

Ατομική Διπλωματική Εργασία

**ΚΑΤΑΡΡΙΠΤΟΝΤΑΣ ΕΜΠΟΔΙΑ ΣΤΟΝ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟ: ΜΕΛΕΤΗ  
ΠΕΡΙΠΤΩΣΗΣ ΓΙΑ ΤΗΝ ΤΕΧΝΟΛΟΓΙΑ ΑΝΑΓΝΩΡΙΣΗΣ ΦΩΝΗΣ ΓΙΑ  
ΠΡΟΓΡΑΜΜΑΤΙΣΤΕΣ ΜΕ ΚΙΝΗΤΙΚΕΣ ΔΥΣΚΟΛΙΕΣ**

**Μάριος Φλουρέντζου**

**ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ**



**ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**

**Μάιος 2024**

# **ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ**

## **ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**

**Καταρρίπτοντας Εμπόδια στον Προγραμματισμό: Μελέτη Περίπτωσης  
για την Τεχνολογία Αναγνώρισης Φωνής για Προγραμματιστές με  
Κινητικές Δυσκολίες**

**Μάριος Φλουρέντζου**

Επιβλέπων Καθηγητής  
Δρ. Κωνσταντίνος Παττίχης

Η Ατομική Διπλωματική Εργασία υποβλήθηκε προς μερική εκπλήρωση των  
απαιτήσεων απόκτησης του πτυχίου Πληροφορικής του Τμήματος Πληροφορικής του  
Πανεπιστημίου Κύπρου

Μάιος 2024

## **Ευχαριστίες**

Θα ήθελα να εκφράσω τις βαθύτατες ευχαριστίες μου προς τον επιβλέποντα καθηγητή μου, Δρ. Κωνσταντίνο Παττίχη, για την αμέριστη υποστήριξη, την καθοδήγηση και τις πολύτιμες συμβουλές που μου προσέφερε καθ' όλη τη διάρκεια της εκπόνησης της παρούσας διπλωματικής εργασίας. Η συνεχής παρουσία του και η εμπιστοσύνη που μου έδειξε υπήρξαν καθοριστικά για την πραγματοποίηση αυτού του εγχειρήματος καθώς, και τον καθηγητή μου Δρ. Αργύρη Κωνσταντινίδη που μου έκανε το Μάθημα ΕΠΛ 435: Αλληλεπίδραση Ανθρώπου Υπολογιστή ο οποίος με βοήθησε σημαντικά να αποκομίσω καινούργιες γνώσεις οι οποίες με βοήθησαν και στην παρούσα διπλωματική εργασία.

Θα ήθελα να πω ένα τεράστιο ευχαριστώ σε όλους τους φίλους μου που εδώ και τέσσερα χρόνια είναι πάντα δίπλα μου σε ό,τι κι αν τους χρειαστώ και με στηρίζουν απεριόριστα στο καθετί, σε όλη την διάρκεια αυτής της Ακαδημαϊκής πορείας . Αυτή η διπλωματική εργασία φέρει γνώσεις και δεξιότητες που αποκόμισα σε όλο αυτό το διάστημα των σπουδών μου στο Πανεπιστήμιο Κύπρου οι οποίες με βοήθησαν να φέρω εις πέρας αυτή την αποστολή.

Τέλος, θα ήθελα να ευχαριστήσω θερμά την οικογένεια μου και την Ψυχολόγο μου Δρ. Ρεβέκκα Μουγή για όλη την ψυχολογική υποστήριξη και εμπύχωση που μου παρείχαν.

## Περίληψη

Η Διπλωματική μου εργασία βασίζεται στην φωνητική αναγνώριση με σκοπό να διευκολύνει τα άτομα που έχουν κάποιου είδους αναπηρία και δεν μπορούν να πληκτρολογήσουν με ευκολία στον υπολογιστή ή να αλληλοεπιδράσουν γρήγορα με αυτόν, κάνοντας χρήση της φωνής τους με σκοπό να εκτελέσουν τις διάφορες δραστηριότητες που έχουν γρήγορα αποδοτικά και αποτελεσματικά χρησιμοποιώντας ένα εργαλείο φωνητικής αναγνώρισης.

Έχει γίνει έρευνα και έχουμε βρει τρία εργαλεία που υποστηρίζουν φωνητική αναγνώριση και έχουμε καταγράψει τα χαρακτηριστικά για το κάθε εργαλείο σε πίνακα έτσι ώστε να αποφασίσουμε ποιο είναι το ιδανικό που θα μας διευκολύνει καλύτερα για να το αξιολογήσουμε. Τα εργαλεία αυτά είναι το Dragon Professional V16 [11] το οποίο καταλήξαμε να προμηθευτούμε βάση των χαρακτηριστικών του, το Speechlo [15] που είναι αντίστοιχο εργαλείο με παρόμοια χαρακτηριστικά με αυτά του Dragon Professional V16 και τον λογογράφο Office 365 [16] που παρέχεται μέσα στα προϊόντα Office 365 της Microsoft.

Μέσα από την έρευνά μας βρήκαμε δύο μεθόδους αξιολόγησης: το Word Error Rate (WER) [19] και τη μέθοδο Automatic Speech Recognition (ASR) [18]. Εφαρμόσαμε την πρώτη μέθοδο αξιολόγησης, τον υπολογισμό του ποσοστού σφάλματος λέξης (WER), κατά την εκτέλεση του Δεύτερου Σεναρίου Χρήσης. Αυτό το σενάριο αφορούσε τη συγγραφή κώδικα σε τρεις γλώσσες προγραμματισμού, C, Java, Python με στόχο την αξιολόγηση της αποδοτικότητας και της αξιοπιστίας του Dragon Professional V16. Τα ποσοστά του WER που προέκυψαν κατά την μεταγραφή του προγράμματος στις γλώσσες C, Java και Python κυμαίνονται στο εύρος 13-18%, 20-27% και 3-5% αντίστοιχα και αποδεικνύουν μια αρκετά καλή απόδοση του εργαλείου στην μεταγραφή κώδικα ιδιαίτερα στην γλώσσα Python. Το Πρώτο Σενάριο Χρήσης που σχεδιάσαμε αποσκοπούσε στη χρήση του εργαλείου για την εκτέλεση διάφορων δραστηριοτήτων στον υπολογιστή, καθώς και για την πλοήγηση σε διάφορα προγράμματα και ιστοσελίδες χρησιμοποιώντας τη φωνή ως μέσο για αλληλεπίδραση με τον υπολογιστή.

Μέσα από την παρούσα Διπλωματική Εργασία διευκολύνθηκα και εγώ ως άμεσα ενδιαφερόμενος σχετικά με αυτό το θέμα αφού αντιμετωπίζω δυσκολία στην Πληκτρολόγηση και γενικότερα στην αλληλεπίδραση με τον υπολογιστή στην καθημερινότητα μου κατά την διεκπεραίωση διάφορων δραστηριοτήτων.

## **Ακρόνυμα**

- WER – Word Error Rate
- ASR – Automatic Speech Recognition

## Πίνακας Περιεχομένων

Ευχαριστίες .....	2
Ακρόνυμα .....	4
<b>Κεφάλαιο 1 Εξερευνώντας την Τεχνολογία .....</b>	<b>7</b>
1.1 Εισαγωγή .....	7
1.2 Προσωπικές Εμπειρίες στον Τομέα της Τεχνολογίας .....	10
1.3 Σκοπός της Παρούσας Διπλωματικής Εργασίας.....	11
1.4 Παρόμοιες Μελέτες που έχουν Πραγματοποιηθεί στον Τομέα της Φωνητικής Αναγνώρισης .....	12
1.5 Σύνομη Περιγραφή Περιεχομένων Διπλωματικής Εργασίας .....	15
<b>Κεφάλαιο 2 Βασικές Τεχνολογίες Εργαλείων Φωνητικής Αναγνώρισης.....</b>	<b>16</b>
2.1 Μετατροπή Ηχογράφησης σε Κείμενο (Transcription) .....	17
2.2 Ορισμός και Σημασία.....	17
2.3 Τεχνικές Μεταγραφής .....	17
2.4 Εργαλεία και Τεχνολογίες.....	18
2.5 Προκλήσεις που Επέρχονται .....	18
2.6 Η Δύναμη του Transkriptor.....	18
2.6.1 Χαρακτηριστικά και Δυνατότητες .....	19
2.6.2 Προκλήσεις.....	19
2.7 Χαρακτηριστικά των διαθέσιμων λογισμικών και αξιολόγηση τους .....	19
2.7.1 Dragon Professional V16 .....	20
2.7.2 Speechelo .....	22
2.7.3 Λογογράφος Microsoft Office 365 (Speech-to-text) .....	25
2.7.4 Συγκεντρωτικός Πίνακας Εργαλείων .....	28
2.7.5 Σύγκριση των Εργαλείων Dragon Professional V16, Speechelo και Microsoft Office 365 (Speech-to-text).....	29
2.8 Μετρικές Αξιολόγησης των λογισμικών Φωνητικής Αναγνώρισης.....	30
2.8.1 Word Error Rate - WER .....	31
2.8.2 Κανονικοποίηση και Προεπεξεργασία των Μεταγραφών .....	33
2.8.3 Έξι βήματα για την εκτέλεση αξιολόγησης συστήματος ASR.....	36
<b>Κεφάλαιο 3: Πρώτο Σενάριο Χρήσης: Πλοήγηση με τη χρήση του Εργαλείου Dragon Professional V16 .....</b>	<b>38</b>
3.1 Αναβαθμίσεις από προηγούμενες εκδόσεις.....	39
3.2 Πρώτο Σενάριο Χρήσης: Πλοήγηση και διαχείριση αρχείων σε υπολογιστή με χρήση του Dragon Professional V16.....	40
3.2.1 Άνοιγμα/ Κλείσιμο Εφαρμογών .....	41
3.2.2 Στιγμιότυπα Οθονών από τις διάφορες εκτελέσεις .....	41

3.2.3 Πλοήγηση αρχείων .....	44
3.2.4 Στιγμιότυπα Οθωνών από την εκτέλεση των πιο πάνω βημάτων .....	45
3.2.5 Πλοήγηση στο διαδίκτυο.....	48
3.2.6 Αναζήτηση πτήσεων προς Ζυρίχη μέσω της ιστοσελίδας booking.com .....	48
3.2.7 Στιγμιότυπα Οθωνών από την εκτέλεση των πιο πάνω βημάτων. ....	50
<b>Κεφάλαιο 4: Δεύτερο Σενάριο Χρήσης- Δημιουργία προγράμματος στις γλώσσες προγραμματισμού C, Java και Python με χρήση του εργαλείου Dragon Professional V16 .....</b>	<b>56</b>
4.1 Μεθοδολογία του Σεναρίου Χρήσης.....	58
4.2 Πρωτότυπο Πρόγραμμα στην γλώσσα C.....	59
4.3 Μεταγραφή του προγράμματος με χρήση του Dragon Professional (7 <sup>η</sup> Προσπάθεια).....	60
4.4 Υπολογισμός του ποσοστού σφάλματος λέξης Word Error Rate- WER.....	61
4.5 Σχολιασμός ποσοστού σφάλματος λέξης WER .....	62
4.6 Πρωτότυπο πρόγραμμα στην γλώσσα Java .....	63
4.7 Μεταγραφή του προγράμματος με χρήση του Dragon Professional (9 <sup>η</sup> Προσπάθεια).....	64
4.8 Υπολογισμός του ποσοστού σφάλματος λέξης Word Error Rate- WER.....	65
4.9 Σχολιασμός ποσοστού σφάλματος λέξης WER .....	66
4.10 Πρωτότυπο πρόγραμμα στην γλώσσα Python.....	67
4.11 Μεταγραφή του προγράμματος με χρήση του Dragon Professional (6 <sup>η</sup> Προσπάθεια).....	68
4.12 Υπολογισμός του ποσοστού σφάλματος λέξης Word Error Rate- WER.....	69
4.13 Σχολιασμός ποσοστού σφάλματος λέξης WER .....	69
<b>Κεφάλαιο 5: Αποτελέσματα από την Εκτέλεση των δυο Σεναρίων Χρήσης.....</b>	<b>71</b>
5.1 Αξιολόγηση των Επιδόσεων του Dragon Professional .....	71
5.2 Πίνακες Αποτελεσμάτων Προγραμματιστικού Σεναρίου Χρήσης.....	72
5.3 Σχολιασμός Παραγόμενου Μεταγραφόμενου Προγράμματος στην Γλώσσα C από το Dragon Professional.....	74
5.4 Σχολιασμός Παραγόμενου Μεταγραφόμενου Προγράμματος στην Γλώσσα Java από το Dragon Professional.....	75
<b>Κεφάλαιο 6: Συμπεράσματα και Μελλοντικές Επιδιώξεις.....</b>	<b>77</b>
6.1 Συμπεράσματα .....	77
6.2 Μελλοντικές Επιδιώξεις .....	78
<b>Αναφορές .....</b>	<b>80</b>
<b>Παράρτημα Α .....</b>	<b>83</b>

## Κεφάλαιο 1 Εξερευνώντας την Τεχνολογία

---

1.1 Εισαγωγή	7
1.2 Προσωπικές Εμπειρίες στον Τομέα της Τεχνολογίας	10
1.3 Σκοπός της Παρούσας Διπλωματικής Εργασίας	11
1.4 Παρόμοιες Μελέτες που έχουν Πραγματοποιηθεί στον Τομέα της Φωνητικής Αναγνώρισης	12
1.5 Σύντομη Περιγραφή Περιεχομένων Διπλωματικής Εργασίας	15

---

### 1.1 Εισαγωγή

Στις μέρες μας, η φωνητική αναγνώριση είναι πολύ σημαντική στη ζωή του ανθρώπου, καθώς επηρεάζει πολλούς τομείς της ζωής του. Αρχικά, η φωνητική αναγνώριση διευκολύνει την προσβασιμότητα του ανθρώπου σε διάφορες τεχνολογικές συσκευές που χρησιμοποιεί στην καθημερινότητά του για να εκτελεί τις δραστηριότητές του εύκολα και γρήγορα, χρησιμοποιώντας ως μέσο επικοινωνίας τη φωνή του. Κάποιες ευρέως διαδεδομένες εφαρμογές που υποστηρίζουν φωνητική αναγνώριση σε έξυπνα τηλέφωνα (smartphones) είναι η Bixby, που δημιουργός της είναι η Samsung, και η Siri της Apple. Αυτές οι δυο εφαρμογές έχουν διευκολύνει σε τεράστιο βαθμό τον σύγχρονο άνθρωπο, αφού χρησιμοποιώντας μόνο τη φωνή του, μπορεί να έχει τον πλήρη έλεγχο του έξυπνου κινητού του [20][21].

Παράλληλα, υπάρχουν και κάποιες άλλες εφαρμογές διαθέσιμες στην αγορά, εξίσου αξιόλογες, που υποστηρίζουν φωνητική αναγνώριση, όπως είναι το Dragon, το οποίο ήταν το πρώτο λογισμικό που δημιουργήθηκε το 1977 από την Dr. Janet Baker και τον Dr. James Baker και υποστήριζε την αγγλική γλώσσα [21][22]. Το λογισμικό Talk and Write δημιουργήθηκε περίπου το 2002 από την Philips και μετέπειτα μεταφράστηκε και υποστηρίζεται και στην ελληνική γλώσσα [20][21]. Υπάρχει επίσης και το λογισμικό Λογογράφος, το οποίο δημιουργήθηκε από την Ελλάδα τον Ιανουάριο του 2002 με τη συμμετοχή ειδικών σε θέματα τεχνολογίας αναγνώρισης φωνής [21]. Διαθέσιμο στην αγορά είναι και το λογισμικό Speechelo, το οποίο είναι αρκετά διαδεδομένο αλλά και αρκετά ακριβό [15]. Τέλος, η Microsoft Office πρόσφατα έχει αναπτύξει έναν εξαιρετικό λογογράφο, τον οποίο χρησιμοποιώ και εγώ στη συγγραφή της παρούσας διπλωματικής εργασίας, ο οποίος είναι απίστευτα αξιόπιστος και πολύ



αποτελεσματικός, και υποστηρίζει μια πληθώρα από γλώσσες για φωνητική αναγνώριση [16][17].

Αξίζει να αναφερθεί ότι υπάρχουν λογισμικά φωνητικής αναγνώρισης που υποστηρίζονται σε ξένες γλώσσες, αλλά υπάρχουν και κάποια άλλα τα οποία υποστηρίζουν και την ελληνική γλώσσα. Υπάρχουν επίσης λογισμικά φωνητικής αναγνώρισης που είναι ανεπτυγμένα για το διαδίκτυο ή ακόμη και κάποια άλλα τα οποία είναι ανεπτυγμένα για υπολογιστές ανεξάρτητων συστημάτων, όπως είναι τα συστήματα των αυτοκινήτων στις μέρες μας, όπου μπορείς να επικοινωνήσεις με αυτά διαμέσου της φωνής σου και να τους δώσεις διάφορες εντολές προς εκτέλεση, όπως για παράδειγμα την εντολή «open the door» έτσι ώστε να ανοίξει η πόρτα. Παρόμοιου είδους συστήματα χρησιμοποιούν οι πιλότοι ή ακόμη και άτομα στο στρατό [22].

Επιπλέον, η φωνητική αναγνώριση συμβάλλει καταλυτικά στα άτομα με αναπηρίες, αφού μέσω αυτής μπορούν να αποκτήσουν πρόσβαση σε τεχνολογία και υπηρεσίες που θα μπορούσαν να είναι δυσπρόσιτες για αυτά. Αρχικά, τα άτομα με κινητικές δυσκολίες ή προβλήματα όρασης μπορούν να χρησιμοποιήσουν τη φωνή τους για να επικοινωνήσουν με υπολογιστές, κινητά τηλέφωνα, ηλεκτρονικές συσκευές και άλλες τεχνολογίες, που θα τους βοηθήσουν ακόμα και στο γράψιμο χρησιμοποιώντας ως είσοδο τη φωνή τους, κάτι που για αυτά τα άτομα είναι πολύ σημαντικό, αφού επιταχύνει σημαντικά τους ρυθμούς γραφής και διεκπεραίωσης διαφόρων δραστηριοτήτων σε ηλεκτρονικό υπολογιστή ή κάποιο άλλο τεχνολογικό μέσο [21][23]. Επίσης, γίνεται χρήση προγνωστικού λεξικού το οποίο είναι ορθογραφημένο, κάτι που προσφέρει την αποφυγή πολλών ορθογραφικών λαθών σε σύγκριση με τη γραφή στο χέρι. Όταν γράφουμε στον ηλεκτρονικό υπολογιστή ή σε κάποιο άλλο ηλεκτρονικό μέσο, υπάρχει και η έννοια της καλλιγραφίας. Για παράδειγμα, τα παιδιά με δυσλεξία δυσκολεύονται σε μεγάλο βαθμό να γράψουν στο χέρι. Η μόνη φορά που καταφέρνουν αυτά τα παιδιά να «γράψουν» τα πιο ωραία τους γράμματα είναι όταν γράφουν σε κάποιο ηλεκτρονικό μέσο [3].

Η φωνητική αναγνώριση, όμως, έχει και κάποιες αρνητικές πτυχές που αξίζει να αναφερθούν. Καταρχάς, τα πλείστα συστήματα αναγνώρισης φωνής που υπάρχουν επιβάλλεται να γίνουν πιο αποδοτικά και αξιόπιστα, αφού υπάρχουν περιπτώσεις που δεν αναγνωρίζουν επακριβώς κάποιες λέξεις ή φράσεις από τον χρήστη, με αποτέλεσμα να χρειάζεται να παρέμβει και να κάνει τις απαραίτητες διορθώσεις [21]. Επίσης,

κάποια λογισμικά φωνητικής αναγνώρισης είναι αρκετά ακριβά, με αποτέλεσμα κάποια άτομα που τα χρειάζονται άμεσα να μην έχουν την οικονομική ευχέρεια να τα αποκτήσουν [22][23]. Υπάρχουν περιπτώσεις όπου τα λογισμικά φωνητικής αναγνώρισης επιβάλλεται να εκπαιδευτούν από τον χρήστη, έτσι ώστε να αναγνωρίζουν τη φωνή του, κάτι που είναι μια χρονοβόρα και δύσκολη διαδικασία για τα άτομα τα οποία δεν έχουν καθαρή άρθρωση [3]. Επιπρόσθετα, η αποδοτικότητα αυτών των λογισμικών εξαρτάται από το λεξιλόγιο που τους καταχωρεί ο κατασκευαστής τους. Αν το λεξικό κάποιου λογισμικού περιλαμβάνει 100.000 λέξεις της ελληνικής γλώσσας, πολύ πιθανόν να ανταποκρίνεται μόνο στις ανάγκες των παιδιών του δημοτικού. Κάποια λογισμικά υποστηρίζουν και πιο εξειδικευμένο λεξιλόγιο, όπως είναι το Talk and Write, που περιέχει ειδικό λεξικό για δικηγόρους [21].

Συμπερασματικά, η φωνητική αναγνώριση είναι ένα εργαλείο που διευκολύνει απίστευτα τη ζωή του σύγχρονου ανθρώπου όταν τη χρησιμοποιεί. Μας δίνει τη δυνατότητα να χρησιμοποιούμε τη φωνή μας για να εκτελέσουμε διάφορες εργασίες και εντολές, προσφέροντας ευκολία και άνεση στην καθημερινότητα μας. Παρόλα αυτά, χρήζει βελτίωσης, έτσι ώστε η ζωή των ανθρώπων να γίνει πιο εύκολη, παραγωγική και αποτελεσματική.

## 1.2 Προσωπικές Εμπειρίες στον Τομέα της Τεχνολογίας

Ανατρέχοντας στα παιδικά μου χρόνια στο Δημοτικό, θυμάμαι έντονα τις προσπάθειες των δασκάλων μου να βρουν τρόπους που θα με βοηθούσαν να χρησιμοποιήσω τον ηλεκτρονικό υπολογιστή με μεγαλύτερη ευκολία. Ανάμεσα στις λύσεις που δοκιμάστηκαν ήταν ένα ειδικό πληκτρολόγιο με μεγαλύτερα πλήκτρα, τα οποία ήταν διαχωρισμένα μεταξύ τους, επιτρέποντάς μου να τα πιέζω ένα- ένα με τον δείκτη του αριστερού μου χεριού. Επίσης, χρησιμοποιούσα μια ιχνόσφαιρα (trackball), καθώς δυσκολευόμουν σε μεγάλο βαθμό να χρησιμοποιήσω το ποντίκι (mouse).

Στην πρώτη μας συνάντηση μέσω Teams με τον Επιβλέποντα Καθηγητή μου, Δρ. Κωνσταντίνο Παττίχη, μοιράστηκα τις δυσκολίες που αντιμετωπίζω σχετικά με την αλληλεπίδραση στον ηλεκτρονικό υπολογιστή, έτσι μου πρότεινε να εστιάσουμε την έρευνα μας στον τομέα της αναγνώρισης φωνής. Αυτή η εισήγηση έγινε με σκοπό να διευκολυνθώ προσωπικά μέσω της χρήσης κάποιου εργαλείου φωνητικής αναγνώρισης, καθώς και να παρέχουμε χρήσιμες πληροφορίες και σε άλλα άτομα με παρόμοιες ανάγκες.

Αυτή η ιδέα με ενθουσίασε από την πρώτη στιγμή και έτσι αποφασίσαμε να προχωρήσουμε με μια ενδεδειγμένη έρευνα, ερευνώντας βαθιά την θεματολογία της φωνητικής αναγνώρισης στον τομέα της πληροφορικής. Το ταξίδι αυτής της έρευνας δεν αποτέλεσε μόνο μια ευκαιρία για ακαδημαϊκή ανάπτυξη αλλά και μια προσωπική διαδρομή αυτογνωσίας και εξέλιξης, προσφέροντας πρακτικές λύσεις σε άτομα που αντιμετωπίζουν δυσκολίες στην αλληλεπίδραση με τον ψηφιακό κόσμο.

### **1.3 Σκοπός της Παρούσας Διπλωματικής Εργασίας**

Ο κύριος σκοπός της παρούσας διπλωματικής εργασίας είναι η διεξοδική έρευνα και κριτική αξιολόγηση μερικών εργαλείων φωνητικής αναγνώρισης που είναι διαθέσιμα στην αγορά, με σκοπό την επιλογή του πλέον κατάλληλου εργαλείου βάσει των χαρακτηριστικών και των δυνατοτήτων του. Η αξιολόγηση θα επικεντρωθεί στην αποδοτικότητα του επιλεγμένου εργαλείου μέσα από την εκτέλεση σεναρίων χρήσης που θα αντικατοπτρίζουν πραγματικές ανάγκες, με ιδιαίτερη έμφαση στην υποστήριξη ατόμων με αναπηρίες.

Αυτός ο στόχος θα υλοποιηθεί μέσω ενός συνδυασμού θεωρητικής έρευνας και πρακτικής δοκιμής, αρχίζοντας με μια ανάλυση των διαθέσιμων τεχνολογιών φωνητικής αναγνώρισης, καθώς και των προδιαγραφών, των δυνατοτήτων και των περιορισμών τους. Στη συνέχεια, θα πραγματοποιηθεί επιλογή του εργαλείου που αναμένεται να προσφέρει την καλύτερη απόδοση και ευχρηστία για την συγκεκριμένη ομάδα χρηστών.

Η αξιολόγηση της αποδοτικότητας του επιλεγμένου εργαλείου θα βασίζεται στην εκτέλεση σεναρίων χρήσης που θα σχεδιαστούν για να αποτυπώσουν τις πραγματικές ανάγκες αλληλεπίδρασης των ατόμων με κινητικές και άλλες δυσκολίες με τον υπολογιστή. Αυτά τα σενάρια θα περιλαμβάνουν εργασίες όπως είναι η πλοήγηση στο διαδίκτυο, η δημιουργία κάποιου απλού προγράμματος στις πλέον πιο διαδεδομένες γλώσσες προγραμματισμού και η διαχείριση αρχείων, με στόχο την αξιολόγηση της ευκολίας χρήσης, της ακρίβειας αναγνώρισης και της γενικής εμπειρίας του χρήστη με το εργαλείο.

Εν κατακλείδι, η παρούσα διπλωματική εργασία έχει ως στόχο να προσφέρει μια εξαντλητική επισκόπηση των δυνατοτήτων που προσφέρουν κάποια από τα εργαλεία φωνητικής αναγνώρισης στον σύγχρονο ψηφιακό κόσμο. Επιδιώκει να αναδείξει τη σημαντικότητα τους ως μέσων που μπορούν να βελτιώσουν ουσιαστικά την προσβασιμότητα και να ενισχύσουν την αλληλεπίδραση μεταξύ των ατόμων με αναπηρίες και της τεχνολογίας. Μέσα από αυτή την προσέγγιση, στοχεύει επίσης να θέσει τις βάσεις για μελλοντική έρευνα και ανάπτυξη στον εν λόγω τομέα, ανοίγοντας δρόμους για την περαιτέρω βελτίωση και προσαρμογή αυτών των τεχνολογιών σε ένα ευρύ φάσμα αναγκών και εφαρμογών.

#### **1.4 Παρόμοιες Μελέτες που έχουν Πραγματοποιηθεί στον Τομέα της Φωνητικής Αναγνώρισης**

Το άρθρο με τίτλο **"Talking Instead of Typing: Alternate Access to Computers Via Speech Recognition Technology"** των **Albert R. Cavalier και Ralph P. Ferretti** εξετάζει τις δυνατότητες που προσφέρει η τεχνολογία αναγνώρισης ομιλίας στα άτομα με αναπηρίες, βελτιώνοντας την αλληλεπίδρασή τους με τους υπολογιστές. Τονίζεται η σημασία της ομιλίας ως φυσικότερου και ευκολότερου μέσου εισόδου σε σύγκριση με την παραδοσιακή πληκτρολόγηση, ειδικά για όσους αντιμετωπίζουν δυσκολίες στην πληκτρολόγηση.

Η μελέτη αναφέρει πρακτικές εφαρμογές της τεχνολογίας σε άτομα με διάφορους τύπους αναπηριών, επισημαίνοντας παραδείγματα όπου η τεχνολογία αυτή έχει ενισχύσει την αυτονομία και την επικοινωνιακή δυνατότητα αυτών των χρηστών. Για τα άτομα με σωματικές αναπηρίες, τα οφέλη περιλαμβάνουν τη δυνατότητα ελέγχου του περιβάλλοντος μέσω φωνητικών εντολών και τη βελτίωση της πρόσβασης στο ψηφιακό περιεχόμενο. Στα άτομα με προβλήματα ακοής, οι τεχνολογίες αναγνώρισης ομιλίας προσφέρουν λύσεις για πιο ακριβή αναγνώριση και απόδοση της ομιλίας τους σε σύγκριση με την ανθρώπινη ακρόαση. Για τα άτομα με γνωστικές αναπηρίες, η τεχνολογία αναγνώρισης ομιλίας βοηθά στην εκμάθηση και την αλληλεπίδραση με το περιβάλλον, αυξάνοντας την αίσθηση ελέγχου και την κοινωνική αλληλεπίδραση.

Το άρθρο επισημαίνει επίσης τις προκλήσεις και τις ιδιαιτερότητες της χρήσης της τεχνολογίας, όπως είναι η ανάγκη για ακριβή εγγραφή και το προσαρμοσμένο λεξιλόγιο. Η σωστή εγγραφή και η συνεχής βελτίωση και προσαρμογή των συστημάτων αναγνώρισης ομιλίας είναι κρίσιμες για την αποδοτικότητα και την ευκολία χρήσης τους από τα άτομα με αναπηρίες. Η διαρκής ανάπτυξη των τεχνολογιών και η προσαρμογή τους στις ανάγκες των χρηστών θεωρούνται ουσιώδεις για την πλήρη εκμετάλλευση του δυναμικού τους στην υποστήριξη των ατόμων με αναπηρίες [1].

Το άρθρο με τίτλο **“Speech Recognition Technology for Disabilities Education”** των **K. Wendy Tang** σε συνεργασία με το Πανεπιστήμιο **Stony Brook** εξετάζει τη σημαντική επίδραση της τεχνολογίας αναγνώρισης ομιλίας στην υποστήριξη ατόμων με αναπηρίες. Το άρθρο παρέχει μια επισκόπηση των τρεχουσών τεχνολογιών αναγνώρισης ομιλίας, των εφαρμογών τους στην εκπαίδευση, ιδιαίτερα για άτομα με σωματικές, ακουστικές και οπτικές αναπηρίες και παρουσιάζει ένα φοιτητικό έργο που ενσωματώνει αυτές τις τεχνολογίες σε Προσωπικούς Ψηφιακούς Βοηθούς (PDAs) με σκοπό να δημιουργήσει ένα οικονομικά αποδοτικό και φορητό σύστημα παρακολούθησης υγείας.

Το έργο αυτό παρουσιάζεται ως ένα εξαιρετικό παράδειγμα που αξιοποιεί την τεχνολογία αναγνώρισης ομιλίας για να διευκολύνει την ευκολότερη πρόσβαση σε ηλεκτρονικές συσκευές για τα άτομα με αναπηρίες, προσφέροντας έλεγχο χωρίς χρήση των χεριών, ο οποίος είναι ιδιαίτερα ωφέλιμος για τα άτομα με σωματικές αναπηρίες. Το σύστημα υποστηρίζει εργασίες όπως ο έλεγχος συστημάτων, η εκκίνηση εφαρμογών και η δημιουργία εγγράφων, καθιστώντας την τεχνολογία πιο προσιτή σε όσους έχουν ακουστικές και οπτικές αναπηρίες.

Επιπλέον, το άρθρο παραθέτει διάφορα εμπορικά λογισμικά αναγνώρισης ομιλίας που είναι διαθέσιμα για desktop και φορητούς υπολογιστές, όπως το Dragon Naturally Speaking και το IBM ViaVoice, σημειώνοντας τις δυνατότητες τους, τους απαιτούμενους χρόνους εκπαίδευσης και τις απαιτήσεις του συστήματος. Αυτή η προσέγγιση επεκτείνεται σε ενσωματωμένες τεχνολογίες ομιλίας σε μικρότερες συσκευές, ενισχύοντας τη χρηστικότητα των PDAs στα άτομα με ακουστικές και οπτικές αναπηρίες μέσω φωνητικών εντολών.

Το έργο που περιγράφεται στο άρθρο εστιάζει στις πρακτικές εφαρμογές αυτής της τεχνολογίας στη δημιουργία ενός συστήματος παρακολούθησης υγείας ειδικά σχεδιασμένο για άτομα με αναπηρίες, τονίζοντας το δυναμικό του να βελτιώσει τους χρόνους ανταπόκρισης σε έκτακτες ανάγκες και γενικά την ποιότητα ζωής των χρηστών. Το σύστημα συνδυάζει ένα σύστημα παρακολούθησης υγείας που παρακολουθεί συνεχώς την καρδιά του χρήστη με έναν βοηθό ομιλίας που διευκολύνει την αλληλεπίδραση του χρήστη με το σύστημα. Αυτή η συνεχής παρακολούθηση θα μειώσει το χρόνο πρόσβασης στα 911s και θα αυξήσει το ποσοστό επιβίωσης του

χρήστη. Αυτή η δυνατότητα είναι πολύ σημαντική για τη χρήση των PDAs από άτομα με αναπηρίες [2].

Το άρθρο με τίτλο **"Dysarthric Speech Recognition Using Dysarthria-Severity-Dependent and Speaker-Adaptive Models"** των **Myung Jong Kim, Joohong Yoo** και **Hoirin Kim** από το **KAIST**, επικεντρώνεται στη βελτίωση των συστημάτων αναγνώρισης ομιλίας για άτομα με δυσαρθρία, μια νευροκινητική διαταραχή ομιλίας που επηρεάζει σημαντικά την καθαρότητα του λόγου. Τα παραδοσιακά αυτόματα συστήματα αναγνώρισης ομιλίας δεν είναι αποτελεσματικά για δυσλεκτικούς ομιλητές λόγω μεγάλης αναντιστοιχίας στα ακουστικά χαρακτηριστικά.

Η μελέτη παρουσιάζει ένα νέο σχήμα προσαρμογής που σχεδιάστηκε για να προσαρμόζει το μοντέλο αναγνώρισης ομιλίας πιο κοντά στο συγκεκριμένο επίπεδο δυσαρθρίας του ομιλητή. Αρχικά, οι ομιλητές κατηγοριοποιούνται σε προκαθορισμένα επίπεδα σοβαρότητας με βάση τα χαρακτηριστικά τους. Ανάλογα με την κατηγοριοποίηση τους, επιλέγεται ένα αρχικό μοντέλο για προσαρμογή, το οποίο κατασκευάζεται από δεδομένα του δυσαρθρικού λόγου που αντιστοιχούν σε παρόμοια επίπεδα σοβαρότητας. Αυτά τα μοντέλα βελτιώνονται περαιτέρω χρησιμοποιώντας τεχνικές προσαρμογής ομιλητή όπως η Μέγιστη Πιθανοφάνεια Γραμμικής Παλινδρόμησης (MLLR) και η Μέγιστη Απόδειξη (MAP), που προσαρμόζουν τις παραμέτρους του μοντέλου για να ταιριάζουν καλύτερα στα συγκεκριμένα μοτίβα ομιλίας του ατόμου.

Η αποτελεσματικότητα αυτής της προσέγγισης επιβεβαιώθηκε μέσω πειραμάτων που χρησιμοποιούν μια βάση δεδομένων η οποία περιέχει αρκετές εκατοντάδες λέξεις που εκφέρονται από 31 διαφορετικούς ομιλητές με ήπια έως μέτρια δυσαρθρία. Τα αποτελέσματα έδειξαν ότι αυτή η μέθοδος βελτιώνει σημαντικά την ακρίβεια αναγνώρισης, ιδιαίτερα όταν διατίθεται μικρή ποσότητα δεδομένων προσαρμογής, σε σύγκριση με τα συμβατικά προσαρμοσμένα συστήματα. Η έρευνα τονίζει τη δυνατότητα των μοντέλων που εξαρτώνται από τον βαθμό σοβαρότητας της δυσαρθρίας και προσαρμόζονται στα χαρακτηριστικά κάθε ομιλητή, στην ανάπτυξη πιο προσβάσιμων τεχνολογιών αναγνώρισης ομιλίας για άτομα με διάφορα επίπεδα δυσαρθρίας, βελτιώνοντας έτσι την ικανότητα τους να επικοινωνούν αποτελεσματικά μέσω τεχνολογικών συσκευών [3].

### **1.5 Σύντομη Περιγραφή Περιεχομένων Διπλωματικής Εργασίας**

Στο **Κεφάλαιο 2** θα μελετήσουμε τις βασικές τεχνολογίες μερικών εργαλείων φωνητικής αναγνώρισης που είναι διαθέσιμα στην αγορά και θα αναφέρουμε κάποιες σημαντικές μετρικές αξιολόγησης τους.

Στο **Κεφάλαιο 3** θα πραγματοποιηθεί αξιολόγηση του εργαλείου Dragon Professional V16 μέσα από την εκτέλεση του πρώτου σεναρίου χρήσης που αφορά στην πλοήγηση αρχείων στον ηλεκτρονικό υπολογιστή όπως επίσης και στο διαδίκτυο με χρήση φωνητικών εντολών που υποστηρίζει το εργαλείο.

Στο **Κεφάλαιο 4** θα διεξαχθεί το δεύτερο σενάριο χρήσης που αποσκοπεί στην δημιουργία ενός απλού προγράμματος στις γλώσσες Προγραμματισμού C, Java και Python με σκοπό την αξιολόγηση της αποδοτικότητας του εργαλείου στην μεταγραφή κώδικα κάνοντας χρήση του ποσοστού σφάλματος λέξης WER για την αξιολόγηση των αποτελεσμάτων.

Στο **Κεφάλαιο 5** θα γίνει σχολιασμός των αποτελεσμάτων που έχουν παραχθεί μέσα από την εκτέλεση των δύο διακριτών σεναρίων χρήσης.

Στο **Κεφάλαιο 6** θα αναφερθούν κάποιες μελλοντικές επιδιώξεις που αποσκοπούν στην βελτίωση της αποδοτικότητας των εργαλείων φωνητικής αναγνώρισης.



## Κεφάλαιο 2 Βασικές Τεχνολογίες Εργαλείων Φωνητικής Αναγνώρισης

---

2.1 Μετατροπή Ηχογράφησης σε Κείμενο (Transcription)	17
2.2 Ορισμός και Σημασία	17
2.3 Τεχνικές Μεταγραφής	17
2.4 Εργαλεία και Τεχνολογίες	18
2.5 Προκλήσεις που Επέρχονται	18
2.6 Η Δύναμη του Transkriptor	18
2.6.1 Χαρακτηριστικά και Δυνατότητες	19
2.6.2 Προκλήσεις	19
2.7 Χαρακτηριστικά των διαθέσιμων λογισμικών και αξιολόγηση τους	19
2.7.1 Dragon Professional V16	20
2.7.2 Speechelo	22
2.7.3 Λογογράφος Microsoft Office 365 (Speech-to-text)	25
2.7.4 Συγκεντρωτικός Πίνακας Εργαλείων	28
2.7.5 Σύγκριση των Εργαλείων Dragon Professional V16, Speechelo και Microsoft Office 365 (Speech-to-text)	29
2.8 Μετρικές Αξιολόγησης των λογισμικών Φωνητικής Αναγνώρισης	30
2.8.1 Word Error Rate – WER	31
2.8.2 Κανονικοποίηση και Προεπεξεργασία των Μεταγραφών	33
2.8.3 Έξι βήματα για την εκτέλεση αξιολόγησης συστήματος ASR	36

---

Σε αυτό το κεφάλαιο θα αναφέρουμε κάποιες βασικές τεχνολογίες των εργαλείων φωνητικής αναγνώρισης καθώς και κάποιες βασικές μετρικές που χρησιμοποιούμε για να αξιολογήσουμε την αποδοτικότητα τους.

## 2.1 Μετατροπή Ηχογράφησης σε Κείμενο (Transcription)

Στην εποχή της ψηφιακής τεχνολογίας, η δυνατότητα μετατροπής της φωνής σε κείμενο διαδραματίζει σημαντικό ρόλο στους τομείς της επικοινωνίας, της εκπαίδευσης, της έρευνας και πολλών άλλων. Η διαδικασία αυτή, είναι γνωστή ως μεταγραφή (transcribing), που επιτρέπει την ακριβή καταγραφή του προφορικού περιεχομένου σε γραπτή μορφή, προσφέροντας μια ανεκτίμητη πηγή πληροφοριών και δεδομένων.

## 2.2 Ορισμός και Σημασία

Μεταγραφή ορίζουμε την διαδικασία κατά την οποία ο ηχογραφημένος λόγος μετατρέπεται σε αναλυτικό γραπτό κείμενο. Αυτή η διαδικασία μπορεί να είναι χειροκίνητη, με έναν ακροατή να ακούει την ηχογράφιση και να την μεταγράφει, ή αυτόματη, μέσω της χρήσης λογισμικού φωνητικής αναγνώρισης. Η μεταγραφή είναι ιδιαίτερα σημαντική στον τομέα της δικαστικής αναφοράς, της ιατρικής, της δημοσιογραφίας, καθώς και στην επιστημονική έρευνα, αφού προσφέρει τη δυνατότητα μιας ακριβούς, έγκυρης και προσβάσιμης ανάλυσης του προφορικού λόγου σε γραπτό κείμενο [20].

## 2.3 Τεχνικές Μεταγραφής

Η μεταγραφή μπορεί να εκτελεστεί μέσω διαφόρων τεχνικών, είτε χειροκίνητα είτε αυτόματα:

- Χειροκίνητη Μεταγραφή: Απαιτεί την προσεκτική ακρόαση της ηχογράφησης από ένα άτομο, το οποίο μεταγράφει αυτά που έχουν λεχθεί σε κείμενο. Αυτή η μέθοδος είναι χρονοβόρα αλλά προσφέρει υψηλή ακρίβεια.
- Αυτόματη Μεταγραφή: Γίνεται χρήση λογισμικού φωνητικής αναγνώρισης για την αυτοματοποιημένη μετατροπή του ηχητικού σήματος σε κείμενο. Ενώ είναι γρηγορότερη, μπορεί να επηρεαστεί από λάθη αναγνώρισης [18].

## 2.4 Εργαλεία και Τεχνολογίες

Για την αυτόματη μεταγραφή, υπάρχουν διάφορα εργαλεία και πλατφόρμες διαθέσιμες, με τα πιο δημοφιλή να περιλαμβάνουν:

- ✓ Google Docs Voice Typing
- ✓ Transkriptor
- ✓ Dragon Naturally Speaking
- ✓ IBM Watson Speech to Text
- ✓ Otter.ai
- ✓ Sonix

Κάθε ένα από αυτά τα εργαλεία έχει τα δικά του πλεονεκτήματα και περιορισμούς, με κάποια να είναι καλύτερα για επαγγελματική χρήση και κάποια άλλα να είναι πιο προσιτά για καθημερινή χρήση [4], [5], [6], [7], [8], [9].

## 2.5 Προκλήσεις που Επέρχονται

Η ακρίβεια της μεταγραφής μπορεί να επηρεαστεί από πολλούς και διάφορους παράγοντες, όπως ο θόρυβος του περιβάλλοντος, οι διάλεκτοι και η ποιότητα της ηχογράφησης. Η βελτιστοποίηση της ποιότητας της ηχογράφησης και η επιλογή ενός κατάλληλου εργαλείου για το είδος της ηχογράφησης είναι κρίσιμη για την μείωση των λαθών και την αύξηση της αποδοτικότητας της μεταγραφής.

Συμπερασματικά, η μετατροπή ηχογραφήσεων σε κείμενο αποτελεί μια σημαντική διαδικασία που επιτρέπει την πρόσβαση σε πληροφορίες και δεδομένα με νέους και πιο αποτελεσματικούς τρόπους, δίνοντας το έναυσμα για περαιτέρω αναλύσεις και εφαρμογές σε διάφορους τομείς.

## 2.6 Η Δύναμη του Transkriptor

Το Transkriptor αποτελεί ένα προηγμένο εργαλείο αυτόματης μεταγραφής που χρησιμοποιεί τεχνολογία φωνητικής αναγνώρισης για να μετατρέπει ηχογραφήσεις σε γραπτό κείμενο. Σχεδιάστηκε για να εξυπηρετεί μια πληθώρα χρηστών, από δημοσιογράφους και ερευνητές έως φοιτητές και επαγγελματίες, το Transkriptor προσφέρει μια εύχρηστη και αποδοτική λύση για την γρήγορη μετατροπή του προφορικού λόγου σε κείμενο.[4]

### 2.6.1 Χαρακτηριστικά και Δυνατότητες

- Ακρίβεια Μεταγραφής Χρησιμοποιώντας προηγμένους αλγόριθμους φωνητικής αναγνώρισης, το Transkriptor διασφαλίζει υψηλά επίπεδα ακρίβειας στη μεταγραφή, ακόμα και σε ηχογραφήσεις με περίπλοκη ονοματολογία σε διάφορες γλώσσες.
- Υποστήριξη Πολλαπλών Γλωσσών Το Transkriptor μπορεί να χειριστεί ηχογραφήσεις σε πολλαπλές γλώσσες, καθιστώντας το ιδανικό για διεθνείς χρήστες ή για εργασίες που απαιτούν μεταγραφή σε πολύγλωσσο περιβάλλον.
- Ευχρηστία Με ένα απλό και διαισθητικό περιβάλλον, οι χρήστες μπορούν εύκολα να ανεβάζουν τις ηχογραφήσεις τους και να λαμβάνουν τις μεταγραφές τους σε λίγα μόνο λεπτά.
- Ευελιξία Μορφοποίησης Το Transkriptor επιτρέπει στους χρήστες να προσαρμόζουν τη μορφοποίηση του τελικού κειμένου, παρέχοντας ευελιξία για διάφορες ανάγκες και προτιμήσεις [4].

### 2.6.2 Προκλήσεις

Παρά των πολυάριθμων δυνατοτήτων που διαθέτει το Trascriptor, η διαδικασία αυτόματης μεταγραφής μπορεί να συναντήσει εμπόδια όπως η παρουσία θορύβου στο περιβάλλον ηχογράφησης ή η χρήση διαλέκτων και ιδιοματισμών. Το Transkriptor αντιμετωπίζει αυτές τις προκλήσεις με τη συνεχή βελτίωση των αλγορίθμων του και την παροχή επιλογών για χειροκίνητη επεξεργασία και διόρθωση των μεταγραφών με την συμβολή του χρήστη, εξασφαλίζοντας έτσι τη μέγιστη δυνατή ακρίβεια και ποιότητα.

Συνοψίζοντας, το Transkriptor αποτελεί μια ισχυρή λύση για την αυτόματη μεταγραφή ηχογραφήσεων σε κείμενο, προσφέροντας ταχύτητα, ακρίβεια και ευελιξία στους χρήστες που επιθυμούν να μετατρέψουν τον προφορικό λόγο σε γραπτό περιεχόμενο καταβάλλοντας την ελάχιστη δυνατή προσπάθεια.

### 2.7 Χαρακτηριστικά των διαθέσιμων λογισμικών και αξιολόγησή τους

Σε αυτό το υπό κεφάλαιο θα βρούμε κάποια εργαλεία φωνητικής αναγνώρισης, θα παραθέσουμε τα χαρακτηριστικά τους και στο τέλος θα γίνει μια αξιολόγηση αυτών των εργαλείων για να επιλέγει το κατάλληλο που θα μας βοηθήσει καλύτερα.

### 2.7.1 Dragon Professional V16

Το Dragon Professional V16, γνωστό και ως Dragon Naturally Speaking ή απλώς Dragon, είναι ένα ισχυρό λογισμικό αναγνώρισης φωνής που προσφέρει πολλά πλεονεκτήματα στους χρήστες. Μερικά βασικά χαρακτηριστικά παρατίθενται στον Πίνακα 2.1 [10], [11], [12].

Πίνακας 2.1: Χαρακτηριστικά εργαλείου φωνητικής αναγνώρισης Dragon Professional V16

<u>Χαρακτηριστικό</u>	<u>Περιγραφή</u>
Παραγωγικότητα Χωρίς Χέρια	Το Dragon Professional V16 μας δίνει την δυνατότητα να ελέγχουμε τον υπολογιστή μας και να εκτελούμε διάφορες εργασίες χρησιμοποιώντας τη φωνή μας, μειώνοντας σε μεγάλο βαθμό την χρήση των χεριών για πληκτρολόγηση ή ακόμα και την χρήση ποντικιού. Αυτό μπορεί να αυξήσει σημαντικά την παραγωγικότητά στους χρήστες που έχουν περιορισμένη κινητικότητα.
Ταχύτερη Καταχώρηση Κειμένου	Η διάταξη κειμένου με το Dragon Professional V16 είναι συχνά ταχύτερη από την πληκτρολόγηση, καθώς μπορεί να αναγνώσει την εκφορά των λέξεων σε πραγματικό χρόνο. Αυτό είναι ιδιαίτερα χρήσιμο για τη σύνταξη email, εγγράφων, αναφορών και άλλου είδους κειμενικού περιεχομένου.
Βελτιωμένη Ακρίβεια	Το Dragon Professional V16 χρησιμοποιεί προηγμένη τεχνολογία αναγνώρισης φωνής που μαθαίνει συνεχώς και προσαρμόζεται στη φωνή μας και τα πρότυπα ομιλίας μας κατά τη διάρκεια που το χρησιμοποιούμε. Έτσι,

	αυτό οδηγεί σε βελτιωμένη ακρίβεια και μείωση των σφαλμάτων στις αναγραφές.
Προσβασιμότητα	Το Dragon Professional V16 είναι ένα απαραίτητο εργαλείο για άτομα με αναπηρίες που είναι δύσκολο ή αδύνατο για αυτά να πληκτρολογήσουν. Το Dragon Professional V16 επίσης, μπορεί να παρέχει ένα μέσο επικοινωνίας και πρόσβασης στον υπολογιστή στα άτομα που έχουν προβλήματα όρασης.
Πολλαπλές Εργασίες	Οι φωνητικές εντολές στο Dragon Professional V16 μας επιτρέπουν να εκτελούμε πολλαπλές εργασίες ταυτόχρονα. Για παράδειγμα, έχουμε την δυνατότητα να συντάσσουμε ένα email ενώ ψάχνουμε πληροφορίες για κάτι που μας ενδιαφέρει στο διαδίκτυο. Επομένως αυτό μας βοηθά να γίνουμε πιο παραγωγικοί.
Προσαρμογή	Το Dragon Professional V16 μας δίνει την ευκαιρία να δημιουργήσουμε προσαρμοσμένες φωνητικές εντολές και μακροεντολές για την αυτοματοποίηση επαναλαμβανόμενων εργασιών, όπως την εκκίνηση εφαρμογών ή την εκτέλεση συγκεκριμένων λειτουργιών. Αυτή η προσαρμογή μπορεί να μας εξοικονομήσει χρόνο και προσπάθεια.
Υποστήριξη Πολλαπλών Γλωσσών	Το Dragon Professional V16 υποστηρίζει πολλές γλώσσες και διαλέκτους, καθιστώντας το ένα ευέλικτο εργαλείο για χρήστες σε όλον τον κόσμο.

Αυξημένη Παραγωγικότητα για Επαγγελματίες	Επαγγελματίες σε διάφορους τομείς, όπως συγγραφείς, δημοσιογράφοι, πράκτορες υποστήριξης πελατών και άλλοι μπορούν να επωφεληθούν από το Dragon Professional V16, επιταχύνοντας τη ροή εργασίας τους και μειώνοντας τον χρόνο που αφιερώνουν σε διοικητικές εργασίες.
Κόστος	\$699

### 2.7.2 Speechelo

Το Speechelo είναι ένα λογισμικό μετατροπής κειμένου σε ομιλία (TTS) που μετατρέπει το γραπτό κείμενο σε ήχο που ακούγεται φυσικά σαν ανθρώπινη ομιλία. Το Speechelo προσφέρει ένα νέο επίπεδο προσβασιμότητας για την διαχείριση του περιεχομένου του χρήστη. Δεν πρόκειται απλά για τη μετατροπή λέξεων, αλλά για τον ενισχυτικό τρόπο που αξιοποιείται το περιεχόμενό του. Ανεξάρτητα από τον σκοπό για τον οποίο χρησιμοποιείται ο μετατροπέας κειμένου σε ήχο προσφέρει μια καινοτόμο προσέγγιση στην παρουσίαση περιεχομένου. Το Speechelo αφήνει ελεύθερη την δυναμικότητα της φωνητικής δημιουργίας με χρήση ΑΙ και μεταμορφώνει με ένα μοναδικό έξυπνο τρόπο την αλληλεπίδραση του χρήστη με το κείμενο μέσα από τις δυνατότητες που προσφέρει. Μερικά από τα χαρακτηριστικά του εργαλείου παρατίθενται στον Πίνακα 2.2 [15].

Πίνακας 2.2 Χαρακτηριστικά εργαλείου φωνητικής αναγνώρισης Speechelo

<u>Χαρακτηριστικό</u>	<u>Περιγραφή</u>
Δημιουργία Περιεχομένου	Το Speechelo μπορεί να βοηθήσει τους δημιουργούς περιεχομένου, όπως YouTubers, podcasters και bloggers, να μετατρέπουν εύκολα γρήγορα και αποτελεσματικά το γραπτό περιεχόμενο σε ηχητική μορφή. Αυτό μπορεί να εξοικονομήσει χρόνο και να ικανοποιήσει ένα ευρύτερο κοινό, συμπεριλαμβανομένων των ατόμων που προτιμούν να ακούν αντί να διαβάζουν.
Προσβασιμότητα	Το Speechelo καθιστά το ψηφιακό περιεχόμενο πιο προσβάσιμο για άτομα με προβλήματα όρασης ή δυσκολίες ανάγνωσης, παρέχοντας μια ηχητική εμπειρία του κειμενικού περιεχομένου διευκολύνοντας με αυτό τον τρόπο την διεκπεραίωση διαφόρων δραστηριοτήτων .
Παραγωγή Voiceover	Το Speechelo έχει την δυνατότητα να δημιουργήσει voiceovers για την παραγωγή βίντεο, διαφημίσεων, παρουσιάσεων ή μαθήματα e-learning κάνοντας χρήση μιας πληθώρας επιλογής φωνών.
Υποστήριξη Πολλαπλών Γλωσσών	Το Speechelo υποστηρίζει πολλές γλώσσες και αυτό το κάνει αρκετά εύχρηστο. Επίσης, μας επιτρέπει να δημιουργήσουμε voiceovers σε διάφορες γλώσσες για να καλύψουμε τις ανάγκες σε ένα ευρύ κοινό.



Προσαρμογή	Το Speechelo, προσφέρει μια γκάμα επιλογών προσαρμογής όσο αφορά την προσαρμογή του τόνου, του ύψους, της ταχύτητας και της προφοράς της παραγόμενης φωνής. Αυτό δίνει την δυνατότητα στον χρήστη να προσαρμόσει τη φωνή του ικανοποιώντας έτσι κάθε του ανάγκη.
Φυσικός Ήχος	Η ακμάζουσα πρόοδος στην τεχνολογία TTS (Text-To-Speech) έχει δημιουργήσει φωνές που ακούγονται φυσικές, καθιστώντας δύσκολο για τους ακροατές να διακρίνουν μεταξύ της υπολογιστικής ομιλίας που παράγεται και της ανθρώπινης ομιλίας.
Εκπαίδευση	Το Speechelo είναι ιδανικό εργαλείο για τη δημιουργία voiceovers e-learning modules, εκπαιδευτικά βίντεο και εκπαιδευτικό υλικό, παρέχοντας στους μαθητές μια πιο προηγμένη και ενημερωτική εμπειρία.
Κόστος	\$47

### 2.7.3 Λογογράφος Microsoft Office 365 (Speech-to-text)

Η Microsoft Office 365 πρόσφατα έχει αναπτύξει έναν πολύ αξιόπιστο λογογράφο. Χρησιμοποιώντας ως μέσο τη φωνή μας μπορούμε μέσω του εργαλείου αυτού να συντάξουμε πιο εύκολα γρήγορα και αποτελεσματικά διάφορα έγγραφα, email, σημειώσεις, παρουσιάσεις ή ακόμα και σημειώσεις διαφανειών. Επιταχύνοντας έτσι σε αρκετά μεγάλο βαθμό την διεκπεραίωση των δραστηριοτήτων μας. Μερικά από τα χαρακτηριστικά του εργαλείου παρατίθενται στον Πίνακα 2.3 [16], [17].

Πίνακας 2.3 Χαρακτηριστικά εργαλείου φωνητικής αναγνώρισης Microsoft Office 365 (Speech-to-text)

<u>Χαρακτηριστικό</u>	<u>Περιγραφή</u>
Αποτελεσματικότητα και Ταχύτητα	Ο λογογράφος του office 365 μας επιτρέπει να εισάγουμε κείμενο πολύ πιο γρήγορα παρά να πληκτρολογούμε. Αυτό μπορεί να είναι ιδιαίτερα χρήσιμο για εργασίες που περιλαμβάνουν πολύ κείμενο, όπως η σύνταξη αναφορών, email ή εγγραφων.
Μείωση Κόπωσης από την Πληκτρολόγηση	Η πληκτρολόγηση για εκτεταμένα χρονικά διαστήματα μπορεί να οδηγήσει σε φυσική κόπωση και δυσφορία, ιδίως στα χέρια και στους καρπούς. Ο λογογράφος του office 365 μπορεί να βοηθήσει καταλυτικά στην ανακούφιση αυτών των προβλημάτων μειώνοντας σε αρκετά μεγάλο βαθμό την ανάγκη για πληκτρολόγηση.
Βελτιωμένη Ακρίβεια	Οι σύγχρονες τεχνολογίες αναγνώρισης φωνής έχουν γίνει πολύ ακριβείς, με αποτέλεσμα να γίνονται λιγότερα λάθη παρά όταν πληκτρολογούμε. Αυτό μπορεί να οδηγήσει σαφέστατα σε πιο

	ολοκληρωμένα και απαλλαγμένα από λάθη έγγραφα.
Προσβασιμότητα	Η φωνητική υπαγόρευση καθιστά το ψηφιακό περιεχόμενο πιο προσβάσιμο σε άτομα με περιορισμένη κινητικότητα ή δεξιότητες. Επιτρέπει σε ένα ευρύτερο πληθυσμό ανθρώπων να αλληλοεπιδράσει με την τεχνολογία και να δημιουργήσει τις δραστηριότητες που επιθυμεί με μεγαλύτερη ανεξαρτησία και ευελιξία με την βοήθεια αυτού του λογογράφου.
Πολλαπλές Εργασίες	Ο λογογράφος office 365 μας επιτρέπει να εκτελούμε πολλαπλές εργασίες ταυτόχρονα με αποτελεσματικό τρόπο. Μας δίνει την δυνατότητα να του υπαγορεύουμε τις σκέψεις μας ή τις ιδέες μας ενώ εκτελούμε ταυτόχρονα άλλες εργασίες.
Λειτουργία Χωρίς Χέρια	Ο λογογράφος office 365 μας δίνει την δυνατότητα να αλληλοεπιδράσουμε με τον υπολογιστή χωρίς να καθίσταται αναγκαίο να γίνει χρήση των χεριών μας. Αυτό μπορεί να είναι χρήσιμο σε περιπτώσεις όπου η χρήση πληκτρολόγιου είναι αναποτελεσματική ή επικίνδυνη, όπως για παράδειγμα όταν οδηγούμε ή φοράμε γάντια.
Φυσική Ομιλία	Ο λογογράφος office 365 μας επιτρέπει συνήθως να καταγράψουμε αυτά τα οποία επιθυμούμε με ένα πιο φυσικό τρόπο στην γραφή μας, η οποία μπορεί

	να είναι χρήσιμη για περιεχόμενο όπως αναρτήσεις σε blogs, ομιλίες ή αφηγήσεις.
Βελτίωση της Παραγωγικότητας	Πολλοί άνθρωποι είναι πιο παραγωγικοί όταν έχουν την ευκαιρία να εκφράζουν τις σκέψεις τους ελεύθερα, χωρίς να υπάρχουν διακοπές κατά την διάρκεια της πληκτρολόγησης. Αυτό επομένως, μπορεί να συντελέσει σε σημαντικό βαθμό στη βελτιστοποίηση της δημιουργικής δραστηριότητας.
Υποστήριξη Γλωσσών	Συνήθως, τα εργαλεία φωνητικής υπαγόρευσης υποστηρίζουν πολλές γλώσσες και προφορές, καθιστώντας τα ευέλικτα για ένα ευρύ κοινό.
Ενσωμάτωση	Ο λογογράφος του office 365 ενσωματώνεται απευθείας στις εφαρμογές που υποστηρίζει η Microsoft όπως το Word, το Outlook και το PowerPoint, καθιστώντας την εύκολη χρήση στην υφιστάμενη ροή εργασίας του χρήστη.
Βελτίωση της Δημιουργίας Περιεχομένου	Η φωνητική υπαγόρευση μπορεί να γίνει ιδιαίτερα χρήσιμη για τους δημιουργούς περιεχομένου, καθώς τους επιτρέπει να επινοούν γρήγορα νέες ιδέες και να καταγράφουν με μεγάλη ευκολία τις σκέψεις τους.
Χαρακτηριστικά Προσβασιμότητας	Πολλά εργαλεία φωνητικής υπαγόρευσης όπως και ο λογογράφος office 365 υποστηρίζουν φωνητικές εντολές που αφορούν σε θέματα

	μορφοποίησης , επεξεργασίας και της πλοήγησης στα έγγραφα, βελτιώνοντας την προσβασιμότητα και την ευχρηστία.
--	---

#### 2.7.4 Συγκεντρωτικός Πίνακας Εργαλείων

##### Πίνακας 2.4 Συγκεντρωτικός Πίνακας Εργαλείων

[10], [11], [12], [15], [16], [17].

Χαρακτηριστικό	Dragon Professional V16	Speechelo	Λογογράφος Microsoft Office 365
Παραγωγικότητα Χωρίς Χέρια	✓	✓	✓
Ταχύτερη Καταχώρηση Κειμένου	✓	✓	✓
Βελτιωμένη Ακρίβεια	✓		✓
Προσβασιμότητα	✓	✓	✓
Πολλαπλές Εργασίες	✓		✓
Προσαρμογή	✓	✓	
Υποστήριξη Πολλαπλών Γλωσσών	✓	✓	✓
Αυξημένη Παραγωγικότητα για Επαγγελματίες	✓	✓	
Δημιουργία Περιεχομένου		✓	

Παραγωγή Voiceover		✓	
Φυσικός Ήχος		✓	
Εκπαίδευση	✓	✓	
Μείωση Κόπωσης από την Πληκτρολόγηση	✓	✓	✓
Φυσική Ομιλία	✓	✓	✓
Βελτίωση της Παραγωγικότητας	✓	✓	✓
Ενσωμάτωση			✓
Βελτίωση της Δημιουργίας Περιεχομένου		✓	✓

### 2.7.5 Σύγκριση των Εργαλείων Dragon Professional V16, Speechelo και Microsoft Office 365 (Speech-to-text)

Η επιλογή μεταξύ του Dragon Professional V16, του Speechelo και του λογογράφου Microsoft Office 365 εξαρτάται από τις ανάγκες και προτιμήσεις του κάθε χρήστη, καθώς κάθε εργαλείο προσφέρει μοναδικά χαρακτηριστικά και δυνατότητες.

Το **Dragon Professional V16** είναι ένα εξαιρετικά αναγνωρισμένο λογισμικό δικτύωσης, στην φωνητική αναγνώριση γνωστό για την ακρίβεια και τις προηγμένες του δυνατότητες. Θεωρείται βιομηχανικό πρότυπο στην μεταγραφή, με μέσο ρυθμό λάθους μία λέξη ανά 100. Η δύναμη του Dragon βρίσκεται στη μεγάλη βάση δεδομένων λεξιλογίου που διαθέτει και στη δυνατότητα προσαρμογής εντολών, ιδιαίτερα στην επαγγελματική έκδοση, για να διευκολύνει ακόμα περισσότερο τους χρήστες που το χρησιμοποιούν. Επίσης, προσφέρει τη δυνατότητα μεταγραφής αρχείων ήχου σε διάφορες μορφές. Ωστόσο, η αγορά του Dragon απαιτεί σημαντική αρχική επένδυση, πράγμα που μπορεί να αποτελέσει σκέψη για τους χρήστες που διαθέτουν περιορισμένο προϋπολογισμό [10], [11], [12].

Ο λογογράφος του **Office 365 (Speech-to-text)**, που περιέχεται μέσα στα προϊόντα της Microsoft, προσφέρει μια σταθερή, δωρεάν εναλλακτική λύση για τους χρήστες των

Windows. Προσφέρει καλή ακρίβεια μεταγραφής με λιγότερα από δύο λάθη ανά 100 λέξεις κατά μέσο όρο. Το εργαλείο φωνητικής αναγνώρισης του Office 365 είναι ενσωματωμένο με το ευρύτερο σύστημα φωνητικής αναγνώρισης των Windows, κάνοντας το βολικό για τους χρήστες που ήδη χρησιμοποιούν το περιβάλλον των Windows [16], [17].

Το **Speechelo** αποτελεί ένα πρωτοποριακό λογισμικό μετατροπής κειμένου σε ομιλία (text-to-speech), σε αντίθεση από τα εργαλεία μετατροπής ομιλίας σε κείμενο (speech-to-text) όπως λειτουργεί το Dragon Professional V16 και ο λογογράφος του Office 365. Το Speechelo έχει σχεδιαστεί ειδικά για την παραγωγή φωνητικών μεταγραφών από κείμενα (voicerecorders) προσφέροντας δύο εκδόσεις την τυπική και την επαγγελματική έκδοση. Η επαγγελματική έκδοση, Speechelo Pro, διακρίνεται για την ευρεία γκάμα φωνητικών επιλογών που προσφέρει και τη δυνατότητα διαχείρισης μεγαλύτερων αρχείων ήχου, καθιστώντας το ιδανικό για την παραγωγή voicerecorders για βίντεο και για αντίστοιχα παρόμοια περιεχόμενα [15].

Συμπερασματικά, αν αναζητούμε ένα αξιόπιστο και επαγγελματικού επιπέδου εργαλείο για τη μετατροπή της ομιλίας σε κείμενο και είμαστε διατεθειμένοι να επενδύσουμε σε αυτό, το Dragon Professional V16 αποτελεί μια εκπληκτική επιλογή. Αντιθέτως, για βασικές ανάγκες χωρίς την ανάγκη επιπλέον δαπανών, ο λογογράφος του Office 365 αποτελεί μια πρακτική και αποδοτική λύση. Τέλος, αν η προτεραιότητα μας εστιάζει στη δημιουργία voicerecorders από κείμενα, τότε το Speechelo, ιδίως η επαγγελματική του έκδοση Pro, είναι η ιδανικότερη επιλογή για να ικανοποιήσει τις ανάγκες μας σχετικά με αυτή την δραστηριότητα, την δημιουργία voicerecorders.

## **2.8 Μετρικές Αξιολόγησης των λογισμικών Φωνητικής Αναγνώρισης**

Σε αυτό το σημείο του κεφαλαίου, θα βρούμε κάποιες μετρικές αξιολόγησης τις οποίες θα εφαρμόσουμε στο επόμενο κεφάλαιο στο Λογισμικό Φωνητικής Αναγνώρισης που θα χρησιμοποιήσουμε για να αξιολογήσουμε την απόδοση του μέσα από την εκτέλεση δύο διακριτών σεναρίων χρήσης.

### 2.8.1 Word Error Rate - WER

Το WER αποτελεί την πιο διαδεδομένη μετρική για την αξιολόγηση της ακρίβειας συστημάτων αναγνώρισης ομιλίας και ανακαλύφθηκε από το Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας των ΗΠΑ. Το WER μετρά την ακρίβεια ενός συστήματος αναγνώρισης ομιλίας μέσω της σύγκρισης του αριθμού των λανθασμένων λέξεων στην αναγνωρισμένη έξοδο με τον συνολικό αριθμό των λέξεων στο σωστά μεταγραφόμενο κείμενο αναφοράς. Αυτή η μετρική συνήθως εκφράζεται ως ποσοστό, προσφέροντας μια άμεση ενδεικτική αξιολόγηση της επίδοσης του λογισμικού στο πεδίο της φωνητικής αναγνώρισης, ειδικά όταν το κύριο ενδιαφέρον είναι η μετατροπή προφορικής γλώσσας σε κείμενο.

Μέσω της εφαρμογής και της συγκριτικής ανάλυσης του WER, θα μπορέσουμε να διακρίνουμε την αποδοτικότητα και την αξιοπιστία των διαφόρων λογισμικών φωνητικής αναγνώρισης, καθοδηγώντας την επιλογή προς το λογισμικό που παρουσιάζει την υψηλότερη ακρίβεια και, κατ' επέκταση, την μεγαλύτερη αξιοπιστία στην ερμηνεία και την μεταγραφή προφορικών δεδομένων σε κείμενο.

Ο τύπος με τον οποίο μπορεί να υπολογιστεί το WER είναι:

$$\text{WER} = (S + D + I) / N$$

Όπου:

Το S αντιπροσωπεύει τον αριθμό των αντικαταστάσεων (λέξεις που αναγνωρίζονται εσφαλμένα).

Το D αντιπροσωπεύει τον αριθμό των διαγραφών (λέξεις που υπάρχουν στο αναφερόμενο κείμενο αλλά όχι στην αναγνωρισμένη έξοδο).

Το I αντιπροσωπεύει τον αριθμό των εισαγωγών (λέξεις που υπάρχουν στην αναγνωρισμένη έξοδο αλλά όχι στο αναφερόμενο κείμενο).

Το N αντιπροσωπεύει τον συνολικό αριθμό των λέξεων στο αναφερόμενο κείμενο (reference text) [18], [19]

Ας δούμε ένα παράδειγμα για να κατανοήσουμε καλύτερα πώς λειτουργεί το WER:



Κείμενο που ειπώθηκε από τον χρήστη: « Μια αλεπού χρώματος καφέ πήδηξε πάνω από ένα τρελό σκύλο»

Μεταγραφόμενο Κείμενο: «Μια που καφέ πήδηξε στον τρελό σκύλο μας»

Σε αυτό το απλό παράδειγμα βλέπουμε ότι υπάρχουν κάποιες διαφορές μεταξύ του μεταγραφόμενου κειμένου και του κειμένου που ειπώθηκε από τον χρήστη.

Πιο συγκεκριμένα, σύμφωνα με τις τρεις αυτές κατηγορίες λαθών προκύπτουν τα εξής σε κάθε κατηγορία:

- Αντικαταστάσεις(S): Η λέξη «αλεπού» αναγνωρίστηκε εσφαλμένα ως η λέξη « που» , η λέξη «πάνω» αναγνωριστικέ εσφαλμένα ως η λέξη «στον».
- Διαγραφές(D): Οι λέξεις «χρώματος», «ένα» και «από» δεν αναγράφηκαν στο μεταγραφόμενο κείμενο.
- Εισαγωγές(I) Η λέξη «μας» έχει εισαχθεί επιπρόσθετα στο μεταγραφόμενο κείμενο.

Επομένως προκύπτει:

- $S = 2$ .
- $D = 4$ .
- $I = 1$ .
- $N = 10$ . (πλήθος λέξεων του κειμένου αναφοράς, δηλαδή του κειμένου που ειπώθηκε από τον χρήστη.)

Θα αντικαταστήσουμε τις πιο πάνω τιμές στον τύπο που υπολογίζει το ποσοστό σφάλματος λέξης (WER)

$$WER = (S + D + I) / N$$

$$WER = (2 + 4 + 1) / 10 = 7 / 10 = 0.7$$

Έτσι, σε αυτό το παράδειγμα, ο ρυθμός σφάλματος λέξης είναι περίπου 70%, που υποδηλώνει ότι περίπου 70% των λέξεων στην μεταγραφόμενη έξοδο διαφέρουν από το κείμενο που ειπώθηκε από τον χρήστη. Χαμηλότερες τιμές WER υποδεικνύουν καλύτερη ακρίβεια στο σύστημα αναγνώρισης ομιλίας.

### **2.8.2 Κανονικοποίηση και Προεπεξεργασία των Μεταγραφών**

Στη διαδικασία υπολογισμού του Word Error Rate (WER) και άλλων σημαντικών μετρικών, η κανονικοποίηση του κειμένου αναδεικνύεται ως ένας κρίσιμος παράγοντας που επηρεάζει σημαντικά την ακρίβεια των αποτελεσμάτων των μεταγραφών από τα συστήματα φωνητικής αναγνώρισης. Η διαδικασία αυτή περιλαμβάνει την τροποποίηση του αρχικού κειμένου με στόχο τη δημιουργία ενός ενιαίου προτύπου που θα διευκολύνει τη σύγκριση μεταξύ των αυτόματα παραχθέντων μεταγραφών και των αντίστοιχων μεταγραφών αναφοράς.

Η κανονικοποίηση μπορεί να περιλαμβάνει τη μετατροπή όλων των γραμμάτων σε πεζά, την αφαίρεση ή την προσθήκη σημείων στίξης, την ενοποίηση των αριθμών (είτε ως λέξεων είτε ως αριθμητικών συμβόλων), καθώς και την τροποποίηση ή την αντικατάσταση ειδικών όρων ή ονομάτων. Αυτή η διαδικασία διασφαλίζει ότι οι συγκρίσεις που γίνονται μεταξύ των μεταγραφών είναι δίκαιες και ακριβείς, ελαχιστοποιώντας τα σφάλματα που μπορεί να προκύψουν λόγω διαφορών στη μορφοποίηση ή στην παρουσίαση του κειμένου.

Η σημασία της κανονικοποίησης κειμένου είναι αδιαμφισβήτητη, καθώς μπορεί να επηρεάσει άμεσα την ποιότητα και την αξιοπιστία των συμπερασμάτων που εξάγονται από την αξιολόγηση των συστημάτων ASR. Έτσι, η προσεκτική εφαρμογή της κανονικοποίησης αποτελεί ένα θεμελιώδες βήμα στη διαδικασία αξιολόγησης, επιτρέποντας τη συλλογή πιο ακριβών και συγκρίσιμων δεδομένων που θα βοηθήσουν στη βελτίωση και στην ανάπτυξη των τεχνολογιών φωνητικής αναγνώρισης. [18]

Ας δούμε ένα παράδειγμα

#### **Αρχική Μεταγραφή:**

Ομιλητής 1: Γεια σου, με λένε Τάσο! Πώς είσαι σήμερα;

Ομιλητής 2: Καλά είμαι, Τάσο. Χαίρομαι που σε ξανά βλέπω. Ο αριθμός τηλεφώνου μου είναι 777-1234.

## **Κανονικοποιημένη και Προεπεξεργασμένη Μεταγραφή:**

ομιλητής 1 γεια σου με λένε τάσο πώς είσαι σήμερα

ομιλητής 2 καλά είμαι τάσο χαίρομαι που σε ξανά βλέπω ο αριθμός τηλεφώνου μου είναι <αριθμός\_τηλεφώνου>

Η διαδικασία που ακολουθείται για την κανονικοποίηση μιας μεταγραφής είναι η εξής:

1. **Μετατροπή σε πεζά:** Μετατρέπουμε ολόκληρο το κείμενο σε πεζά γράμματα για να εξασφαλίσουμε ομοιομορφία και για να είναι ανεξάρτητο από την πεζογραφία.
  - Αρχικό: "Γεια σου, με λένε Τάσο! Πώς είσαι σήμερα;"
  - Προεπεξεργασμένο: "ομιλητής 1 γεια σου με λένε τάσο πώς είσαι σήμερα"
2. **Αφαίρεση Σημείων Στίξης:** Αφαιρούμε όλα τα σημεία στίξης, τα οποία συχνά δεν είναι σχετικά για ορισμένες εργασίες επεξεργασίας φυσικής γλώσσας.
  - Αρχικό: "Γεια σου, με λένε Τάσο! Πώς είσαι σήμερα;"
  - Προεπεξεργασμένο: "ομιλητής 1 γεια σου με λένε τάσο πώς είσαι σήμερα"
3. **Χειρισμός Αριθμών:** Αντικαθιστούμε αριθμούς τηλεφώνου ή οποιαδήποτε ευαίσθητη αριθμητική πληροφορία με ένα σύμβολο αντικατάστασης όπως <αριθμός\_τηλεφώνου> για να διατηρήσουμε το απόρρητο του χρήστη.
  - Αρχικό: "Ο αριθμός τηλεφώνου μου είναι 777-1234."
  - Προεπεξεργασμένο: "ομιλητής 2 καλά είμαι τάσο χαίρομαι που σε ξανά βλέπω ο αριθμός τηλεφώνου μου είναι <αριθμός\_τηλεφώνου>"
4. **Ετικέτες Ομιλητή:** Αν υπάρχουν ετικέτες ομιλητή, καλό θα ήταν να τις διατηρήσουμε για περαιτέρω ανάλυση ή για διευκρίνιση, όμως μπορεί να επιθυμούμε να τις μετατρέψουμε πεζά για να διατηρήσουμε την ομοιομορφία.
  - Αρχικό: "Ομιλητής 1: Γεια σου, με λένε Τάσο!"
  - Προεπεξεργασμένο: "ομιλητής 1 γεια σου με λένε τάσο"

5. **Διακριτική Διάσπαση:** Διαχωρίζουμε το κείμενο σε μεμονωμένες λέξεις ή τμήματα (tokens). Σε αυτό το παράδειγμα, διαχωρίσαμε το κείμενο σε λέξεις για περαιτέρω ανάλυση. [18]

Συνοψίζοντας, τα βήματα προεπεξεργασίας που προτείναμε μπορεί να χρειαστεί να προσαρμοστούν ανάλογα με τους συγκεκριμένους στόχους και τις απαιτήσεις της εκάστοτε δραστηριότητας. Ενδεχομένως, σε κάποιες περιπτώσεις, να απαιτηθεί η διατήρηση ορισμένων σημείων στίξης ή η διαφορετική διαχείριση των αριθμών, ενδεικτικά. Το ουσιώδες είναι η ευελιξία και η προσαρμοστικότητα της προεπεξεργασίας κειμένου, η οποία θα πρέπει να σχεδιάζεται και να εφαρμόζεται με γνώμονα τις ειδικές ανάγκες κάθε πρότζεκτ, καθώς και τον τύπο και τη μορφή των δεδομένων που είναι διαθέσιμα. Αυτή η προσαρμοστική προσέγγιση εξασφαλίζει την βέλτιστη αξιοποίηση των δυνατοτήτων του συστήματος ASR και την πιο ακριβή ερμηνεία των αποτελεσμάτων, ανταποκρινόμενη αποτελεσματικά στις συγκεκριμένες προκλήσεις κάθε εφαρμογής.

### **2.8.3 Έξι βήματα για την εκτέλεση αξιολόγησης συστήματος ASR**

Σε αυτή την ενότητα, θα εξετάσουμε τη διαδικασία αξιολόγησης ενός συστήματος Αυτόνομης Φωνητικής Αναγνώρισης (ASR) μέσω ενός δομημένου πλαισίου που ακολουθεί έξι βασικά βήματα. Αυτή η διαδικασία είναι κρίσιμη για την κατανόηση της αξιοπιστίας και της ακρίβειας του λογισμικού ASR, προσφέροντας μια ολοκληρωμένη εικόνα της επίδοσής του. [18]

#### **Βήμα 1: Επιλογή Δείγματος Ηχογραφημένης Ομιλίας**

Ξεκινάμε με την επιλογή ενός κατάλληλου δείγματος ηχογραφημένης ομιλίας. Αυτό πρέπει να περιλαμβάνει ένα εύρος από 25 έως 50 λέξεις ανά εκφώνηση, δίνοντας προσοχή στην εκφώνηση, η οποία θα πρέπει να είναι σαφής και χωρίς σιωπηλές παύσεις. Είναι σημαντικό να επιλεγεί ένα δείγμα που αντιπροσωπεύει την ποικιλομορφία της φυσικής ομιλίας, συμπεριλαμβανομένων των διακοπών, επανεκκινήσεων και μεταβάσεων μεταξύ διαφορετικών ιδεών.

#### **Βήμα 2: Χειροκίνητη Μεταγραφή**

Στη συνέχεια, πραγματοποιούμε μια προσεκτική χειροκίνητη μεταγραφή του επιλεγμένου ηχητικού δείγματος. Αυτή η διαδικασία, παρόλο που είναι χρονοβόρα, κρίνεται απαραίτητη για την δημιουργία αξιόπιστων μεταγραφών αναφοράς που θα χρησιμοποιηθούν στην σύγκριση με τις αυτόματες μεταγραφές.

#### **Βήμα 3: Εκτέλεση της Υπηρεσίας ASR**

Το επόμενο βήμα αφορά την εισαγωγή του ηχητικού αρχείου στο σύστημα ASR για την παραγωγή της αυτόματης μεταγραφής. Η χρήση εργαλείων όπως το Amazon Transcribe μπορεί να διευκολύνει την διαχείριση μεγάλων όγκων δεδομένων.

#### **Βήμα 4: Κανονικοποίηση των Αποτελεσμάτων του ASR**

Οι αυτόματες μεταγραφές πρέπει να κανονικοποιηθούν, προκειμένου να είναι ενιαίος ο τρόπος που παρουσιάζονται τα αποτελέσματα. Αυτό περιλαμβάνει τη μετατροπή σε πεζά γράμματα, την αφαίρεση σημείων στίξης και την αντικατάσταση ορισμένων στοιχείων με συμβολικές αναπαραστάσεις.

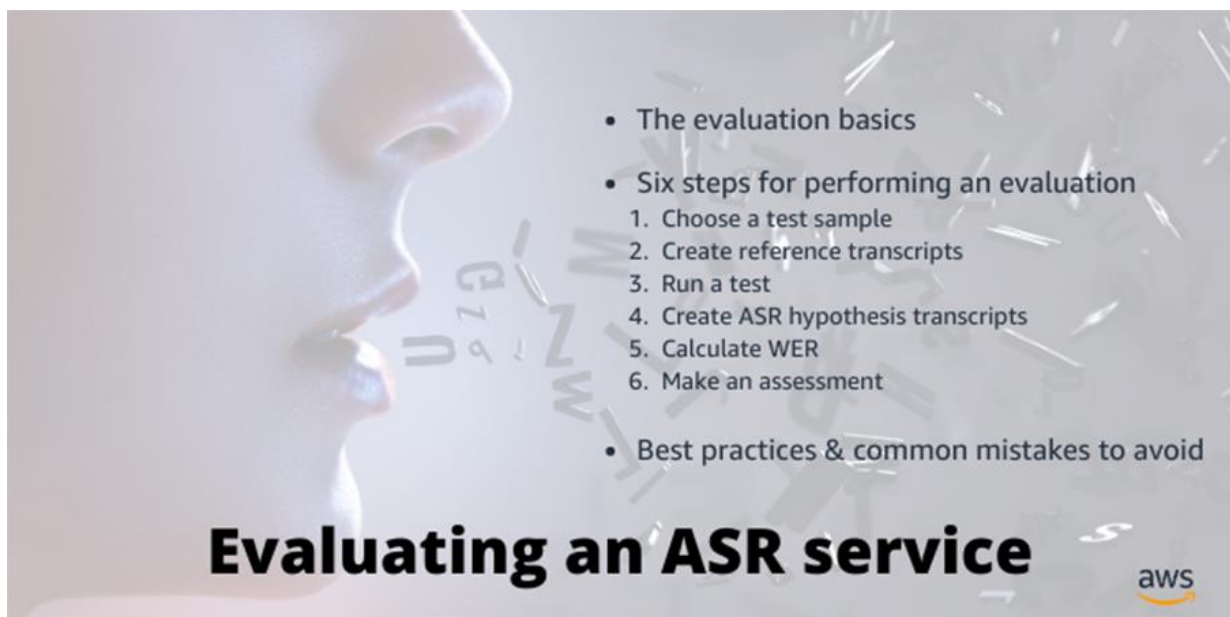
## **Βήμα 5: Υπολογισμός του WER με Εργαλείο Ανοιχτού Κώδικα**

Χρησιμοποιούμε ένα εργαλείο ανοιχτού κώδικα, όπως το Python πακέτο `jwer`, για να υπολογίσουμε το Word Error Rate (WER) μεταξύ των μεταγραφών αναφοράς και των αυτόματων μεταγραφών.

## **Βήμα 6: Αξιολόγηση χρησιμοποιώντας το WER**

Το WER μας παρέχει μια καθοριστική μετρική για την αξιολόγηση της απόδοσης του συστήματος ASR. Ένα χαμηλό WER σημαίνει υψηλότερη ακρίβεια του συστήματος, ενώ τα αποτελέσματα θα πρέπει να αναλυθούν με βάση τις ανάγκες του χρήστη και τα χαρακτηριστικά του υπολογιστικού περιβάλλοντος.

Καταλήγοντας, η εφαρμογή αυτού του πλαισίου αξιολόγησης επιτρέπει μια διεξοδική και ακριβή εκτίμηση της απόδοσης ενός συστήματος ASR, παρέχοντας πολύτιμες πληροφορίες για τη βελτίωση και την περαιτέρω ανάπτυξή του.



Πηγή: [18] Walker, S., 2019. "Evaluating an Automatic Speech Recognition Service". Available at: <https://aws.amazon.com/blogs/machine-learning/evaluating-an-automatic-speech-recognition-service/>

## Κεφάλαιο 3: Πρώτο Σενάριο Χρήσης: Πλοήγηση με τη χρήση του Εργαλείου Dragon Professional V16

---

3.1 Αναβαθμίσεις από προηγούμενες εκδόσεις	39
3.2 Πρώτο Σενάριο Χρήσης: Πλοήγηση και διαχείριση αρχείων σε υπολογιστή με χρήση του Dragon Professional V16	40
3.2.1 Άνοιγμα/ Κλείσιμο Εφαρμογών	41
3.2.2 Στιγμιότυπα Οθωνών από τις διάφορες εκτελέσεις	41
3.2.3 Πλοήγηση αρχείων	44
3.2.4 Στιγμιότυπα Οθωνών από την εκτέλεση των πιο πάνω βημάτων	45
3.2.5 Πλοήγηση στο διαδίκτυο	48
3.2.6 Αναζήτηση πτήσεων προς Ζυρίχη μέσω της ιστοσελίδας booking.com	48
3.2.7 Στιγμιότυπα Οθωνών από την εκτέλεση των πιο πάνω βημάτων.	49

---

Σε αυτό το κεφάλαιο, θα διεξάγουμε μια ενδελεχή αξιολόγηση του λογισμικού φωνητικής αναγνώρισης Dragon Professional V16 μέσω ειδικά σχεδιασμένων σεναρίων χρήσης. Ο σκοπός μας είναι να κατανοήσουμε τον βαθμό ευχρηστίας και αποδοτικότητας του εργαλείου για άτομα με δυσκολίες στην εύκολη και γρήγορη αλληλεπίδραση με τον υπολογιστή κατά την εκτέλεση των δραστηριοτήτων τους.

Η αξιολόγηση θα διενεργηθεί μέσω δύο διακριτών σεναρίων χρήσης:

1. **Σενάριο Αλληλεπίδρασης με τον Υπολογιστή:** Θα χρησιμοποιήσουμε το Dragon Professional V16 για την εκτέλεση διαφόρων δραστηριοτήτων στον υπολογιστή, καθώς και για την πλοήγηση σε διάφορα προγράμματα και ιστοσελίδες. Αυτό το σενάριο έχει στόχο να διαπιστώσει πώς ανταποκρίνεται το λογισμικό στην εκτέλεση φωνητικών εντολών που σχετίζονται με την αλληλεπίδραση του χρήστη με τον υπολογιστή, χρησιμοποιώντας αποκλειστικά την φωνή.

2. **Σενάριο Προγραμματισμού:** Στο κεφάλαιο 4 θα αναπτύξουμε ένα σύντομο προγραμματιστικό κώδικα με σκοπό την υπαγόρευση αυτού στο Dragon Professional V16. Αυτό θα μας επιτρέψει να αξιολογήσουμε την αξιοπιστία και την ακρίβεια του λογισμικού στον χειρισμό μεταγραφών υπαγόρευσης εντολών σε τρεις διαφορετικές γλώσσες προγραμματισμού: C, Java και Python. Αυτό το σενάριο αποσκοπεί στη διερεύνηση της δυνατότητας του εργαλείου να διευκολύνει την συγγραφή κώδικα μέσω φωνητικών εντολών, προσφέροντας σημαντική εξοικονόμηση χρόνου και μειωμένη χρήση του πληκτρολογίου για διορθώσεις.

Μέσω αυτών των σεναρίων, στοχεύουμε να αποκτήσουμε μια ολοκληρωμένη εικόνα της πρακτικής αξίας και της ευρωστίας του Dragon Professional V16 σε πραγματικές συνθήκες χρήσης, επικεντρώνοντας τόσο στην προγραμματιστική συγγραφή όσο και στην καθημερινή αλληλεπίδραση μας με τον υπολογιστή.

### **3.1 Αναβαθμίσεις από προηγούμενες εκδόσεις**

Το Dragon Professional V16 αντιπροσωπεύει ένα σημαντικό άλμα προς τα εμπρός από τις προηγούμενες εκδόσεις του, με πολλές βελτιώσεις που ανεβάζουν την απόδοσή του. Βασική αναβάθμιση είναι η βελτιωμένη ακρίβεια στην αναγνώριση ομιλίας. Το λογισμικό καταγράφει και μεταγράφει τώρα τις προφορικές λέξεις με αυξημένη ακρίβεια, αντιμετωπίζοντας αποτελεσματικά μια ποικιλία προκλήσεων ομιλίας. Επιπλέον, επεξεργάζεται ομιλία σε κείμενο πιο γρήγορα, βελτιώνοντας την αποτελεσματικότητα στις ροές εργασίας. Το λογισμικό έχει επίσης εξελιχθεί ώστε να προσαρμόζεται καλύτερα και να μαθαίνει από τη συμβολή του χρήστη, βελτιώνοντας συνεχώς την ακρίβειά του. Η ενσωμάτωση του με διάφορες εφαρμογές και λειτουργικά συστήματα έχει βελτιωθεί για πιο ομαλή, πιο φιλική προς το χρήστη αλληλεπίδραση. Αυτή η έκδοση διευρύνει τις γλωσσικές δυνατότητες του εργαλείου, προσελκύοντας μια πιο διαφοροποιημένη, διεθνή βάση χρηστών. Η επέκταση των δυνατοτήτων των φωνητικών εντολών προσφέρει μεγαλύτερο έλεγχο και ευκολία στην πλοήγηση στον υπολογιστή. Τα ενισχυμένα μέτρα ασφαλείας είναι μια άλλη σημαντική βελτιστοποίηση, που διασφαλίζει την εμπιστευτικότητα των ευαίσθητων προσωπικών δεδομένων. Αυτές οι βελτιώσεις καθιστούν το Dragon Professional V16 πιο ικανό, ευέλικτο και ασφαλές εργαλείο αναγνώρισης φωνής, ξεπερνώντας σημαντικά τις



δυνατότητες των προκατόχων του όσον αφορά την εμπειρία χρήστη και τη λειτουργικότητα σε σύγκριση με άλλα παρόμοια εργαλεία [11], [12].

### **3.2 Πρώτο Σενάριο Χρήσης: Πλοήγηση και διαχείριση αρχείων σε υπολογιστή με χρήση του Dragon Professional V16**

**Στόχος:** Αξιολόγηση της αποδοτικότητας του Dragon Professional V16 στην πλοήγηση σε λειτουργικό σύστημα Windows, στη διαχείριση αρχείων και στην εκτέλεση κοινών εργασιών χρησιμοποιώντας φωνητικές εντολές.

#### **Περιβάλλον δοκιμής:**

- ✓ Φορητός Υπολογιστής με λειτουργικό σύστημα Windows 11 pro
- ✓ Το Dragon Professional V16 εγκαταστάθηκε και ρυθμίστηκε κατάλληλα
- ✓ Χρήση μικροφώνου υπολογιστή
- ✓ Γλώσσα Εκτέλεσης Φωνητικών Εντολών Αγγλική

#### **Βήματα που απαιτούνται πριν την εκτέλεση του σεναρίου χρήσης:**

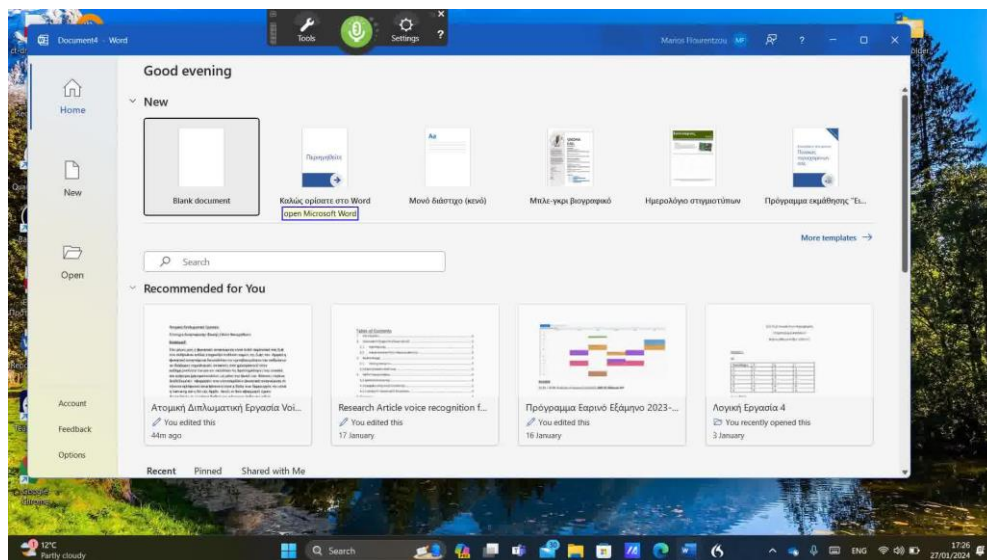
1. Βεβαιωνόμαστε ότι το Dragon Professional V16 είναι σωστά εγκατεστημένο και κατάλληλα ρυθμισμένο για αναγνώριση φωνής.
2. Εξοικειωνόμαστε με τις συγκεκριμένες φωνητικές εντολές για πλοήγηση και διαχείριση αρχείων χρησιμοποιώντας τον οδηγό χρήστη που μας παρέχει το Dragon [13].
3. Ανοίγουμε μια ποικιλία εφαρμογών που χρησιμοποιούνται συνήθως, όπως το Microsoft Word, ένα πρόγραμμα περιήγησης ιστού και ένα παράθυρο εξερεύνησης αρχείων για μια γρήγορη δοκιμή.

### 3.2.1 Άνοιγμα/ Κλείσιμο Εφαρμογών

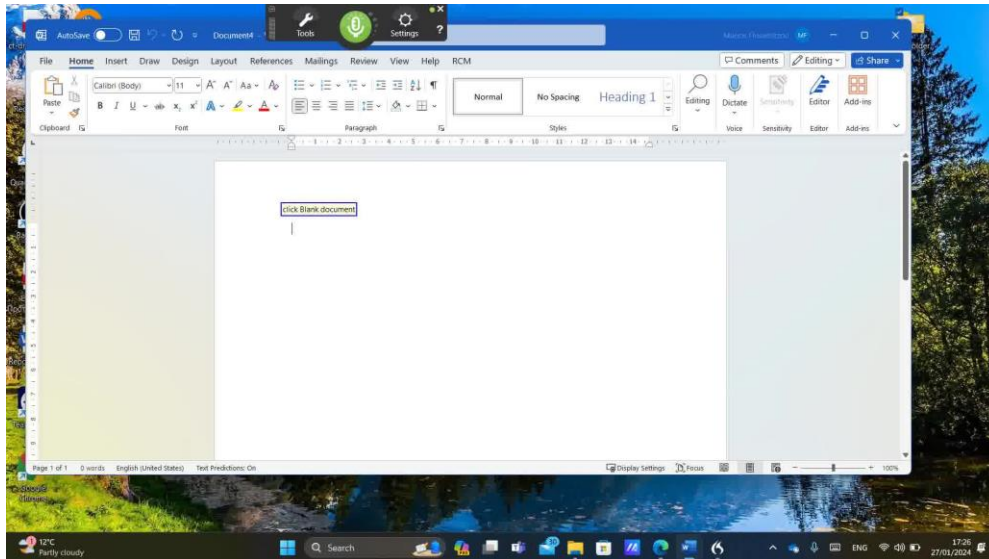
#### Μεθοδολογία

Όταν βρεθούμε στην επιφάνεια εργασίας, χρησιμοποιούμε μια φωνητική εντολή όπως "Open Microsoft Word" για να εκκινήσουμε την εφαρμογή και την φωνητική εντολή close window, για να κλείσει το παράθυρο που είναι ενεργό. Στην συνέχεια εκτελούμε άλλες παρόμοιες φωνητικές εντολές για να εκκινήσουμε και να τερματίσουμε κι άλλες εφαρμογές. [13]

### 3.2.2 Στιγμιότυπα Οθωνών από τις διάφορες εκτελέσεις



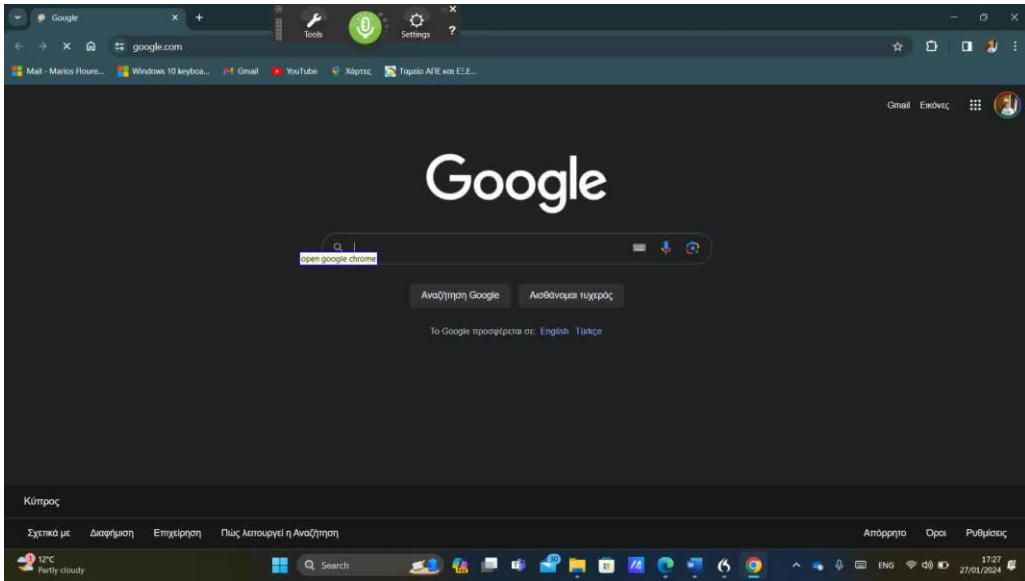
Εικόνα 3.2.2.1: Opening Microsoft Word via voice activated command - "Open Microsoft Word"



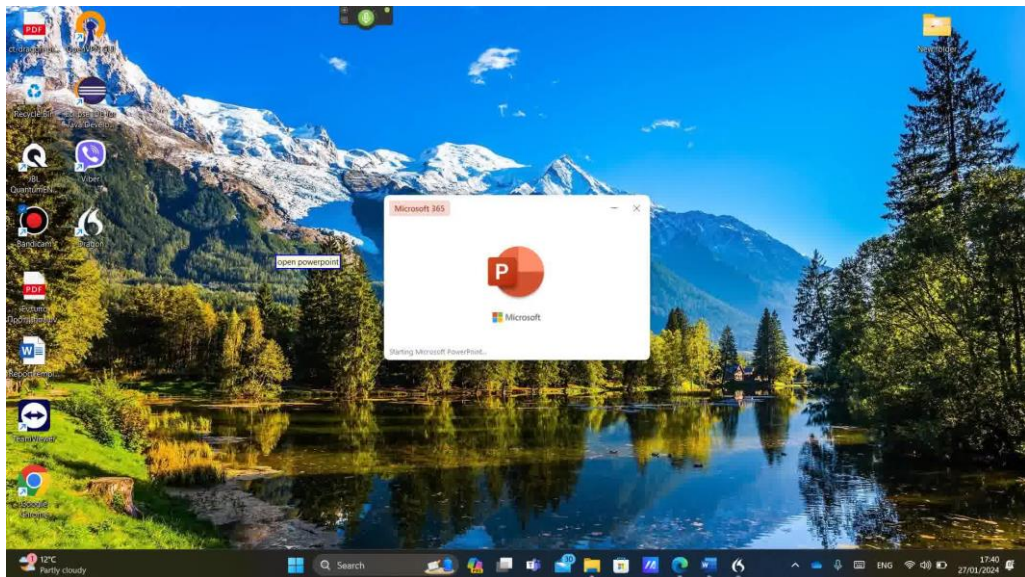
*Εικόνα 3.2.2.2: Microsoft Word launched via voice activated command*



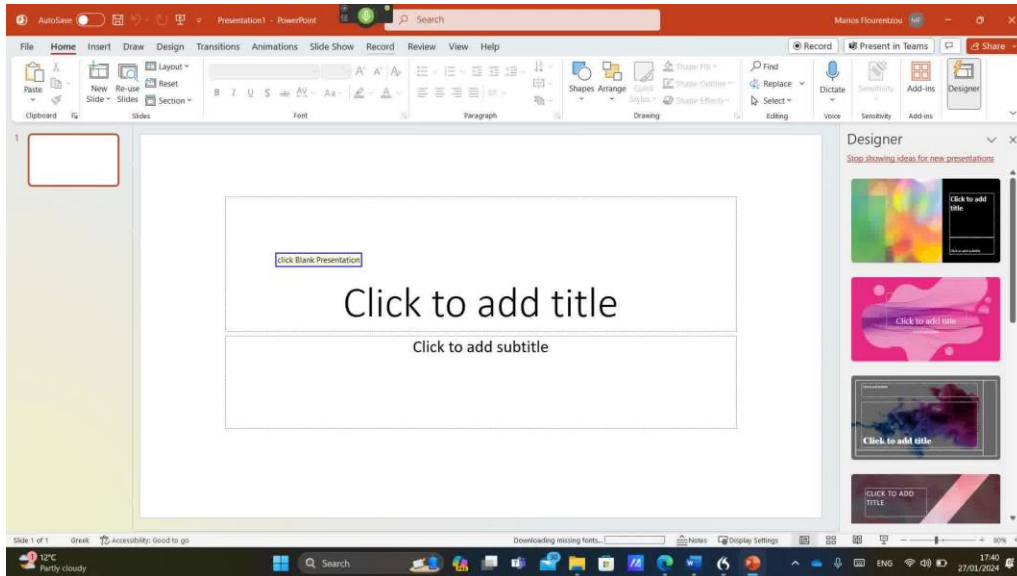
*Εικόνα 3.2.2.3: Microsoft Word terminated via voice activated command – “Close Window”*



*Εικόνα 3.2.2.4: Launching Google Chrome via voice activated command – “Open Google Chrome”*



*Εικόνα 3.2.2.5: Launching Microsoft Powerpoint – “Open Microsoft Powerpoint”*



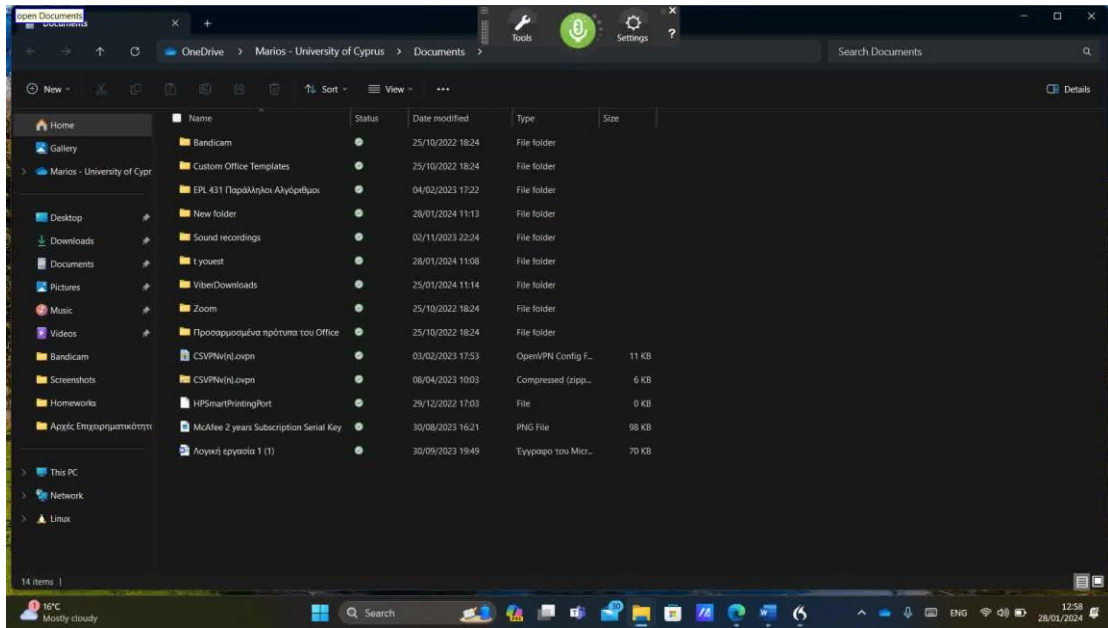
Εικόνα 3.2.2.6: Launching a blank presentation – “Click Blank Presentation”

### 3.2.3 Πλοήγηση αρχείων

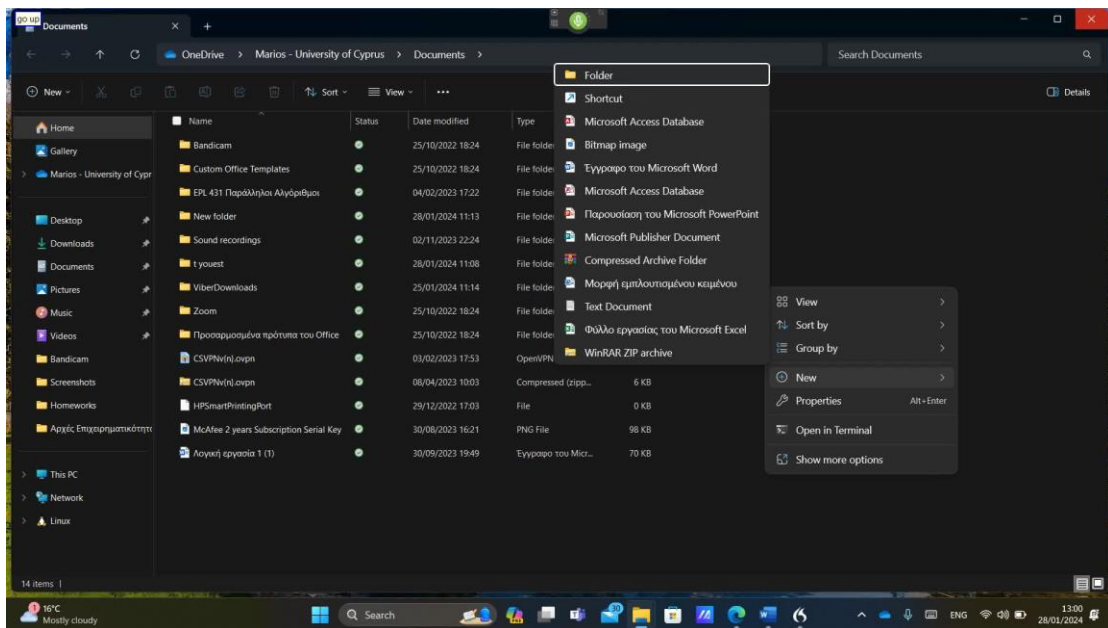
Θα χρησιμοποιήσουμε φωνητικές εντολές για να πλοηγηθούμε σε έναν συγκεκριμένο φάκελο του file explorer. [13]

1. Θα χρησιμοποιήσουμε την φωνητική εντολή «**Open Documents**» για να μεταβούμε στον φάκελο **Documents**.
2. Θα δημιουργήσουμε ένα καινούργιο φάκελο χρησιμοποιώντας φωνητικές εντολές για το mouse όπως: **mouse right click** → **mouse click New** → **mouse click folder**. Για να δημιουργήσουμε τον καινούργιο φάκελο.
3. Θα δώσουμε το όνομα «**TestFolder**» στον φάκελο.
4. Θα μεταβούμε σε αυτόν κάνοντας χρήση της εντολής «**mouse double click.**»
5. Θα μεταβούμε έξω από τον φάκελο για να τον διαγράψουμε κάνοντας χρήση της εντολής «**go back.**»
6. Θα διαγράψουμε τον φάκελο «**TestFolder**» χρησιμοποιώντας φωνητικές εντολές όπως **mouse right click** on «**TestFolder**» → **mouse click delete** και μετέπειτα **mouse click** «**yes**» στο προειδοποιητικό μήνυμα που θα εμφανιστεί για να ολοκληρωθεί η διαγραφή του φακέλου.

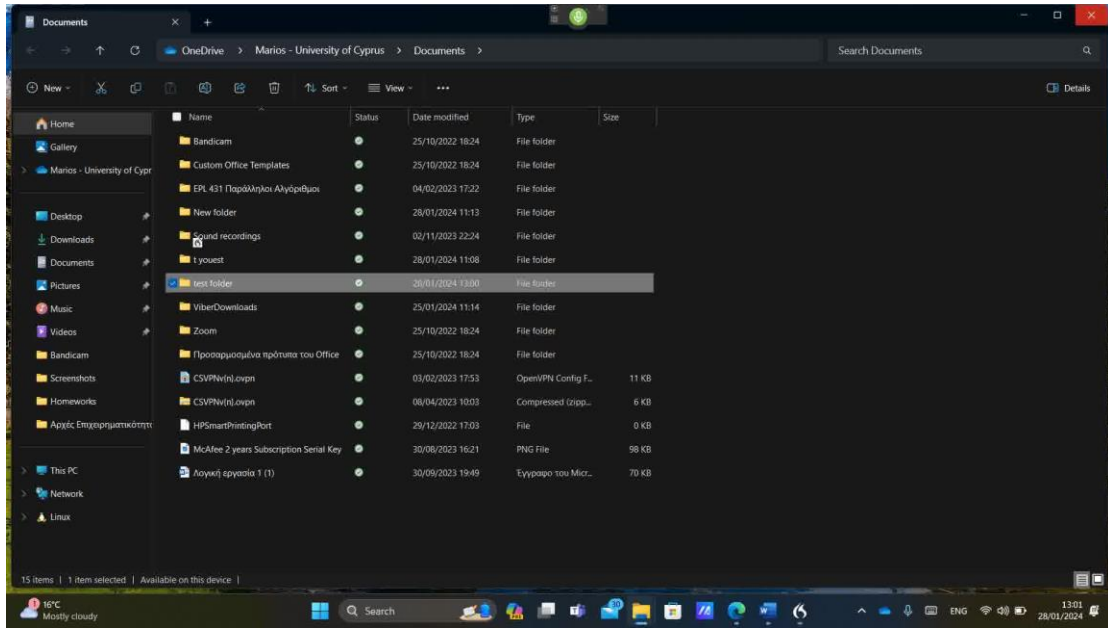
### 3.2.4 Στιγμιότυπα Οθονών από την εκτέλεση των πιο πάνω βημάτων



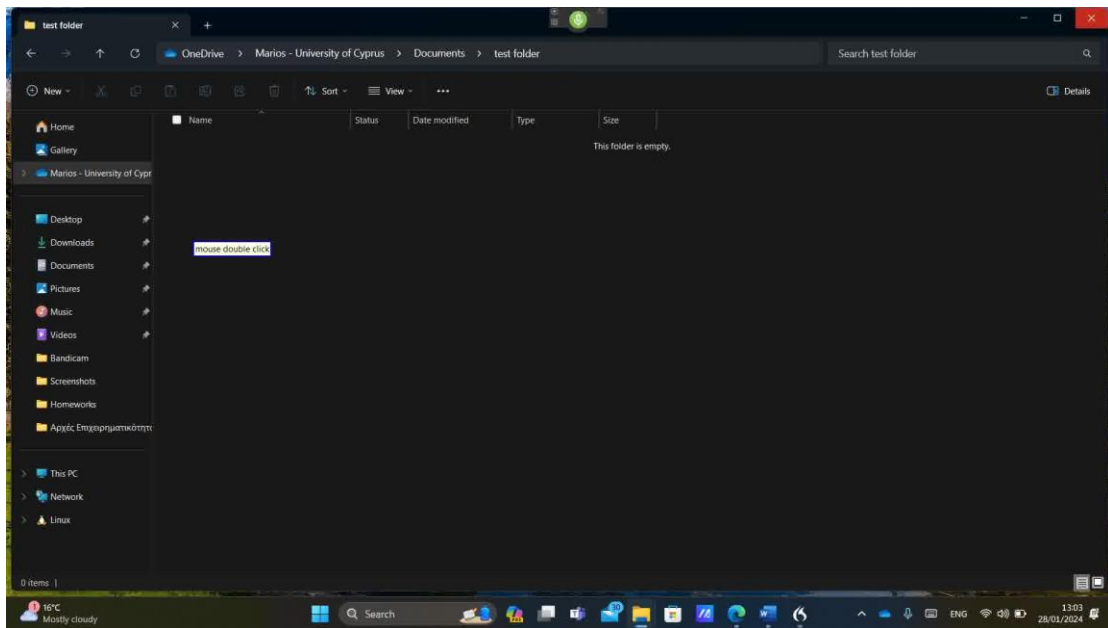
Εικόνα 3.2.4.1: Launching Documents – “Open Documents”



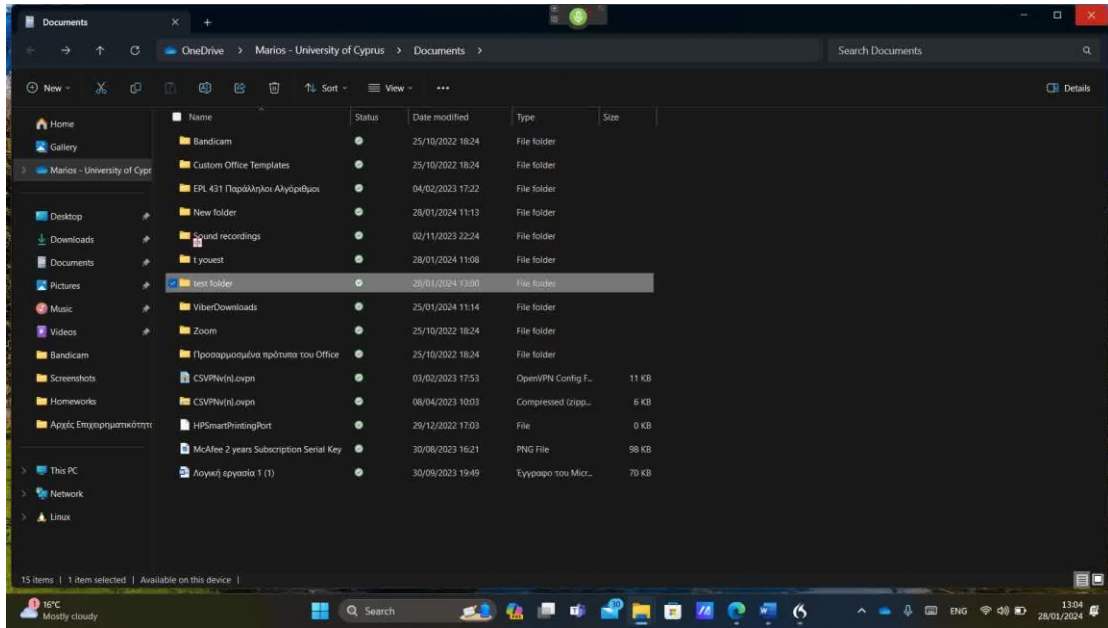
Εικόνα 3.2.4.2: Create a New Folder – Using Mouse Voice Commands



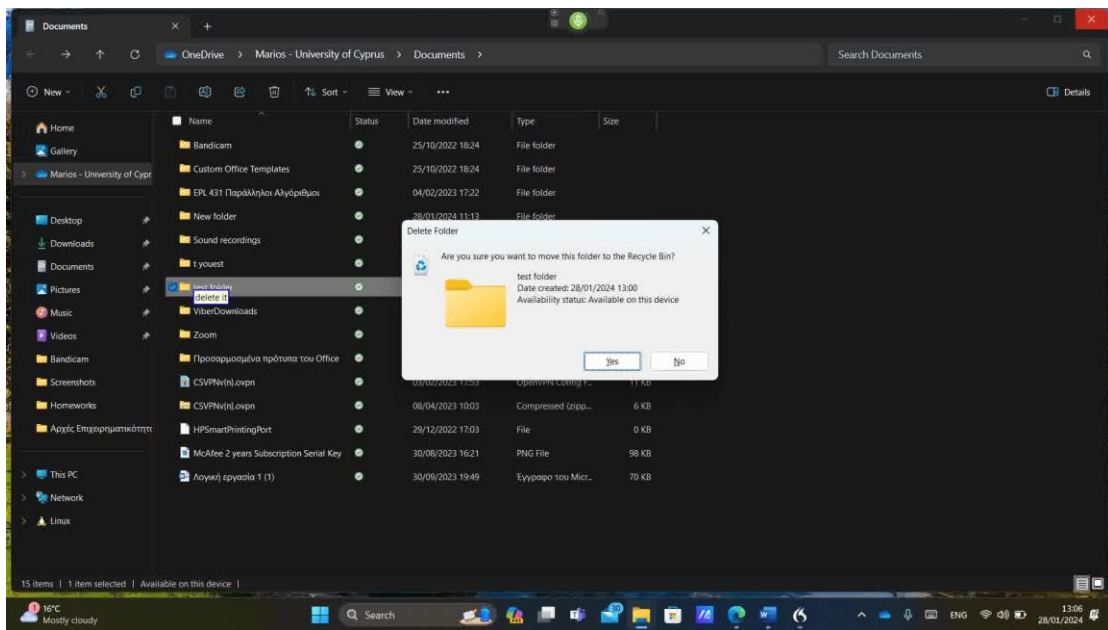
*Εικόνα 3.2.4.3: Rename the Folder – “Test Folder”*



*Εικόνα 3.2.4.4: Launching Test Folder via Voice Command – “mouse double click”*

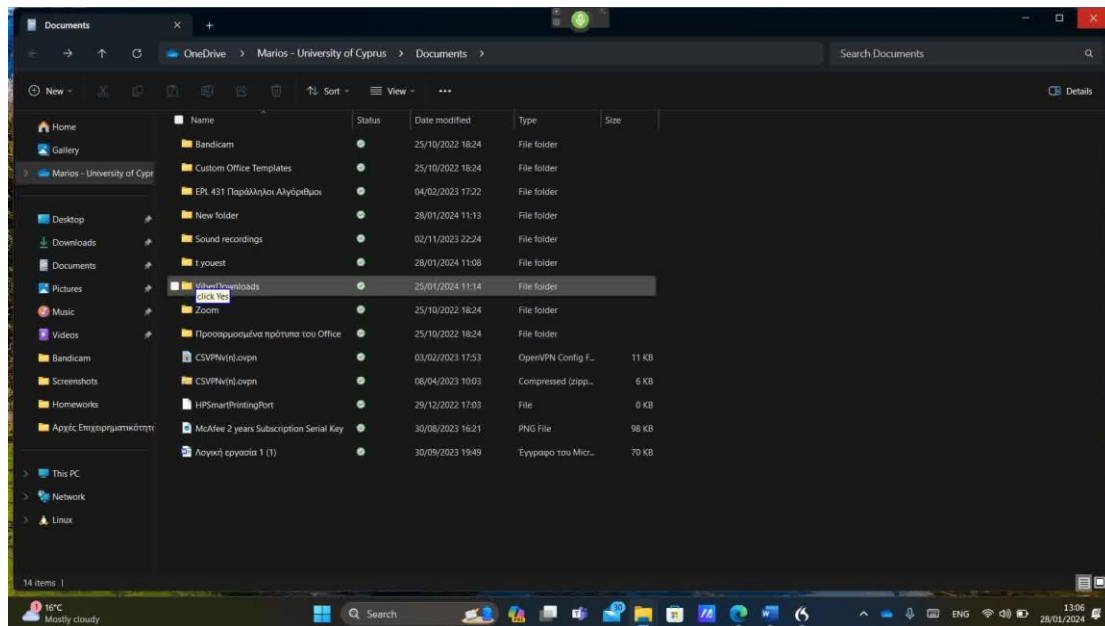


Εικόνα 3.2.4.5: Go back from Test Folder via Voice Command – “Go back”



Εικόνα 3.2.4.6: Delete Test Folder – Using Mouse Voice Commands





Εικόνα 3.2.4.7: Test Folder deleted from Documents.

### 3.2.5 Πλοήγηση στο διαδίκτυο

Το Dragon Professional μας παρέχει την δυνατότητα μέσω της επέκτασης που διαθέτει για τους web browsers να πλοηγούμαστε με ευκολία σε διάφορες ιστοσελίδες στο διαδίκτυο κάνοντας χρήση των φωνητικών εντολών που αφορούν στην πλοήγηση ειδικά για το διαδίκτυο. [13]

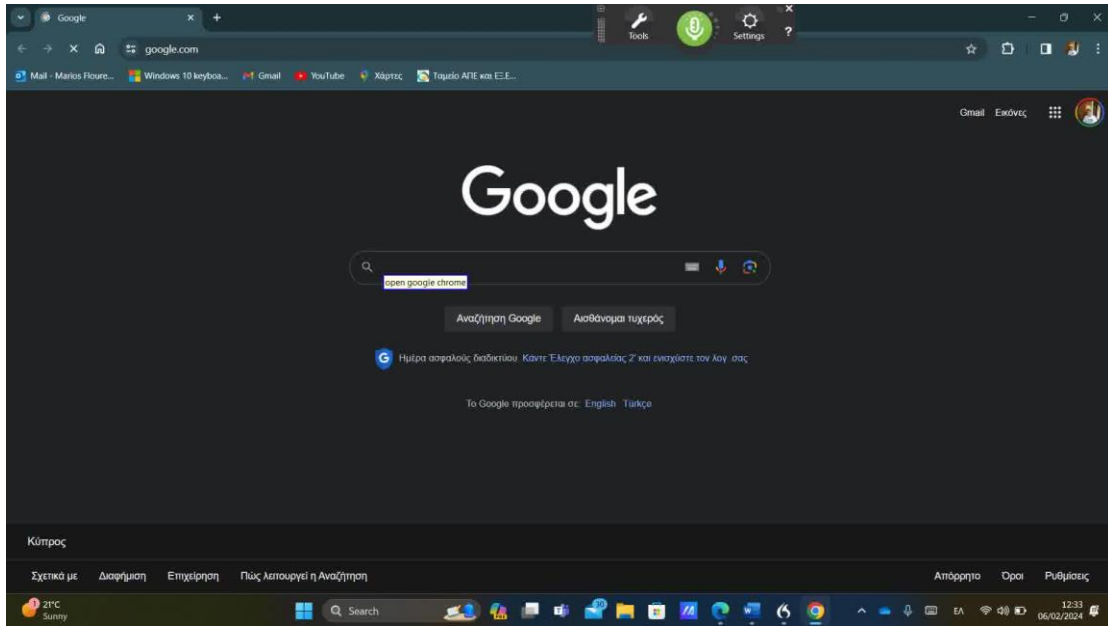
### 3.2.6 Αναζήτηση πτήσεων προς Ζυρίχη μέσω της ιστοσελίδας booking.com

#### Μεθοδολογία Εκτέλεσης

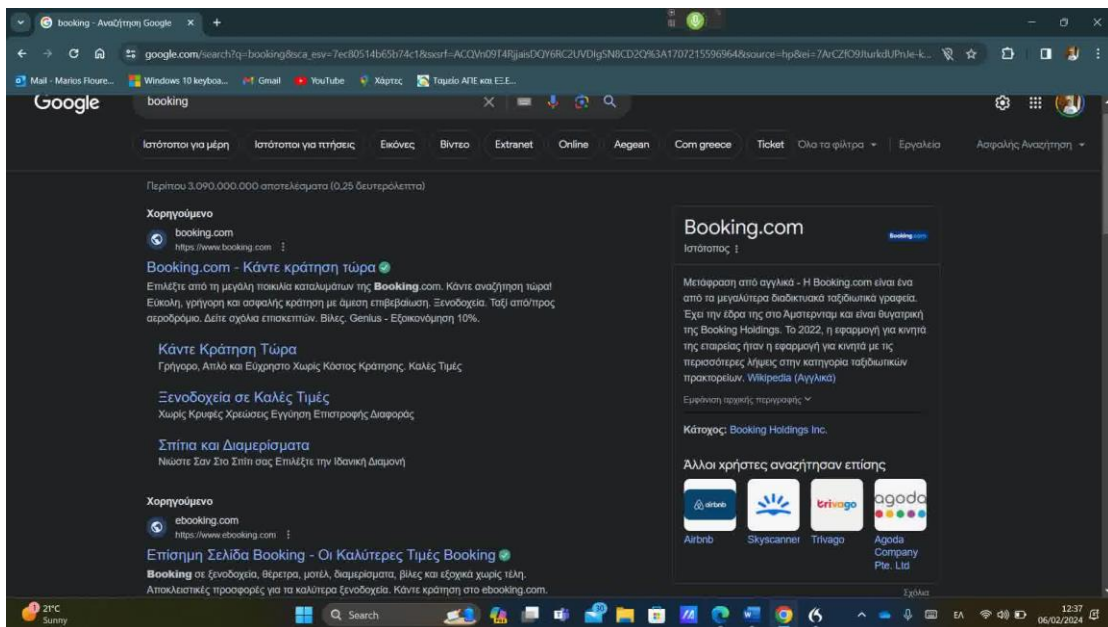
1. Θα ανοίξουμε το Google Chrome κάνοντας χρήση της φωνητικής εντολής «**Open Google Chrome**».
2. Θα αναζητήσουμε στην Μηχανή αναζήτησης του Google με την βοήθεια του Dragon Professional την ιστοσελίδα booking.com
3. Θα μετάρουμε στην ιστοσελίδα με χρήση της φωνητικής εντολής «**Click booking.com**».

4. Θα πατήσουμε στο κουμπί «**flights**» χρησιμοποιώντας την φωνητική εντολή «**Click flights**» για να αναζητήσουμε διαθέσιμες πτήσεις από Λάρνακα προς Ζυρίχη καταχωρώντας τυχαία κάποιες ημερομηνίες.
5. Θα κάνουμε αναζήτηση αποτελεσμάτων χρησιμοποιώντας την φωνητική εντολή «**Click Search**».
6. Από τα αποτελέσματα αναζήτησης επιλέγουμε την πτήση που επιθυμούμε πατώντας στο αντίστοιχο κουμπί με χρήση της εντολής «**Click View Details**».
7. Στην ιστοσελίδα υπάρχουν περισσότερα από ένα κουμπιά με την ονομασία «**View Details**» έτσι το Dragon Professional για να μας διευκολύνει να προσδιορίσουμε σε ποιο θέλουμε να μετάβουμε θα μας εμφανίσει ένα αριθμημένο τετραγωνάκι δίπλα από κάθε κουμπί. Θα επιλέξουμε το τετραγωνάκι με τον αριθμό 2 λέγοντας του την φωνητική εντολή «**Choose 2**» για να μετάβουμε στην συγκεκριμένη επιλογή.
8. Θα μας εμφανιστεί ένα παράθυρο στο οποίο αναγράφονται περισσότερες πληροφορίες για την συγκεκριμένη πτήση. Θα πατήσουμε το κουμπί «**Select**» που βρίσκεται στο τέλος του παραθύρου, με χρήση της φωνητικής εντολής «**Click Select**».
9. Με το πάτημα αυτού του κουμπιού θα μετάβουμε στην περιοχή της ιστοσελίδας που αφορά στην αγορά της συγκεκριμένης πτήσης προς Ζυρίχη. Επιλέγουμε την επιλογή «**Flexible ticket**» λέγοντας του την φωνητική εντολή «**Click Flexible ticket** » για να επιλέγει.
10. Τέλος, θα μετάβουμε στην αρχική σελίδα του Google Chrome με χρήση της φωνητικής εντολής «**Click Home**» και θα τερματίσουμε την λειτουργία του.

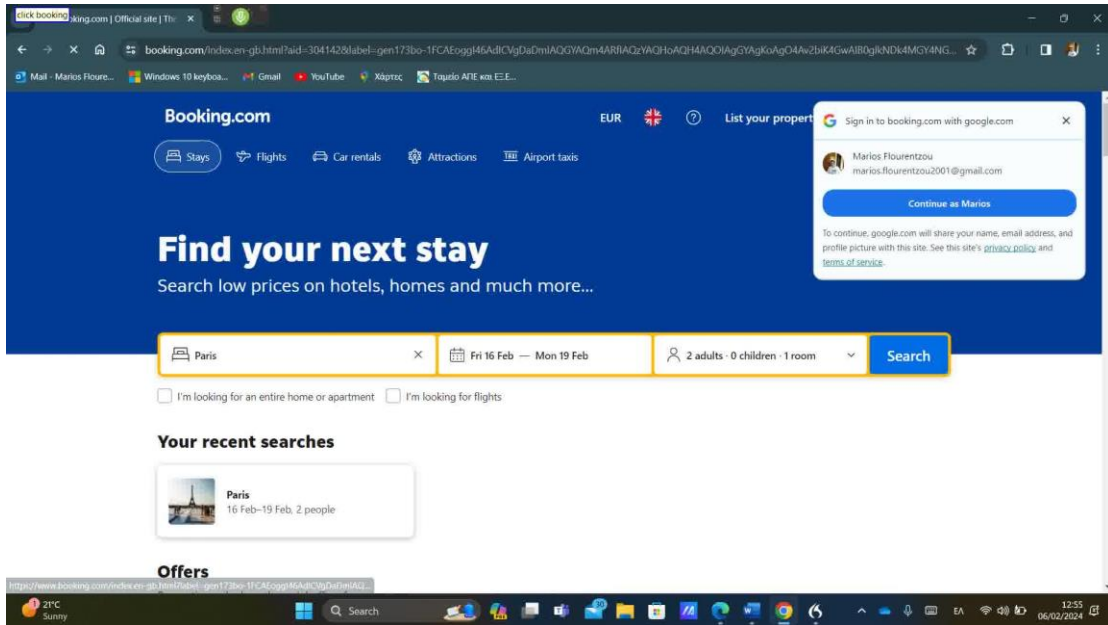
### 3.2.7 Στιγμιότυπα Οθονών από την εκτέλεση των πιο πάνω βημάτων.



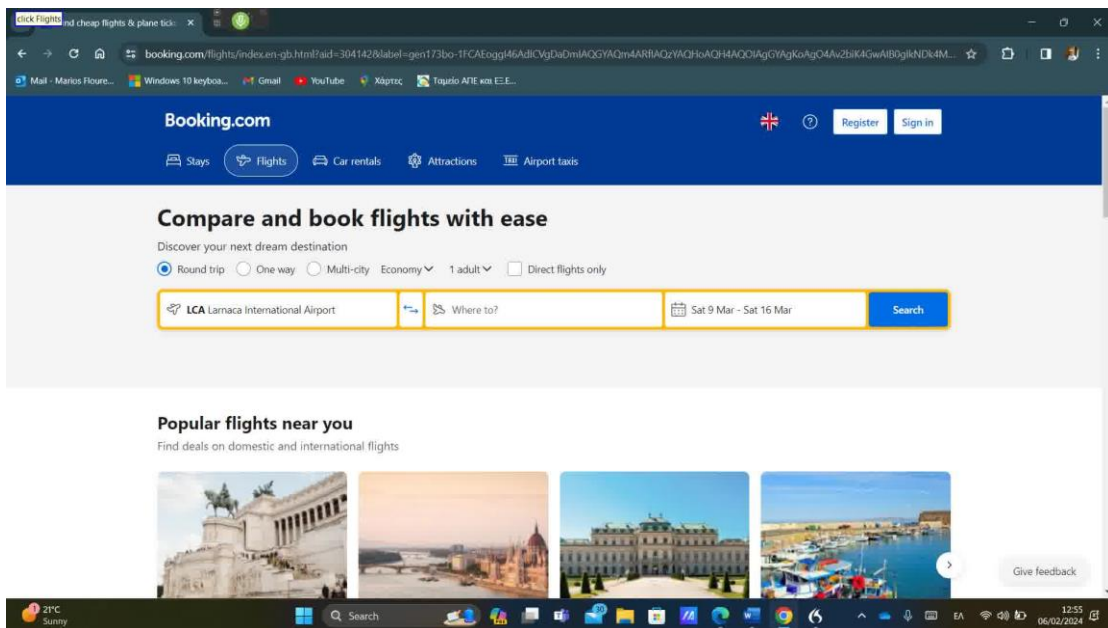
Εικόνα 3.2.7.1: Launching Google Chrome – “Open Google Chrome”



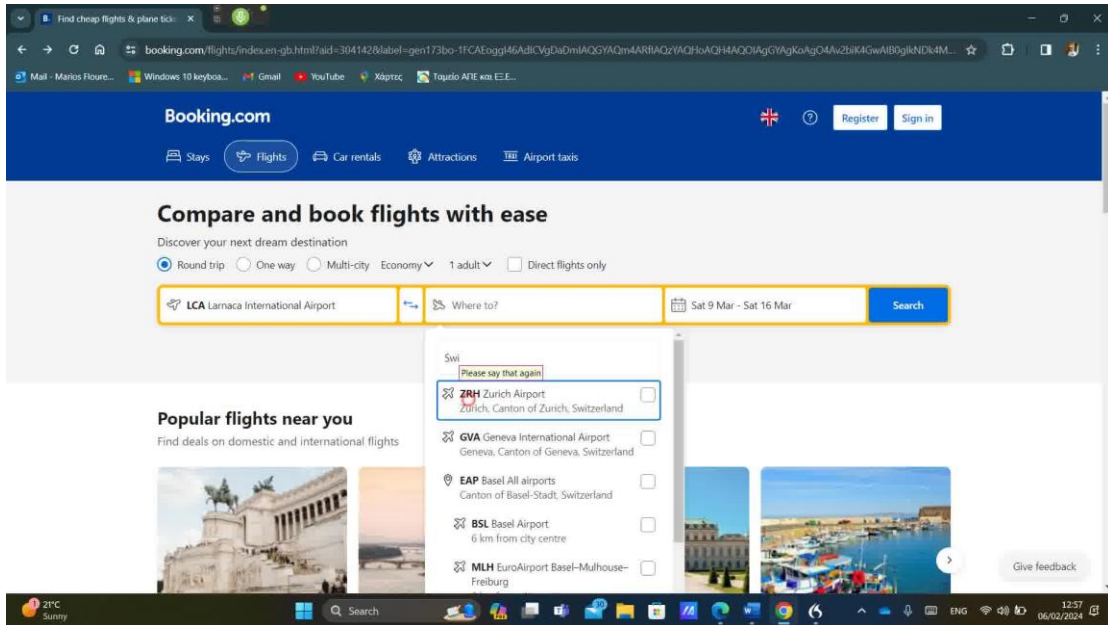
Εικόνα 3.2.7.2 Search the word «booking» Using Voice on Search Engine



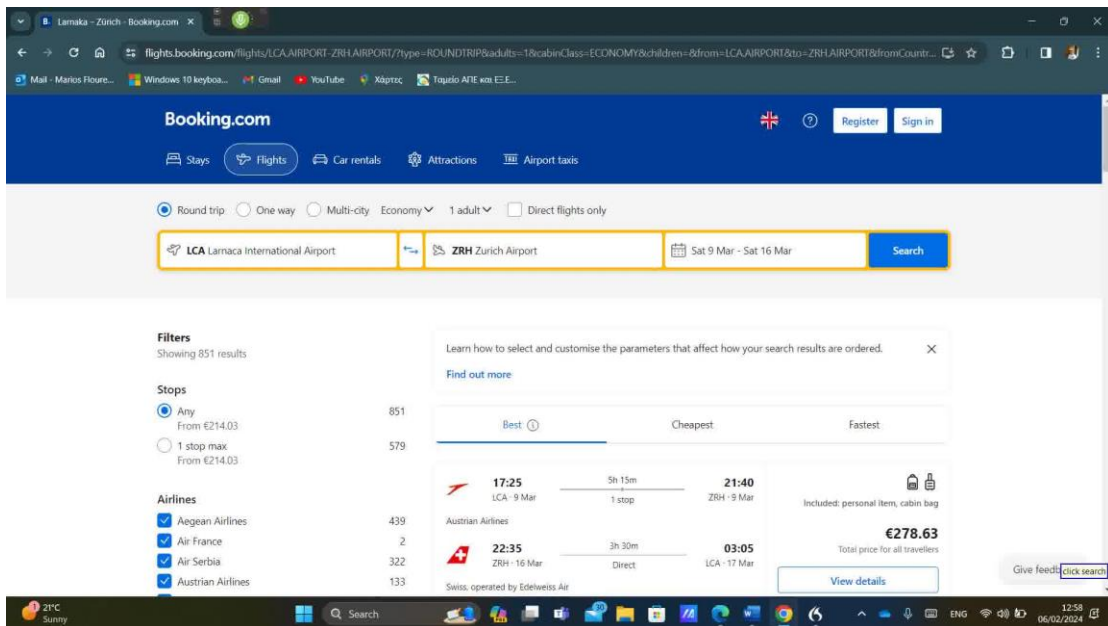
Εικόνα 3.2.7.3 Go to « booking.com» - Click booking.



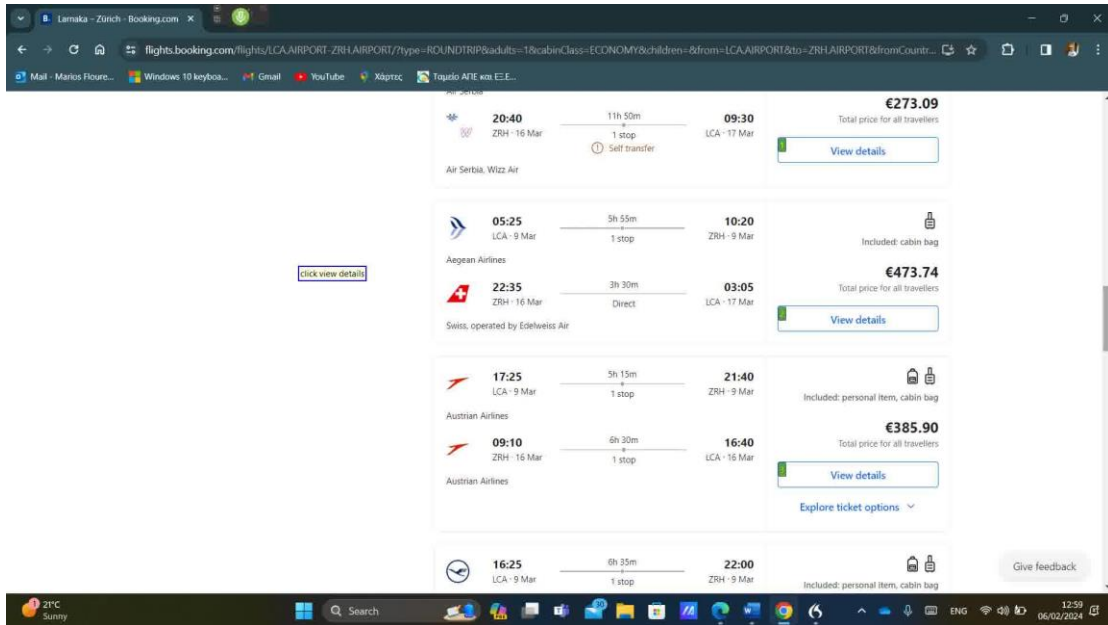
Εικόνα 3.2.7.4 Select Flights Using mouse command- «Click flights»



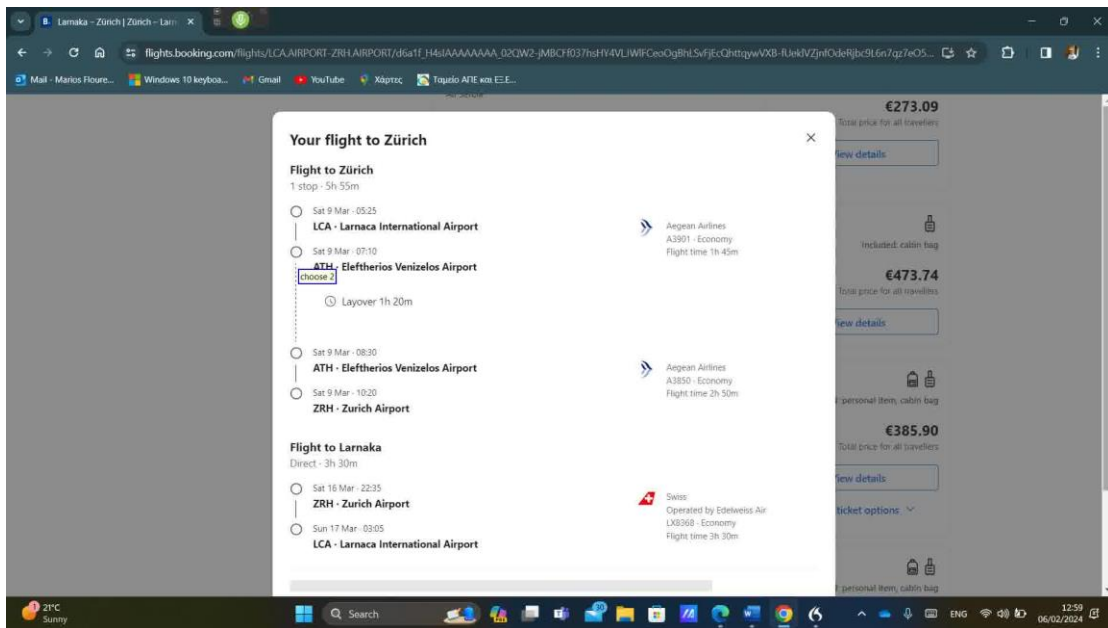
Εικόνα 3.2.7.5 Insert Destination



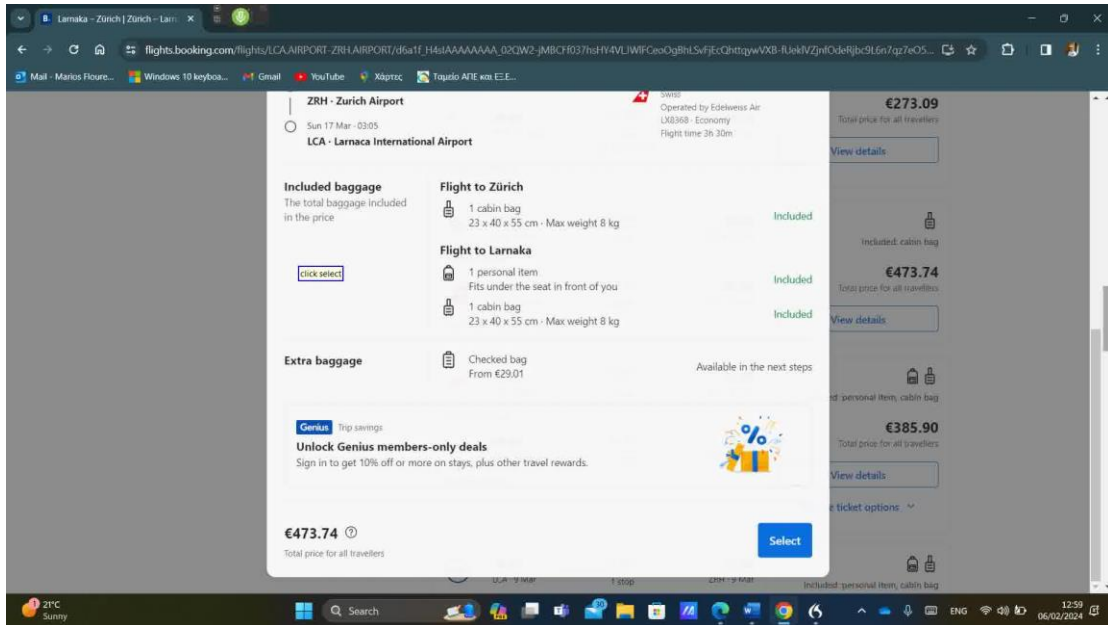
Εικόνα 3.2.7.6 Select Search Using mouse command- «Click Search»



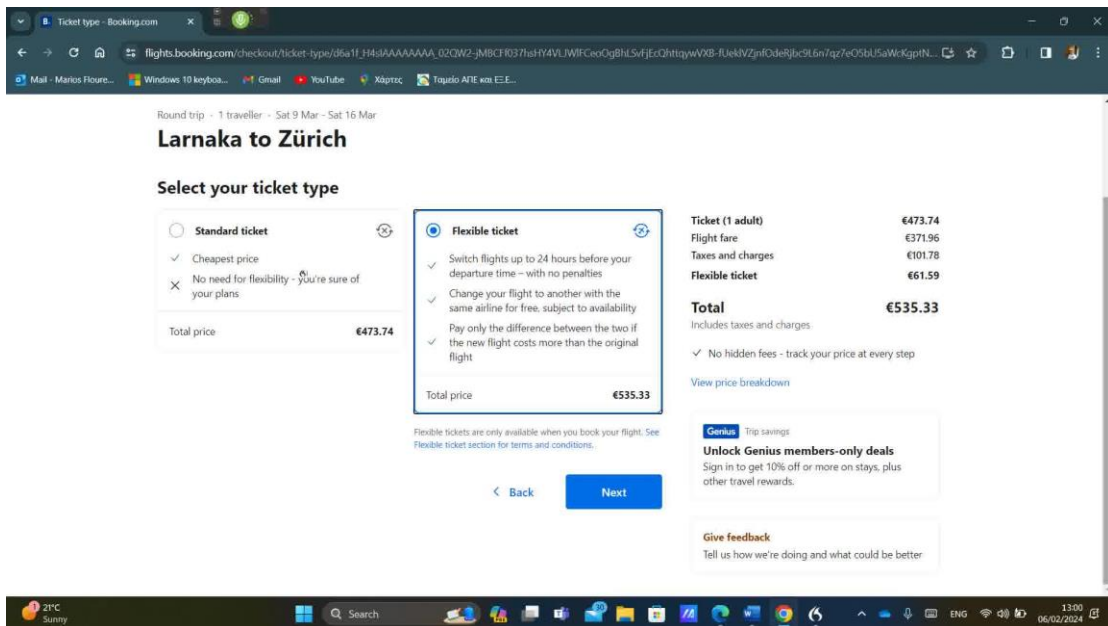
Εικόνα 3.2.7.7 Select View Details Using mouse command- «Click View details»



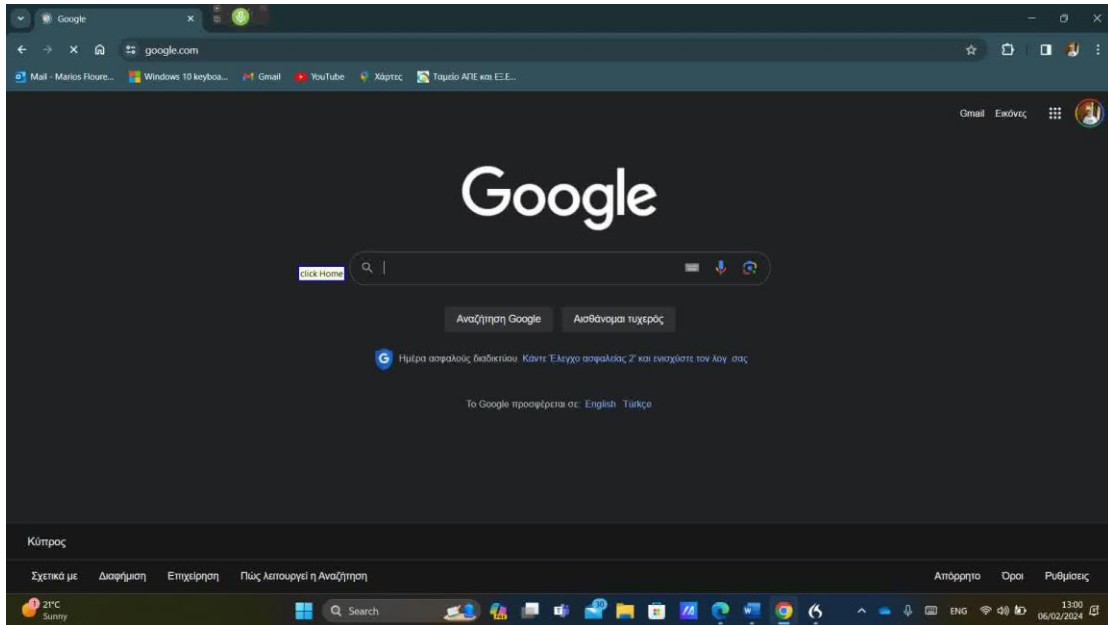
Εικόνα 3.2.7.8 Select the second green box generated from the voice Dragon tool Using mouse command- «Choose 2»



Εικόνα 3.2.7.9 Click the Select button Using mouse command- «Click Select»



Εικόνα 3.2.7.10 Select Flexible Ticket Using mouse command- «Click Flexible Ticket»



*Εικόνα 3.2.7.11 Go back to Home Page- «Click Home»*



*Εικόνα 3.2.7.12 Close Google Chrome- «Close Window»*



## Κεφάλαιο 4: Δεύτερο Σενάριο Χρήσης- Δημιουργία προγράμματος στις γλώσσες προγραμματισμού C, Java και Python με χρήση του εργαλείου Dragon Professional V16

---

4.1 Μεθοδολογία του Σεναρίου Χρήσης	58
4.2 Πρωτότυπο Πρόγραμμα στην γλώσσα C	59
4.3 Μεταγραφή του προγράμματος με χρήση του Dragon Professional (7η Προσπάθεια)	60
4.4 Υπολογισμός του ποσοστού σφάλματος λέξης Word Error Rate- WER	61
4.5 Σχολιασμός ποσοστού σφάλματος λέξης WER	62
4.6 Πρωτότυπο πρόγραμμα στην γλώσσα Java	63
4.7 Μεταγραφή του προγράμματος με χρήση του Dragon Professional (9η Προσπάθεια)	64
4.8 Υπολογισμός του ποσοστού σφάλματος λέξης Word Error Rate- WER	65
4.9 Σχολιασμός ποσοστού σφάλματος λέξης WER	66
4.10 Πρωτότυπο πρόγραμμα στην γλώσσα Python	67
4.11 Μεταγραφή του προγράμματος με χρήση του Dragon Professional (6η Προσπάθεια)	68
4.12 Υπολογισμός του ποσοστού σφάλματος λέξης Word Error Rate- WER	69
4.13 Σχολιασμός ποσοστού σφάλματος λέξης WER	69

---

**Στόχος:** Η αξιολόγηση της αποδοτικότητας και της αξιοπιστίας του Dragon Professional V16 στην υπαγόρευση και τη σύνταξη ενός προγράμματος στις γλώσσες προγραμματισμού C, Java και Python.

### **Περιβάλλον δοκιμής:**

- ✓ Φορητός Υπολογιστής με λειτουργικό σύστημα Windows 11 pro
- ✓ Το Dragon Professional V16 εγκαταστάθηκε και ρυθμίστηκε κατάλληλα
- ✓ Ενσωματωμένο περιβάλλον ανάπτυξης (IDE) ή επεξεργαστές κειμένου για C, Java και Python (π.χ. Visual Studio Code)
- ✓ Χρήση μικροφώνου υπολογιστή
- ✓ Γλώσσα Εκτέλεσης Φωνητικών Εντολών Αγγλική

### **Βήματα που απαιτούνται πριν την εκτέλεση του σεναρίου χρήσης:**

Πριν προχωρήσουμε στην εκτέλεση του προτεινόμενου σεναρίου χρήσης, απαιτούνται ορισμένα προκαταρκτικά βήματα για να διασφαλιστεί η ομαλή λειτουργία και η μέγιστη απόδοση του λογισμικού υπαγόρευσης Dragon Professional V16. Ακολουθούν τα βήματα που πρέπει να εκτελεστούν:

1. Εγκατάσταση και Ρύθμιση του Dragon Professional: Αρχικά, επιβεβαιώνουμε ότι το Dragon Professional V16 έχει εγκατασταθεί σωστά και ρυθμίζεται καταλλήλως για την αναγνώριση φωνής. Αυτό σημαίνει την προσαρμογή των ρυθμίσεων για την βελτιστοποίηση της ακρίβειας στην αναγνώριση της φωνής του χρήστη.
2. Εξοικείωση με Φωνητικές Εντολές: Συνεχίζουμε με την εξοικείωση με τις φωνητικές εντολές που σχετίζονται με την μεταγραφή κειμένου, καθώς και με εκείνες που αφορούν τα σύμβολα και τα σημεία στίξης που χρησιμοποιούνται στις γλώσσες προγραμματισμού, όπως η φωνητική εντολή «open parentheses» για παράδειγμα, για την εισαγωγή ανοιχτής παρένθεσης. Η σωστή χρήση των φωνητικών εντολών θα εξασφαλιστεί μέσα από την ανατροφοδότηση από τον οδηγό χρήσης που παρέχεται από το Dragon. [13]
3. Πρακτική Εκπαίδευση με το Visual Studio Code: Τέλος, ανοίγουμε το Visual Studio Code με φωνητική εντολή και αρχίζουμε να γράφουμε απλά προγράμματα, όπως το κλασικό «hello world», καθώς και άλλες βασικές εφαρμογές όπως ο υπολογισμός αθροίσματος αριθμών σε πίνακα ή το παραγοντικό ενός αριθμού. Αυτή η διαδικασία θα μας επιτρέψει να εξοικειωθούμε περαιτέρω με τις φωνητικές εντολές και να εκπαιδύσουμε το

Dragon Professional ώστε να αναγνωρίζει με ακρίβεια τον τόνο της φωνής και τις διατυπώσεις μας.

Ακολουθώντας αυτά τα βήματα, εξασφαλίζουμε ότι το λογισμικό υπαγόρευσης θα λειτουργήσει με την μέγιστη δυνατή απόδοση κατά τη διάρκεια της μεταγραφής των προγραμμάτων μας, διευκολύνοντας την αξιολόγηση της ακρίβειας και της αποδοτικότητας του στα διαφορετικά προγραμματιστικά περιβάλλοντα.

#### **4.1 Μεθοδολογία του Σεναρίου Χρήσης**

Θα αναπτύξουμε ένα πρόγραμμα το οποίο θα ζητά από τον χρήστη να εισάγει δέκα πραγματικούς αριθμούς, οι οποίοι θα αποθηκεύονται σε έναν πίνακα με την ονομασία arr. Κατόπιν, μια συνάρτηση με την ονομασία calculate θα είναι υπεύθυνη για τον υπολογισμό του αθροίσματος, του γινομένου, της διαφοράς και του πηλίκου των δέκα αυτών αριθμών. Τα αποτελέσματα αυτών των υπολογισμών θα παρουσιάζονται στον χρήστη με την χρήση κατάλληλων μηνυμάτων στην οθόνη.

Η συγκεκριμένη υλοποίηση του προγράμματος θα γίνει σε τρεις διαφορετικές γλώσσες προγραμματισμού: C, Java και Python. Αυτό θα μας επιτρέψει να αξιολογήσουμε την απόδοση και την αξιοπιστία της μεταγραφής που θα πραγματοποιηθεί από το λογισμικό υπαγόρευσης Dragon Professional, συγκρίνοντας την με την αρχική εκδοχή του προγράμματος για κάθε μια από τις γλώσσες προγραμματισμού. Μετά την ολοκλήρωση κάθε μεταγραφής, θα υπολογίζουμε το ποσοστό σφάλματος λέξεων (Word Error Rate – WER) και θα καταγράφουμε τις παρατηρήσεις μας, επιτρέποντας μια ολοκληρωμένη αξιολόγηση των επιδόσεων του λογισμικού σε διαφορετικά προγραμματιστικά περιβάλλοντα.

## 4.2 Πρωτότυπο Πρόγραμμα στην γλώσσα C

```
1 #include <stdio.h>
2
3 void calculate(float arr[], int n, float *sum, float *product, float *difference, float
*quotient) {
4     *sum = 0;
5     *product = 1;
6     *difference = arr[0];
7     *quotient = arr[0];
8
9     for (int i = 0; i < n; i++) {
10        *sum += arr[i];
11        *product *= arr[i];
12
13        if (i != 0) {
14            *difference -= arr[i];
15
16            if (arr[i] != 0) {
17                *quotient /= arr[i];
18            } else {
19                printf("Division by zero encountered, quotient calculation stopped.\n");
20                break;
21            }
22        }
23    }
24 }
25
26 int main() {
27     float numbers[10];
28     float sum, product, difference, quotient;
29
30     printf("Enter 10 float numbers:\n");
31     for (int i = 0; i < 10; i++) {
32         scanf("%f", &numbers[i]);
33     }
34
35     calculate(numbers, 10, &sum, &product, &difference, &quotient);
36
37     printf("Sum of array elements: %f\n", sum);
38     printf("Product of array elements: %f\n", product);
39     printf("Difference of array elements: %f\n", difference);
40     printf("Quotient of array elements: %f\n", quotient);
41
42     return 0;
43 }
```

### 4.3 Μεταγραφή του προγράμματος με χρήση του Dragon Professional (7<sup>η</sup> Προσπάθεια)

```
1 #include <stdio.h>
2
3 void calculate(flute arr[], int n, flute *sum, flute *product, flute *difference, flute
*quotient) {
4     *sum = 0;
5     *product = 1;
6     *difference = arr[0];
7     *quotient = arr[0];
8
9     for (int i = 0; i < n; i++) {
10        *sum += arr[i]
11        *product *= arr[i]
12
13        if (i != 0) {
14            *difference -= arr[i];
15
16            if (arr[i] != 0) {
17                *quotient /= arr[i];
18            } else {
19                printf("Division by zero quotient calculation stopped.\n");
20                brake;
21            }
22        }
23    }
24 }
25
26 int main() {
27     flute numbers[10];
28     flute sum, product, difference, quotient;
29
30     printf("Enter 10 flute numbers:");
31     for (int i = 0; i < 10; i++) {
32         scan F("%f", &numbers[i]);
33     }
34
35     calculate(numbers, 10, &sum, &product, &difference, &quotient);
36
37     printf("Sum of array elements: \n", sum);
38     printf("Product of array elements: f\n", product);
39     printf("Difference of array elements: %f\n", difference);
40     printf("Quotient of array elements: %f\n", quotient);
41
42     return 0;
43 }
```

#### 4.4 Υπολογισμός του ποσοστού σφάλματος λέξης Word Error Rate- WER

Ας θυμηθούμε τον μαθηματικό τύπο με τον οποίο μπορεί να υπολογιστεί το WER:

$$\text{WER} = (\text{S} + \text{D} + \text{I}) / \text{N}$$

Όπου:

- Το S αντιπροσωπεύει τον αριθμό των αντικαταστάσεων (λέξεις που αναγνωρίζονται εσφαλμένα από το Dragon Professional).
  - Το D αντιπροσωπεύει τον αριθμό των διαγραφών (λέξεις που υπάρχουν στο πρωτότυπο πρόγραμμα αλλά όχι στο μεταγραφόμενο πρόγραμμα που παράχθηκε από το Dragon Professional.)
  - Το I αντιπροσωπεύει τον αριθμό των εισαγωγών (λέξεις που υπάρχουν στο μεταγραφόμενο πρόγραμμα που παράχθηκε από το Dragon Professional αλλά δεν περιέχονται στο πρωτότυπο πρόγραμμα).
  - Το N αντιπροσωπεύει τον συνολικό αριθμό των λέξεων που υπάρχουν στο πρωτότυπο πρόγραμμα [19].
- 
- Substitutions (S): 15
  - Deletions (D): 1
  - Insertions (I): 1
  - N = 135. (πλήθος λέξεων του πρωτότυπου προγράμματος)
  - Τροποποιήσεις που περιέχει το μεταγραφόμενο πρόγραμμα σε σύγκριση με το πρωτότυπο (S+D+I): 17

Θα αντικαταστήσουμε τις πιο πάνω τιμές στον τύπο που υπολογίζει το ποσοστό σφάλματος λέξης (WER)

$$\text{WER} = (15+1+1) / 135$$

Το ποσοστό σφάλματος είναι το WER X 100 που προκύπτει 12.59 %

**Επισημάνση:** Ο υπολογισμός του ποσοστού σφάλματος λέξης WER έγινε με την χρήση ενός προγράμματος στην γλώσσα Python το οποίο κατασκευάσαμε.

Επομένως, ο ρυθμός σφάλματος λέξης είναι περίπου 12.59%, που υποδηλώνει ότι περίπου 12.59% των λέξεων στο μεταγραφόμενο πρόγραμμα C με χρήση του Dragon Professional V16 διαφέρουν από το πρωτότυπο πρόγραμμα στην γλώσσα C.

#### **4.5 Σχολιασμός ποσοστού σφάλματος λέξης WER**

Σε επαγγελματικά περιβάλλοντα όπου η ακρίβεια έχει ζωτική σημασία, όπως στους τομείς της Ιατρικής, της Πληροφορικής και της Νομικής, λογισμικά υπαγόρευσης όπως το Dragon Professional αναμένεται να έχουν ένα ποσοστό λαθών στην αναγνώριση λέξεων (WER) περίπου 5% ή και χαμηλότερο. Έτσι, ένα WER της τάξης του 12,59% μπορεί να μην κρίνεται ικανοποιητικό για τέτοιες εφαρμογές, καθώς ακόμα και ασήμαντα σφάλματα μπορούν να επιφέρουν σημαντικές παρερμηνείες ή ανακρίβειες.

Η απαίτηση για υψηλή ακρίβεια γίνεται ακόμα πιο κρίσιμη όταν πρόκειται για την υπαγόρευση κώδικα προγραμματισμού. Σε αυτό το πλαίσιο, το WER θα πρέπει ιδανικά να είναι όσο το δυνατόν πλησιέστερο στο 0%, διότι ακόμη και μικρά λάθη μπορούν να προκαλέσουν σημαντικά προβλήματα στη μεταγλώττιση ή την εκτέλεση των προγραμμάτων. Ένα WER του 12,59% είναι εξαιρετικά υψηλό για τέτοιες εφαρμογές, με τα λάθη να μπορούν να οδηγήσουν σε σοβαρά συντακτικά ή λειτουργικά προβλήματα στον κώδικα.

Συμπερασματικά, ενώ ένα WER του 12,59% μπορεί να είναι αποδεκτό για γενικές εφαρμογές αναγνώρισης ομιλίας, για επαγγελματικές εφαρμογές υπαγόρευσης και ειδικά για την προγραμματιστική κωδικοποίηση, ένα τέτοιο ποσοστό λαθών θεωρείται σημαντικά υψηλό. Η επιδίωξη υψηλότερων επιδόσεων και μικρότερου WER είναι κρίσιμη για να αποφευχθούν τα λάθη που μπορούν να επηρεάσουν την ακρίβεια και την αποδοτικότητα των εφαρμογών, καθιστώντας το λογισμικό πιο αξιόπιστο και αποτελεσματικό σε αυτούς τους τομείς.

## 4.6 Πρωτότυπο πρόγραμμα στην γλώσσα Java

```
1 import java.util.Scanner;
2
3 public class NumberOperations {
4
5     // Method to perform the calculations
6     public static float[] calculate(float[] arr) {
7         float sum = 0;
8         float product = 1;
9         float difference = arr[0];
10        float quotient = arr[0];
11
12        for (int i = 0; i < arr.length; i++) {
13            sum += arr[i];
14            product *= arr[i];
15
16            if (i != 0) {
17                difference -= arr[i];
18
19                if (arr[i] != 0) {
20                    quotient /= arr[i];
21                } else {
22                    System.out.println("Division by zero encountered, quotient calculation
stopped.\n");
23                    return new float[]{sum, product, difference,
Float.NaN}; // Using NaN to indicate division by zero
24                }
25            }
26        }
27
28        return new float[]{sum, product, difference, quotient};
29    }
30
31    public static void main(String[] args) {
32        Scanner scanner = new Scanner(System.in);
33        float[] numbers = new float[10];
34
35        System.out.println("Enter 10 float numbers:");
36        for (int i = 0; i < 10; i++) {
37            numbers[i] = scanner.nextFloat();
38        }
39
40        float[] results = calculate(numbers);
41
42        System.out.println("\nSum of array elements: " + results[0]);
43        System.out.println("\nProduct of array elements: " + results[1]);
```



```

44     System.out.println("\nDifference of array elements: " + results[2]);
45     System.out.println("\nQuotient of array elements: " + (Float.isNaN(results[3])
? "Undefined (division by zero)" : results[3]));
46
47     scanner.close();
48 }
49 }

```

#### 4.7 Μεταγραφή του προγράμματος με χρήση του Dragon Professional (9<sup>η</sup> Προσπάθεια)

```

1 import java.util.Scanner;
2
3 public class NumberOperations {
4
5     // Method to perform the calculations
6     public static flute[] calculate(flute[] arr) {
7         flute sum = 0;
8         flute product = 1;
9         flute difference = arr[0];
10        flute quotient = arr[0];
11
12        for (int i = 0; i < arr.length; i++) {
13            sum += arr[i]
14            product *= arr[i]
15
16            if (i != 0) {
17                difference -= arr[i];
18
19                if (arr[i] != 0) {
20                    quotient /= arr[i];
21                } else {
22                    System.out.println("Division by zero, calculation stopped.");
23                    return new flute[]{sum, product, difference, Float.NaN}; // Using NaN
to indicate division by zero
24                }
25            }
26        }
27
28        return new flute[]{sum, product, difference, quotient};
29    }
30
31    public static void main(String[] args) {

```

```

32 Scanner scanner = new Scanner(System.in);
33 flute[] numbers = new flute[10];
34
35 System.out.println("Enter ten flute numbers:");
36 for (i = 0; i < 10; i++) {
37     numbers[i] = scanner.nextFlute();
38 }
39
40 flute[] results = calculate(numbers);
41
42 System.out.println("Sum of the array elements: " + results[0]);
43 System.out.println("Product of the array elements: " + results[1]);
44 System.out.println("Difference of the array elements: " + results[2]);
45 System.out.println("Quotient of the array elements: " +
(Float.isNaN(results[3]) ? "Undefined due to division by zero" : results[3]));
46 }
47 }

```

#### 4.8 Υπολογισμός του ποσοστού σφάλματος λέξης Word Error Rate- WER

$$WER = (S + D + I) / N$$

- Substitutions (S): 24
- Deletions (D): 5
- Insertions (I): 6
- N: 176 (πλήθος λέξεων του πρωτότυπου προγράμματος)
- Τροποποιήσεις που περιέχει το μεταγραφόμενο πρόγραμμα σε σύγκριση με το πρωτότυπο (S+D+I): 35

Θα αντικαταστήσουμε τις πιο πάνω τιμές στον τύπο που υπολογίζει το ποσοστό σφάλματος λέξης (WER)

$$WER = (S+D+I) / N = (24+5+6) / 176$$

Το ποσοστό σφάλματος είναι το WER X 100 που προκύπτει: 19.89%

**Επισημάνση:** Ο υπολογισμός του ποσοστού σφάλματος λέξης WER έγινε με την χρήση ενός προγράμματος στην γλώσσα Python το οποίο κατασκευάσαμε.

Επομένως, ο ρυθμός σφάλματος λέξης είναι περίπου 19.89%, που υποδηλώνει ότι περίπου 19.89% των λέξεων στο μεταγραφόμενο πρόγραμμα Java με χρήση του Dragon Professional V16 διαφέρουν από το πρωτότυπο πρόγραμμα στην γλώσσα Java.

#### **4.9 Σχολιασμός ποσοστού σφάλματος λέξης WER**

Δεδομένου ότι ασχολούμαστε με μεταγραφή κώδικα στην γλώσσα προγραμματισμού Java, με χρήση του Dragon Professional το WER 19.89% είναι αρκετά υψηλό. Η μεταγραφή κώδικα απαιτεί υψηλό βαθμό ακρίβειας, επειδή ακόμη και ένα μόνο σφάλμα (όπως ένα τυπογραφικό λάθος σε ένα όνομα μεταβλητής ή ένα ερωτηματικό που λείπει) μπορεί να οδηγήσει σε συντακτικά σφάλματα που εμποδίζουν τη σωστή μεταγλώττιση ή εκτέλεση του κώδικα.

Δεδομένου ότι τα σφάλματα περιλαμβάνουν κρίσιμα στοιχεία του κώδικα Java (λανθασμένοι τύποι δεδομένων και ονόματα μεθόδων), ένα WER 19.89% υποδηλώνει σημαντικό πρόβλημα ακρίβειας. Στον προγραμματισμό, η ακρίβεια είναι πρωταρχικής σημασίας επειδή οι ανακρίβειες μπορεί να οδηγήσουν σε σφάλματα μεταγλώττισης ή ακούσια συμπεριφορά. Επομένως, στο πλαίσιο της υπαγόρευσης κώδικα στην γλώσσα Java, επιβάλλει ποσοστό σφάλματος λέξης WER όσο το δυνατόν πλησιέστερο στο 0% που είναι το ιδανικό.

#### 4.10 Πρωτότυπο πρόγραμμα στην γλώσσα Python

```
1 def calculate(arr):
2     sum = 0.0
3     product = 1.0
4     difference = arr[0]
5     quotient = arr[0]
6
7     for i in range(len(arr)):
8         sum += arr[i]
9         product *= arr[i]
10
11        if i != 0:
12            difference -= arr[i]
13
14            if arr[i] != 0:
15                quotient /= arr[i]
16            else:
17                print("Division by zero encountered, quotient calculation stopped.")
18                return sum, product, difference, float('nan') # Using 'nan' to indicate
19                division by zero
20
21    return sum, product, difference, quotient
22
23 def main():
24     numbers = [] # For 10 numbers
25
26     print("Enter 10 float numbers:")
27     for _ in range(10):
28         numbers.append(float(input()))
29
30     results = calculate(numbers)
31
32     print("\nSum of array elements:", results[0])
33     print("Product of array elements:", results[1])
34     print("Difference of array elements:", results[2])
35     print("Quotient of array elements:", "Undefined (division by zero)" if
36     str(results[3]) == 'nan' else results[3])
37
38 if __name__ == "__main__":
39     main()
```

#### 4.11 Μεταγραφή του προγράμματος με χρήση του Dragon Professional (6<sup>η</sup> Προσπάθεια)

```
1 def calculate(arr):
2     some = 0.0
3     product = 1.0
4     difference = arr[0]
5     quotient = arr[0]
6
7     for i in range(len(arr)):
8         some += arr[i]
9         product *= arr[i]
10
11         if i != 0:
12             difference -= arr[i]
13
14             if arr[i] != 0:
15                 quotient /= arr[i]
16             else:
17                 print("Division by zero calculation stopped.")
18                 return sum, product, difference, float('nan') # Using 'nan' to indicate
19                 division by zero
20
21     return sum, product, difference, quotient
22
23 def main():
24     numbers = [] # For 10 numbers
25
26     print("Enter 10 float numbers:")
27     for _ in range(10):
28         numbers.append(float(input()))
29
30     results = calculate(numbers)
31
32     print("\nSum of array elements:", results[0])
33     print("Product of array elements:", results[1])
34     print("Difference of array elements:", results[2])
35     print("Quotient of array elements:", "Undefined (division by zero)" if
36     str(results[3]) == 'nan' else results[3])
37
38 if __name__ == "__main__":
39     main()
```

#### 4.12 Υπολογισμός του ποσοστού σφάλματος λέξης Word Error Rate- WER

$$\text{WER} = (S + D + I) / N$$

- Substitutions (S): 2
- Deletions (D): 2
- Insertions (I): 0
- N: 119 (πλήθος λέξεων του πρωτότυπου προγράμματος)
- Τροποποιήσεις που περιέχει το μεταγραφόμενο πρόγραμμα σε σύγκριση με το πρωτότυπο (S+D+I): 4

Θα αντικαταστήσουμε τις πιο πάνω τιμές στον τύπο που υπολογίζει το ποσοστό σφάλματος λέξης (WER)

$$\text{WER} = (S+D+I) / N = (2+2+0) / 119$$

Το ποσοστό σφάλματος είναι το WER X 100 που προκύπτει: 3.36%

**Επισήμανση:** Ο υπολογισμός του ποσοστού σφάλματος λέξης WER έγινε με την χρήση ενός προγράμματος στην γλώσσα Python το οποίο κατασκευάσαμε.

Επομένως, ο ρυθμός σφάλματος λέξης είναι περίπου 3.36%, που υποδηλώνει ότι περίπου 3.36 % των λέξεων στο μεταγραφόμενο πρόγραμμα Python με χρήση του Dragon Professional V16 διαφέρουν από το πρωτότυπο πρόγραμμα στην γλώσσα Python.

#### 4.13 Σχολιασμός ποσοστού σφάλματος λέξης WER

Η επίτευξη ενός ποσοστού σφάλματος λέξεων (WER) 3,36% κατά τη διαδικασία υπαγόρευσης κώδικα Python μέσω του Dragon Professional, αποτελεί μια ξεκάθαρη ένδειξη της εξαιρετικής αποδοτικότητας και ακριβείας που μπορούν να προσφέρουν οι σύγχρονες τεχνολογίες αναγνώρισης ομιλίας στην γλώσσα Python. Αυτό το στατιστικό δείχνει ότι το λογισμικό είναι ικανό να καταγράψει με σαφήνεια το μεγαλύτερο μέρος του προφορικά διατυπωμένου κώδικα Python, απαιτώντας μόνο οριακές προσαρμογές. Αυτή η επίδοση όχι μόνο βελτιώνει σημαντικά την παραγωγικότητα, εξοικονομώντας χρόνο από τη διαδικασία της πληκτρολόγησης, αλλά καθιστά επίσης τον

προγραμματισμό πιο προσβάσιμο, ιδιαίτερα για όσους αντιμετωπίζουν δυσκολίες στη χρήση πληκτρολογίου ή επιδιώκουν να αποφύγουν την επαναλαμβανόμενη κόπωση.

Η αξιοπιστία που προσδίδεται στο Dragon Professional για εργασίες με απαιτήσεις υψηλής τεχνικής ακρίβειας, όπως ο προγραμματισμός σε Python, επιβεβαιώνεται μέσω αυτού του χαμηλού WER. Παράλληλα, η δυνατότητα του λογισμικού να μαθαίνει από τις διορθώσεις και να προσαρμόζεται στις λεξιλογικές προτιμήσεις του χρήστη, υπόσχεται μια συνεχή βελτίωση στην ακρίβεια με την πάροδο του χρόνου.

Παρ' όλα αυτά, η φύση του προγραμματισμού, που απαιτεί μεγάλη ακρίβεια στη σύνταξη και στη λογική, καθιστά απαραίτητη την προσεκτική ανασκόπηση του κώδικα που έχει παραχθεί μέσω υπαγόρευσης. Ακόμη και ένα σχετικά μικρό ποσοστό λαθών μπορεί να προκαλέσει σημαντικά ζητήματα στη λειτουργικότητα του κώδικα, επισημαίνοντας την ανάγκη για διεξοδικό έλεγχο και διόρθωση των πιθανών σφαλμάτων.

Συνοψίζοντας, ένα WER της τάξης του 3,36% στην υπαγόρευση κώδικα Python με τη χρήση τεχνολογιών όπως το Dragon Professional, αποτελεί μία εντυπωσιακή επίδοση. Αυτό υποδεικνύει το υψηλό επίπεδο αποδοτικότητας και προσβασιμότητας που μπορεί να προσφέρει η φωνητική εισαγωγή στον τομέα της ανάπτυξης λογισμικού, μετά από αρκετή εκπαίδευση του εργαλείου.

## Κεφάλαιο 5: Αποτελέσματα από την Εκτέλεση των δυο Σεναρίων Χρήσης

---

5.1 Αξιολόγηση των Επιδόσεων του Dragon Professional	71
5.2 Πίνακες Αποτελεσμάτων Προγραμματιστικού Σεναρίου Χρήσης	72
5.3 Σχολιασμός Παραγόμενου Μεταγραφόμενου Προγράμματος στην Γλώσσα C από το Dragon Professional	74
5.4 Σχολιασμός Παραγόμενου Μεταγραφόμενου Προγράμματος στην Γλώσσα Java από το Dragon Professional	75
5.5 Σχολιασμός Παραγόμενου Μεταγραφόμενου Προγράμματος στην Γλώσσα Python από το Dragon Professional	76

---

Στο Κεφάλαιο 5, θα προχωρήσουμε στη λεπτομερή ανάλυση των αποτελεσμάτων από την υλοποίηση των δύο σεναρίων χρήσης που επιχειρήσαμε προηγουμένως. Αυτή η ανάλυση θα αποκαλύψει πώς το λογισμικό φωνητικής αναγνώρισης Dragon Professional V16 αντεπεξήλθε στις προκλήσεις των δυο σεναρίων χρήσης που σχεδιάσαμε με σκοπό να αξιολογήσουμε την αποδοτικότητα του.

### 5.1 Αξιολόγηση των Επιδόσεων του Dragon Professional

Η αποδοτικότητα του εργαλείου κατά την εκτέλεση των σεναρίων χρήσης μετά από αρκετή εκπαίδευση ήταν πολύ καλή.

Στο πρώτο σενάριο, το οποίο αφορούσε στην πλοήγηση διαδικτύου και στην διαχείριση αρχείων, η αποδοτικότητα του λογισμικού ήταν επίσης πολύ καλή. Το λογισμικό κατάφερε να αντιληφθεί με ευκολία τις φωνητικές εντολές, παρά την ανάγκη επανάληψης ορισμένων εντολών λόγω πιθανών εξωτερικών παραγόντων, όπως ο θόρυβος, που ενδεχομένως επηρέασαν την αντίχρευσή τους από το μικρόφωνο.

Στο δεύτερο σενάριο, το οποίο εστίαζε στην δημιουργία προγράμματος στις γλώσσες προγραμματισμού C, Java και Python, τα αποτελέσματα από τον υπολογισμό του WER επιβεβαίωσαν την υψηλή αποδοτικότητα του εργαλείου, ιδίως στην Python, με ένα



ποσοστό σφάλματος λέξεων μόλις 3,36%. Αυτό συγκριτικά με τα ποσοστά WER 19,89% και 12,59% για τις γλώσσες Java και C αντίστοιχα, καταδεικνύει μια σαφώς άριστη σχεδόν απόδοση στην γλώσσα Python, αποδεικνύοντας μια πολύ καλή αξιοπιστία του Dragon Professional στη μεταγραφή κώδικα στην γλώσσα Python.

Για την μεταγραφή του προγράμματος για την κάθε γλώσσα προγραμματισμού έγιναν αρκετές προσπάθειες μέχρι να πετύχουμε την καλύτερη δυνατή προσπάθεια μεταγραφής στην κάθε περίπτωση. Ο λόγος είναι διότι η σύνταξη οπουδήποτε προγράμματος σε κάποια γλώσσα προγραμματισμού απαιτεί συγκεκριμένο λεξιλόγιο έτσι χρειάστηκαν αρκετές προσπάθειες μέχρι το εργαλείο να ανταποκριθεί ανάλογα, αφού υπήρχε συγκεκριμένο λεξιλόγιο για την κάθε γλώσσα που δεν υπήρχε καταχωρημένο στο λεξικό του εργαλείου.

Η εμπειρία αυτή επιβεβαιώνει ότι εργαλεία όπως το Dragon Professional απαιτούν συνεχή εκπαίδευση και προσαρμογή στη φωνή του χρήστη για να επιτύχουν την καλύτερη δυνατή απόδοση. Παρά τις προκλήσεις, τα αποτελέσματα μας επέτρεψαν να καταλήξουμε σε μια αξιοπρεπή αξιολόγηση της αποδοτικότητας του Dragon Professional, ανοίγοντας παράλληλα πεδίο για περαιτέρω βελτιώσεις και εξελίξεις στην τεχνολογία φωνητικής αναγνώρισης.

## 5.2 Πίνακες Αποτελεσμάτων Προγραμματιστικού Σεναρίου Χρήσης

Πίνακας 5.1: Μετρικές απόδοσης εργαλείου κατά την μεταγραφή του προγράμματος στην C, Java, Python

Γλώσσα Προγραμματισμού	Αντικαταστάσεις (S)	Διαγραφές (D)	Εισαγωγές (I)	N	WER	Χρόνος Αφιέρωσης (λεπτά)
C	15-18	1-3	1-3	135	13-18%	90-120
Java	24-27	5-7	6-7	176	20-27%	120-150
Python	2-3	2-2	0-1	119	3-5%	60-90

Στην Python, η ακρίβεια μεταγραφής του κώδικα φαίνεται να είναι ιδιαίτερα υψηλή, με το ποσοστό του Word Error Rate (WER) να κυμαίνεται μεταξύ 3-5%, καταδεικνύοντας την άριστη λειτουργία του συστήματος σε αυτή τη γλώσσα προγραμματισμού. Αντίθετα, οι κώδικες στις γλώσσες προγραμματισμού C και Java

παρουσίασαν υψηλότερα ποσοστά WER, στο εύρος του 13-18% και 20-27% αντίστοιχα, υποδεικνύοντας σημαντικές προκλήσεις στην ακριβή μεταγραφή τους. Τα τυπογραφικά λάθη και οι παραλείψεις στίξης επηρέασαν αρνητικά τη συντακτική ορθότητα του κώδικα, ενώ η ακριβής αποτύπωση της σύνθετης σύνταξης και του ειδικού λεξιλογίου στις γλώσσες C και Java αποτέλεσε επιπλέον πρόκληση.

**Πίνακας 5.2: Μετρικές απόδοσης εργαλείου κατά την μεταγραφή του προγράμματος στην C, Java, Python (Εκτέλεση του πειράματος από 3 διαφορετικούς χρήστες)**

Γλώσσα Προγραμματισμού	Αντικαταστάσεις (S)	Διαγραφές (D)	Εισαγωγές (I)	N	WER	Χρόνος Αφίερωσης (λεπτά)
C	15-26	1-3	1-3	135	12-31%	90-120
Java	18-44	0-4	6-9	176	17-38%	120-150
Python	3-16	2-2	0-1	119	3-5%	60-90

Στην Python, η ακρίβεια μεταγραφής του κώδικα φαίνεται να είναι εξαιρετικά υψηλή, με το ποσοστό του Word Error Rate (WER) να διατηρείται σε ένα πολύ χαμηλό εύρος 3-5%, καταδεικνύοντας την άριστη λειτουργία του συστήματος σε αυτή τη γλώσσα προγραμματισμού. Αντίθετα, οι κώδικες στις γλώσσες προγραμματισμού C και Java παρουσίασαν αυξημένα ποσοστά WER, στο εύρος του 12-31% και 17-38% αντίστοιχα, υποδεικνύοντας σημαντικές προκλήσεις στην ακριβή μεταγραφή τους από τους τρεις διαφορετικούς χρήστες που εκτέλεσαν το σενάριο χρήσης. Οι δυσκολίες αυτές ενδέχεται να οφείλονται σε τυπογραφικά λάθη και παραλείψεις στίξης, καθώς και στην ανάγκη αποτύπωσης πιο σύνθετης σύνταξης και ειδικού λεξιλογίου των γλωσσών C και Java, ενισχύοντας την ανάγκη για περαιτέρω βελτίωση του συστήματος σε αυτές τις γλώσσες.

### 5.3 Σχολιασμός Παραγόμενου Μεταγραφόμενου Προγράμματος στην Γλώσσα C από το Dragon Professional

Όπως βλέπουμε στο μεταγραφόμενο πρόγραμμα που δημιουργήθηκε από το εργαλείο υπάρχουν κάποιες διαφορές σε σχέση με το πρωτότυπο πρόγραμμα.

- ✓ Στην γραμμή 3 ο τύπος float μεταγράφηκε ως “flute” για κάθε όρισμα στην συνάρτηση calculate.
- ✓ Στις γραμμές 10, 11 δεν μεταγράφηκε στο τέλος το semicolon (;).
- ✓ Στην γραμμή 19 παραλείφθηκε η λέξη «encountered» στο μήνυμα που θα εκτυπώσει η printf.
- ✓ Στην γραμμή 20 το «break” έχει μεταγραφεί ως «brake”.
- ✓ Στις γραμμές 27 και 28 ο τύπος float μεταγράφηκε ως «flute”.
- ✓ Στην γραμμή 30 το μήνυμα που θα τυπώσει η printf έχει μεταγραφεί ως «Enter 10 **flute** numbers:» αντί **float**.
- ✓ Στην γραμμή 32 η εντολή scanf έχει μεταγραφεί ως **scan F**.
- ✓ Στην γραμμή 38 δεν μεταγράφηκε το σύμβολο % μπροστά από το f (%f).

Σε γενικές γραμμές, η απόδοση του εργαλείου ήταν ικανοποιητική. Αυτά τα εργαλεία χρειάζονται χρόνο για να μάθουν τον τρόπο με τον οποίο μιλάμε. Στην έβδομη προσπάθεια μεταγραφής, το αποτέλεσμα ήταν καλύτερο σε σχέση με τις προηγούμενες προσπάθειες. Υπήρξε δυσκολία στην αντίληψη κάποιων λέξεων της γλώσσας προγραμματισμού C και κάποιων συμβόλων, όπως το «%».

#### 5.4 Σχολιασμός Παραγόμενου Μεταγραφόμενου Προγράμματος στην Γλώσσα Java από το Dragon Professional

Όπως βλέπουμε στο μεταγραφόμενο πρόγραμμα που δημιουργήθηκε από το εργαλείο υπάρχουν κάποιες διαφορές σε σχέση με το πρωτότυπο πρόγραμμα.

- ✓ Στις γραμμές 6 – 10, 23, 28, 33, 37, 40 ο τύπος float μεταγράφηκε ως “flute”
- ✓ Στις γραμμές 13,14 δεν μεταγράφηκε στο τέλος το semicolon (;).
- ✓ Στην γραμμή 22 το μήνυμα που μεταγράφηκε στην println είναι διαφορετικό από αυτό που αναγράφεται στο πρωτότυπο πρόγραμμα.
- ✓ Στην γραμμή 35 το μήνυμα που μεταγράφηκε στην println είναι διαφορετικό από αυτό που αναγράφεται στο πρωτότυπο πρόγραμμα.
- ✓ Στην γραμμή 36 δεν μεταγράφηκε ο τύπος int μπροστά από την μεταβλητή i στον βρόγχο for.
- ✓ Στις γραμμές 42 – 45 το μήνυμα που μεταγράφηκε στην println είναι διαφορετικό από αυτό που αναγράφεται στο πρωτότυπο πρόγραμμα.
- ✓ Στην γραμμή 46 η εντολή κλεισίματος της βιβλιοθήκης scanner δεν μεταγράφηκε (scanner.close)

Η απόδοση του εργαλείου δεν ήταν και πολύ ικανοποιητική. Στην ενάτη προσπάθεια μεταγραφής το αποτέλεσμα ήταν καλύτερο σε σχέση με τις προηγούμενες προσπάθειες. Υπήρξε δυσκολία να αντιληφθεί κάποιες λέξεις όπως για παράδειγμα την λέξη float σε αρκετά σημεία στο μεταγραφόμενο πρόγραμμα.

Η βελτίωση της ακρίβειας των συστημάτων φωνής σε κείμενο για γλώσσες προγραμματισμού ή η διασφάλιση ότι ο υπαγορευμένος κώδικας υφίσταται διεξοδική αναθεώρηση και διαδικασία διόρθωσης θα ήταν απαραίτητη για να καταστεί ένα τέτοιο σύστημα όπως το Dragon Professional πρακτικό βολικό και αξιόπιστο για τους προγραμματιστές.

## 5.5 Σχολιασμός Παραγόμενου Μεταγραφόμενου Προγράμματος στην Γλώσσα Python από το Dragon Professional

Όπως βλέπουμε στο μεταγραφόμενο πρόγραμμα που δημιουργήθηκε από το εργαλείο υπάρχουν κάποιες πολύ **μικρές** διαφορές σε σχέση με το πρωτότυπο πρόγραμμα.

- ✓ Στις γραμμές 2 και 8 η μεταβλητή «sum» μεταγράφηκε ως «some»
- ✓ Στην γραμμή 17 παραλείφθηκε η λέξη «encountered» στο μήνυμα που θα εκτυπώσει η print

Η απόδοση του εργαλείου ήταν σχεδόν άριστη. Στην έκτη κιόλας προσπάθεια μεταγραφής το αποτέλεσμά ήταν εξαιρετικό. Αυτό επιβεβαιώνεται και από το χαμηλό ποσοστό σφάλματος λέξεων WER που προέκυψε.

## Κεφάλαιο 6: Συμπεράσματα και Μελλοντικές Επιδιώξεις

---

6.1 Συμπεράσματα	77
6.2 Μελλοντικές Επιδιώξεις	78

---

### 6.1 Συμπεράσματα

Καθώς φτάνουμε στο τέλος αυτής της υπέροχης ερευνητικής διαδρομής, η οποία ξεκίνησε ως μια αποστολή για την εξερεύνηση και αξιολόγηση των εργαλείων φωνητικής αναγνώρισης, αντιλαμβανόμαστε πλέον την ανεκτίμητη συμβολή τους στην προώθηση της προσβασιμότητας και της τεχνολογικής ένταξης. Αυτή η διπλωματική εργασία δεν αποτέλεσε απλώς μια ακαδημαϊκή άσκηση, αλλά μια προσωπική περιπέτεια γνώσης, κατανόησης και ανακάλυψης.

Η εμπάθунση στις δυνατότητες, τις προκλήσεις και τις προοπτικές των εργαλείων φωνητικής αναγνώρισης, μας έδωσε την ευκαιρία να αντιληφθούμε την επιρροή της τεχνολογίας στην καθημερινή ζωή και την επιστημονική έρευνα. Παράλληλα, η προσπάθεια ενσωμάτωσης και αξιολόγησης του εργαλείου φωνητικής αναγνώρισης Dragon Professional V16 μέσα από πρακτικά σενάρια χρήσης, έφερε στο φως τις ανεξερεύνητες δυνατότητές του στην συγγραφή κώδικα στις τρεις πλέον πιο διαδεδομένες γλώσσες προγραμματισμού στον επιστημονικό τομέα της Πληροφορικής C, Java και Python.

Η εργασία αυτή έχει ως γνώμονα όχι μόνο την ακαδημαϊκή συνεισφορά αλλά και την πρακτική εφαρμογή των ευρημάτων της στην πραγματική ζωή, ιδιαίτερα για τα άτομα με αναπηρίες, τα οποία θα έχουν την δυνατότητα να γνωρίσουν κάποια από τα εργαλεία φωνητικής αναγνώρισης που είναι διαθέσιμα στην σύγχρονη αγορά, αλλά πιο συγκεκριμένα να εξερευνήσουν τις δυνατότητες και λειτουργίες που διαθέτει το Dragon Professional V16 έτσι ώστε να μπορέσουν να βελτιώσουν σημαντικά την ποιότητα της αλληλεπίδρασής τους με τον ψηφιακό κόσμο. Το εργαλείο αυτό αποτελεί μια υπενθύμιση της άρρηκτα συνδεδεμένης σχέσης μεταξύ ανθρώπου και τεχνολογίας, και του ρόλου που μπορεί να διαδραματίσει τα τελευταία χρόνια στη διευκόλυνση της καθημερινότητας και της αυτονομίας του ανθρώπου.

Κλείνοντας αυτό το ερευνητικό ταξίδι στον κόσμο της Πληροφορικής, ελπίζω να έχω συμβάλει καταλυτικά στην ανάδειξη της σημασίας των εργαλείων φωνητικής αναγνώρισης και να έχω εμπνεύσει για περαιτέρω έρευνα και ανάπτυξη στον συγκεκριμένο τομέα. Η τεχνολογία συνεχίζει να εξελίσσεται με αλματώδης ρυθμούς όπως μαζί της και οι δυνατότητες για περαιτέρω βελτίωση της ανθρώπινης εμπειρίας. Ας συμβάλουμε και εμείς ως νέοι ερευνητές στον χώρο με σκοπό να δημιουργηθεί ένα εξίσου αντάξιο εργαλείο φωνητικής αναγνώρισης που να υποστηρίζει την ελληνική γλώσσα έτσι ώστε να μπορούν και τα καθαρά ελληνόφωνα άτομα που το έχουν ανάγκη να το χρησιμοποιούν στην καθημερινότητα τους. Ας προχωρήσουμε λοιπόν με αισιοδοξία και διευρυνμένο μυαλό προς τις νέες προκλήσεις και τις ανεξερεύνητες δυνατότητες που μας περιμένουν να τις ερευνήσουμε.

## **6.2 Μελλοντικές Επιδιώξεις**

Στο επίκεντρο των μελλοντικών βελτιώσεων του λογισμικού Dragon Professional V16 βρίσκεται η ενίσχυση της ακρίβειας της φωνητικής αναγνώρισης σε όλες τις γλώσσες προγραμματισμού. Αυτή η προσπάθεια περιλαμβάνει τη μείωση του ποσοστού σφάλματος στην αναγνώριση λέξεων (WER) σε διάφορα, ειδικά σε θορυβώδη, περιβάλλοντα, όπου η ακρίβεια είναι κρίσιμη. Αυτό απαιτεί την ανάπτυξη νέων τεχνολογικών λύσεων που θα επιτρέπουν στο λογισμικό να απομονώνει και να επεξεργάζεται τα φωνητικά δεδομένα με μεγαλύτερη ακρίβεια, παρέχοντας βελτιωμένη υποστήριξη στους χρήστες.

Η επέκταση του λεξιλογίου του Dragon Professional V16 αποτελεί επίσης προτεραιότητα, με στόχο την ενσωμάτωση εξειδικευμένων τεχνικών όρων και συντομεύσεων για κάθε γλώσσα προγραμματισμού. Αυτό θα διευκολύνει την αναγνώριση και μεταγραφή πιο σύνθετων εντολών και κώδικα, μειώνοντας τον χρόνο επεξεργασίας και βελτιώνοντας τη γενική αποδοτικότητα του λογισμικού. Επιπλέον, η προσαρμοσμένη εκπαίδευση θα ενισχύσει τη δυνατότητα του λογισμικού να μαθαίνει από τη φωνή και τις συνήθειες του χρήστη, επιτρέποντας έτσι την ταχύτερη και ακριβέστερη εκτέλεση εντολών.

Τέλος, η διεύρυνση της συμβατότητας του λογισμικού και η ενσωμάτωση τεχνολογιών τεχνητής νοημοσύνης θα καταστήσουν το Dragon Professional V16 ένα ακόμη πιο ισχυρό εργαλείο. Η υποστήριξη περισσότερων προγραμματιστικών περιβαλλόντων και η ικανότητα διαχείρισης πολλαπλών γλωσσών και διαλέκτων θα επιτρέψουν την ευρεία

χρήση του λογισμικού σε διάφορα τεχνολογικά πεδία. Παράλληλα, η χρήση τεχνητής νοημοσύνης για την πρόβλεψη και αυτόματη διόρθωση λαθών θα ενισχύσει την απόδοση του λογισμικού, καθιστώντας το απαραίτητο εργαλείο για επαγγελματίες και ερευνητές που εξαρτώνται από ακριβή και αποδοτική φωνητική αναγνώριση.



## **Αναφορές**

- [1] Cavalier, A.R. and Ferretti, R.P., 1996. Talking instead of typing: Alternate access to computers via speech recognition technology. Focus on Autism and Other Developmental Disabilities, 11(2), pp.79-85.
- [2] Tang, K.W., Kamoua, R., Sutan, V., Farooq, O., Eng, G., Chu, W.C. and Hou, G., 2004. Speech recognition technology for disabilities education. Journal of Educational Technology Systems, 33(2), pp.173-184.
- [3] Kim, M.J., Yoo, J., and Kim, H., 2013. Dysarthric Speech Recognition Using Dysarthria-Severity-Dependent and Speaker-Adaptive Models. In Proceedings of Interspeech 2013, 25-29 August 2013, Lyon, France, pp. 3622-3626.
- [4] Transkriptor, "Μετατροπή ήχου ή βίντεο σε κείμενο," Transkriptor.  
Available at: <https://transkriptor.com/>
- [5] "5 Best Transcription Services of 2024 Compared (Cheap & Accurate)", 2024  
Available at: <https://wpbeginner.com>
- [6] Google, "Google Docs Voice Typing," Google Docs Help, 2024.  
Available at: <https://support.google.com/docs/answer/4492226?hl=en>
- [7] IBM, "IBM Watson Speech to Text," IBM Watson, 2023.  
Available at: <https://www.ibm.com/cloud/watson-speech-to-text>
- [8] Otter.ai, "Otter.ai: Meeting Notes and Transcription," Otter.ai, 2024.  
Available at: <https://otter.ai/>
- [9] Sonix, "Sonix: Automated Transcription, Translation, & Subtitles," Sonix, 2024.  
Available at: <https://sonix.ai/>
- [10] Voice Recognition Australia, "Top 10 Reasons to Use Dragon voice recognition software", 2021.  
Available at: <https://www.voicerecognition.com.au/blogs/news/top-10-reasons-to-use-dragon-voice-recognition-software>
- [11] Nuance, "Dragon Speech Recognition - Get More Done by Voice", 2023.  
Available at: <https://www.nuance.com/dragon.html>

- [12] Nuance, "Dragon Professional Speech Recognition, v16", 2023.  
Available at: <https://www.nuance.com/dragon/business-solutions/dragon-professional.html>
- [13] Nuance, "Dragon Professional 16 Help", 2023.  
Available at: <https://www.nuance.com/products/help/dragon16/dragon-for-pc/enx/dpg-cp/Content/GetStart1.htm>
- [14] Dragonspechtips.com, "Dragon Professional Commands and Training", 2022.  
Available at: <https://dragonspechtips.com/>
- [15] Speechelo, "Text To Speech Software, Transforming Text to Voice Instantly", 2023. Available at: <https://speechelo.com/>
- [16] Microsoft, "Dictate in Microsoft 365," Microsoft Support, 2023. Available at: <https://support.microsoft.com/en-us/office/dictate-in-microsoft-365-eab203e1-d030-43c1-84ef-999b0b9675fe>
- [17] Adam Cogan, "Going handsfree in Office 365 – You should be using Dictate," 2021. Available at: <https://adamcogan.com/2021/09/10/going-handsfree-in-office-365-you-should-using-dictate/>
- [18] Walker, S., 2019. "Evaluating an Automatic Speech Recognition Service".  
Available at: <https://aws.amazon.com/blogs/machine-learning/evaluating-an-automatic-speech-recognition-service/>
- [19] Rev Blog, "What Word Error Rate (WER) Is & What it Means", 2020. Available at: <https://www.rev.com/blog/resources/what-is-wer-what-does-word-error-rate-mean>
- [20] Robert Koch, Clickworker, "Transcribe Audio to Text: A Guide to Speech Recognition and Transcription", 2023.  
Available at: <https://www.clickworker.com/customer-blog/transcribe-audio-to-text/>
- [21] Total Voice Technologies, "A Brief History of Voice Recognition Technology", 2022.  
Available at: <https://www.totalvoicetech.com/a-brief-history-of-voice-recognition-technology/>

[22] Wikipedia, "Dragon NaturallySpeaking", 2023.

Available at: [https://en.wikipedia.org/wiki/Dragon\\_NaturallySpeaking](https://en.wikipedia.org/wiki/Dragon_NaturallySpeaking)

[23] Rewisdom.ai, "7 Use Cases of Speech To Text Software", 2022.

Available at: <https://www.rewisdom.ai/>

## **Παράρτημα Α**

Στα πλαίσια της παρούσας διπλωματικής εργασίας, συντάξαμε ένα ερευνητικό άρθρο που περιλαμβάνει το δεύτερο σενάριο χρήσης της Διπλωματικής Εργασίας. Το σενάριο αυτό εστιάζει στη χρήση του εργαλείου Dragon Professional V16 για τη συγγραφή κώδικα σε τρεις γλώσσες προγραμματισμού: C, Java και Python. Το άρθρο αξιολογεί την αποδοτικότητα του εργαλείου σε αυτή την απαιτητική πρόκληση στον τομέα της Πληροφορικής, με στόχο να διευκολύνει τα άτομα με κινητικές δυσκολίες να αλληλεπιδρούν αποδοτικά, εύκολα και γρήγορα με τον υπολογιστή, ακόμα και όταν πρόκειται να γράψουν κώδικα.

Η μελέτη περίπτωσης που μελετήθηκε αναλύει τη διαδικασία μεταγραφής κώδικα με τη χρήση της τεχνολογίας φωνητικής αναγνώρισης του Dragon Professional V16 και εξετάζει την απόδοση του μεταξύ των τριών γλωσσών προγραμματισμού.

Το άρθρο, το οποίο δημοσιεύθηκε στο συνέδριο, έχει τίτλο: "A Case Study of Voice Recognition Technology for Developing Programming Skills for Students with Motor Disabilities". Μέσα από αυτή τη μελέτη, επιδιώκουμε να αναδείξουμε τα πλεονεκτήματα και τις προκλήσεις της εφαρμογής της τεχνολογίας φωνητικής αναγνώρισης στον προγραμματισμό. Στόχος μας είναι να συμβάλλουμε στην ανάπτυξη καινοτόμων τεχνολογιών που προάγουν την ισότιμη πρόσβαση στην εκπαίδευση και την εργασία στον τομέα της πληροφορικής, διευκολύνοντας τα άτομα με κινητικές δυσκολίες να αξιοποιήσουν τις δυνατότητες τους και να ενταχθούν ενεργά στο ψηφιακό περιβάλλον.

### **Λεπτομέρειες Διεξαγωγής Συνεδρίου**

Το συνέδριο PETRA 2024 (PErvasive Technologies Related to Assistive Environments) θα πραγματοποιηθεί στην Κρήτη, Ελλάδα, από τις 26 έως τις 28 Ιουνίου 2024, με την άφιξη των συμμετεχόντων στις 25 Ιουνίου. Το συνέδριο επικεντρώνεται σε διεπιστημονικές προσεγγίσεις που στοχεύουν στη βελτίωση της ποιότητας ζωής και της ανθρώπινης απόδοσης μέσω της χρήσης των διάχυτων τεχνολογιών που παρέχονται σε διάφορα περιβάλλοντα, όπως το σπίτι, ο χώρος εργασίας και οι δημόσιοι χώροι. Καλύπτονται θέματα όπως η αλληλεπίδραση ανθρώπου-υπολογιστή, οι τεχνολογίες υγείας και υποστήριξης, τα έξυπνα περιβάλλοντα, η φορητή υπολογιστική και η τεχνητή νοημοσύνη. Τα άρθρα του συνεδρίου θα δημοσιευθούν στη ψηφιακή

βιβλιοθήκη της ACM. Για περισσότερες πληροφορίες, μπορείτε να επισκεφτείτε την ιστοσελίδα του συνεδρίου: <http://www.petrae.org/> .

# A case study of Voice Recognition Technology for Developing Programming Skills for Students with Motor Disabilities

Marios, MF, Flourentzou\*

Department of Computer Science, University of Cyprus, Cyprus flourentzou.marios@ucy.ac.cy

Eirini, ES, Schiza

HealthXR, CYENS, Cyprus, e.schizas@cyens.org.cy

Constantinos, CP, Pattichis

Department of Computer Science, University of Cyprus, HealthXR, CYENS, Cyprus, pattichis@ucy.ac.cy

This paper focuses on evaluating a commercial voice recognition technology (Dragon Professional v16), for developing programming skills with a particular emphasis on its utility for people with disabilities. We have designed a testing scenario that involves dictating the same program in three different programming languages: C, Java, and Python. Following the dictation of these programs, the study employs the Word Error Rate (WER) metric to quantify Dragon Professional v16's performance. The WER for C, Java, and Python were in the range of 13-18%, 20-27% and 3-5% respectively. The findings of this study offer insights for programmers, including those with disabilities, who rely on voice recognition technology to engage with their work.

**CCS CONCEPTS** • Information systems applications • Computing platforms

**Additional Keywords and Phrases:** Voice recognition technology, Motor disabilities, Programming

**ACM Reference Format:**

Marios, MF, Flourentzou, Eirini, ES, Schiza, and Constantinos, CP, Pattichis. 2024. A case study of Voice Recognition Technology for Developing Programming Skills for Students with Motor Disabilities: ACM Conference Proceedings Manuscript Submission Template. In Woodstock '18: ACM Symposium on Neural Gaze Detection, June 03–05, 2018, Woodstock, NY. ACM, New York, NY, USA, 10 pages. NOTE: This block will be automatically generated when manuscripts are processed after acceptance.

## 1 Introduction

Voice recognition technology stands as a transformative tool for people with disabilities, offering them an invaluable avenue for communication and independence [1], [2]. For those with physical impairments, particularly affecting motor skills, this technology serves as a bridge, bypassing traditional barriers by enabling control of devices, composition of messages, and navigation of digital spaces through spoken commands. In educational and professional contexts, it levels the playing field, allowing individuals to participate and contribute without the limitations imposed by conventional input methods.

---

\* Place the footnote text for the author (if applicable) here.

The objective of this work was to evaluate a commercial voice recognition technology (Dragon Professional v16), for developing programming skills with a particular emphasis on its utility for people with disabilities. We have designed a testing scenario that involves dictating the same program in three different programming languages: C, Java, and Python. These languages were selected due to their prominence in the software development industry, their unique syntaxes, and their diverse applications.

## 2 Methodology

### 2.1 Case Study

The work presented here was carried out in the context of the course CS400 Diploma Project, at the Department of Computer Science at the University of Cyprus by the first author (MF). MF was born prematurely at six months of gestation and subsequently was diagnosed with spastic quadriplegic cerebral palsy. This condition has impacted both his upper and lower extremities, necessitating the use of a wheelchair for mobility. His computer access is based on a specialized keyboard with larger, spaced keys, enabling him to individually press them with his left hand's index finger, and a trackball, given his significant struggle with using a mouse.

### 2.2 Voice Recognition Technology exploited

The DragonR Professional v16 ( <https://www.nuance.com/dragon.html> ) voice recognition software was used. DragonR integrates seamlessly with popular applications, allowing users to navigate and control their PC entirely by voice commands, thereby significantly reducing physical strain for users, especially those with mobility impairments. It was installed on a laptop (model ZenBook UX535LI, Intel(R) Core(TM) i7-10870H CPU @ 2.20 GHz, 8 Core(s)) with Windows 11 Pro and Visual Studio Code. Training was carried out over three daily sessions for seven days, spending in total 8 hours. The laptop's microphone was used.

### 2.3 Programming skills experiments

A simple program to input ten real numbers and compute their sum, product, difference, and quotient was used to evaluate the performance of DragonR Professional v16 voice recognition in three different programming languages (C, Java and Python). This approach enables us to thoroughly assess the performance and reliability of transcriptions made by the DragonR dictation software by comparing them against the original versions of the program written in each programming language.

The performance of each programming transcription was evaluated using the Word Error Rate (WER) given by:  $WER = (S + D + I) / N$ , where: S represents the number of substitutions (words that are incorrectly recognized); D represents the number of deletions (words that exist in the prototype program but not in the dictated program); I represents the number of insertions (words that exist in the dictated program produced but are not contained in the prototype program); N represents the total number of words that exist in the prototype program.

## 3 Results

To achieve optimal transcription quality across each programming language, multiple attempts were made to refine the process. This iterative approach was essential due to the unique syntax and vocabulary demands of each programming language. The tool needed to familiarize itself with the specific lexicon associated with each language, which initially wasn't part of its built-in dictionary. Through persistent efforts, the tool was tuned to best capture the nuances of each language's syntax, demonstrating its adaptability and learning capacity in recognizing and transcribing programming commands.

[Table 1](#) summarizes the performance evaluation metrics for the best-worst trials for the three programming skills experiments carried out. The outcomes, based on WER underscored the tool's efficiency in Python, achieving low word error rates in the range of 3-5%. However, the WER range for C (13-18%) and Java (20-27%) are

prohibitive for the use of DragonR for the transcription of programs in these programming languages. Errors included typographical errors in variable names or omitted punctuation, that disrupted the syntactic integrity of the code, hindering accurate compilation and execution.

**Table 1: Performance evaluation metrics for the best-worst trials for the three programming skills experiments carried out (see section 2.4 for WER metric definition)**

Programming Language	Substitutions (S)	Deletions (D)	Insertions (I)	N	WER	Time Spent in min.
C	15-18	1-3	1-3	135	13-18%	90-120
Java	24-27	5-7	6-7	176	20-27%	120-150
Python	2-3	2-2	0-1	119	3-5%	60-90

## 4 Concluding Remarks

In conclusion, the exploration and evaluation of DragonR capabilities in transcribing programming languages have demonstrated its potential in streamlining the coding process, particularly in Python where it achieved a low Word Error Rate of 3%. This low error rate, indicative of the software's capability to accurately interpret orally expressed Python code with minimal adjustments, significantly boosts productivity and accessibility, particularly for those challenged by traditional typing methods. Despite this, the precise nature of programming underscores the importance of diligent code review to mitigate any functionality issues caused by minor errors.

However, the WER range for C and Java are prohibitive for the use of DragonR for the transcription of programs in these programming languages. This stark contrast in performance across different languages Python, Java, and C underscores the importance of specialized training and the tool's ability to adapt to the distinct syntax and vocabulary inherent to each programming language.

DragonR not being specialized for coding, necessitated considerable time and multiple attempts to adapt to the specific commands each transcription demanded, especially for C and Java. Certainly, the accuracy of the transcriptions improved with further training, showcasing the software's adaptability and learning capabilities. The journey from initial attempts to achieving optimal transcription accuracy showcases not only the tool's learning curve but also the role of user persistence and adaptation in harnessing technology to meet the precise demands of programming tasks. This study reaffirms the growing significance of voice recognition technologies in making programming more accessible, efficient, and less physically taxing, thereby opening new avenues for innovation and inclusivity in software development.

## ACKNOWLEDGMENTS

This work has received partial funding from the European Union's Horizon 2020 Research and Innovation Programme under Grant Agreement No 739578 and the Government of the Republic of Cyprus through the Deputy Ministry of Research, Innovation and Digital Policy.

## REFERENCES

- < bib id="bib1">< number>[1]</ number> Cavalier, A.R. and Ferretti, R.P., 1996. Talking instead of typing: Alternate access to computers via speech recognition technology. Focus on Autism and Other Developmental Disabilities, 11(2), pp.79-85.</ bib>  
 < bib id="bib2">< number>[2]</ number> Tang, K.W., Kamoua, R., Sutan, V., Farooq, O., Eng, G., Chu, W.C. and Hou, G., 2004. Speech recognition technology for disabilities education. Journal of Educational Technology Systems, 33(2), pp.173-184.</ bib>