πεδία ενημερώνονται στο WordPress μέσω του REST API με πιστοποιημένα HTTP αιτήματα. Τα δεδομένα JSON περιλαμβάνουν ζεύγη κλειδιού-τιμής για τα προσαρμοσμένα πεδία.	Εξασφαλίζει την ακριβή και οργανωμένη αποθήκευση των αποτελεσμάτων των αλγορίθμων.



Εικόνα 5-1 Διάγραμμα Ροής αυτοματοποιημένων διαδικασιών-αλγορίθμων επεξεργασίας για ένα πολυμεσικό αρχείο

5.4. Επεκτασιμότητα και Ενσωμάτωση Νέων Αλγορίθμων

Η αρθρωτή αρχιτεκτονική που επιλέχτηκε για την ανάπτυξη του συστήματος επιτρέπει την εύκολη ενσωμάτωση νέων αλγορίθμων με ελάχιστες τροποποιήσεις στην υπάρχουσα δομή. Αυτή η ευελιξία διευκολύνεται, κυρίως, από τον συνδυασμό των Celery και RabbitMQ, οι οποίοι απομονώνουν την εκτέλεση των αλγορίθμων από την κύρια λογική της εφαρμογής. Το σύστημα έχει σχεδιαστεί για να επεξεργάζεται πολυμεσικά αρχεία εκτελώντας μια προκαθορισμένη ακολουθία αλγορίθμων, καθένας από τους οποίους είναι υπεύθυνος για συγκεκριμένη εργασία. Η ροή των εργασιών εκτέλεσης αλγορίθμων υπόκειται σε διαχείριση από το Celery, το οποίο διανέμει τις εργασίες ασύγχρονα, ενώ το RabbitMQ χρησιμεύει ως διαμεσολαβητής μηνυμάτων, εξασφαλίζοντας την αποτελεσματική ανάθεση και εκτέλεση των εργασιών για κάθε πολυμεσικό αρχείο αναρτάται στο σύστημα.

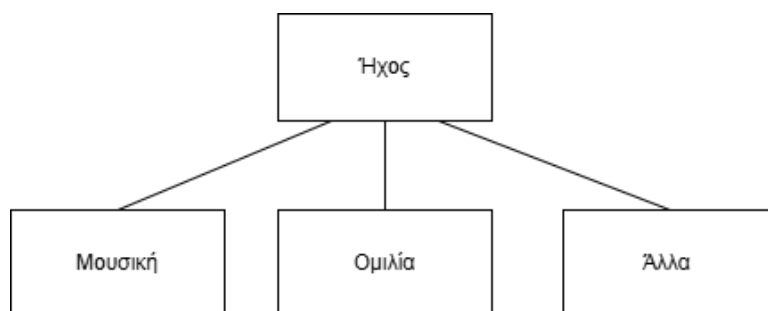
Η ενσωμάτωση ενός νέου αλγορίθμου στο σύστημα απαιτεί ελάχιστη προσπάθεια. Αρχικά χρειάζεται η υλοποίηση ενός νέου αλγορίθμου σε ένα πηγαίο αρχείο κώδικα (script) σε γλώσσα προγραμματισμού python, έχοντας τις αναμενόμενες παραμέτρους ως εισόδους, δηλαδή, ένα μοναδικό κλειδί και την αντίστοιχη διαδρομή πολυμεσικού αρχείου που πρόκειται να υποβληθεί σε επεξεργασία. Έπειτα, το αρχείο που δημιουργείται μπορεί να ενσωματωθεί στο σύστημα, προσθέτοντας το όνομα του αρχείου στη λίστα

αλγορίθμων που εκτελούνται από τη συνάρτηση του Celery. Δεδομένου ότι η διαδικασία εκτέλεσης διεκπεραιώνεται διαδοχικά από τον εργάτη του Celery, ο νέος αλγόριθμος συμπεριλαμβάνεται αυτόματα στον αγωγό επεξεργασίας χωρίς να απαιτούνται περαιτέρω τροποποιήσεις στην κεντρική υποδομή. Επιπλέον, εάν ο αλγόριθμος χρειάζεται ένα συγκεκριμένο περιβάλλον εκτέλεσης, μπορεί να ρυθμιστεί ώστε να εκτελείται σε ένα δικό του εικονικό περιβάλλον, παρόμοια με την υπάρχουσα υλοποίηση για την αναγνώριση ονομαστικών οντοτήτων. Σε περίπτωση κατά την οποία ο νεοπροστιθέμενος αλγόριθμος παράγει έξοδο που πρέπει να αποθηκευτεί, απαιτείται και η δημιουργία μιας συνάρτησης που θα αναλαμβάνει το χειρισμό του αποτελέσματος, ανάλογα με την περίπτωση, είτε μέσω του Api του Wordpress είτε με απευθείας αποθήκευση στο σύστημα αρχείου του διακομιστή. Ωστόσο, για να αναγνωρίσει το σύστημα τον νέο αλγόριθμο, όπως και κάθε αλλαγή μπορεί να συνέβη στα αρχεία κώδικα του συστήματος, απαιτείται η επανεκκίνηση της υπηρεσίας (systemd) του Celery. Δεδομένου ότι το Celery φορτώνει την ουρά εργασιών του κατά την εκκίνηση, οποιεσδήποτε τροποποιήσεις στην ακολουθία εκτέλεσης εργασιών, όπως η προσθήκη ή τροποποίηση αλγορίθμων, απαιτούν επανεκκίνηση της υπηρεσίας για να τεθούν σε ισχύ. Με τον τρόπο αυτό διασφαλίζεται ότι ο ενημερωμένος αγωγός εκτέλεσης φορτώνεται σωστά και είναι λειτουργικός.

Η δυνατότητα εύκολης ενσωμάτωσης νέων αλγορίθμων χωρίς αλλαγή της βασικής δομής του συστήματος αναδεικνύει την ευελιξία της πλατφόρμας. Αυτός ο σχεδιασμός επιτρέπει τη συνεχή βελτίωση των δυνατοτήτων της εφαρμογής μέσω της προσθήκης βελτιωμένων μοντέλων ταξινόμησης, εναλλακτικών τεχνικών επεξεργασίας και άλλων χρήσιμων εργαλείων αυτοματοποίησης. Η αρχιτεκτονική του συστήματος διασφαλίζει ότι στο σύστημα μπορούν, συνεχώς, να εντάσσονται νέες λειτουργίες χωρίς να επηρεάζονται οι υπάρχουσες ροές εργασίας.

5.5. Ανάπτυξη Αλγορίθμου Ταξινόμησης Ήχου

Στο πλαίσιο της παρούσας διατριβής, υλοποιείται ένας καινοτόμος αλγόριθμος ταξινόμησης ήχου, ο οποίος στοχεύει στην αποτελεσματική κατηγοριοποίηση των ηχητικών σημάτων σε τρεις βασικές κατηγορίες: μουσική, ομιλία και άλλα (περιβαλλοντικοί ήχοι), όπως φαίνεται και στην Εικόνα 5-2.



Εικόνα 5-2 Σχήμα ταξινόμησης των κλάσεων του αλγόριθμου ταξινόμησης ήχου

Η καινοτομία της προσέγγισης έγκειται στη χρήση μιας μεθόδου εξαγωγής χαρακτηριστικών βασισμένη στο μετασχηματισμό των κυματιδίων, η οποία παρέχει μια αναπαράσταση χρόνου-συχνότητας του ηχητικού σήματος, σε αντίθεση με τις παραδοσιακές προσεγγίσεις που βασίζονται σε φασματογραφήματα με βάση το μετασχηματισμό Fourier (STFT) ή σε συντελεστές Mel-Frequency Cepstral Coefficients (MFCC). Τα χαρακτηριστικά αυτά συλλαμβάνουν κρίσιμα φασματικά και χρονικά χαρακτηριστικά, καθιστώντας τα ιδιαίτερα αποτελεσματικά για εργασίες ταξινόμησης. Επιπρόσθετα, η επιλογή της συγκεκριμένης μεθόδου βασίζεται στο γεγονός ότι, μέχρι

σήμερα, ο μετασχηματισμός των κυματιδίων δεν έχει αξιοποιηθεί με αυτόν τον τρόπο σε παρόμοια προβλήματα, γεγονός που αναμένεται να προσφέρει σημαντικές βελτιώσεις στην απόδοση και την ακρίβεια του αλγορίθμου. Τα εξαγόμενα χαρακτηριστικά τροφοδοτούνται σε ένα συνελκτικό νευρωνικό δίκτυο, το οποίο έχει σχεδιαστεί ώστε να μαθαίνει αυτόματα τις πιο διακριτικές ιδιότητες μεταξύ των κατηγοριών. Με αυτές τις επιλογές, προσδοκάται ότι η μεθοδολογία θα αποδειχθεί αποτελεσματική και θα συμβάλει σημαντικά στη βελτίωση της ταξινόμησης ήχου σε περιβάλλοντα πολυμέσων. Στη συνέχεια, θα παρουσιαστούν λεπτομερώς οι διαδικασίες που ακολουθήθηκαν για την ανάπτυξη του αλγορίθμου.

5.5.1. Σετ Δεδομένων και Εξαγωγή Χαρακτηριστικών

Για την υλοποίηση του αλγορίθμου χρησιμοποιήθηκε το σετ δεδομένων LVLib-SMO v2¹, το οποίο περιλαμβάνει κατηγοριοποιημένα δείγματα ήχου σε τρεις κύριες κατηγορίες: την ομιλία, την μουσική και άλλα (περιβαλλοντικοί ήχοι). Το σύνολο δεδομένων περιλαμβάνει δεδομένα από τα σύνολα GTZAN Music-Speech², GTZAN Genre Collection³ και AESDD⁴ (Acted Emotional Speech Dynamic Database). Τα δεδομένα αποτελούνται από περίπου 450 λεπτά ασυμπίεστου ήχου, που είναι ηχογραφημένος στα 22.050 Hz, 16-bit, μονοφωνικά, και έχει υποστεί προεπεξεργασία σε τμήματα μήκους 750ms, με αποτέλεσμα να προκύπτουν πάνω από 35.000 δείγματα ήχου. Για να ξεπεραστεί το πρόβλημα της έλλειψης δεδομένων, συνιστώνται επίσης δύο τεχνικές επαύξησης των δεδομένων στο σύνολο των δεδομένων που έχουν αναφερθεί ότι αυξάνουν την ακρίβεια του μοντέλου για εργασίες ταξινόμησης ήχων. Αυτές είναι η χρονική επιμήκυνση με συντελεστές 0.85 και 1.15 και μετατόπιση του τόνου με τιμές (σε ημιτόνια) -4 και 4. Οι επαυξήσεις των δεδομένων πραγματοποιήθηκαν χρησιμοποιώντας το εργαλείο γραμμής εντολών Sound eXchange (SoX).

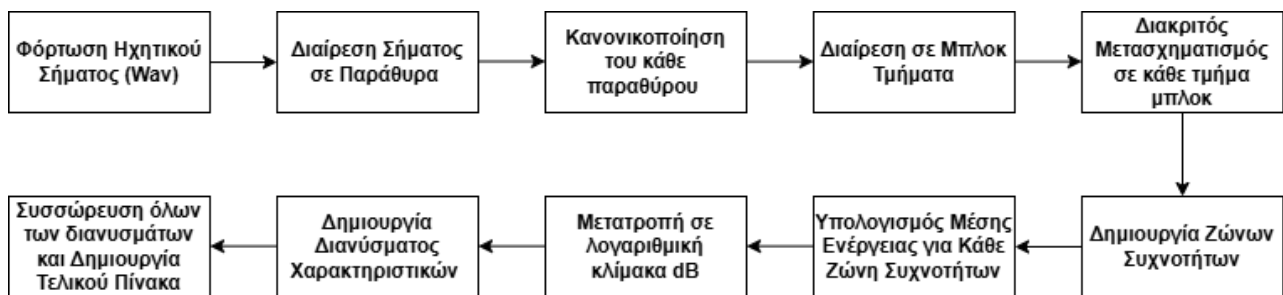
Η εξαγωγή χαρακτηριστικών πραγματοποιήθηκε μέσω της χρήσης του διακριτού μετασχηματισμού κυματιδίων, ο οποίος προσφέρει μια αναπαράσταση του σήματος στον χρόνο και στη συχνότητα. Αρχικά, κάθε ηχητικό σήμα χωρίστηκε σε τμήματα διάρκειας 3 δευτερολέπτων με 50% επικάλυψη. Συγκεκριμένα, κάθε νέο τμήμα ξεκινά από τη μέση του προηγούμενου, ώστε να διατηρηθεί η χρονική συνέχεια και να αυξηθεί ο αριθμός των δειγμάτων για ανάλυση. Κατόπιν, κάθε τμήμα κανονικοποιήθηκε στο διάστημα [-1, 1] για να εξασφαλιστεί η ομοιομορφία των δεδομένων και να διευκολυνθεί η σύγκλιση κατά την εκπαίδευση των νευρωνικών δικτύων. Στη συνέχεια, τα κανονικοποιημένα τμήματα διαιρέθηκαν σε μη επικαλυπτόμενα μπλοκ των 512 δειγμάτων. Το μέγεθος αυτό επιλέχθηκε λόγω της ιδιότητας ότι αποτελεί δύναμη του 2, γεγονός που βελτιστοποιεί τους υπολογισμούς σε ψηφιακά φίλτρα και μετασχηματισμούς. Για κάθε μπλοκ εφαρμόστηκε ο διακριτός μετασχηματισμός κυματιδίων με χρήση του κυματιδίου Daubechies-4 (db4) και έξι επίπεδα αποσύνθεσης, ενώ για κάθε ζώνη συχνότητας (επίπεδο αποσύνθεσης) υπολογίστηκε η μέση ισχύς των συντελεστών και οι τιμές αυτές μετατράπηκαν σε λογαριθμική κλίμακα (dB), προκειμένου να αποτυπωθεί το ενεργειακό περιβάλλον του σήματος σε ένα δισδιάστατο πίνακα που χρησιμοποιείται ως είσοδος σε συνελκτικό νευρωνικό δίκτυο. Η παραπάνω διαδικασία παρουσιάζεται στην εικόνα 5-3, ενώ ο δισδιάστατος αυτός πίνακας που προκύπτει παρουσιάζεται στην Εικόνα 5-4 και αποτελεί τη χρονική – συχνотική αναπαράσταση των ήχων για κάθε κλάση.

¹ <https://m3c.web.auth.gr/research/datasets/lvlib-smo-v2-for-general-audio-classification/>

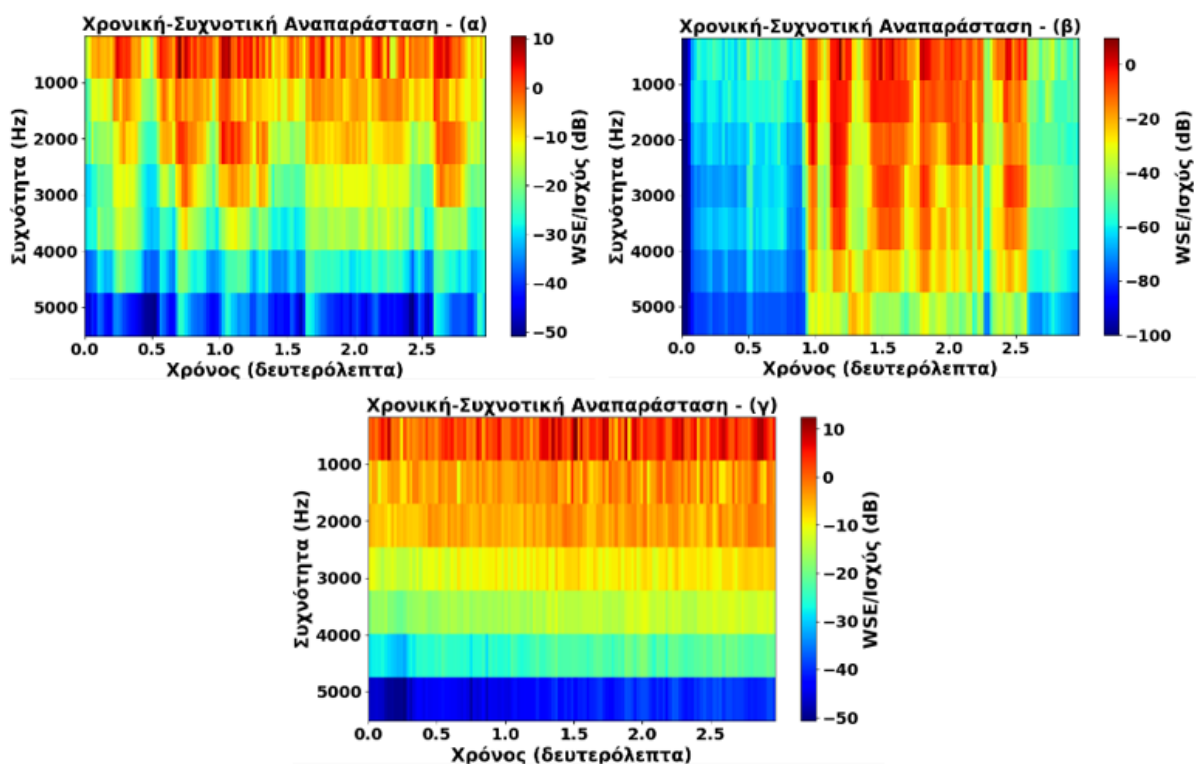
² <https://www.kaggle.com/datasets/lnicalo/gtzan-musicspeech-collection>

³ <https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification>

⁴ <https://m3c.web.auth.gr/research/aesdd-speech-emotion-recognition/>



Εικόνα 5-3 Αναπαράσταση των βημάτων εξαγωγής χαρακτηριστικών από δεδομένα ήχου για τη δημιουργία χρονικής-συχνοτικής αναπαράστασης



Εικόνα 5-4 Φασματογραφήματα με βάση το κυματίδιο Daubechies-4 και 6 επίπεδα αποσύνθεσης για καθεμία από τις 3 κλάσεις, (α) Μουσική, (β) Ομιλία, (γ) Άλλα (περιβαλλοντικοί ήχοι)

Για την επεξεργασία του ηχητικού σήματος, επιλέχθηκαν τμήματα διάρκειας 3 δευτερολέπτων. Η χρονική αυτή περίοδος προσφέρει ένα επαρκές παράθυρο για την πλήρη καταγραφή των χαρακτηριστικών και των μεταβατικών ιδιοτήτων του σήματος, αποφεύγοντας την ανάμειξη πολλαπλών ηχητικών χαρακτηριστικών σε ένα μόνο τμήμα. Αυτή η επιλογή επιτυγχάνει μια ισορροπία μεταξύ της πολυπλοκότητας και της πληροφορίας, καθώς τα τμήματα παραμένουν διαχειρίσιμα από τον επεξεργαστή, ενώ ταυτόχρονα περιέχουν επαρκή πληροφορία για την ανίχνευση και τη ταξινόμηση των χαρακτηριστικών. Επιπρόσθετα, εφαρμόστηκε επικάλυψη 50% μεταξύ των τμημάτων, εξασφαλίζοντας τη διατήρηση της χρονικής συνέχειας και την πλήρη κάλυψη των σημαντικών μεταβατικών φαινομένων που ενδέχεται να συμβαίνουν στα όρια των τμημάτων. Η συγκεκριμένη επικάλυψη οδηγεί, επίσης, σε αύξηση του αριθμού των δειγμάτων χωρίς την ανάγκη νέας ηχογράφησης, ενισχύοντας, κατά συνέπεια, και την

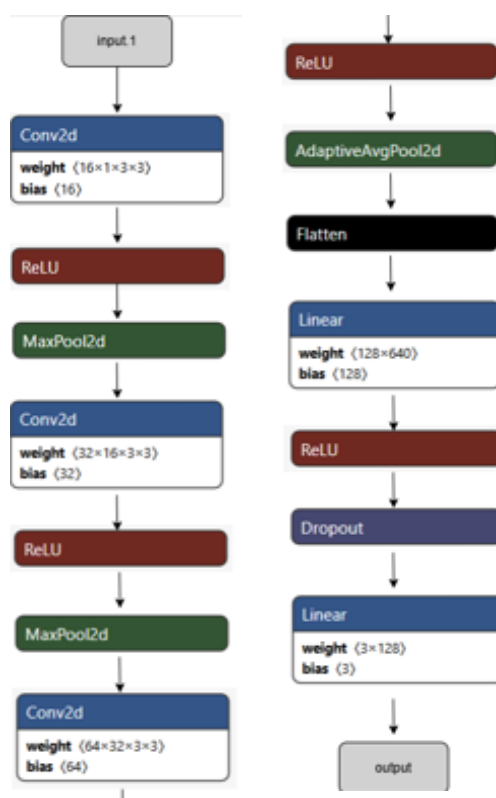
ακρίβεια στην εκπαίδευση του μοντέλου. Παράλληλα, πραγματοποιήθηκε κανονικοποίηση των τιμών του σήματος στο διάστημα $[-1, 1]$, με στόχο την επίτευξη ομοιομορφίας και σταθερότητας στα δεδομένα, καθώς και τη μείωση των διακυμάνσεων που ενδέχεται να προκαλέσουν αριθμητικές δυσκολίες κατά την επεξεργασία. Η κοινή κλίμακα των δεδομένων συμβάλλει, επίσης, στη βελτίωση της σύγκλισης των αλγορίθμων μηχανικής μάθησης. Στη συνέχεια, τα δεδομένα διαχωρίστηκαν σε μη επικαλυπτόμενα μπλοκ των 512 δειγμάτων, επιλογή που βασίζεται στο γεγονός ότι το 512 είναι δύναμη του 2, γεγονός που διευκολύνει τους αλγορίθμους επεξεργασίας και βελτιώνει την ταχύτητα υπολογισμού. Επιπλέον, το μέγεθος αυτό παρέχει μια καλή ισορροπία μεταξύ των τοπικών χρονικών χαρακτηριστικών και της συχνοτικής ανάλυσης, επιτρέποντας την καταγραφή τοπικών μεταβολών στο σήμα χωρίς να θυσιάζεται η ανάλυση της κατανομής της ενέργειας στις συχνότητες. Για τον διακριτό μετασχηματισμό των κυματιδίων χρησιμοποιήθηκε το κυματίδιο Daubechies-4 (db4). Ο συγκεκριμένος μετασχηματισμός επιτρέπει την ταυτόχρονη ανάλυση του σήματος στον χρόνο και στη συχνότητα, καθιστώντας τον ιδανικό για μη στάσιμα σήματα όπως ηχητικά. Το Daubechies-4 το οποίο είναι ένα ορθογώνιο κυματίδιο, επιλέχθηκε λόγω της καλής ισορροπίας μεταξύ της χρονικής και συχνοτικής ανάλυσης. Επίσης, είναι ικανό να ανιχνεύσει αποτελεσματικά τόσο τις μεταβατικές όσο και τις πιο σταθερές ιδιότητες του σήματος και αποτελεί μία από τις πλέον διαδεδομένες επιλογές στην ανάλυση ηχητικών σημάτων. Ο μετασχηματισμός υλοποιήθηκε σε έξι επίπεδα αποσύνθεσης, επιτρέποντας την εξαγωγή χαρακτηριστικών σε πολλαπλές κλίμακες. Τα αρχικά επίπεδα ανιχνεύουν τις υψηλές συχνότητες, ενώ τα τελικά επικεντρώνονται στις χαμηλές συχνότητες. Τα 6 επίπεδα παρέχουν αρκετή λεπτομέρεια για να καταγράψουν τα σημαντικά χαρακτηριστικά του σήματος, χωρίς να εισάγουν υπερβολική υπολογιστική πολυπλοκότητα ή να δημιουργούν υπερβολική διάσταση στα δεδομένα. Περισσότερα επίπεδα αυξάνουν την ανάλυση συχνότητας αλλά μειώνουν την ανάλυση χρόνου, κάτι που μπορεί να μην είναι ιδανικό για σύντομα, μεταβατικά σήματα. Από την άλλη, λιγότερα επίπεδα μπορεί να μην παρέχουν αρκετή πληροφορία στις χαμηλές συχνότητες. Τέλος, υπολογίστηκε η μέση ισχύς για κάθε ζώνη συχνότητων, η οποία, στη συνέχεια, μετατράπηκε σε λογαριθμική κλίμακα (dB). Ο υπολογισμός της μέσης ισχύος συνοψίζει την ενεργειακή πληροφορία του σήματος, καθιστώντας τα δεδομένα πιο εύχρηστα για ανάλυση, ενώ η χρήση της λογαριθμικής κλίμακας συνάδει με την ανθρώπινη αντίληψη των διαφορών στην ένταση του ήχου, με ιδιαίτερη έμφαση στις μεταβολές στα χαμηλότερα επίπεδα ισχύος. Συνολικά, ο συνδυασμός των παραπάνω επιλογών οδηγεί σε μια ισορροπημένη και αποτελεσματική μέθοδο εξαγωγής χαρακτηριστικών που ικανοποιεί την απαίτηση για λεπτομερή ανάλυση χρόνου-συχνότητας με πρακτική εφαρμογή, ενώ παράλληλα εξασφαλίζει επαρκείς πληροφορίες για την ταξινόμηση και την ανάλυση των ηχητικών σημάτων. Οι παραπάνω επιλογές επιβεβαιώνονται ως κατάλληλες και από τα διάφορα πειράματα που πραγματοποιήθηκαν, τα αποτελέσματα των οποίων παρατίθενται στο κεφάλαιο της αξιολόγησης.

5.5.2. Μοντέλο Ταξινόμησης Ήχου και Εκπαίδευση

Το μοντέλο που χρησιμοποιήθηκε είναι ένα συνελικτικό νευρωνικό δίκτυο που λαμβάνει ως είσοδο μια δυσδιάστατη εικόνα, δηλαδή το φασματογράφημα με βάση τα κυματίδια που υπολογίστηκε από τα δεδομένα. Η χρήση των συνελικτικών νευρωνικών δικτύων επιλέγεται λόγω της ικανότητάς τους να εξάγουν τοπικά χαρακτηριστικά από δεδομένα που έχουν δομή εικόνας. Στην παρούσα εργασία, η μετατροπή του ηχητικού σήματος σε φασματογράφημα μέσω του μετασχηματισμού κυματιδίων δημιουργεί μια δισδιάστατη αναπαράσταση, η οποία επιτρέπει στα δίκτυα αυτά να ανιχνεύουν χωρικά μοτίβα και δομές. Επίσης, μαθαίνουν ιεραρχικά χαρακτηριστικά, από απλά τοπικά μοτίβα μέχρι σύνθετες διακριτικές δομές, διευκολύνοντας έτσι την ακριβή διάκριση μεταξύ διαφορετικών κατηγοριών ήχου. Επιπλέον, η αποδοτική επεξεργασία μεγάλων συνόλων

δεδομένων και η δυνατότητα εξαγωγής χαρακτηριστικών με χαμηλό υπολογιστικό κόστος καθιστούν τα συνελκτικά νευρωνικά δίκτυα την ιδανική επιλογή για εφαρμογές ταξινόμησης σε περιβάλλοντα πολυμέσων.

Η αρχιτεκτονική του μοντέλου περιλαμβάνει επίπεδα συνελίξεων (Conv2D), επίπεδα δειγματοληψίας (MaxPool2d) και πλήρως συνδεδεμένα επίπεδα. Το συνελκτικό τμήμα του μοντέλου αποτελείται από τρία στρώματα, καθένα από τα οποία χρησιμοποιεί συναρτήσεις ενεργοποίησης ReLU για την εισαγωγή μη γραμμικότητας και λειτουργίες μέγιστης δειγματοληψίας για τη σταδιακή μείωση των χωρικών διαστάσεων. Το πρώτο στρώμα συνελίξεως εφαρμόζει 16 φίλτρα, το δεύτερο στρώμα χρησιμοποιεί 32 φίλτρα και το τρίτο στρώμα χρησιμοποιεί 64 φίλτρα, όλα με μεγέθη πυρήνα 3×3 . Για να διασφαλιστεί η συμβατότητα με ποικίλα μεγέθη εισόδου και να αποφευχθούν σφάλματα μηδενικού μεγέθους, το τελευταίο επίπεδο συνελκτικής επεξεργασίας ενσωματώνει μια προσαρμοστική λειτουργία δειγματοληψίας του μέσου όρου (AdaptiveAvgPool2d), τυποποιώντας την έξοδο σε μια σταθερή διάσταση των 1×10 , ανεξαρτήτως του αρχικού μεγέθους του εισερχόμενου σήματος, αποφεύγοντας σφάλματα μηδενικού μεγέθους. Η έξοδος από τα συνελκτικά στρώματα τοποθετείται σε ένα άνυσμα στήλης (flattening), όπου περνάει στα πλήρως συνδεδεμένα στρώματα, ξεκινώντας με ένα πυκνό στρώμα που περιέχει 128 νευρώνες που ενεργοποιούνται από την ReLU, ακολουθούμενο από ένα στρώμα εγκατάλειψης (Dropout) με ποσοστό εγκατάλειψης 50% για τον μετριασμό της υπερπροσαρμογής. Το τελικό πυκνό στρώμα εξάγει προβλέψεις που αντιστοιχούν στον αριθμό των κλάσεων της εργασίας ταξινόμησης. Ο σχεδιασμός του μοντέλου είναι υπολογιστικά αποδοτικός, καθιστώντας το κατάλληλο για περιβάλλοντα με περιορισμένους πόρους, και εξισορροπεί αποτελεσματικά την εξαγωγή χαρακτηριστικών με την απόδοση ταξινόμησης. Στην Εικόνα 5-5 παρουσιάζεται η βασική αρχιτεκτονική που χρησιμοποιήθηκε.



Εικόνα 5-5 Σχηματική αναπαράσταση της βασικής αρχιτεκτονικής του μοντέλου εκπαίδευσης. Δημιουργήθηκε με τη βοήθεια του εργαλείου Netron (<https://netron.app/>)

Ενώ στον Πίνακα 5-3 παρουσιάζεται μια λεπτομερής ανάλυση του μοντέλου που χρησιμοποιήθηκε για την εκπαίδευση δεδομένων με σχήμα εισόδου (7,109).

Πίνακας 5-3 Παρουσίαση αρχιτεκτονικής του μοντέλου εκπαίδευσης

Στρώση (Layer)	Σχήμα Εισόδου	Σχήμα Εξόδου	Μέγεθος Φίλτρου	Βήμα (Stride)	Συμπλήρωση (Padding)	Περιγραφή
Conv2d (16 φίλτρα)	(7, 129)	(7, 129)	(3,3)	(1,1)	(1,1)	16 φίλτρα, πυρήνας 3×3
ReLU	(7, 129)	(7, 129)	-	-	-	Συνάρτηση ενεργοποίησης
MaxPool2d	(7, 129)	(3, 64)	(2,2)	(2,2)	-	Υποδειγματοληψία
Conv2d (32 φίλτρα)	(3, 64)	(3, 64)	(3,3)	(1,1)	(1,1)	32 φίλτρα, πυρήνας 3×3
ReLU	(3, 64)	(3, 64)	-	-	-	Συνάρτηση ενεργοποίησης
MaxPool2d	(3, 64)	(1, 32)	(2,2)	(2,2)	-	Υποδειγματοληψία
Conv2d (64 φίλτρα)	(1, 32)	(1, 32)	(3,3)	(1,1)	(1,1)	64 φίλτρα, πυρήνας 3×3
ReLU	(1, 32)	(1, 32)	-	-	-	Συνάρτηση ενεργοποίησης
AdaptiveAvgPool2d	(1, 32)	(1, 10)	-	-	-	Εξασφαλίζει σταθερό πλάτος εξόδου (1,10)
Flatten	(1, 10)	(10)	-	-	-	Ισοπέδωση των χωρικών διαστάσεων
Linear	(10)	(128)	-	-	-	Πλήρως συνδεδεμένο επίπεδο
ReLU	(128)	(128)	-	-	-	Συνάρτηση ενεργοποίησης
Dropout (0.5)	(128)	(128)	-	-	-	Κανονικοποίηση (Dropout)
Linear (Έξοδος)	(128)	(3)	-	-	-	Τελική έξοδος για τις κατηγορίες

Το παραπάνω μοντέλο εκπαιδεύτηκε χρησιμοποιώντας μια προσέγγιση διαστρωματικής διασταυρούμενης επικύρωσης με 3 διαμερίσεις (Stratified K-Fold Cross-Validation), η οποία είναι μια τεχνική που χωρίζει το σύνολο δεδομένων σε διαμερίσεις ίσου μεγέθους, διατηρώντας παράλληλα την ίδια κατανομή των κλάσεων. Σε κάθε διαμέριση, το σύνολο δεδομένων χωρίστηκε σε σύνολα εκπαίδευσης και δοκιμής, ενώ ένα επιπλέον 10% των δειγμάτων της εκπαίδευσης χρησιμοποιήθηκε ως σύνολο επικύρωσης για τον προσδιορισμό της εποχής που αποδίδει το καλύτερο μοντέλο κατά τη διάρκεια της εκπαίδευσης. Το μοντέλο εκπαιδεύτηκε για 30 εποχές με μέγεθος παρτίδας 32, χρησιμοποιώντας τον βελτιστοποιητή (optimizer) τον Adam με ρυθμό μάθησης 0,0001 και μείωση βάρους 0,0001 για καλύτερη γενίκευση. Ως συνάρτηση απώλειας (loss function) χρησιμοποιήθηκε η Cross-Entropy Loss. Σε κάθε εποχή, παρακολουθούνταν η απώλεια επικύρωσης του μοντέλου και αποθηκευόταν το μοντέλο με τις καλύτερες επιδόσεις για κάθε αναδίπλωση. Αυτή η στρατηγική διασταυρούμενης επικύρωσης διασφαλίζει ότι το μοντέλο μαθαίνει ισχυρά χαρακτηριστικά και γενικεύει καλά σε διαφορετικά υποσύνολα δεδομένων.

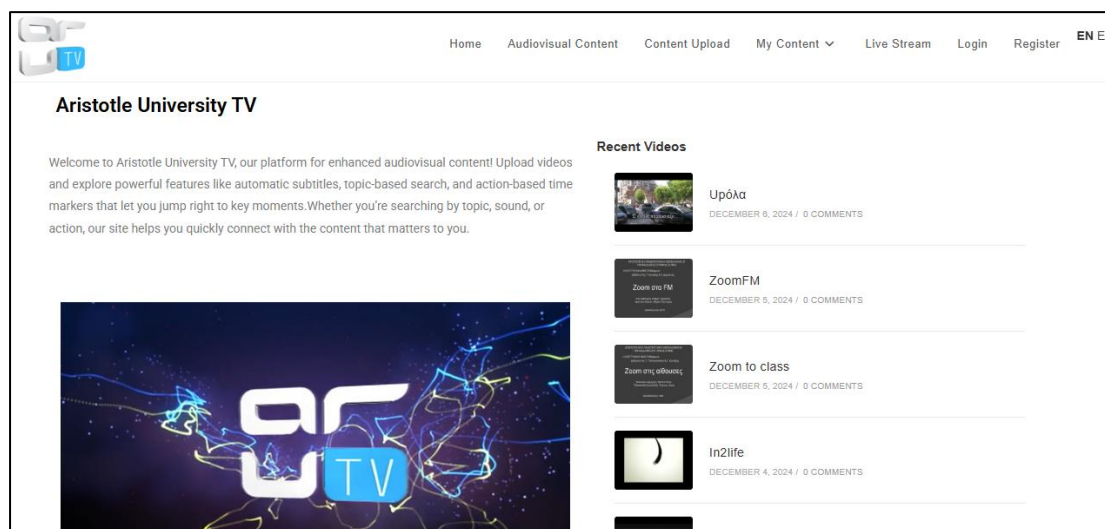
5.6. Γραφική Προβολή Διεπαφής Χρήστη

Η γραφική προβολή του ιστοτόπου αποτελεί έναν από τους βασικούς πυλώνες της εφαρμογής, καθώς προσφέρει στους χρήστες τη δυνατότητα να αλληλοεπιδρούν με το σύστημα μέσω ενός φιλικού και κατανοητού περιβάλλοντος. Η σχεδίαση του ιστοτόπου έχει επικεντρωθεί στη διασφάλιση μιας ομαλής εμπειρίας του χρήστη, με χρήση σύγχρονων εργαλείων και τεχνολογιών για την παρουσίαση των δεδομένων. Κύριοι στόχοι είναι η εύκολη πλοήγηση, η παρουσίαση των αποτελεσμάτων με σαφή και κατανοητό τρόπο, καθώς και η υποστήριξη όλων των απαιτούμενων λειτουργιών της εφαρμογής μέσω του διαδικτυακού περιβάλλοντος. Στην ενότητα αυτή, θα παρουσιαστεί η δομή και η λειτουργικότητα του ιστοτόπου, καθώς και τα χαρακτηριστικά που κάνουν την αλληλεπίδραση με το σύστημα αποτελεσματική και αποδοτική για τον χρήστη.

Τα βασικά χαρακτηριστικά της ιστοσελίδας είναι η ύπαρξη ενός κεντρικού μενού στο επάνω μέρος της σελίδας, μέσω του οποίου ο χρήστης μπορεί να περιηγηθεί στην ιστοσελίδα. Διατίθεται, επίσης, και η επιλογή γλώσσας στο σύστημα, είτε της ελληνικής είτε της αγγλικής. Στις εικόνες που θα παρουσιαστούν, εκτός από αυτές που αφορούν στη δημιουργία λογαριασμού και την σύνδεση του χρήστη στο σύστημα, υποθέτουμε ότι ο χρήστης είναι συνδεδεμένος στο σύστημα.

5.6.1. Αρχική Σελίδα

Η αρχική σελίδα της εφαρμογής διαφέρει από τη συνηθισμένη εικόνα των υπόλοιπων εφαρμογών οπτικοακουστικού περιεχομένου, με σκοπό να καθοδηγήσει τους χρήστες του συστήματος σχετικά με το τι ακριβώς πραγματεύεται η ιστοσελίδα και τι πρωτοπόρες δυνατότητες προσφέρει. Στην Εικόνα 5-6, παρατίθεται ένα στιγμιότυπο της αρχικής σελίδας του ιστοτόπου.



Εικόνα 5-6 Στιγμιότυπο αρχικής σελίδας του ιστοτόπου

5.6.2. Σελίδα Σύνδεσης

Ένα ένας χρήστης που διαθέτει λογαριασμό στην ιστοσελίδα θα μπορεί να συνδεθεί σε αυτήν, χρησιμοποιώντας το όνομα χρήστη και τον κωδικό του. Διαφορετικά θα πρέπει να μεταβεί στην σελίδα δημιουργίας λογαριασμού, όπως φαίνεται και στην Εικόνα 5-7.

Εικόνα 5-7 Στιγμιότυπο σελίδας σύνδεσης του χρήστη

5.6.3. Σελίδα Δημιουργίας Λογαριασμού

Στην σελίδα δημιουργίας λογαριασμού ο χρήστης μπορεί να δημιουργήσει ένα λογαριασμό, εφόσον διαθέτει ένα έγκυρο ηλεκτρονικό ταχυδρομείο. Η φόρμα υποβολής εγγραφής παρουσιάζεται στην Εικόνα 5-8.

Εικόνα 5-8 Στιγμιότυπο σελίδας δημιουργίας λογαριασμού χρήστη

Σε αυτήν την σελίδα, μόνο ο χρήστης που είναι συνδεδεμένος στο σύστημα θα έχει τη δυνατότητα να αναρτήσει ένα πολυμεσικό αρχείο. Στην Εικόνα 5-9 αποτυπώνεται η φόρμα μεταφόρτωσης, στην οποία ο χρήστης μπορεί να συμπληρώσει κάποια μεταδεδομένα, όπως κατηγορία και όνομα. Επίσης, αυτόματα συμπληρώνεται το όνομα του συνδεδεμένου χρήστη.

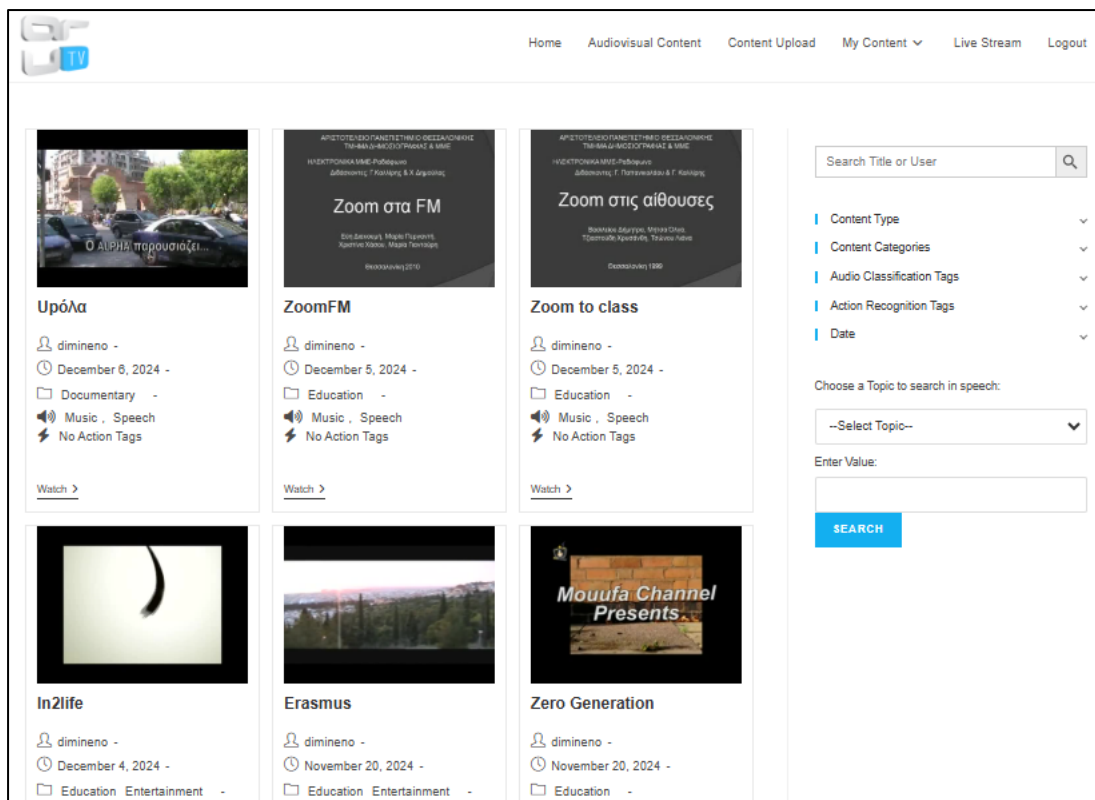
Εικόνα 5-9 Στιγμιότυπο σελίδας μεταφόρτωσης οπτικοακουστικού υλικού

5.6.4. Σελίδα Προβολής Οπτικοακουστικού Περιεχομένου

Σε αυτήν τη σελίδα, η οποία είναι ορατή για όλους του χρήστες ασχέτως αν είναι εγγεγραμμένοι ή όχι, υπάρχει αναρτημένο όλο το οπτικοακουστικό υλικό του συστήματος. Οι χρήστες έχουν την δυνατότητα να αναζητούν το περιεχόμενο που θέλουν να παρακολουθήσουν με ποικίλους τρόπους.

Ένας από αυτούς είναι η απλή αναζήτηση στους τίτλους του αναρτημένου υλικού ή στα ονόματα των χρηστών. Επίσης, διατίθεται η δυνατότητα αναζήτησης περιεχομένου με τη χρήση διαφόρων φίλτρων. Τα διαθέσιμα φίλτρα χωρίζονται σε 2 κατηγορίες, εξαρτώνται από τα μεταδεδομένα και εισέρχονται χειροκίνητα από τους χρήστες (όπως η γενική κατηγορία και η ημερομηνία μεταφόρτωσης). Τα υπόλοιπα φίλτρα αφορούν τα αποτελέσματα αλγορίθμων και αφορούν στην αναγνώριση δράσεων και στην κατηγοριοποίηση του ήχου. Καθώς στο σύστημα μπορεί να υπάρχουν διαφορετικά είδη αρχείων όπως ήχος και βίντεο, ο χρήστης μπορεί να επιλέξει τι είδος θέλει να παρακολουθήσει ή ακούσει. Στα αρχεία ήχου τοποθετούνται και τα στατικά βίντεο.

Επιπρόσθετα, διατίθεται η δυνατότητα αναζήτησης στο περιεχόμενο της ομιλίας. Δηλαδή, υπάρχουν έξι θέματα με βάση τα οποία μπορεί να γίνει η αναζήτηση. Τα θέματα αυτά αφορούν σε όνομα ανθρώπου, γεγονόςτος, γεωπολιτικής τοποθεσίας, τοποθεσίας, προϊόντος και οργανισμού. Κατ' αυτόν τον τρόπο, ο χρήστης επιλέγει το θέμα και έπειτα συμπληρώνει την λέξη ή φράση σε ελληνικούς ή αγγλικούς χαρακτήρες. Τέλος, προσφέρεται η επιλογή αναζήτησης σε όλο το κείμενο της ομιλίας, σε περίπτωση που ο χρήστης επιθυμεί να αναζητήσει κάτι διαφορετικό. Όλα τα παραπάνω είδη αναζήτησης λειτουργούν και συνδυαστικά (Εικόνα 5-10).



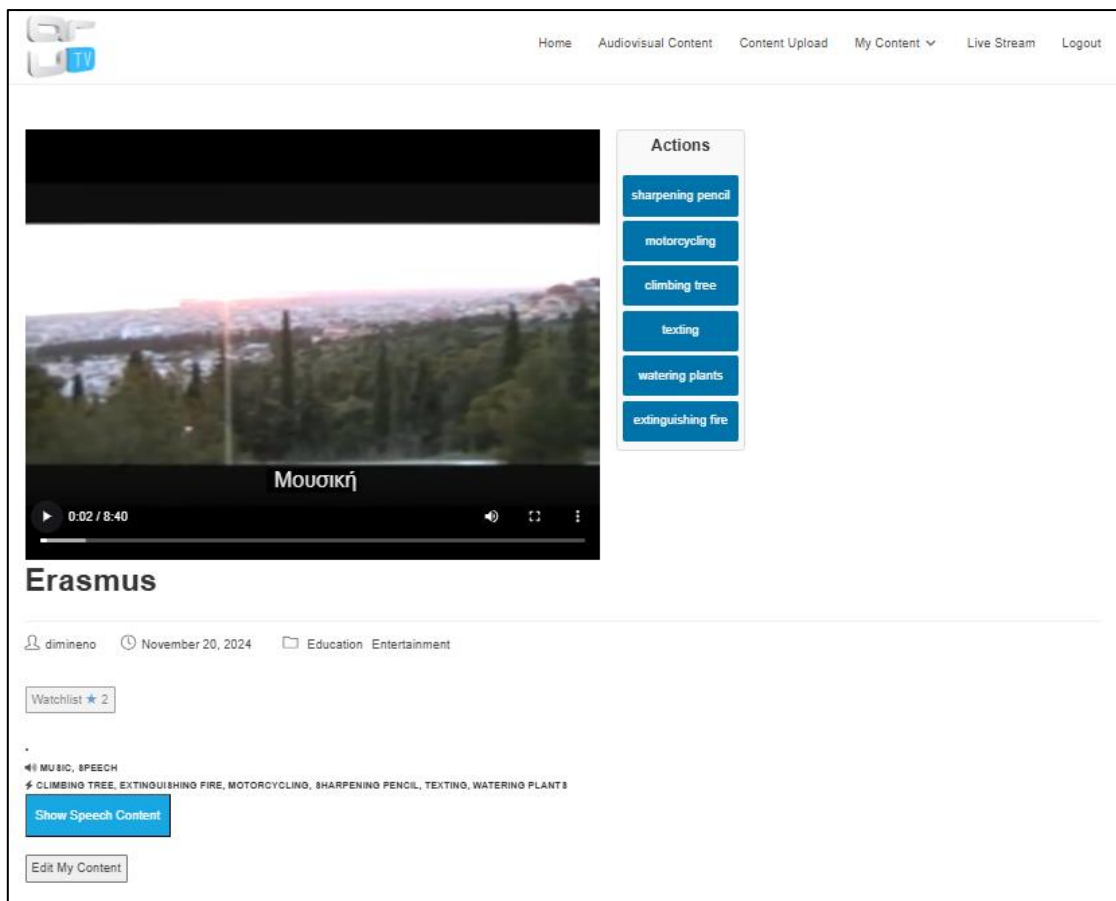
Εικόνα 5-10 Στιγμιότυπο σελίδας προβολής οπτικοακουστικού υλικού

5.6.5. Σελίδα Αναπαραγωγής περιεχομένου

Στην σελίδα αναπαραγωγής περιεχομένου ο χρήστης έχει την δυνατότητα να παρακολουθήσει το αρχείο που επιθυμεί. Ανεξαρτήτως αρχείου γίνονται διακριτές οι διάφορες κατηγορίες και οι ετικέτες που αντιστοιχούν στο αρχείο. Εάν ο χρήστης το επιθυμεί, μπορεί να το προσθέσει σε μια λίστα, ώστε να το παρακολουθήσει αργότερα. Εάν το αρχείο προβολής ανήκει στο χρήστη τότε υπάρχει η δυνατότητα επεξεργασίας του. Επίσης, διατίθεται η επιλογή, κατά την οποία ο χρήστης μπορεί να διαβάσει, εφόσον το επιθυμεί, το κείμενο της ομιλίας του αρχείου.

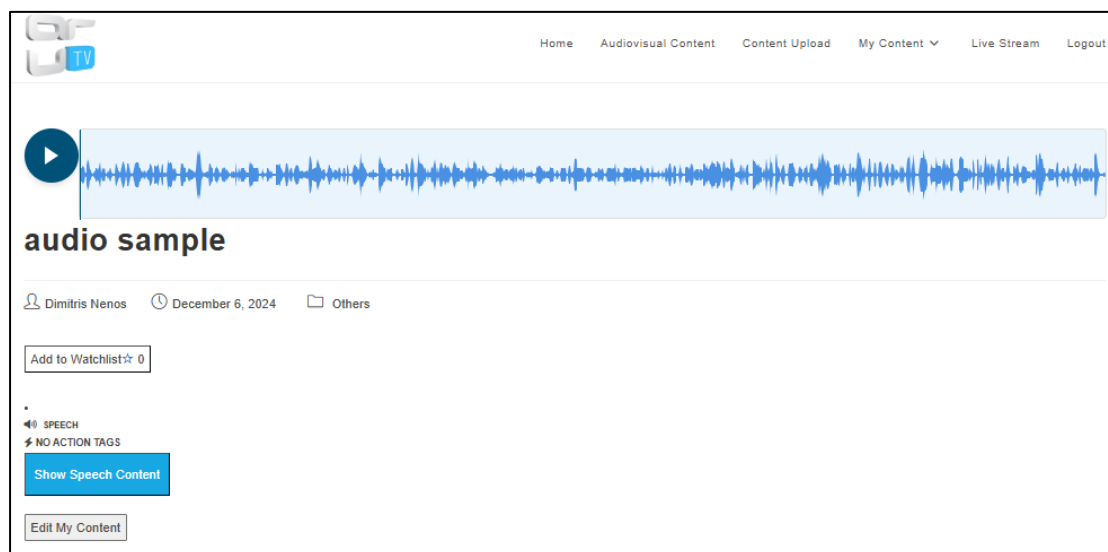
Στη περίπτωση που το αρχείο είναι βίντεο και υπάρχει κάποιο αποτέλεσμα του αλγορίθμου αναγνώρισης δράσης, ο χρήστης έχει τη δυνατότητα μέσω της δημιουργίας δυναμικών κουμπιών να μεταβεί στο αντίστοιχο δευτερόλεπτο που βρέθηκε κάποια δράση.

Επιπλέον, ο χρήστης έχει τη δυνατότητα να προσθέσει ελληνικούς ή αγγλικούς υπότιτλους εφόσον το επιθυμεί. Στη Εικόνα 5-11 παρουσιάζεται ένα στιγμιότυπο σελίδας αναπαραγωγής ενός αρχείου βίντεο.



Εικόνα 5-11 Στιγμιότυπο σελίδας αναπαραγωγής βίντεο

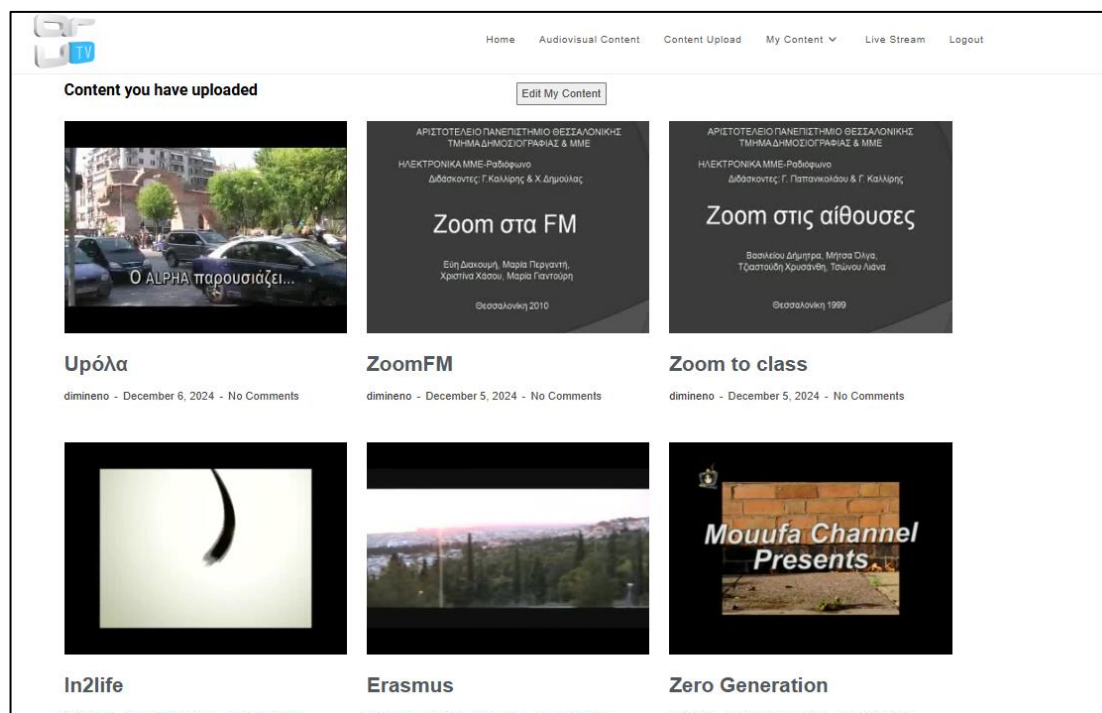
Στην περίπτωση που το αρχείο είναι ήχος τότε τροποποιείται ελάχιστα η μορφή της σελίδας αναπαραγωγής, όπως φαίνεται στην Εικόνα 5-12.



Εικόνα 5-12 Στιγμιότυπο σελίδας αναπαραγωγής ήχου

5.6.6. Σελίδα Περιεχομένου εκάστοτε χρήστη

Σε αυτή την σελίδα εμφανίζεται όλο το υλικό που έχει αναρτήσει ο εκάστοτε χρήστης που είναι συνδεδεμένος στην εφαρμογή (Εικόνα 5-13). Επίσης, υπάρχει μια φόρμα επεξεργασίας περιεχομένου σε περίπτωση λανθασμένης εισόδου, είτε λόγω λάθους του χρήστη ή του αλγόριθμου.



Εικόνα 5-13 Σελίδα περιεχομένου του εκάστοτε χρήστη

5.6.7. Προφίλ Χρηστών

Όπως κάθε πλατφόρμα, έτσι και η εφαρμογή της παρούσας διπλωματικής διατριβής διαθέτει σελίδα προφίλ χρηστών. Η συγκεκριμένη αφορά στις πληροφορίες του λογαριασμού του κάθε χρήστη. Σε αυτήν, ο χρήστης μπορεί να αλλάξει τα προσωπικά του στοιχεία, όπως αυτό αποτυπώνεται στην Εικόνα 5-14.

Home Audiovisual Content Content Upload My Content Live Stream Logout

Account

Username *
neniak

First Name
Dimitris

Last Name
Nenos

Email Address
neniak1997@gmail.com

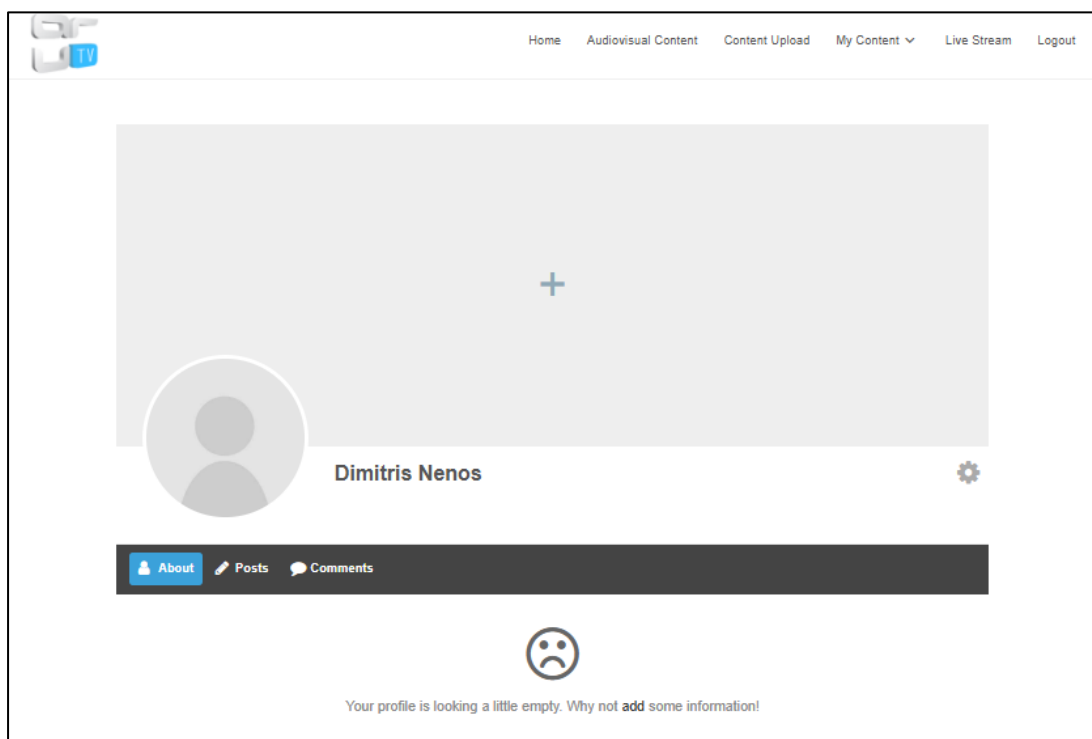
Update Account

Account >
Change Password >
Privacy >
Delete Account >

Dimitris Nenos
[View profile](#)

Εικόνα 5-14 Στιγμιότυπο σελίδας λογαριασμού του χρήστη

Τέλος, στην εφαρμογή εντοπίζεται και το προφίλ του χρήστη, το οποίο είναι δημόσιο και φανερό σε όλους (Εικόνα 5-15). Σε αυτό υπάρχουν σε μορφή λίστας και όλα τα αρχεία που έχουν αναρτηθεί από τον εκάστοτε χρήστη.

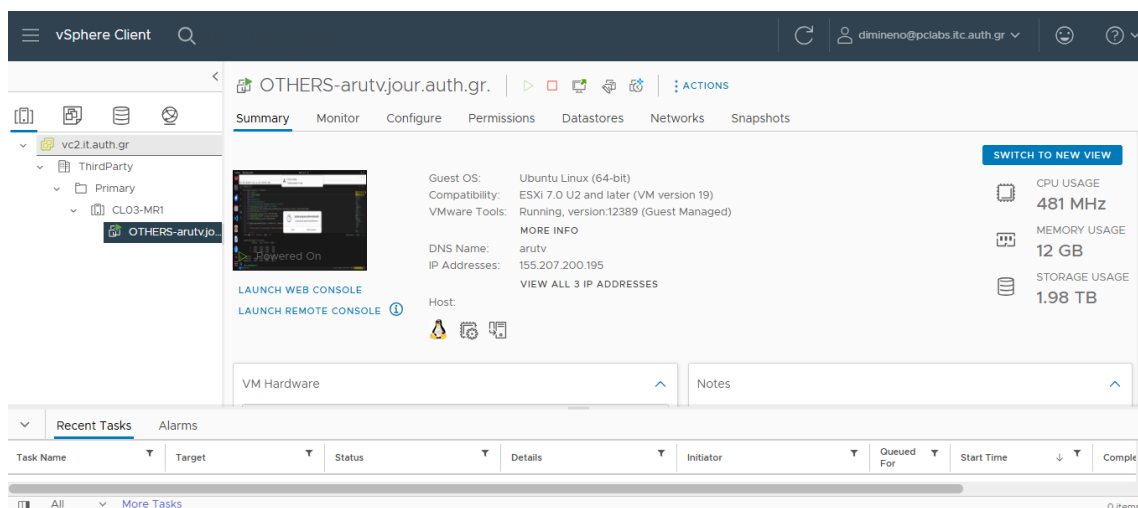


Εικόνα 5-15 Στιγμιότυπο δημοσίου προφίλ χρηστών

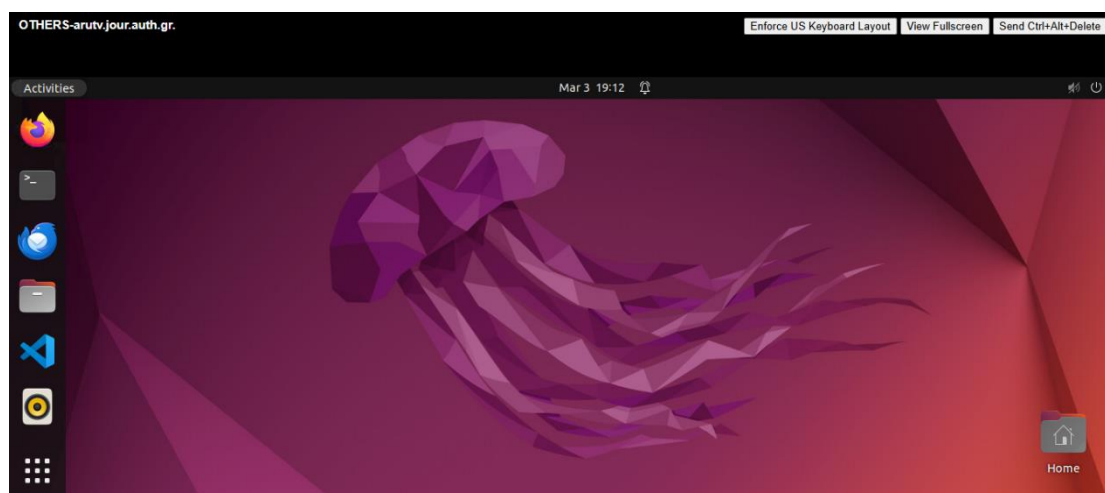
5.7. Γραφική Προβολή Περιβάλλοντος Διαχειριστή

5.7.1. Πλατφόρμα Διαχείρισης Εικονικής Μηχανής

Στην Εικόνα 5-16 απεικονίζεται το vSphere Client, το οποίο είναι ένα περιβάλλον εργασίας για τη διαχείριση εικονικών μηχανών. Από εδώ γίνεται η παρακολούθηση της εικονικής μηχανής (που παραχωρήθηκε από το ΚΗΔ του Α.Π.Θ.) και των πόρων της σε πραγματικό χρόνο. Επιπλέον από εδώ πραγματοποιείται και η είσοδος στην εικονική μηχανή και στο περιβάλλον της (Εικόνα 5-17).



Εικόνα 5-16 Εργαλείο για την παρακολούθηση και διαχείριση εικονικών μηχανών και του εικονικού περιβάλλοντος σε πραγματικό χρόνο.

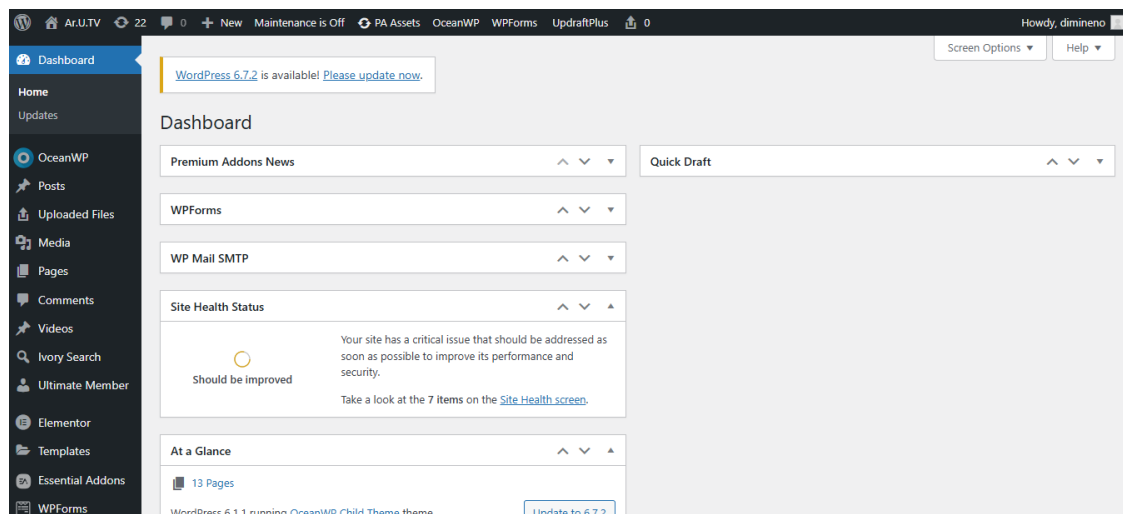


Εικόνα 5-17 Το περιβάλλον της εικονικής μηχανής του συστήματος

5.7.2. Πίνακας Διαχείρισης του Συστήματος Διαχείρισης Περιεχομένου

Η εικόνα 5-18 παρουσιάζει το κεντρικό περιβάλλον διαχείρισης ενός WordPress ιστοτόπου. Ο διαχειριστής μπορεί να προσθέτει ή να επεξεργάζεται περιεχόμενο (αναρτήσεις, σελίδες), να εγκαθιστά ή να ρυθμίζει πρόσθετα, να ελέγχει τη

λειτουργικότητα της ιστοσελίδας, καθώς και να διαχειρίζεται ενημερώσεις ή ρυθμίσεις ασφαλείας.



Εικόνα 5-18 Ο Πίνακας Ελέγχου (Dashboard) του WordPress

5.8. Σύνοψη Πέμπτου Κεφαλαίου

Στο πέμπτο κεφάλαιο πραγματοποιείται η υλοποίηση του συστήματος. Αρχικά περιγράφεται η ανάπτυξη της βασικής υποδομής και της εγκατάσταση του συστήματος διαχείρισης περιεχομένου. Έπειτα αναλύεται η ενσωμάτωση των απαραίτητων μηχανισμών επικοινωνίας και διαχείρισης εργασιών, όπως το RabbitMQ για τη διαχείριση μηνυμάτων και το Celery για την εκτέλεση ασύγχρονων διεργασιών. Γίνεται περιγραφή των αλγορίθμων ανάλυσης πολυμεσικού περιεχομένου και παρουσιάζεται η ευκολία προσθήκης ή αφαίρεσης αλγορίθμων λόγω της αρθρωτής αρχιτεκτονικής που δημιουργήθηκε. Επιπλέον, παρουσιάζεται λεπτομερώς η ανάπτυξη του αλγορίθμου ταξινόμησης ήχου με τη χρήση ενός φασματογραφήματος με βάση τα κυματίδια ως είσοδος σε ένα νευρωνικό δίκτυο. Τέλος, παρατίθεται η γραφική διεπαφή του ιστοτόπου, η δομή των σελίδων και η λειτουργικότητα της εφαρμογής αλλά και το περιβάλλον του διαχειριστή. Στο επόμενο κεφάλαιο θα γίνει η περιγραφή της αξιολόγησης των επιδόσεων, της ευχρηστίας και όλων των αλγορίθμων του συστήματος.

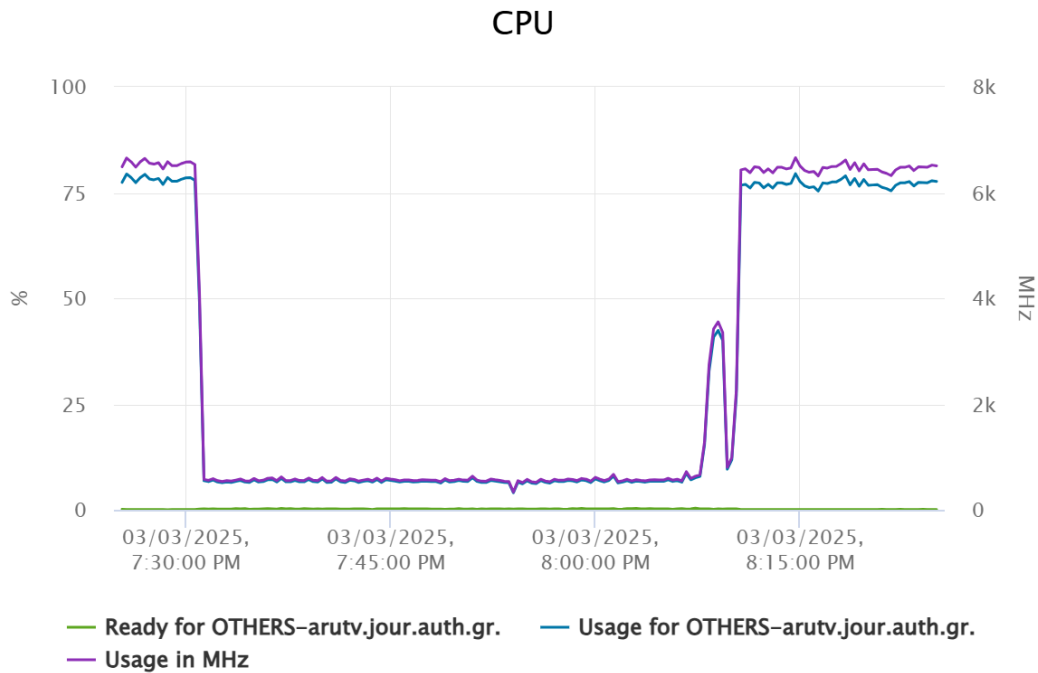
Κεφάλαιο 6ο

6. Αξιολόγηση

Το διαδικτυακό περιβάλλον αναπτύχθηκε με βασικό στόχο τη δημιουργία μιας φιλικής προς το χρήστη πλατφόρμας που θα είναι ικανή να διαχειρίζεται, να αναλύει και να ανακτά μεγάλο όγκο οπτικοακουστικού υλικού. Το σύστημα διευκολύνει τη μεταφόρτωση, την αναζήτηση και την αλληλεπίδραση με το οπτικοακουστικό υλικό και έχει σχεδιαστεί για να εξυπηρετεί αρχάριους χρήστες έως έμπειρους επαγγελματίες. Το σύστημα αυτό διαθέτει διάφορες βασικές καινοτομίες. Ενσωματώνει αλγορίθμους για τη μετατροπή ομιλίας σε κείμενο, την αναγνώριση ονομαστικών οντοτήτων, την αναγνώριση ενεργειών και την ταξινόμηση ήχου με τη χρήση φασματογραφημάτων βασισμένων σε κυματίδια, με αποτέλεσμα να επιτρέπει την εξαγωγή πλούσιων σημασιολογικών μεταδεδομένων και να ξεπερνά τους περιορισμούς των παραδοσιακών προσεγγίσεων που εξαρτώνται από τα μεταδεδομένα. Επιπλέον, η αρχιτεκτονική βασίζεται σε ένα κλιμακούμενο και ασύγχρονο πλαίσιο που αξιοποιεί τεχνολογίες όπως το RabbitMQ και το Celery, εξασφαλίζοντας τον αποδοτικό χειρισμό των υπολογιστικά απαιτητικών εργασιών. Ένα επιπλέον πλεονέκτημα του συστήματος έγκειται στον αρθρωτό και επεκτάσιμο σχεδιασμό του, ο οποίος διευκολύνει την απλή προσθήκη, αφαίρεση ή αντικατάσταση αλγορίθμων μηχανικής μάθησης. Αυτή η προσαρμοστικότητα όχι μόνο απλοποιεί τις μελλοντικές αναβαθμίσεις, αλλά διασφαλίζει, επίσης, ότι το σύστημα παραμένει ευέλικτο στις νέες τεχνολογίες και στις εξελισσόμενες απαιτήσεις των χρηστών. Συλλογικά, αυτά τα χαρακτηριστικά ενισχύουν σημαντικά την ακρίβεια αναζήτησης θέτοντας θεμέλια για διαρκή καινοτομία και μακροπρόθεσμη βιωσιμότητα.

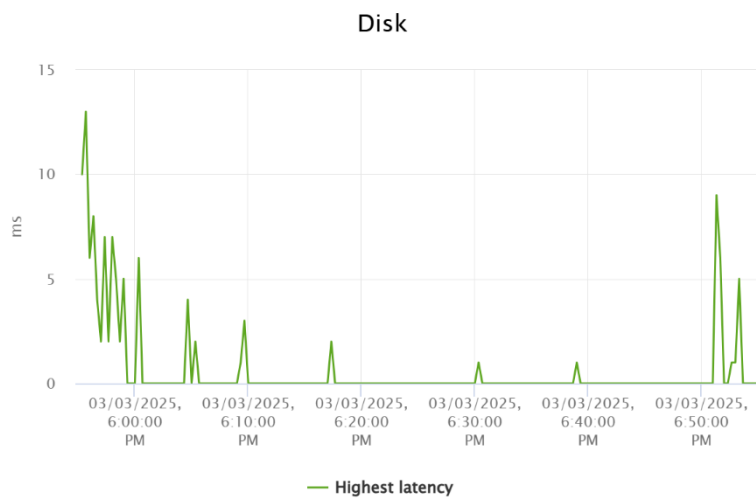
6.1. Αξιολόγηση Απόδοσης του Συστήματος

Σε αυτήν την ενότητα παρουσιάζεται μια ολοκληρωμένη αξιολόγηση του συστήματος κατά την εκτέλεση των αλγορίθμων, βασισμένη στα διαγράμματα που εξάγει το εργαλείο vSphere, καθώς και στα δεδομένα από τη χρήση των nvidia-smi και htop μέσω της γραμμής εντολών της εικονικής μηχανής και τα logs της. Στην Εικόνα 6-1 παρουσιάζεται τα διάγραμμα χρήσης CPU που αποκαλύπτει την κατανομή των vCPU και την ένταση χρήσης τους σε πραγματικό χρόνο, δείχνοντας ότι υπό συνθήκες εκτέλεσης αλγορίθμων η χρήση κυμαίνεται στο εύρος του 40% έως 80%. Σε περιόδους αυξημένης ζήτησης, η αύξηση της χρήσης της CPU παραμένει εντός των αποδεκτών ορίων, επιβεβαιώνοντας την αποτελεσματική διαχείριση των εργασιών.

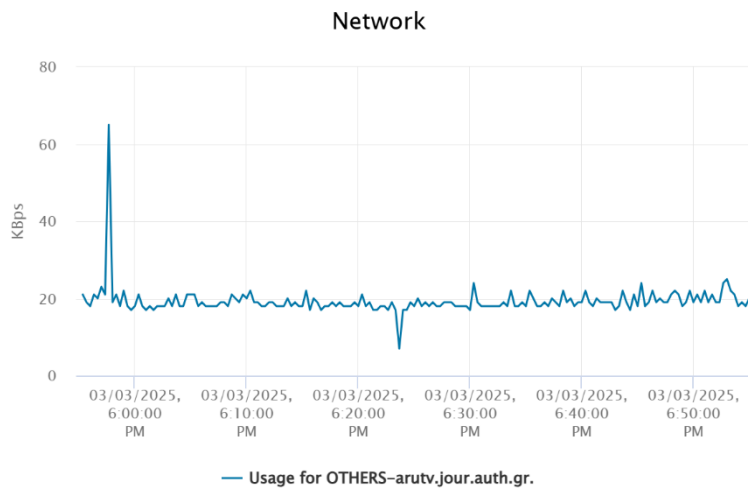


Εικόνα 6-1 Διάγραμμα χρήσης της CPU του συστήματος κατά την εκτέλεση αλγορίθμων (vSphere)

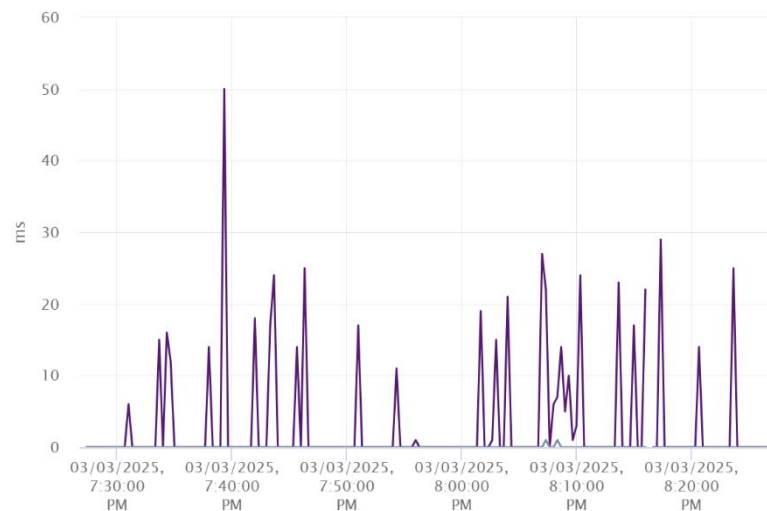
Το διάγραμμα λανθάνουσας κατάστασης των δίσκων (disk latency) (Εικόνα 6-2) δείχνει ότι οι καθυστερήσεις στις αναγνώσεις και εγγραφές δεδομένων παραμένουν χαμηλές κατά τη διάρκεια της εκτέλεσης των αλγορίθμων, διασφαλίζοντας την ομαλή λειτουργία του συστήματος. Παράλληλα, η δικτυακή κυκλοφορία, όπως αποτυπώνεται στην Εικόνα 6-3, εμφανίζει σταθερούς ρυθμούς μεταφοράς δεδομένων με ελάχιστες απώλειες ή σφάλματα, εξασφαλίζοντας την απρόσκοπτη επικοινωνία μεταξύ των στοιχείων του συστήματος. Η απόδοση του εικονικού δίσκου (virtual disk) παραμένει εντός των βέλτιστων ορίων, αποδεικνύοντας ότι το σύστημα μπορεί να διαχειριστεί απαιτητικές διεργασίες χωρίς προβλήματα συμφόρησης (Εικόνα 6-4)



Εικόνα 6-2 Διάγραμμα λανθάνουσας κατάστασης δίσκων (Disk Latency) του συστήματος κατά την εκτέλεση αλγορίθμων (vSphere).

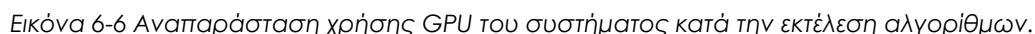


Εικόνα 6-3 Διάγραμμα δικτυακής κυκλοφορίας του συστήματος κατά την εκτέλεση αλγορίθμων (vSphere).



Εικόνα 6-4 Διάγραμμα απόδοσης εικονικού δίσκου (Virtual Disk Performance) του συστήματος κατά την εκτέλεση αλγορίθμων (vSphere).

Η χρήση της GPU κυμαίνεται μεταξύ 40% και 90% κατά τη διάρκεια εκτέλεσης αλγορίθμων, με αύξηση της δραστηριότητας σε απαιτητικές διεργασίες, όπως η αναγνώριση ενεργειών. Αντίστοιχα, η μνήμη του συστήματος παρακολουθείται μέσω της γραμμής εντολών (htop), το οποίο επιβεβαιώνει ότι δεν παρατηρούνται φαινόμενα υπερφόρτωσης, ενώ η διαχείριση της διαθέσιμης μνήμης παραμένει ομαλή χωρίς προβλήματα. Στη Εικόνα 6-5, 6-6 φαίνονται τα στιγμιότυπα των εντολών htop και nvidia-smi αντίστοιχα.



Πίνακας 6-1 Σύνοψη Μετρικών Απόδοσης του Συστήματος κατά την Εκτέλεση Αλγορίθμων

77

Τα παραπάνω δεδομένα αποδεικνύουν ότι το σύστημα λειτουργεί εντός των βέλτιστων ορίων για όλους τους βασικούς πόρους κατά τη διάρκεια εκτέλεσης αλγορίθμων. Η σταθερή χρήση της CPU, η χαμηλή λανθάνουσα κατάσταση των δίσκων, η αξιόπιστη απόδοση του εικονικού δίσκου και η συνεπής δικτυακή μεταφορά δεδομένων επιβεβαιώνουν ότι η εικονική μηχανή βρίσκεται σε καλή κατάσταση και διαχειρίζεται αποτελεσματικά τις απαιτήσεις του συστήματος. Συνολικά, η αξιολόγηση καταδεικνύει ότι το σύστημα είναι ικανό να υποστηρίξει εφαρμογές υψηλών απαιτήσεων, παρέχοντας αξιόπιστη και σταθερή λειτουργία ακόμη και υπό αυξημένο φόρτο εργασίας.

6.2. Αξιολόγηση Περιβάλλοντος Εφαρμογής

6.2.1. Ποιοτική Αξιολόγηση

Η αξιολόγηση του συστήματος πραγματοποιήθηκε μέσω ποιοτικής ανάλυσης, η οποία βασίζεται τόσο σε ημιδομημένες συζητήσεις όσο και σε άτυπες δοκιμές χρήσης από συμμετέχοντες με διαφορετικά επίπεδα εμπειρίας. Ο στόχος ήταν η εκτίμηση της χρηστικότητας, της απόδοσης και της συνολικής εμπειρίας του χρήστη, καθώς και η ανίχνευση πιθανών σημείων προς βελτίωση. Δεδομένου ότι το σύστημα απευθύνεται τόσο σε έμπειρους όσο και σε αρχάριους χρήστες, κρίθηκε απαραίτητο να συμμετάσχουν άτομα που ανήκουν και στα δύο σύνολα. Η μεθοδολογία αξιολόγησης του συστήματος βασίστηκε στην αλληλεπίδραση των χρηστών με το σύστημα και στη συλλογή σχολίων αναφορικά με την εμπειρία τους. Πραγματοποιήθηκαν παρατηρήσεις και συνεντεύξεις με χρήστες που δοκίμασαν το σύστημα σε πραγματικές συνθήκες. Οι συμμετέχοντες αποτελούνταν από 2 ομάδες, η πρώτη ομάδα αποτελούνταν από χρήστες με εμπειρία στη διαχείριση οπτικοακουστικού υλικού, ενώ η δεύτερη περιλάμβανε χρήστες χωρίς ιδιαίτερη εμπειρία. Οι χρήστες κλήθηκαν να αλληλοεπιδράσουν με το σύστημα εκτελώντας συγκεκριμένες εργασίες, όπως σύνδεση στο σύστημα μεταφόρτωσης αρχείου, αναζήτηση αρχείου στο σύστημα χρησιμοποιώντας τα διάφορα φίλτρα αναζήτησης, καθώς και την προβολή διαφόρων πολυμεσικών αρχείων. Τα χαρακτηριστικά των συμμετεχόντων βρίσκονται συνοπτικά στον Πίνακα 6-2.

Πίνακας 6-2 Χαρακτηριστικά Συμμετεχόντων στην αξιολόγηση του περιβάλλοντος της εφαρμογής

Συμμετέχων	Ομάδα	Εμπειρία στη διαχείριση οπτικοακουστικού υλικού	Προσωπικά/Επαγγελματικά Χαρακτηριστικά
1	Έμπειρος	Ναι	30 ετών, επαγγελματίας στον τομέα των ΜΜΕ
2	Αρχάριος	Όχι	26 ετών, φοιτητής
3	Έμπειρος	Ναι	35 ετών, δημιουργός περιεχομένου
4	Αρχάριος	Όχι	28 ετών, φοιτητής
5	Έμπειρος	Ναι	40 ετών, επαγγελματίας στην παραγωγή βίντεο
6	Αρχάριος	Όχι	27 ετών, φοιτητής

Η ανάλυση των αποτελεσμάτων βασίστηκε στις παρατηρήσεις και τα σχόλια των χρηστών. Οι έμπειροι χρήστες αξιολόγησαν το σύστημα ως αποδοτικό, λειτουργικό και εύχρηστο. Επισημάνθηκε ότι η ταχύτητα και η απόδοση ήταν ικανοποιητικές, με εξαίρεση κάποιες περιπτώσεις όπως η ανάρτηση μεγάλων αρχείων. Η διεπαφή χρήστη

χαρακτηρίστηκε ως σαφής, διευκολύνοντας την πλοήγηση στις διάφορες λειτουργίες του συστήματος. Επιπλέον, οι αλγόριθμοι αναγνώρισης και κατηγοριοποίησης λειτούργησαν με μεγάλη ακρίβεια, παρέχοντας αξιόπιστα αποτελέσματα. Ωστόσο, προτάθηκαν ορισμένες βελτιώσεις, όπως ήταν η προσθήκη κουμπιών διεπαφής για να μπορεί ο χρήστης να δει με ευκολία τα αποτελέσματα του αλγορίθμου αναγνώρισης δράσης. Μια άλλη πρόταση ήταν ο διαχωρισμός των αρχείων ήχου και βίντεο, ορίζοντας τα στατικά βίντεο ως αρχεία ήχου. Οι χρήστες χωρίς εμπειρία χαρακτήρισαν το σύστημα φιλικό ως προς χρήστη, αν και αντιμετώπισαν ορισμένες δυσκολίες κατά την αρχική εξοικείωση. Η πλοήγηση στις βασικές λειτουργίες, όπως η αναπαραγωγή και η αναζήτηση περιεχομένου, κρίθηκε εύκολη, ενώ η διεπαφή θεωρήθηκε οπτικά καθαρή και καλοσχεδιασμένη. Μια πρόταση βελτίωσης ήταν η δημιουργία ενός καλύτερου προφίλ χρήστη. Συνολικά η αξιολόγηση έδειξε ότι το σύστημα είναι αποτελεσματικό και λειτουργικό για χρήστες τόσο με εμπειρία όσο και χωρίς, με μικρά περιθώρια βελτίωσης στην κατανόηση της ακριβούς λειτουργίας των αλγορίθμων από αρχάριους χρήστες. Οι προτάσεις για την βελτίωση της ιστοσελίδας λήφθηκαν υπόψιν και διορθώθηκαν, με αποτέλεσμα τη δημιουργία μιας ιστοσελίδας με υψηλή απόδοση και ευχρηστία για την πανεπιστημιακή κοινότητα.

6.2.2. Ποσοτική Αξιολόγηση

Η ποσοτική αξιολόγηση βασίστηκε στους συμμετέχοντες που κλήθηκαν να πραγματοποιήσουν διάφορα σενάρια χρήσης, όπως αναφέρθηκε στη προηγούμενη ενότητα. Επίσης ζητήθηκε από τους χρήστες να συμπληρώσουν ένα ερωτηματολόγιο 10 ερωτήσεων, το οποίο βασίστηκε στο σύστημα SUS (System Usability Scale). Η κλίμακα ευχρηστίας συστήματος (SUS) είναι ένα ευρέως διαδεδομένο εργαλείο που χρησιμοποιείται για την αξιολόγηση της ευχρηστίας ενός συστήματος. Αποτελείται από 10 δηλώσεις στις οποίες οι χρήστες υποδεικνύουν το επίπεδο συμφωνίας τους σε μια 5-βάθμια κλίμακα (1: Διαφωνώ απόλυτα, 5: Συμφωνώ απόλυτα) (Brooke, 1996). Στο Πίνακα 6-3 παρουσιάζονται οι ερωτήσεις που τέθηκαν στους χρήστες.

Πίνακας 6-3 Ερωτήσεις του ερωτηματολογίου της Κλίμακας Ευχρηστίας Συστήματος (System Usability Scale)

Αριθμός	Ερώτηση
1	Νομίζω ότι θα ήθελα να χρησιμοποιώ αυτό το σύστημα συχνά.
2	Βρήκα το σύστημα πιο πολύπλοκο από όσο θα ήθελα.
3	Θεώρησα ότι το σύστημα ήταν εύκολο στη χρήση.
4	Νομίζω ότι θα χρειαζόμουν την υποστήριξη ενός τεχνικού προσώπου για να μπορέσω να χρησιμοποιήσω αυτό το σύστημα.
5	Βρήκα ότι οι διάφορες λειτουργίες αυτού του συστήματος ήταν καλά ενσωματωμένες.
6	Νομίζω ότι υπήρχε πολύ μεγάλη ασυνέπεια σε αυτό το σύστημα.
7	Φαντάζομαι ότι οι περισσότεροι άνθρωποι θα μάθαιναν να χρησιμοποιούν αυτό το σύστημα πολύ γρήγορα.
8	Βρήκα το σύστημα πολύ δυσκίνητο στη χρήση.
9	Αισθάνθηκα πολύ σίγουρος για τη χρήση του συστήματος.
10	Χρειάστηκε να μάθω πολλά πράγματα προτού μπορέσω να ξεκινήσω με αυτό το σύστημα.

Για τη βαθμολόγηση, οι απαντήσεις στα θετικά διατυπωμένα στοιχεία προσαρμόζονται αφαιρώντας το 1, και οι απαντήσεις στα αρνητικά διατυπωμένα στοιχεία αφαιρούνται από το 5. Το άθροισμα αυτών των προσαρμοσμένων βαθμολογιών πολλαπλασιάζεται στη συνέχεια με το 2,5, με αποτέλεσμα μια τελική βαθμολογία μεταξύ 0 και 100. Μια

υψηλότερη βαθμολογία SUS υποδηλώνει καλύτερη χρηστικότητα. Στο Πίνακα 6-4 παρουσιάζονται τα αποτελέσματα του ερωτηματολογίου για κάθε χρήστη, ενώ στην Εικόνα 6-7 παρουσιάζεται ένα ραβδόγραμμα όπου οι μέσες βαθμολογίες για καθεμία από τις δέκα ερωτήσεις του SUS.

Πίνακας 6-4 Τα αποτελέσματα του ερωτηματολογίου της Κλίμακας Ευχρηστίας Συστήματος για κάθε χρήστη και ο συνολικός μέσος όρος.

Χρήστης	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	Τελική Βαθμολογία SUS
1	5	1	5	1	5	1	5	1	5	1	$40 \times 2,5 = 100$
2	4	2	4	2	4	2	4	2	4	2	$30 \times 2,5 = 75$
3	5	2	4	1	5	2	4	2	5	2	$34 \times 2,5 = 85$
4	4	1	5	2	4	2	4	1	4	2	$33 \times 2,5 = 82,5$
5	5	1	4	2	5	2	5	1	4	2	$35 \times 2,5 = 87,5$
6	4	1	5	2	5	2	4	2	5	2	$34 \times 2,5 = 85$
Μέσος Όρος	-	-	-	-	-	-	-	-	-	-	$34 \times 2,5 = 85$



Εικόνα 6-7 Αναπαράσταση ραβδογράμματος που δείχνει τις μέσες βαθμολογίες για καθεμία από τις δέκα ερωτήσεις του SUS (System Usability Scale)

Τα αποτελέσματα του ερωτηματολογίου System Usability Scale (SUS) δείχνουν ότι το σύστημα παρουσιάζει υψηλή ευχρηστία, με συνολικό μέσο σκορ περίπου 85. Οι περισσότεροι χρήστες έδωσαν πολύ θετικές αξιολογήσεις, ειδικά στις ερωτήσεις που αφορούν την ευκολία χρήσης (E3), την ενσωμάτωση λειτουργιών (E5) και την ταχύτητα εκμάθησης (E7). Οι αρνητικές ερωτήσεις, όπως η E2 (περίπλοκο σύστημα) και η E6 (ασυνέπεια), είχαν πολύ χαμηλές τιμές, υποδεικνύοντας ότι οι χρήστες δεν αντιμετώπισαν σοβαρά προβλήματα. Η ποσοτική ανάλυση δείχνει ότι το σύστημα είναι φιλικό προς τον χρήστη και κατάλληλο για χρήστες διαφορετικών επιπέδων εμπειρίας. Οι υψηλές βαθμολογίες SUS επιβεβαιώνουν ότι η διεπαφή είναι σαφής και αποτελεσματική, με

ελάχιστα σημεία προς βελτίωση. Οι μικρές διαφορές μεταξύ έμπειρων και αρχάριων χρηστών δείχνουν ότι με λίγη εξοικείωση, το σύστημα μπορεί να χρησιμοποιηθεί εξίσου εύκολα από όλους.

6.3. Αξιολόγηση των Μοντέλων του Συστήματος

Σε αυτή την ενότητα γίνεται η αξιολόγηση της απόδοσης των τριών βασικών μοντέλων που χρησιμοποιήθηκαν στο υπάρχον σύστημα. Αρχικά χρησιμοποιήθηκε το μοντέλο `medium` του `Whisper`, που όπως έχει ήδη ειπωθεί είναι ένα σύστημα αναγνώρισης ομιλίας σχεδιασμένο να μετατρέπει την ομιλία σε κείμενο σε πολλές γλώσσες συμπεριλαμβανομένης και της ελληνικής. Η απόδοση του μοντέλου αξιολογείται κυρίως με τη χρήση του ποσοστού σφάλματος λέξης (`Word Error Rate`), το οποίο μετρά τον αριθμό των εισαγωγών, διαγραφών και των αντικαταστάσεων που απαιτούνται για την αντιστοίχιση της εξόδου του αλγορίθμου με το κείμενο αναφοράς, κανονικοποιημένο με βάση τον συνολικό αριθμό λέξεων. Για την ελληνική γλώσσα το παραπάνω μοντέλο έχει ποσοστό σφάλματος λέξης περίπου 15%, ανάλογα με την ποιότητα του ήχου και τη μεταβλητότητα του ομιλητή. Το ποσοστό αυτό δηλώνει ένα σχετικά υψηλό επίπεδο ακριβείας, καθώς όσο μικρότερο είναι το ποσοστό σφάλματος λέξης τόσο καλύτερο είναι το μοντέλο που αξιολογείται. Έπειτα, χρησιμοποιείται το μοντέλο `el_core_news_lg`, μέρος της βιβλιοθήκης `SpaCy` και έχει σχεδιαστεί για την επεξεργασία ελληνικών κειμένων ώστε να αναγνωρίζει ονομαστικές οντότητες. Το μοντέλο επιδεικνύει ακρίβεια 75,4%, ανάκληση 78,6% και `F1-Score` 77%. Οι τιμές αυτές υποδεικνύουν την ευρωστία του μοντέλου στον χειρισμό εργασιών αναγνώρισης οντοτήτων στην ελληνική γλώσσα. Τέλος, το μοντέλο `ResNet-50`, που υλοποιήθηκε στο πλαίσιο `MMAction2`, υπερέχει στην αναγνώριση ανθρώπινων ενεργειών σε βίντεο. Η αξιολόγησή του βασίζεται στην ακρίβεια της κορυφαίας κλάσης, η οποία μετρά το ποσοστό των δειγμάτων όπου η κορυφαία προβλεπόμενη κλάση ταιριάζει με την βασική αλήθεια. Το μοντέλο επιτυγχάνει ακρίβεια 75,1% στο σύνολο δεδομένων `Kinetics-400`, αναδεικνύοντας την αποτελεσματικότητά του στην ακριβή ταξινόμηση ενεργειών. Επιπλέον, η μετρική ακριβείας για τις πρώτες 5 κορυφαίες κλάσεις, δείχνει τη συχνότητα με την οποία η σωστή κλάση εμφανίζεται στις πέντε κορυφαίες προβλέψεις του μοντέλου, υποστηρίζει περαιτέρω την αξιοπιστία του μοντέλου σε σενάρια πολλαπλών κλάσεων. Το μοντέλο `ResNet-50`, αν και ελαφρώς λιγότερο ακριβές από πιο προηγμένες αρχιτεκτονικές, όπως η `SlowFast`, προσφέρει μια καλή ισορροπία μεταξύ ακριβείας και υπολογιστικής απόδοσης. Όλες οι παραπάνω μετρικές παρέχουν μια ολοκληρωμένη κατανόηση των δυνατοτήτων κάθε μοντέλου, καθιστώντας τα κατάλληλα για τα ολοκληρωμένα συστήματα που αναπτύχθηκαν στην παρούσα διατριβή. Στον Πίνακα 6-5 παρουσιάζονται συνοπτικά οι μετρικές του κάθε αλγορίθμου που χρησιμοποιήθηκε.

Πίνακας 6-5 Βασικές μετρικές απόδοσης των μοντέλων των αλγορίθμων του συστήματος

Μοντέλο	Αλγόριθμος	Δείκτης	Τιμή
Whisper (medium)	Μετατροπής ομιλίας σε κείμενο	Word Error Rate (WER)	~15%
SpaCy (el_core_news_lg)	Αναγνώρισης ονομαστικών οντοτήτων	Ακρίβεια	75,4%
		Ανάκληση	78,6%
		F1-Score	77%
ResNet-50 (MMAction2)	Αναγνώριση ανθρώπινων ενεργειών σε βίντεο	Top-1 Accuracy (Κορυφαία κλάση)	75,1%

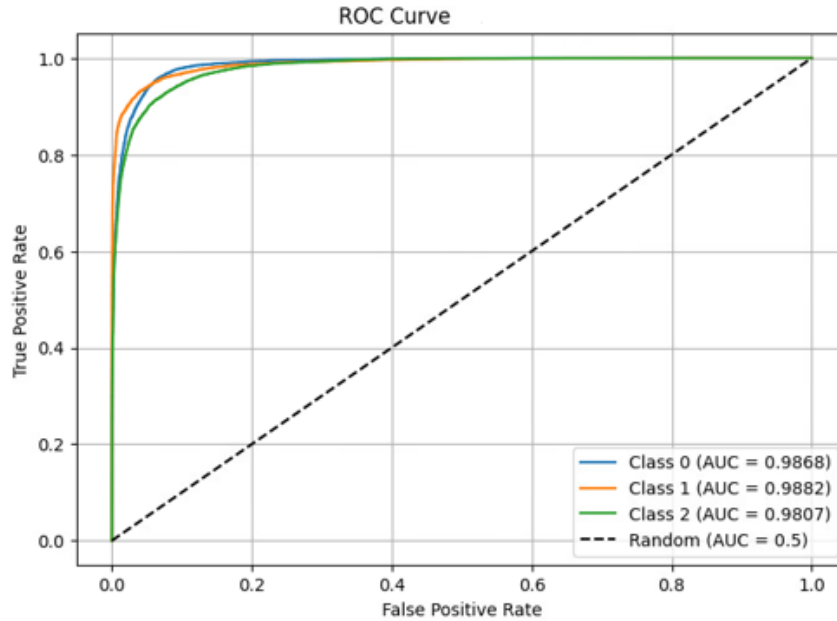
6.3.1. Αξιολόγηση του Αλγορίθμου Ταξινόμησης Ήχου

Για την αξιολόγηση των εκπαιδευσεων χρησιμοποιήθηκαν μετρικές αξιολόγησης όπως η accuracy, precision, recall και η f1-score για κάθε διαμέριση της επικύρωσης διασταύρωσης. Η ακρίβεια (accuracy) μετρά τη συνολική ορθότητα του μοντέλου υπολογίζοντας την αναλογία των σωστά ταξινομημένων περιπτώσεων προς το σύνολο των περιπτώσεων. Η ακρίβεια (precision) ποσοτικοποιεί το ποσοστό των σωστά προβλεπόμενων θετικών περιπτώσεων επί του συνόλου των προβλεπόμενων θετικών περιπτώσεων, ενώ η ανάκληση (recall) αντιπροσωπεύει το ποσοστό των σωστών θετικών περιπτώσεων που αναγνωρίστηκαν σωστά. Το F1-score είναι ο αρμονικός μέσος όρος της precision και της recall, παρέχοντας ένα ισορροπημένο μέτρο, ιδίως όταν πρόκειται για μη ισορροπημένα σύνολα δεδομένων. Οι παραπάνω μετρικές προσφέρουν συλλογικά μια ολοκληρωμένη αξιολόγηση της απόδοσης ενός μοντέλου. Οι μετρικές υπολογίζονται ξεχωριστά για κάθε διαμέριση, ενώ η συνολική ακρίβεια προκύπτει ως ο μέσος όρος όλων των διαμερίσεων. Στα τελικά αποτελέσματα, εκτός από την αναφορά της μέσης ακρίβειας (mean accuracy), παρουσιάζονται επίσης οι μετρικές από τη τελευταία διαμέριση της εκπαίδευσης, ώστε να δοθεί μια αντιπροσωπευτική εικόνα της απόδοσης του μοντέλου. Επιπλέον, πραγματοποιήθηκε ανάλυση μέσω των καμπυλών ROC-AUC (Receiver Operating Characteristic - Area Under the Curve), με τη χρήση ενός σετ δεδομένων ανεξάρτητο από αυτό της εκπαίδευσης, για την οπτική αξιολόγηση της ικανότητας του μοντέλου να διαχωρίζει σωστά τις θετικές και αρνητικές κατηγορίες. Η καμπύλη ROC-AUC αποτελεί γραφική απεικόνιση της απόδοσης ενός μοντέλου, βασισμένη στον υπολογισμό των τιμών του TPR (True Positive Rate) και του FPR (False Positive Rate). Ο δείκτης AUC (Area Under the Curve) χρησιμοποιείται για την εκτίμηση της συνολικής ικανότητας του μοντέλου να διαχωρίζει μεταξύ θετικών και αρνητικών κατηγοριών.

Αρχικά πραγματοποιήθηκαν πειράματα χρησιμοποιώντας μπλοκ μεγέθους 512 δειγμάτων και μήκος σήματος 0,75 δευτερολέπτων, εφαρμόζοντας διαφορετικές οικογένειες κυματιδίων (Daubechies, Symlet, Coiflet και Biorthogonal) σε διαφορετικά επίπεδα αποσύνθεσης. Για κάθε οικογένεια κυματιδίων υπολογίζεται και η καμπύλη ROC-AUC μόνο για το επίπεδο αποσύνθεσης με την μεγαλύτερη ακρίβεια. Πρώτα χρησιμοποιήθηκε το κυματίδιο Daubechies-4 (Db4). Όπως φαίνεται και στον Πίνακα 6-6, δοκιμάστηκε σε 3 διαφορετικά επίπεδα και είχε ισχυρή απόδοση ταξινόμησης με το υψηλότερο μέσο accuracy στα 91% στο επίπεδο 6, ενώ όπως φαίνεται και στην Εικόνα 6-8 το μοντέλο έχει πολύ καλή ικανότητα διαχωρισμού μεταξύ θετικών και αρνητικών κατηγοριών.

Πίνακας 6-6 Πίνακας αποτελεσμάτων για το κυματίδιο Daubechies-4 σε 3 διαφορετικά επίπεδα αποσύνθεσης με μήκος σήματος 0,75 δευτερόλεπτα.

Επίπεδο	Κλάση	Precision	Recall	F1-score	Mean Accuracy
4	Μουσική	0,9177	0,8922	0,9047	89,19 ± 0,5%
	Ομιλία	0,9035	0,9059	0,9047	
	Άλλο	0,8657	0,8901	0,8777	
5	Μουσική	0,9269	0,9214	0,9242	91,03 ± 0,24%
	Ομιλία	0,9395	0,8907	0,9144	
	Άλλο	0,8627	0,9140	0,8876	
6	Μουσική	0,9332	0,9154	0,9243	91,36 ± 0,48%
	Ομιλία	0,9203	0,9240	0,9221	
	Άλλο	0,8848	0,8999	0,8923	

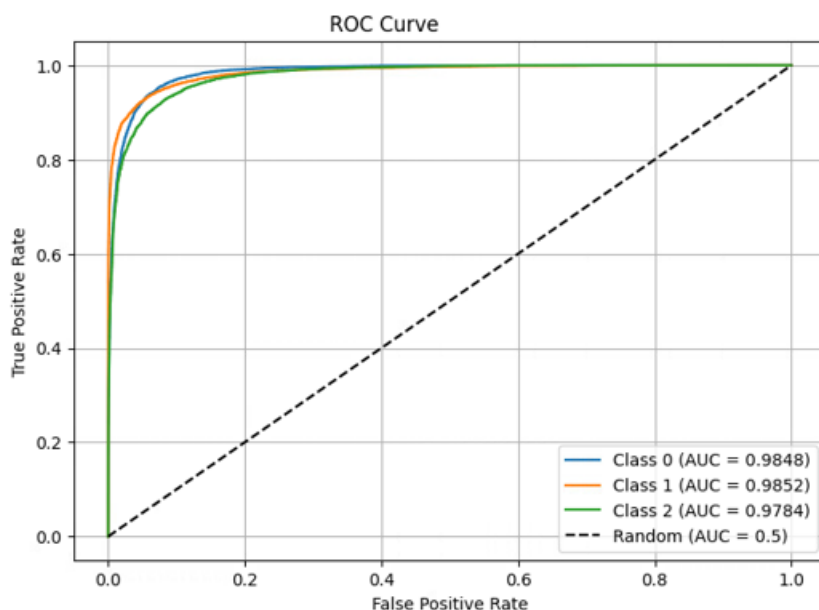


Εικόνα 6-8 Καμπύλη ROC-AUC για το κυματίδιο Daubechies-4 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 0,75 δευτερόλεπτα.

Έπειτα χρησιμοποιήθηκε το κυματίδιο Symlet-4 (Sym4) που παρουσίασε παρόμοιες επιδόσεις με το Daubechies-4, με μεγαλύτερο accuracy στο 91% στο επίπεδο 6 όπως φαίνεται και στον Πίνακα 6-7, ενώ στην Εικόνα 6-9 φαίνεται η καλή επίδοση όσο αφορά τη καμπύλη roc.

Πίνακας 6-7 Πίνακας αποτελεσμάτων για το κυματίδιο Symlet-4 σε 3 διαφορετικά επίπεδα αποσύνθεσης με μήκος σήματος 0,75 δευτερόλεπτα.

Επίπεδο	Κλάση	Precision	Recall	F1-score	Mean Accuracy
4	Μουσική	0,8656	0,9531	0,9073	89,65 ± 0,07%
	Ομιλία	0,9440	0,8650	0,9028	
	Άλλο	0,8884	0,8640	0,8760	
5	Μουσική	0,9217	0,9256	0,9236	91,13 ± 0,49%
	Ομιλία	0,9417	0,8980	0,9193	
	Άλλο	0,8738	0,9109	0,8920	
6	Μουσική	0,8963	0,9518	0,9232	91,11 ± 0,44%
	Ομιλία	0,9283	0,9164	0,9223	
	Άλλο	0,9186	0,8672	0,8921	

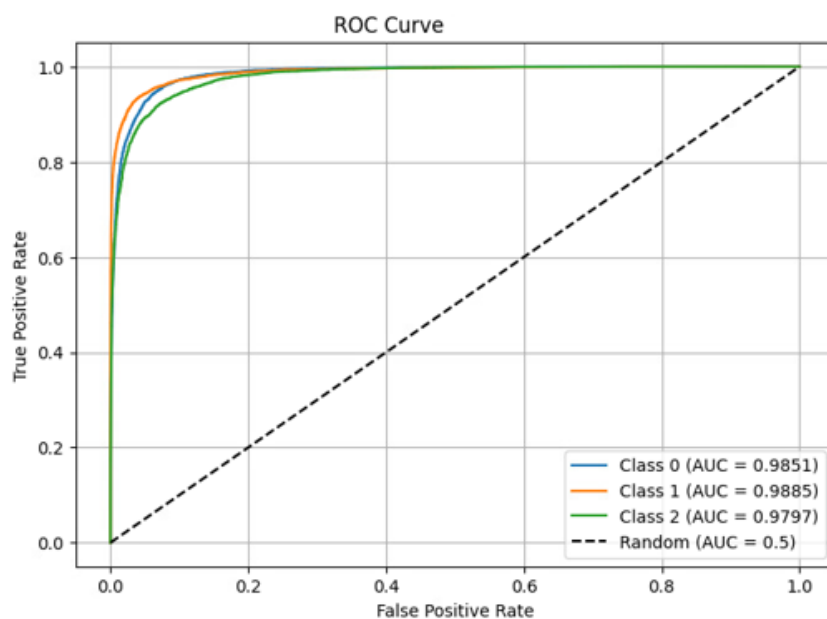


Εικόνα 6-9 Καμπύλη ROC-AUC για το κυματίδιο *Symlet-4* με επίπεδο αποσύνθεσης 6 και μήκος σήματος 0,75 δευτερόλεπτα.

Τέλος χρησιμοποιήθηκε το κυματίδιο *Coiflet-5* (*Coif5*) που πέτυχε ελαφρώς χαμηλότερη απόδοση από τα 2 προηγούμενα και το *Biorthogonal-4.4* (*Bior4.4*) που και αυτό είχε παρόμοια απόδοση. Τα αποτελέσματα των κυματιδίων αυτών παρουσιάζονται αντίστοιχα στους Πίνακες 6-8 και 6-9, ενώ αντίστοιχα παρουσιάζονται και στις εικόνες 6-10 και 6-11 οι καμπύλες roc.

Πίνακας 6-8 Πίνακας αποτελεσμάτων για το κυματίδιο *Coiflet-5* σε 3 διαφορετικά επίπεδα αποσύνθεσης με μήκος σήματος 0,75 δευτερόλεπτα.

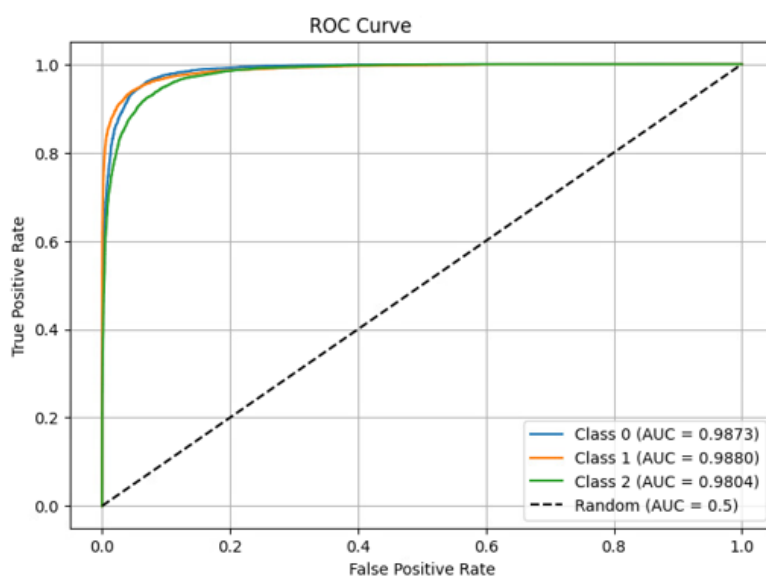
Επίπεδο	Κλάση	Precision	Recall	F1-score	Mean Accuracy
4	Μουσική	0,8758	0,9371	0,9054	89,70 ± 0,08%
	Ομιλία	0,9191	0,9004	0,9096	
	Άλλο	0,8975	0,8460	0,8710	
5	Μουσική	0,9221	0,8919	0,9068	90,05 ± 0,89%
	Ομιλία	0,9470	0,8582	0,9005	
	Άλλο	0,8135	0,9208	0,8638	
6	Μουσική	0,9207	0,9094	0,9150	90,94 ± 0,59%
	Ομιλία	0,9513	0,8827	0,9157	
	Άλλο	0,8429	0,9163	0,8781	



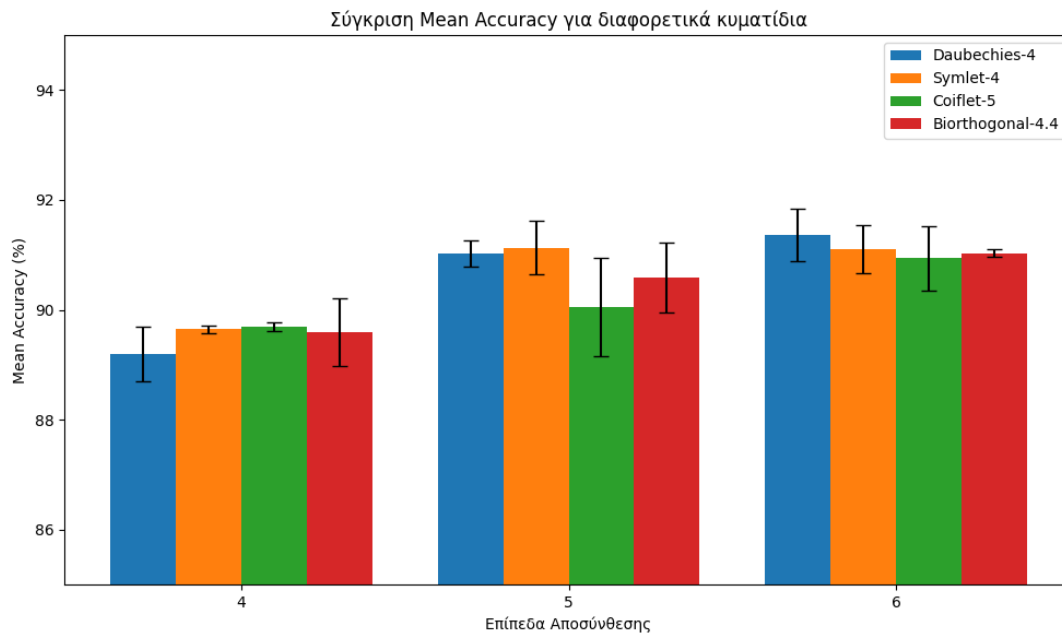
Εικόνα 6-10 Καμπύλη ROC-AUC για το κυματίδιο Coiflet-5 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 0,75 δευτερόλεπτα.

Πίνακας 6-9 Πίνακας αποτελεσμάτων για το κυματίδιο Biorthogonal-4.4 σε 3 διαφορετικά επίπεδα αποσύνθεσης με μήκος σήματος 0,75 δευτερόλεπτα.

Επίπεδο	Κλάση	Precision	Recall	F1-score	Mean Accuracy
4	Μουσική	0,8947	0,9354	0,9146	89,59 ± 0,61%
	Ομιλία	0,9359	0,8761	0,9050	
	Άλλο	0,8764	0,8888	0,8826	
5	Μουσική	0,9218	0,9139	0,9179	90,59 ± 0,63%
	Ομιλία	0,9298	0,9107	0,9201	
	Άλλο	0,8752	0,9019	0,8884	
6	Μουσική	0,9163	0,9231	0,9197	91,03 ± 0,07%
	Ομιλία	0,9276	0,9141	0,9208	
	Άλλο	0,8882	0,8940	0,8911	



Εικόνα 6-11 Καμπύλη ROC-AUC για το κυματίδιο Biorthogonal-4.4 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 0,75 δευτερόλεπτα.

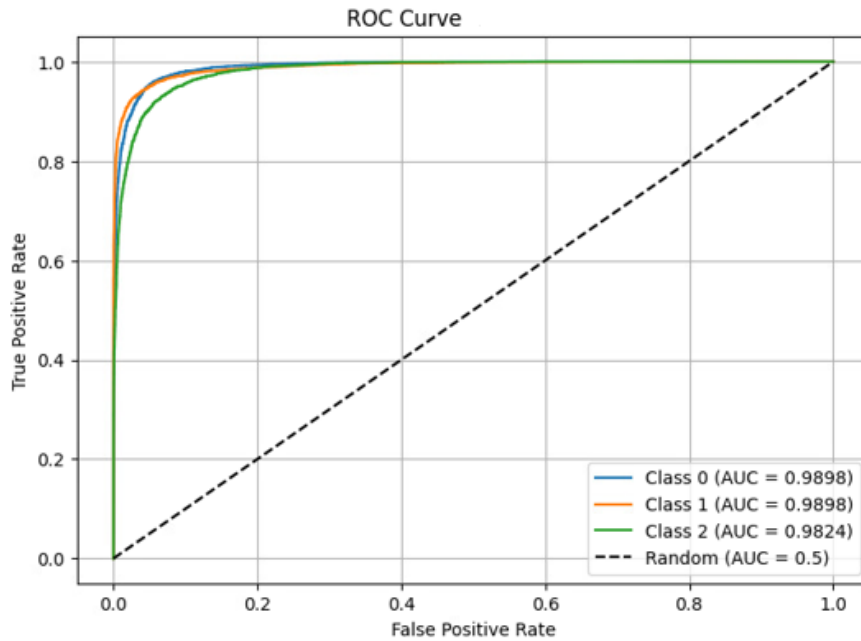


Εικόνα 6-12 Σύγκριση της μέσης ακρίβειας για τις διαφορετικές οικογένειες κυματιδίων σε επίπεδα αποσύνθεσης 4,5 και 6

Με βάση τα αποτελέσματα των προηγούμενων πειραμάτων επιλέχθηκε το κυματίδιο Daubechies-4 ως καλύτερη επιλογή για την εξαγωγή δεδομένων με ελάχιστες διαφορές με τις άλλες οικογένειες κυματιδίων όπως φαίνεται και στην Εικόνα 6-12. Στη συνέχεια πραγματοποιήθηκαν εκ νέου πειράματα, με βάση το κυματίδιο Daubechies-4, με σκοπό να βρεθεί το βέλτιστο μήκος σήματος και του επιπέδου αποσύνθεσης. Για τον λόγο αυτό ακολουθεί μια σειρά πειραμάτων με μήκος σήματος 1, 2 και 3 δευτερόλεπτων για 3 επίπεδα αποσύνθεσης, τα αποτελέσματα των οποίων παρουσιάζονται αντίστοιχα στους Πίνακες 6-10, 6-11, 6-12. Επιπλέον παρουσιάζονται στις εικόνες 6-13, 6-14 και 6-15 οι καμπύλες roc μόνο για το 6 επίπεδο αποσύνθεσης για τα πειράματα με μήκος σήματος 1,2 και 3 δευτερολέπτων αντίστοιχα.

Πίνακας 6-10 Πίνακας αποτελεσμάτων για το κυματίδιο Daubechies-4 σε 3 διαφορετικά επίπεδα αποσύνθεσης με μήκος σήματος 1 δευτερολέπτου.

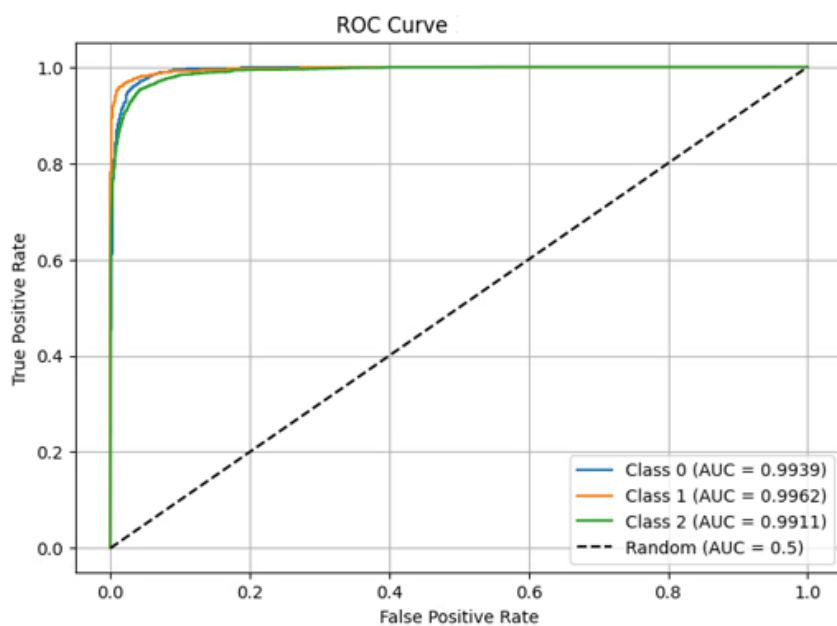
Επίπεδο	Κλάση	Precision	Recall	F1-score	Mean Accuracy
4	Μουσική	0,9293	0,9081	0,9186	91 ± 0,72%
	Ομιλία	0,9671	0,8536	0,9068	
	Άλλο	0,8240	0,9436	0,8797	
5	Μουσική	0,9266	0,9081	0,9173	91,41 ± 0,74%
	Ομιλία	0,9675	0,8784	0,9208	
	Άλλο	0,8369	0,9341	0,8828	
6	Μουσική	0,9313	0,9331	0,9322	91,61 ± 0,42%
	Ομιλία	0,9672	0,8961	0,9303	
	Άλλο	0,8678	0,9310	0,8983	



Εικόνα 6-13 Καμπύλη ROC-AUC για το κυματίδιο Daubechies-4 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 1 δευτερολέπτου.

Πίνακας 6-11 Πίνακας αποτελεσμάτων για το κυματίδιο Daubechies-4 σε 3 διαφορετικά επίπεδα αποσύνθεσης με μήκος σήματος 2 δευτερολέπτων.

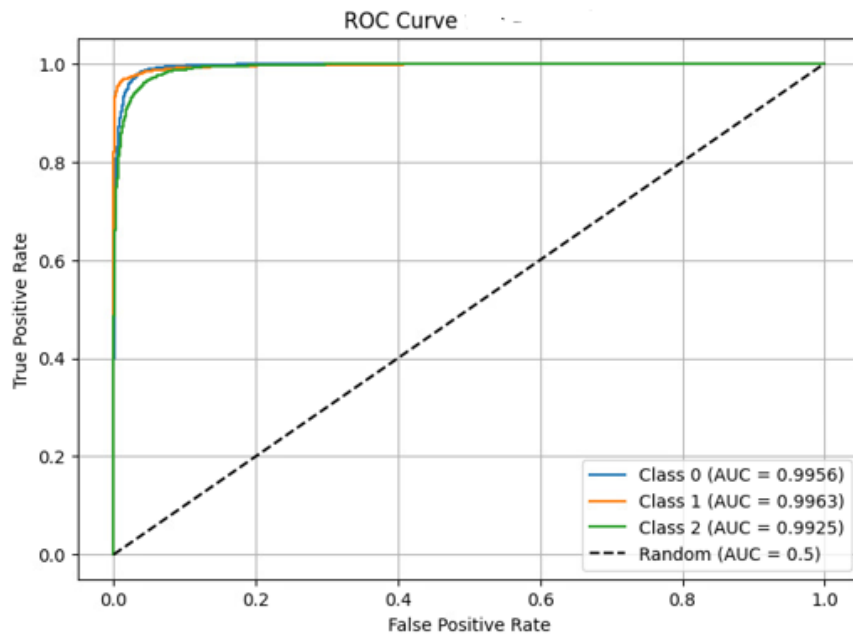
Επίπεδο	Κλάση	Precision	Recall	F1-score	Mean Accuracy
4	Μουσική	0,9055	0,9701	0,9367	93,55 ± 0,15%
	Ομιλία	0,9479	0,9607	0,9543	
	Άλλο	0,9628	0,8731	0,9158	
5	Μουσική	0,9372	0,9676	0,9522	94,66 ± 0,34%
	Ομιλία	0,9758	0,9479	0,9616	
	Άλλο	0,9391	0,9326	0,9358	
6	Μουσική	0,9485	0,9482	0,9483	93,3 ± 0,81%
	Ομιλία	0,9571	0,9738	0,9654	
	Άλλο	0,9384	0,9219	0,9301	



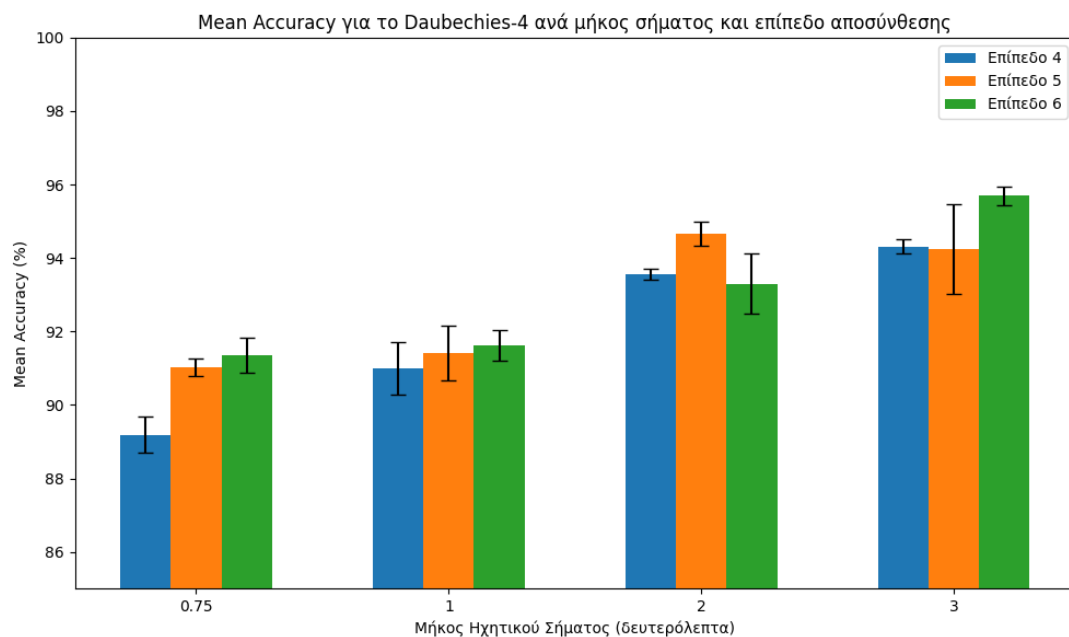
Εικόνα 6-14 Καμπύλη ROC-AUC για το κυματίδιο Daubechies-4 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 2 δευτερολέπτων.

Πίνακας 6-12 Πίνακας αποτελεσμάτων για το κυματίδιο Daubechies-4 σε 3 διαφορετικά επίπεδα αποσύνθεσης με μήκος σήματος 3 δευτερολέπτων

Επίπεδο	Κλάση	Precision	Recall	F1-score	Mean Accuracy
4	Μουσική	0,9523	0,9426	0,9474	94,31 ± 0,2%
	Ομιλία	0,9408	0,9654	0,9529	
	Άλλο	0,9355	0,9211	0,9282	
5	Μουσική	0,9383	0,9500	0,9441	94,24 ± 1,22%
	Ομιλία	0,9682	0,9629	0,9655	
	Άλλο	0,9339	0,9262	0,9301	
6	Μουσική	0,9586	0,9657	0,9622	95,7 ± 0,25%
	Ομιλία	0,9728	0,9709	0,9719	
	Άλλο	0,9481	0,9422	0,9451	



Εικόνα 6-15 Καμπύλη ROC-AUC για το κυματίδιο Daubechies-4 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 3 δευτερολέπτων.



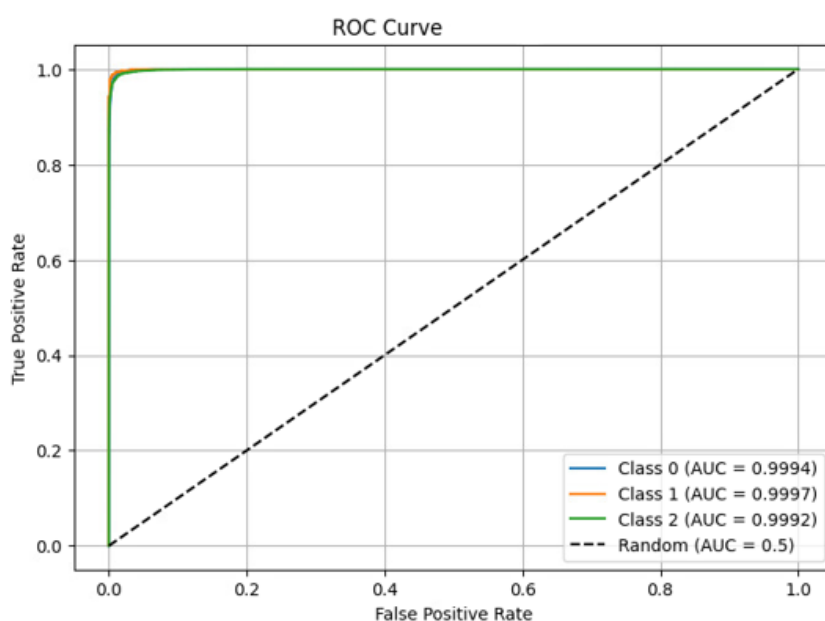
Εικόνα 6-16 Σύγκριση της μέσης ακρίβειας για το κυματίδιο Daubechies-4 για διαφορετικά μήκη σήματος και επιπέδων αποσύνθεσης

Τα αποτελέσματα, όπως φαίνονται συνοπτικά και στην Εικόνα 6-16, αποδεικνύουν ότι ο βέλτιστος συνδυασμός για την εξαγωγή χαρακτηριστικών είναι το κυματίδιο Daubechies-4 με μήκος σήματος 3 δευτερολέπτων και επίπεδο αποσύνθεσης 6, επιτυγχάνοντας μέσο accuracy 95,7%. Στη συνέχεια, εφαρμόστηκαν τεχνικές επαύξησης δεδομένων, όπως περιγράφεται στην ενότητα 5.8.1, για την περαιτέρω ενίσχυση της ικανότητας του μοντέλου να γενικεύει σε δεδομένα που δεν έχει εκπαιδευτεί. Το τελικό μοντέλο, που εκπαιδεύτηκε στο επαυξημένο σύνολο δεδομένων χρησιμοποιώντας το

κυματίδιο Db4 στα 3 δευτερόλεπτα και επίπεδο αποσύνθεσης 6, επέδειξε σημαντική βελτίωση στην απόδοση ταξινόμησης, επιτυγχάνοντας μέσο accuracy 98,28%, με υψηλό precision, recall και F1-score σε όλες τις κλάσεις. Τα αποτελέσματα αυτά, όπως φαίνονται στον Πίνακα 6-13 και στην Εικόνα 6-17, επιβεβαιώνουν την αποτελεσματικότητα της επιλεγμένης προσέγγισης και προ επεξεργασίας στην ακριβή ταξινόμηση διαφορετικών ηχητικών σημάτων, αποδεικνύοντας ότι τα φασματογραφήματα με βάση τα κυματίδια είναι αποτελεσματικά για εργασίες ταξινόμησης ήχου.

Πίνακας 6-13 Πίνακας αποτελεσμάτων τελικού μοντέλου για το κυματίδιο Daubechies-4 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 3 δευτερολέπτων στο σύνολο των δεδομένων μετά τη χρήση τεχνικών επαύξησης.

Κλάση	Precision	Recall	F1-score	Mean Accuracy
Μουσική	0,9839	0,9879	0,9859	98,28 ± 0,38%
Ομιλία	0,9911	0,9896	0,9903	
Άλλο	0,9827	0,9798	0,9812	



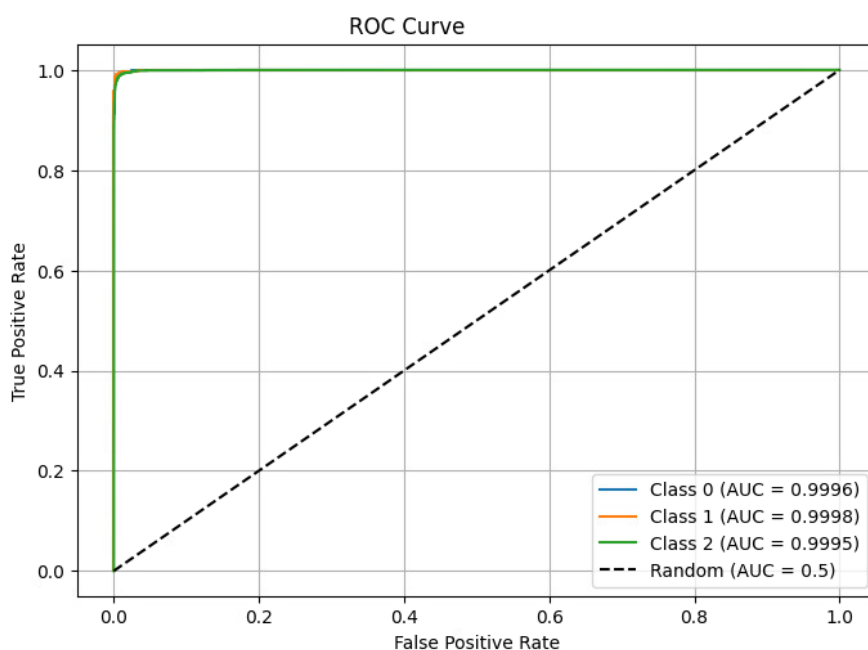
Εικόνα 6-17 Καμπύλη ROC-AUC για το κυματίδιο Daubechies-4 με επίπεδο αποσύνθεσης 6 και μήκος σήματος 3 δευτερολέπτων στο σύνολο των δεδομένων μετά τη χρήση τεχνικών επαύξησης.

Μετά τον προσδιορισμό των βέλτιστων παραμέτρων (οικογένεια κυματιδίων, μήκος ήχου και υπερπαραμέτροι), πραγματοποιήθηκε μια τελευταία εκπαίδευση, με σκοπό την δημιουργία του τελικού μοντέλου που θα χρησιμοποιηθεί στην εφαρμογή. Συγκεκριμένα, έγινε διαχωρισμός του συνόλου των επαυξημένων δεδομένων σε ένα σύνολο εκπαίδευσης/επικύρωσης 80% και ένα σύνολο δοκιμής 20%, και στη συνέχεια το 10% του τμήματος εκπαίδευσης για επικύρωση. Τα αποτελέσματα της τελευταίας εκπαίδευσης του τελικού μοντέλου ταξινόμησης παρουσιάζονται στον Πίνακα 6-14, η καμπύλη Roc στην

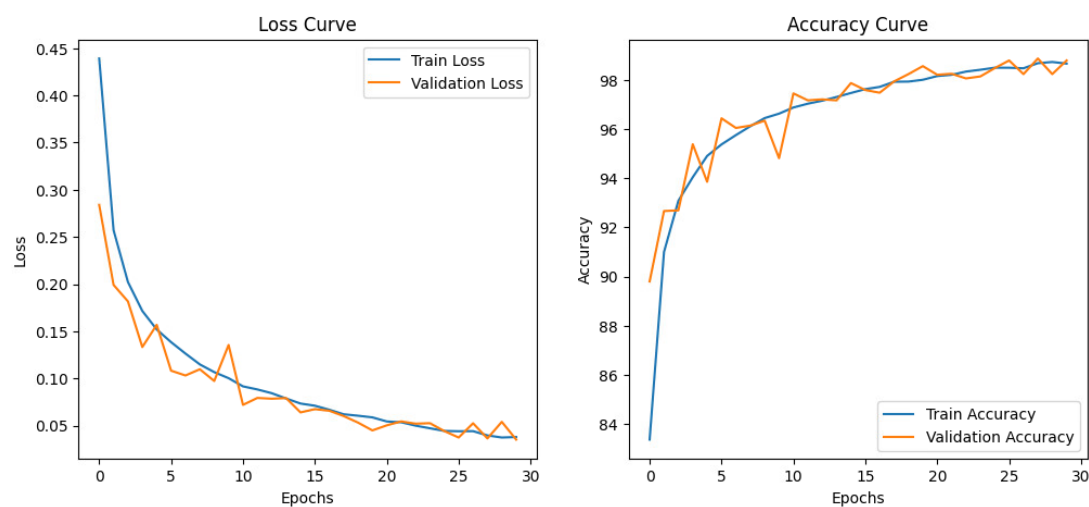
Εικόνα 6-18 ενώ οι καμπύλες απώλειας και ακρίβειας εκπαίδευσης και επικύρωσης στην Εικόνα 6-19.

Πίνακας 6-14 Πίνακας αποτελεσμάτων του τελικού μοντέλου ταξινόμησης του συστήματος

Κλάση	Precision	Recall	F1-score	Accuracy
Μουσική	0,9936	0,9792	0,9864	98,71%
Ομιλία	0,9877	0,9929	0,9903	
Άλλο	0,9794	0,9899	0,9847	



Εικόνα 6-18 Καμπύλη ROC-AUC του τελικού μοντέλου ταξινόμησης του συστήματος



Εικόνα 6-19 Οι καμπύλες απώλειας και ακρίβειας εκπαίδευσης και επικύρωσης του τελικού μοντέλου του συστήματος.

6.4. Σύνοψη Έκτου Κεφαλαίου

Στο έκτο κεφάλαιο παρουσιάστηκε η αξιολόγηση της διαδικτυακής πλατφόρμας, επικεντρώνοντας στην απόδοση του συστήματος, τη χρηστικότητα και την αποτελεσματικότητα των αλγορίθμων. Το σύστημα σχεδιάστηκε για τη διαχείριση και ανάλυση οπτικοακουστικού υλικού, ενσωματώνοντας τεχνολογίες όπως το RabbitMQ και το Celery για την αποδοτική εκτέλεση εργασιών. Η αξιολόγηση της απόδοσης περιλαμβάνει μετρήσεις της χρήσης CPU, GPU, μνήμης και δίσκου, επιβεβαιώνοντας τη σταθερότητα του συστήματος ακόμη και υπό υψηλό φόρτο εργασίας. Η αξιολόγηση για την ευχρηστία της εφαρμογής πραγματοποιήθηκε μέσω ημιδομημένων συζητήσεων με χρήστες διαφορετικών επιπέδων εμπειρίας, ενώ συμπλήρωσαν και ένα ερωτηματολόγιο SUS, το οποίο έδειξε υψηλή ευχρηστία του συστήματος. Τέλος, αναλύθηκε η απόδοση των αλγορίθμων του συστήματος, με έμφαση στον αλγόριθμο ταξινόμησης ήχου που αναπτύχθηκε στα πλαίσια της διπλωματικής. Δοκιμάστηκαν διαφορετικές οικογένειες κυματιδίων με διαφορετικά επίπεδα αποσύνθεσης σε διαφορετικά μήκη ηχητικών σημάτων. Το τελικό μοντέλο που επιλέχθηκε μετά από τα αποτελέσματα των διαφόρων δοκιμών πέτυχε ακρίβεια έως και 98%, χρησιμοποιώντας το κυματίδιο Daubechies-4 με έξι επίπεδα αποσύνθεσης με μήκος ηχητικού σήματος τριών δευτερολέπτων. Συνολικά, η αξιολόγηση κατέδειξε ότι το σύστημα είναι αποδοτικό, ευέλικτο και κατάλληλο για την επεξεργασία πολυμεσικού περιεχομένου.

Κεφάλαιο 7ο

7. Συμπεράσματα

Η παρούσα διατριβή πραγματοποιεί τον κύριο στόχο της, που ήταν η ανάπτυξη ενός διαδικτυακού περιβάλλοντος διαχείρισης, σημασιολογικής επεξεργασίας και αναγνώρισης οπτικοακουστικού περιεχομένου. Μέσω της ενσωμάτωσης εξελιγμένων τεχνολογιών όπως το Whisper για τη μετατροπή ομιλίας σε κείμενο, η αναγνώριση ονομαστικών οντοτήτων και η αναγνώριση ενεργειών και κατηγοριών ήχου, το σύστημα καταφέρνει να παρέχει μια ολοκληρωμένη λύση που υπερβαίνει τα παραδοσιακά μοντέλα που βασίζονται κυρίως στα μεταδεδομένα. Τα αποτελέσματα των δοκιμών του συστήματος ανέδειξαν ότι το σύστημα έχει υψηλή απόδοση και λειτουργικότητα, ενώ ταυτόχρονα απευθύνεται στο ευρύ κοινό, ανεξαρτήτου επιπέδου εμπειρίας. Παρά τις μικρές προκλήσεις που εντοπίστηκαν, όπως προβλήματα στην αρχική εξοικείωση ορισμένων αρχάριων χρηστών, περιορισμένοι πόροι στο σύστημα και καλύτερη εμφάνιση αποτελεσμάτων των αλγορίθμων, οι προτεινόμενες βελτιώσεις έχουν ήδη ενσωματωθεί, υποδηλώνοντας προοπτικές για περαιτέρω επέκταση της πλατφόρμας. Επίσης, στα πλαίσια της εργασίας αναπτύχθηκε και ενσωματώθηκε ένας αλγόριθμος ταξινόμησης ήχου που αξιοποιεί τα φασματογραφήματα με βάση τα κυματίδια. Για την εξαγωγή χαρακτηριστικών χρησιμοποιήθηκαν διάφορες οικογένειες κυματιδίων, οι οποίες δεν είχαν πολλές διαφορές μεταξύ τους όσον αφορά την ακρίβεια. Τα αποτελέσματα επιβεβαιώνουν την αποτελεσματικότητα της προσέγγισης για την ακριβή ταξινόμηση διαφόρων ηχητικών σημάτων, ενώ το τελικό μοντέλο επέδειξε υψηλή ακρίβεια ακόμα και σε δεδομένα διαφορετικά από αυτά της εκπαίδευσης.

7.1. Καινοτομία και Συνεισφορά της Εργασίας

Η παρούσα εργασία διαθέτει καινοτόμα χαρακτηριστικά τόσο στην ανάπτυξη της εφαρμογής όσο και στην ανάπτυξη του αλγορίθμου ταξινόμησης ήχου. Η εφαρμογή Ar.U.TV διαθέτει αναζήτηση σημασιολογικού περιεχομένου, ενσωματώνει χαρακτηριστικά με βάση την τεχνητή νοημοσύνη με τη χρήση των αλγορίθμων που έχουν συζητηθεί εκτενώς και διαθέτει αυτοματοποιημένες ροές εργασιών, με τη χρήση τεχνολογιών ασύγχρονης επικοινωνίας. Η αρχιτεκτονική του συστήματος αναπτύχθηκε με τέτοιο τρόπο ώστε να καθιστά εύκολη την προσθήκη, αφαίρεση ή τροποποίηση των υπάρχοντων αλγορίθμων χωρίς να επηρεάζεται το υπόλοιπο σύστημα. Έτσι το σύστημα θα έχει τη δυνατότητα συνεχώς να ανανεώνεται και να βελτιώνεται. Από την άλλη, αναπτύχθηκε μια νέα προσέγγιση για τη γενική ταξινόμηση ήχου, χρησιμοποιώντας φασματογραφήματα βασισμένα σε κυματίδια ως είσοδο σε ένα συνελκτικό νευρωνικό δίκτυο. Σε αντίθεση με τις παραδοσιακές αναπαραστάσεις ο μετασχηματισμός των κυματιδίων προσφέρει μια καλύτερη λύση από το μετασχηματισμό Fourier, δεδομένου της παράλληλης μετατόπισης και διαστολής του μητρικού κυματιδίου, επιτρέποντας μεταβλητά μεγέθη και θέσεις παραθύρων, καθώς και μια συνάρτηση που είναι πιο κατάλληλη για το δεδομένο σήμα. Τα κυματίδια είναι χρήσιμα για δυναμικά σήματα όπως η ομιλία και η μουσική. Συλλαμβάνουν πιο αποτελεσματικά την τοπική πληροφορία χρόνου-συχνότητας, η οποία ενισχύει τη διακριτική ικανότητα των εξαγόμενων χαρακτηριστικών. Αυτή η αναπαράσταση όχι μόνο ενισχύει την απόδοση ταξινόμησης αλλά διατηρεί και την υπολογιστική αποδοτικότητα. Η προτεινόμενη μέθοδος πέτυχε ακρίβεια ταξινόμησης σχεδόν 99% στις κατηγορίες μουσικής, ομιλίας και άλλων κατηγοριών. Η επίδοση αυτή υπερβαίνει σημαντικά εκείνη των σύγχρονων μεθόδων χρονικής ολοκλήρωσης των χαρακτηριστικών όπως αυτές που παρουσιάστηκαν από τους Vrysis et al. (2020), όπου τα καλύτερα μοντέλα έφτασαν σε ακρίβεια περίπου 88% στο ίδιο σύνολο δεδομένων. Εκτός από την πολύ καλή απόδοση ο αλγόριθμος ταξινόμησης ήχου προσφέρει σημαντικά οφέλη στο σύστημα διαχείρισης

οπτικοακουστικού υλικού. Στο πλαίσιο των συστημάτων διαχείρισης πολυμέσων, όπου η αποτελεσματική ανάκτηση και κατηγοριοποίηση ποικίλου ηχητικού περιεχομένου είναι ζωτικής σημασίας, η προσέγγισή της παρούσας διατριβής προσφέρει σημαντικά οφέλη. Η υψηλή ακρίβεια και η υπολογιστική αποδοτικότητα που επιτυγχάνονται από τη μέθοδό αυτή μπορούν να βελτιώσουν τη συνολική απόδοση των συστημάτων διαχείρισης πολυμέσων και να εξασφαλίσουν γρήγορη και αξιόπιστη ανάλυση ήχου στο παρόν σύστημα. Τέλος, η εφαρμογή του συστήματος αναμένεται να βελτιώσει σημαντικά τη διαχείριση και αξιοποίηση οπτικοακουστικού υλικού εντός του πανεπιστημίου, ενώ ταυτόχρονα προωθεί την καινοτομία και τη διεπιστημονική έρευνα, ανοίγοντας νέες προοπτικές στον τομέα των πολυμέσων.

7.2. Μελλοντικές Επεκτάσεις

Το σύστημα που αναπτύχθηκε περιλαμβάνει πολλούς τομείς στους οποίους υπάρχει χώρος για βελτίωση ή περαιτέρω έρευνα, παρά τα θετικά αποτελέσματα που επιτεύχθηκαν. Μελλοντικές εργασίες θα μπορούσαν να εστιάσουν στην δοκιμή του αλγορίθμου ταξινόμησης, με τα ίδια ή και με πρόσθετα χαρακτηριστικά αν κρίνεται απαραίτητο, σε κατηγορίες όπως αναγνώριση είδους μουσικής ή κατηγοριοποίηση περιβαλλοντικών ήχων και πολλά άλλα. Όσον αφορά το σύστημα, μπορούν να προστεθούν επιπλέον αλγόριθμοι που θα μπορούν να ενισχύσουν τη σημασιολογική ανάλυση και την ανάκτηση πληροφοριών. Άλλες προοπτικές περιλαμβάνουν την ανάπτυξη ενός προσαρμοστικού περιβάλλοντος χρήστη με συστήματα συστάσεων, που θα βελτιώσουν την εμπειρία χρήσης, καθώς και την ενσωμάτωση τεχνολογιών real-time για την άμεση επεξεργασία και ανάλυση περιεχομένου. Τέλος, η πιο εύκολη επέκταση της εφαρμογής είναι οι αλγόριθμοι του συστήματος να εκτελούνται όλοι παράλληλα, το οποίο μπορεί εύκολα να πραγματοποιηθεί λόγω του τρόπου που αναπτύχθηκε το σύστημα και κυρίως λόγω του Celery. Βέβαια, αυτό απαιτεί καλύτερους πόρους στο σύστημα. Οι παραπάνω εξελίξεις θα συμβάλλουν στην περαιτέρω ενίσχυση της λειτουργικότητας και της αποτελεσματικότητας των συστημάτων διαχείρισης οπτικοακουστικού περιεχομένου ανοίγοντας νέες προοπτικές για την εφαρμογή προηγμένων τεχνολογιών τεχνητής νοημοσύνης στον τομέα αυτό.

Βιβλιογραφικές Αναφορές

Βιβλιογραφικές αναφορές

- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford university press.
- Blog Tool, Publishing Platform, and CMS. (2024, November 21). WordPress.Org. <https://wordpress.org/>
- Boiko, B. (2005). *Content management bible*. John Wiley & Sons.
- Boudreau, M., Loch, K., Robey, D., & Straub, D. (1998). Going Global: Using Information Technology to Advance the Competitiveness Of the Virtual Transnational Organization. *Academy of Management Perspectives*, 12. <https://doi.org/10.5465/AME.1998.1334008>
- Branch, R. (2010). Instructional design: The ADDIE approach. In *Instructional Design: The ADDIE Approach* (p. 203). <https://doi.org/10.1007/978-0-387-09506-6>
- Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). Taylor & Francis.
- Celery—Distributed Task Queue—Celery 5.4.0 documentation. (n.d.). Retrieved December 4, 2024, from <https://docs.celeryq.dev/en/stable/index.html>
- Dimoulas, C. A., & Symeonidis, A. L. (2015). Syncing Shared Multimedia through Audiovisual Bimodal Segmentation. *IEEE MultiMedia*, 22(3), 26–42. <https://doi.org/10.1109/MMUL.2015.33>
- Dimoulas, C., Kalliris, G., Papanikolaou, G., & Kalampakas, A. (2007). Long-term signal detection, segmentation and summarization using wavelets and fractal dimension: A bioacoustics application in gastrointestinal-motility monitoring. *Wavelet-Based Algorithms for Medical Problems*, 37(4), 438–462. <https://doi.org/10.1016/j.compbiomed.2006.08.013>
- Dimoulas, C., & Veglis, A. (2023). Theory and Applications of Web 3.0 in the Media Sector. *Future Internet*, 15, 165. <https://doi.org/10.3390/fi15050165>
- Dimoulas, C., Veglis, A., & Kalliris, G. (2015). Audiovisual Hypermedia in the Semantic Web. In *Encyclopedia of Information Science and Technology, Third Edition* (pp. 7594–7604). IGI Global Scientific Publishing. <https://doi.org/10.4018/978-1-4666-5888-2.ch748>
- Dobbelaere, P., & Sheykh Esmaili, K. (2017). Kafka versus RabbitMQ: A comparative study of two industry reference publish/subscribe implementations: *Industry Paper* (p. 238). <https://doi.org/10.1145/3093742.3093908>
- Du, K.-L., & Swamy, M. N. s. (2013). Neural Networks and Statistical Learning. In *Neural Networks and Statistical Learning*. <https://doi.org/10.1007/978-1-4471-5571-3>
- François-Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., & Pineau, J. (2018). An Introduction to Deep Reinforcement Learning. *Foundations and Trends® in Machine Learning*, 11(3–4), 219–354. <https://doi.org/10.1561/22000000071>

- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. The MIT Press.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*, Waltham: Morgan Kaufmann Publishers.
- Harshjeet, C. Gogoi, N. Snehalatha, & S. Amudha. (2024). Enhancing Visual Creativity with Neural Style Transfer using Celery Backend Architecture. *2024 3rd International Conference on Applied Artificial Intelligence and Computing (ICAAIC)*, 966–974. <https://doi.org/10.1109/ICAAIC60222.2024.10575471>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition (Springer Series in Statistics)*.
- Heitmann, J. (1999). Content Management Systems for television production. *EBU Technical Review*, 24–34. https://tech.ebu.ch/docs/techreview/trev_280-heitmann.pdf
- Herath, S., Harandi, M., & Porikli, F. (2017). Going deeper into action recognition: A survey. *Regularization Techniques for High-Dimensional Data Analysis*, 60, 4–21. <https://doi.org/10.1016/j.imavis.2017.01.010>
- Hadjileontiadis, J., L., & M. Panas, S. (1998). A wavelet-based reduction of heart sound noise from lung sounds. *International Journal of Medical Informatics*, 52(1), 183–190. [https://doi.org/10.1016/S1386-5056\(98\)00137-3](https://doi.org/10.1016/S1386-5056(98)00137-3)
- Hong, X. J., Sik Yang, H., & Kim Y. H. (2018). Performance Analysis of RESTful API and RabbitMQ for Microservice Web Application. *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, 257–259. <https://doi.org/10.1109/ICTC.2018.8539409>
- Honnibal Matthew, Ines Montani, Sofie Van Landeghem, & Andriane Boynd. (2020). spaCy: Industrial-strength Natural Language Processing in Python. <https://doi.org/10.5281/zenodo.1212303>
- Ionescu, V. M. (2015). The analysis of the performance of RabbitMQ and ActiveMQ. *2015 14th RoEduNet International Conference - Networking in Education and Research (RoEduNet NER)*, 132–137.
- Jabeen, S., Li, X., Amin, M., Bourahla, O., Li, S., & Jabbar, A. (2022). A Review on Methods and Applications in Multimodal Deep Learning. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 19. <https://doi.org/10.1145/3545572>
- Jehangir, B., Radhakrishnan, S., & Agarwal, R. (2023). A survey on Named Entity Recognition—Datasets, tools, and methodologies. *Natural Language Processing Journal*, 3, 100017. <https://doi.org/10.1016/j.nlp.2023.100017>
- Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science (New York, N.Y.)*, 349(6245), 255–260. <https://doi.org/10.1126/science.aaa8415>
- Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Neural Information Processing Systems*, 25. <https://doi.org/10.1145/3065386>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>

- LeCun Y., B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, & L. D. Jackel. (1989). Backpropagation Applied to Handwritten Zip Code Recognition. *Neural Computation*, 1(4), 541–551. <https://doi.org/10.1162/neco.1989.1.4.541>
- Martinez-Caro, J.-M., Aledo-Hernández, A.-J., Guillen-Perez, A., Sanchez-Iborra, R., & Cano, M.-D. (2018). A Comparative Study of Web Content Management Systems. *Information*, 9, 27. <https://doi.org/10.3390/info9020027>
- Mallat S. G. (1989). A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7), 674–693. <https://doi.org/10.1109/34.192463>
- MMAction2 Contributors. (2020). *OpenMMLab's Next Generation Video Understanding Toolbox and Benchmark* [Computer software]. [MMAction2 Contributors](https://github.com/open-mmlab/mmdetection)
- Montoya-Weiss, M. M., Massey, A. P., & Song, M. (2001). Getting It Together: Temporal Coordination and Conflict Management in Global Virtual Teams. *The Academy of Management Journal*, 44(6), 1251–1262. JSTOR. <https://doi.org/10.2307/3069399>
- Naphade M. R. & Huang T. S. (2002). Extracting semantics from audio-visual content: The final frontier in multimedia retrieval. *IEEE Transactions on Neural Networks*, 13(4), 793–810. <https://doi.org/10.1109/TNN.2002.1021881>
- Nassif, A. B., Shahin, I., Attili, I., Azzeh, M. & Shaalan K. (2019). Speech Recognition Using Deep Neural Networks: A Systematic Review. *IEEE Access*, 7, 19143–19165. <https://doi.org/10.1109/ACCESS.2019.2896880>
- Patel, S., Rathod, V., & Prajapati, J. (2011). Performance Analysis of Content Management Systems Joomla, Drupal and WordPress. *International Journal of Computer Applications*, 21, 39–43. <https://doi.org/10.5120/2496-3373>
- Patel, S., Rathod, V. R., & Parikh, S. (2011). Joomla, Drupal and WordPress—A statistical comparison of open source CMS (p. 187). <https://doi.org/10.1109/TISC.2011.6169111>
- RabbitMQ: One broker to queue them all | RabbitMQ. (n.d.). Retrieved December 4, 2024, from <https://www.rabbitmq.com/>
- Radford, A., Kim, J., Xu, T., Brockman, G., McLeavey, C., & Sutskever, I. (2022). Robust Speech Recognition via Large-Scale Weak Supervision. <https://doi.org/10.48550/arXiv.2212.04356>
- Royce, W. W. (1970). Managing the development of large software systems: Concepts and techniques. *Proceedings of the 9th International Conference on Software Engineering*, 328–338.
- Sarker, I. H., Kayes, A. S. M., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity data science: An overview from machine learning perspective. *Journal of Big Data*, 7(1), 41. <https://doi.org/10.1186/s40537-020-00318-5>
- Schinas, M., Galopoulos, P., & Papadopoulos, S. (2023). MAAM: Media Asset Annotation and Management. *Proceedings of the 2023 ACM International Conference on Multimedia Retrieval*, 659–663. <https://doi.org/10.1145/3591106.3592232>
- Sharma, G., Umapathy, K., & Krishnan, S. (2020). Trends in audio signal feature extraction methods. *Applied Acoustics*, 158, 107020. <https://doi.org/10.1016/j.apacoust.2019.107020>

- Sharvari T. & Sowmya K. (2019). *A study on Modern Messaging Systems- Kafka, RabbitMQ and NATS Streaming*. <https://doi.org/10.48550/arXiv.1912.03715>
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*, 2nd ed. (pp. xxii, 526). The MIT Press.
- Taud, H., & Mas, J. (2018). *Multilayer Perceptron (MLP)* (pp. 451–455). https://doi.org/10.1007/978-3-319-60801-3_27
- Tsipas, N., Zapartas, P., Vrysis, L., Dimoulas, C., & Papanikolaou, G. (2015). *Augmenting Social Multimedia Semantic Interaction Through Audio-Enhanced Web-TV Services*. <https://doi.org/10.1145/2814895.2814907>
- Tzanetakis, G., Essl, G., & Cook, P. (2001). Audio Analysis using the Discrete Wavelet Transform. In *Proceedings of the Conference in Acoustics and Music Theory Applications* (p. 323).
- Vrysis, L., Tsipas, N., Thoidis, I., & Dimoulas, C. (2020). 1D/2D Deep CNNs vs. Temporal Feature Integration for General Audio Classification. *Journal of the Audio Engineering Society*, 68, 66–77. <https://doi.org/10.17743/jaes.2019.0058>
- Wagenpfeil, S., Engel, F., Kevitt, P. M., & Hemmje, M. (2021). AI-Based Semantic Multimedia Indexing and Retrieval for Social Media on Smartphones. *Information*, 12(1). <https://doi.org/10.3390/info12010043>
- Zaman, K., Sah, M., Direkoglu, C. & Unoki M. (2023). A Survey of Audio Classification Using Deep Learning. *IEEE Access*, 11, 106620–106649. <https://doi.org/10.1109/ACCESS.2023.3318015>
- Ζαπάρτας Παναγιώτης & Σαρτσένκο Ιουλία. (2012). *Σχεδιασμός και ανάπτυξη διαδικτυακού περιβάλλοντος διαχείρισης οπτικοακουστικού περιεχομένου*.
- Ομάδα εργασίας Εργαστηρίου Πολυμεσικής Παραγωγής. (2009). *Παραγωγή πολυμεσικού περιεχομένου και δημιουργία ιστοχώρου για τον τηλεοπτικό σταθμό του Αριστοτέλειου Πανεπιστημίου Θεσσαλονίκης*. Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης-Διαπανεπιστημιακό Διατμηματικό Πρόγραμμα Μεταπτυχιακών Σπουδών – Προηγμένα Συστήματα Υπολογιστών & Επικοινωνιών – Τεχνολογίες Πληροφορικής & Επικοινωνιών για τον ήχο και την εικόνα στην παραγωγή και την εκπαίδευση.
- Χατζηλεοντιάδης Λεόντιος. (2021). *The world of Wavelets Part I, Παραδόσεις στα πλαίσια το μαθήματος Προηγμένη Ψηφιακή Επεξεργασία Σημάτων Τμήμα Ηλεκτρολόγων Μηχανικών και Ηλεκτρονικών Υπολογιστών, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης*.