

**ΤΕΧΝΟΛΟΓΙΚΟ ΕΚΠΑΙΔΕΥΤΙΚΟ ΙΔΡΥΜΑ ΚΕΝΤΡΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ**  
**ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΚΩΝ ΕΦΑΡΜΟΓΩΝ**  
**ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΤΕ**

**ΑΝΑΠΤΥΞΗ ΒΙΒΛΙΟΘΗΚΗΣ ΠΟΥ ΕΠΙΤΡΕΠΕΙ**  
**ΤΟΝ ΕΛΕΓΧΟ ΕΦΑΡΜΟΓΩΝ ΜΕΣΩ ΦΩΝΗΤΙΚΩΝ**  
**ΕΝΤΟΛΩΝ**

**Πτυχιακή εργασία του**  
Πετρόπουλος Ευάγγελος (3785)  
Επιβλέπων: Ν. Πεταλίδης

**ΣΕΡΡΕΣ, ΜΑΪΟΣ 2020**

## Υπεύθυνη δήλωση

Υπεύθυνη Δήλωση: Βεβαιώνω ότι είμαι συγγραφέας αυτής της πτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην πτυχιακή εργασία. Επίσης έχω αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επίσης βεβαιώνω ότι αυτή η πτυχιακή εργασία προετοιμάστηκε από εμένα προσωπικά ειδικά για τις απαιτήσεις του προγράμματος σπουδών του Τμήματος Μηχανικών Πληροφορικής ΤΕ του Τ.Ε.Ι. Κεντρικής Μακεδονίας.

## Σύνοψη

Οι μελισσοκόμοι καταγράφουν σημειώσεις για τα μελίσσια που ελέγχουν αλλά πολλές φορές δεν προλαβαίνουν επειδή ο αριθμός των μελισσιών είναι μεγάλος ή δουλεύουν μόνοι τους και πρέπει να σταματάνε την εργασία για να καταγράψουν τις σημειώσεις. Η εφαρμογή που αναπτύξαμε και ονομάζεται ConApi (Contemporary Apiculture) υλοποιεί ένα σύστημα φωνητικών εντολών που βοηθάνε τον μελισσοκόμο στην καταγραφή των σημειώσεων χωρίς την χρήση της συσκευής με τα χέρια παρά μόνο της φωνής.

# Περιεχόμενα

<b>Υπεύθυνη δήλωση</b>	<b>2</b>
<b>Σύνοψη</b>	<b>3</b>
<b>Πρόλογος</b>	<b>8</b>
<b>Ευχαριστίες</b>	<b>9</b>
<b>Ορισμοί</b>	<b>10</b>
<b>1 Εισαγωγή</b>	<b>11</b>
1.1 Δομή της εργασίας . . . . .	11
<b>2 Έλεγχος μελισσιών και χειρόγραφες σημειώσεις</b>	<b>13</b>
2.1 Έλεγχος . . . . .	13
2.2 Χειρόγραφες σημειώσεις . . . . .	13
2.3 Δυσκολίες . . . . .	14
<b>3 Εφαρμογή και φωνητικές εντολές</b>	<b>15</b>
3.1 Εφαρμογή . . . . .	15
3.2 GUI-Γραφικό περιβάλλον διεπαφής χρήστη . . . . .	16
3.3 VUI-Διεπαφή χρήστη φωνής . . . . .	16
3.4 Φωνητικές Εντολές . . . . .	16
<b>4 Ανασκόπηση Speech-to-Text APIs</b>	<b>18</b>
4.1 APIs . . . . .	18
4.1.1 Google Cloud Speech-to-Text API . . . . .	18
4.1.1.1 Περιγραφή . . . . .	18

4.1.1.2	Δυνατότητες . . . . .	18
4.1.2	Microsoft Cognitive Services Speech-to-Text API . . . . .	20
4.1.2.1	Περιγραφή . . . . .	20
4.1.2.2	Δυνατότητες . . . . .	21
4.1.3	IBM Watson API . . . . .	21
4.1.3.1	Περιγραφή . . . . .	21
4.1.3.2	Δυνατότητες . . . . .	22
4.2	Επιλογή API . . . . .	22
<b>5</b>	<b>Σχεδίαση Βιβλιοθήκης</b>	<b>24</b>
5.1	Sequence Diagram . . . . .	26
5.2	State Diagram . . . . .	26
5.3	Component Diagram . . . . .	26
	<b>Γλωσσάρι</b>	<b>28</b>

## Κατάλογος πινάκων

2.1	Ημερολόγιο Μελισσοκόμου. . . . .	14
3.1	Φωνητικές εντολές . . . . .	17

## **Κατάλογος διαγραμμάτων**

# Πρόλογος

Εδώ μπορεί να μπει πρόλογος. (Δεν είναι απαραίτητο).



# Ευχαριστίες

Ευχαριστίες (στο μπαμπά, στη μαμά, κτλ)

# Ορισμοί

Ορισμοί εννοιών που μπορεί να είναι χρήσιμοι. Για παράδειγμα:

**L<sup>A</sup>T<sub>E</sub>X** Σύστημα στοιχειοθεσίας κειμένων

# Κεφάλαιο 1

## Εισαγωγή

Η μελισσοκομία είναι μια επιστήμη που οι άνθρωποι ασχολούνται από τα αρχαία χρόνια και μία από τις δουλειές είναι ο τακτικός έλεγχος για την πρόοδο των μελισσιών. Σε κάθε έλεγχο του μελισσιού καταγράφουν χειρόγραφες σημειώσεις για την πρόοδο του. Οι σημειώσεις είτε γράφονται παράλληλα με τον έλεγχο των μελλισιών είτε μετά το πέρας του έλεγχου.

Δυσκολίες συναντιούνται με την καταγραφή των χειρόγραφων σημειώσεων, όπως στο μεγάλο αριθμό μελισσιών ο μελισσοκόμος να μην προλαβαίνει να καταγράψει τις παρατηρήσεις του από τον έλεγχο, να ξεχνάει τι είχε παρατηρήσει στα μελίσσια στην αρχή του ελέγχου. Επιπλέον ο μελισσοκόμος φοράει ειδική στολή (μάσκα, γάντια) που καθιστούν δύσκολη την καταγραφή των παρατηρήσεων όπως με την μάσκα δεν θα βλέπει καλά λόγο της σίτας που έχει, επίσης με τα γάντια υπάρχει δυσκολία στον να κρατήσει το μολύβι, το τετράδιο και να καταγράψει τις σημειώσεις.

Για τους παραπάνω λόγους είναι ενδιαφέρουσα η ανάπτυξη ενός εργαλείου το οποίο θα λύνει τα χέρια του μελισσοκόμου χρησιμοποιώντας μόνο την φωνή του για την καταγραφή των παρατηρήσεων.

Σκοπός αυτής της εργασίας είναι η ανάπτυξη ενός τέτοιου εργαλείου που θα υποστηρίζει φωνητικές εντολές για την ευκολία καταγραφής των παρατηρήσεων.

### 1.1 Δομή της εργασίας

**Κεφάλαιο 2** Παρουσίαση έλεγχου μελισσιών και χειρόγραφες σημειώσεις.

**Κεφάλαιο 3** Εφαρμογή και φωνητικές εντολές.

**Κεφάλαιο 4** Speech-to-Text API.

## Κεφάλαιο 2

# Έλεγχος μελισσιών και χειρόγραφες σημειώσεις

### 2.1 Έλεγχος

Ο έλεγχος των μελισσιών πραγματοποιείται όλες τις εποχές του χρόνου. Ο μελισσοκόμος μετά απο τον Χειμώνα, στην αρχή της Άνοιξης θα κάνει τον πρώτο έλεγχο των μελισσιών που θα κοιτάξει αν η βασίλισσα του μελισσιού είναι ζωντανή και αν γεννάει γόνο. Στα μέσα της Άνοιξης θα ελέγξει αν ο γόνος αυξήθηκε, αν είναι συμπαγής και πόσα πλαίσια γόνου έχει το μελίσσι.

Την επόμενη εποχή, το καλοκαίρι με τους ελέγχους που θα κάνει θα παρατηρήσει αν τα μελίσσια συλλέγουν μέλι και αν η βασίλισσα σταμάτησε να γεννάει γόνο. Προς το τέλος του καλοκαιριού θέλει να δει αν έχουν συλλέξει αρκετό μέλι για να γίνει ο τρύγος.

Το Φθινόπωρο γίνεται έλεγχος για να διακρίνει αν η βασίλισσα γεννάει συμπαγή γόνο και το Χειμώνα αν τα μελίσσια έχουν ασθένειες.

### 2.2 Χειρόγραφες σημειώσεις

Οι χειρόγραφες σημειώσεις του μελισσοκόμου αποτελούνται από τον αριθμό της κυψέλης και την ηλικία της βασίλισσας. Επιπλέον, σε κάθε έλεγχο καταγράφεται η ημερομηνία, πόσα πλαίσια έχει η κυψέλη και από τα οποία πόσα έχουν πληθυσμό. Επίσης, καταγράφεται απο τα πλαίσια πόσα έχουν γόνο, μέλι και γύρη. Τέλος, σημειώνονται

γενικές παρατηρήσεις/υπενθυμίσεις της κυψέλης.

**Πίνακας 2.1:** Ημερολόγιο Μελισσοκόμου.

Αριθμός Κυψέλης: 6			Ηλικία Βασίλισσας: 05-2019			
Ημερομηνία	Πλαίσια	Πληθυσμός	Γόνος	Μέλι	Γύρη	Παρατηρήσεις
09-04-2019	10	7	4			ΟΚ
18-04-2019	10	10	7			Θέλει όροφο
24-04-2019	15	10	8			2 πλαίσια έχουν βασιλικά κελιά, μπήκε όροφος
03-05-2019	15	10	8			Έκοψα παραφιάδα N10

## 2.3 Δυσκολίες

Οι μελισσοκόμοι συναντάνε δυσκολίες στο να κρατάνε τις σημειώσεις τους. Μια από αυτές τις δυσκολίες είναι η στολή που χρησιμοποιούν, η οποία αποτελείται από την μάσκα, την φόρμα και τα γάντια. Η μάσκα έχει τούλι στο να επιτρέπει τον μελισσοκόμο να βλέπει και να εμποδίζει τις μέλισσες να εισέλθουν μέσα στην μάσκα. Το τούλι της μάσκας δυσκολεύει την διαδικασία καταγραφής των σημειώσεων διότι εμποδίζει στο να βλέπεις καθαρά το τετράδιο. Τα μελισσοκομικά γάντια αποτρέπουν τις μέλισσες από τσιμπήματα στα χέρια αλλά δυσκολεύουν στο κράτημα του μολυβιού ώστε να γράφτούν οι σημειώσεις στο τετράδιο.

## Κεφάλαιο 3

# Εφαρμογή και φωνητικές εντολές

### 3.1 Εφαρμογή

Η εφαρμογή που υλοποιείται σε αυτή την εργασία δύναται να λύσει τις δυσκολίες που αντιμετωπίζει ο μελισσοκόμος. Όπως αναφέραμε παραπάνω η δυσκολία που συναντάει ο μελισσοκόμος είναι η καταγραφή των σημειώσεων που είναι σημαντικές για την πρόοδο των μελισσιών. Η εφαρμογή κρατάει τα δεδομένα όπως παρουσιάζονται στο παραπάνω πίνακα 2.1. Συγκεκριμένα, οι εγγραφές για όλα τα μελίσσια που δουλεύει ο μελισσοκόμος.

Για κάθε μελίσσι τα δεδομένα είναι:

- Αριθμός Κυψέλης
- Ηλικία Βασίλισσας
- Ημερομηνία (κάθε ελέγχου)
- Πλαίσια (αριθμός πλαίστιων κυψέλης)
- Πληθυσμός κυψέλης (αριθμός πλαίστιων με μέλισσες)
- Γόνος (αριθμός πλαίστιων με γόνο)
- Μέλι (αριθμός πλαίστιων με μέλι)
- Γύρη (αριθμός πλαίστιων με γύρη)
- Παρατηρήσεις

Η εφαρμογή με την υλοποίηση ενός γραφικού περιβάλλοντος για την προβολή και επεξεργασία των μελισσοκομικών δεδομένων και την υλοποίηση μιας διεπαφής χρήστη φωνής με φωνητικές εντολές για την καταχώρηση των μελισσοκομικών δεδομένων λύνοντας τις δυσκολίες που αντιμετωπίζει ο μελισσοκόμος κατά τη διάρκεια της εργασίας.

### **3.2 GUI-Γραφικό περιβάλλον διεπαφής χρήστη**

Το γραφικό περιβάλλον της εφαρμογής θα παρουσιάζει με ωραία γραφικά τα μελίσσια και κάθε πληροφορία του μελισσιού και θα δίνει δυνατότητα στον χρήστη να επεξεργαστεί αυτή την πληροφορία.

### **3.3 VUI-Διεπαφή χρήστη φωνής**

### **3.4 Φωνητικές Εντολές**

Δυο κατηγορίες φωνητικών εντολών υποστηρίζονται, οι οποίες είναι:

1. Μελισσιού
2. Έλεγχος μελισσιού

Οι φωνητικές εντολές κατά τον έλεγχο του μελισσιού θα ενεργοποιούνται όταν επιλεγθεί το μελίσσι για έλεγχο. Οι φωνητικές εντολές της εφαρμογής είναι:



**Πίνακας 3.1:** Φωνητικές εντολές

Φωνητικές Εντολές	Περιγραφή	Παράδειγμα
Μελίσσι <αριθμός>	Στο πεδίο <αριθμός> θα λέγεται ο αριθμός του μελισσιού	Μελίσσι 6
Πλαίσια <αριθμός>	Στο πεδίο <αριθμός> θα λέγεται ο αριθμός των πλαίσιων του μελισσιού	Πλαίσια 10
Πληθυσμός <αριθμός>	Στο πεδίο <αριθμός> θα λέγεται ο αριθμός των πλαίσιων με πληθυσμό του μελισσιού	Πληθυσμός 7
Γόνος <αριθμός>	Στο πεδίο <αριθμός> θα λέγεται ο αριθμός των πλαίσιων με γόνο του μελισσιού	Γόνος 4
Μέλι <αριθμός>	Στο πεδίο <αριθμός> θα λέγεται ο αριθμός των πλαίσιων με μέλι του μελισσιού	Μέλι 3
Γύρη <αριθμός>	Στο πεδίο <αριθμός> θα λέγεται ο αριθμός των πλαίσιων με γύρη του μελισσιού	Γύρη 3
Παρατήρηση <κείμενο>	Στο πεδίο <κείμενο> θα λέγεται η παρατήρηση για το μελίσσι	Παρατήρηση      Θέλει όροφο

## Κεφάλαιο 4

# Ανασκόπηση Speech-to-Text APIs

### 4.1 APIs

#### 4.1.1 Google Cloud Speech-to-Text API

##### 4.1.1.1 Περιγραφή

Είναι ένα API μεταγραφής ομιλίας σε κείμενο υποστηρίζοντας πολλές γλώσσες προγραμματισμού, γεγονός που το καθιστά ανεξάρτητο πλατφόρμας. Το API που υποστηρίζεται από τις τεχνολογίες AI της Google και έχει την δυνατότητα χρήσης διαφορετικών μοντέλων μηχανικής εκμάθησης για αιτήματα μεταγραφής ήχου σε Speech-to-Text.

##### 4.1.1.2 Δυνατότητες

- **Παγκόσμιο λεξιλόγιο** - Εκτεταμένη υποστήριξη γλώσσας ομιλίας σε κείμενο σε περισσότερες από 125 γλώσσες και παραλλαγές.
- **Ροή αναγνώριση ομιλίας** - Αποτελέσματα αναγνώρισης ομιλίας σε πραγματικό χρόνο καθώς το API επεξεργάζεται την είσοδο ήχου που μεταδίδεται σε ροή από το μικρόφωνο της εφαρμογής σας ή αποστέλλεται από ένα προκαθορισμένο αρχείο ήχου (inline ή μέσω Cloud Storage).
- **Προσαρμογή ομιλίας** - Προσαρμογή της αναγνώρισης ομιλίας για την μετατροπή ορών για συγκεκριμένους τομείς και σπάνιες λέξεις παρέχοντας συμβουλές και ενισχύση της ακρίβειας της μεταγραφής συγκεκριμένων λέξεων ή φρά-

σεων. Αυτόματη μετατροπή προφορικών αριθμών σε διευθύνσεις, έτη, νομίματα και άλλα χρησιμοποιώντας τάξεις.

- **Πολυκαναλική αναγνώριση** - Το Speech-to-Text μπορεί να αναγνωρίσει διαφορετικά κανάλια σε καταστάσεις πολλαπλών καναλιών (π.χ., τηλεδιάσκεψη) και να σχολιάσει τις μεταγραφές για τη διατήρηση της τάξης.
- **Ανθεκτικότητα θορύβου** - Το Speech-to-Text μπορεί να χειριστεί θορυβώδη ήχο από πολλά περιβάλλοντα χωρίς να απαιτείται επιπλέον ακύρωση θορύβου.
- **Μοντέλα ειδικά για τομέα** - Επιλέξτε από μια επιλογή εκπαιδευμένων μοντέλων για φωνητικό έλεγχο και τηλεφωνική κλήση και μεταγραφή βίντεο βελτιστοποιημένα για απαιτήσεις ποιότητας για συγκεκριμένο τομέα. Για παράδειγμα, το βελτιωμένο μοντέλο τηλεφωνικών κλήσεων είναι συντονισμένο για ήχο που προέρχεται από την τηλεφωνία, όπως οι τηλεφωνικές κλήσεις που έχουν εγγραφεί σε ρυθμό δειγματοληψίας 8khz.
- Υποστηρίζει γλώσσες προγραμματισμού όπως Protocol, C, Go, java, Node.js, PHP, Python και Ruby.

Λαμβάνοντας υπόψη ότι η Google είναι ουσιαστικά το νευρικό σύστημα του Διαδικτύου σε αυτό το σημείο, δεν αποτελεί έκπληξη ότι το API ομιλίας σε κείμενο είναι ένα από τα πιο δημοφιλή - και πιο ισχυρά - API που είναι διαθέσιμα στους προγραμματιστές.

Το Google Speech-To-Text παρουσιάστηκε το 2018, μόλις μία εβδομάδα μετά την ενημέρωση κειμένου σε ομιλία. Το API Speech-To-Text της Google προβάλλει ορισμένες τολμηρές αξιώσεις, μειώνοντας τα λάθη λέξεων κατά 54% σε δοκιμή μετά τη δοκιμή. Σε ορισμένους τομείς, τα αποτελέσματα είναι ακόμη πιο ενθαρρυντικά.

Ένας από τους λόγους για την εντυπωσιακή ακρίβεια των API είναι η δυνατότητα επιλογής μεταξύ διαφορετικών μοντέλων μηχανικής εκμάθησης, ανάλογα με το τι χρησιμοποιείται η εφαρμογή σας. Αυτό καθιστά επίσης το Google Speech-To-Text μια κατάλληλη λύση για εφαρμογές εκτός από τις σύντομες αναζητήσεις ιστού. Μπορεί επίσης να ρυθμιστεί για ήχο από τηλεφωνικές κλήσεις ή βίντεο. Υπάρχει επίσης μια τέταρτη ρύθμιση, την οποία συνιστά η Google να χρησιμοποιείται ως προεπιλογή.

Το API ομιλίας σε κείμενο διαθέτει επίσης μια εντυπωσιακή ενημέρωση για επιλογές εκτεταμένων σημείων στίξης. Αυτό έχει σχεδιαστεί για να κάνει πιο χρήσιμες μεταγραφές, με λιγότερες τρέχουσες προτάσεις ή σφάλματα στίξης.

Η πιο πρόσφατη ενημέρωση επιτρέπει επίσης στους προγραμματιστές να προσθέσουν ετικέτες στον ήχο ή το βίντεο που έχουν μεταγραφεί με βασικά μεταδεδομένα. Αυτό είναι περισσότερο προς όφελος της εταιρείας παρά για τους προγραμματιστές, ωστόσο, καθώς θα επιτρέψει στην Google να αποφασίσει ποιες λειτουργίες είναι πιο χρήσιμες για τους προγραμματιστές.

Το Google Speech-To-Text API δεν είναι, ωστόσο, δωρεάν. Είναι δωρεάν για αναγνώριση ομιλίας για ήχο λιγότερο από 60 λεπτά. Για μεταγραφές ήχου περισσότερο από αυτό, κοστίζει 0,006 ανά 15 δευτερόλεπτα.

### **Πλεονεκτήματα**

- Αναγνωρίζει πάνω από 120 γλώσσες
- Πολλαπλά μοντέλα μηχανικής εκμάθησης για αυξημένη ακρίβεια
- Αυτόματη αναγνώριση γλώσσας
- Μεταγραφή κειμένου
- Σωστή αναγνώριση ουσιαστικών
- Ιδιωτικότητα δεδομένων
- Ακύρωση θορύβου για ήχο από τηλεφωνικές κλήσεις και βίντεο

### **Μειονεκτήματα**

- Κοστίζει χρήματα
- Περιορισμένο πρόγραμμα δημιουργίας λεξιλογίου

## **4.1.2 Microsoft Cognitive Services Speech-to-Text API**

### **4.1.2.1 Περιγραφή**

Microsoft Cognitive Services παρέχουν την μεταγραφή ομιλίας σε κείμενο από την υπηρεσία ομιλίας, γνωστή και ως αναγνώριση ομιλίας, επιτρέπει τη μεταγραφή

ροών ήχου σε κείμενο σε πραγματικό χρόνο. Είναι επίσης ένα μέρος των υπηρεσιών Microsoft Trust που προσφέρουν ασύγκριτες επιλογές ασφάλειας για προγραμματιστές που αναζητούν τα πιο ασφαλή δεδομένα για τις εφαρμογές τους. Το κύριο πράγμα που διαχωρίζει το Microsoft Cognitive Services 'Speech to Text API είναι η λειτουργία Αναγνώρισης Ομιλιτή. Μπορεί να πραγματοποιήσει μεταγραφή σε πραγματικό χρόνο.

#### **4.1.2.2 Δυνατότητες**

##### **Πλεονεκτήματα**

- Βελτιωμένη ασφάλεια δεδομένων μέσω αλγορίθμων αναγνώρισης φωνής
- Μεταγραφή σε πραγματικό χρόνο
- Μετάφραση σε πραγματικό χρόνο
- Προσαρμόσιμο λεξιλόγιο
- Δυνατότητες κειμένου σε ομιλία για φυσικά πρότυπα ομιλίας

##### **Μειονεκτήματα**

- Ενσωματωμένοι περιορισμοί λόγω της δημιουργίας του API για γενικούς σκοπούς
- Χρησιμοποιεί μικροσυσκευές, οι οποίες μπορεί να είναι χρήσιμες για την επίλυση μεμονωμένων προβλημάτων, αλλά δεν ανταποκρίνονται σε μεγαλύτερα προβλήματα

### **4.1.3 IBM Watson API**

#### **4.1.3.1 Πειραγραφή**

Το Watson Speech to Text είναι μια λύση εγγενής στο cloud που χρησιμοποιεί αλγόριθμους AI βαθιάς μάθησης για την εφαρμογή γνώσεων σχετικά με τη γραμματική, τη δομή της γλώσσας και τη σύνθεση ήχου / φωνητικού σήματος για τη δημιουργία προσαρμόσιμης αναγνώρισης ομιλίας για βέλτιστη μεταγραφή κειμένου.

#### 4.1.3.2 Δυνατότητες

##### Πλεονεκτήματα

- Επεξεργάζεται μη δομημένα δεδομένα
- Βοηθά τους ανθρώπους αντί να τους αντικαθιστούν
- Βοηθά να ξεπεραστούν οι ανθρώπινοι περιορισμοί
- Βελτιώνει την παραγωγικότητα παρέχοντας σχετικά δεδομένα
- Βελτιώνει την εμπειρία χρήστη
- Μπορεί να επεξεργαστεί μεγάλες ποσότητες δεδομένων
- Εύκολη εγκατάσταση και έναρξη

##### Μειονεκτήματα

- Δεν υποστηρίζει άμεσα δομημένα δεδομένα
- Ακριβές για μετάβαση σε
- Απαιτείται συντήρηση
- Υποστηρίζει μόνο περιορισμένο αριθμό γλωσσών
- Χρειάζεται χρόνος για πλήρη εφαρμογή
- Απαιτεί εκπαίδευση και κατάρτιση για να αξιοποιήσει πλήρως τους πόρους της

## 4.2 Επιλογή API

Στην προηγούμενη ενότητα περιγράψαμε τρία APIs. Το κάθε ένα είναι ικανό για μετραγραφή ομιλίας σε κείμενο. Όλα τα APIs εφαρμόζουν Machine learning και AI τεχνολογίες. Το καθένα API έχει τα διαχωριστικά σημεία.

Το κύριο πράγμα που διαχωρίζει το Microsoft Cognitive Services 'Speech to Text API είναι η λειτουργία Αναγνώρισης Ομιλητή. Αυτή είναι η ακουστική έκδοση του λογισμικού ασφαλείας, όπως η αναγνώριση προσώπου. Σκεφτείτε το ως σάρωση αμβληστροειδούς για τον ήχο της φωνής του χρήστη. Το καθιστά απίστευτα εύκολο για

διαφορετικά επίπεδα χρηστών. Αυτή η ίδια δυνατότητα αναγνώρισης φωνής επιτρέπει στο λογισμικό να προσαρμόζεται στα συγκεκριμένα στυλ και μοτίβα ομιλίας του χρήστη. Προσφέρει επίσης περισσότερες προσαρμοσμένες επιλογές λεξιλογίου από το Google, ως επιπλέον πλεονέκτημα.

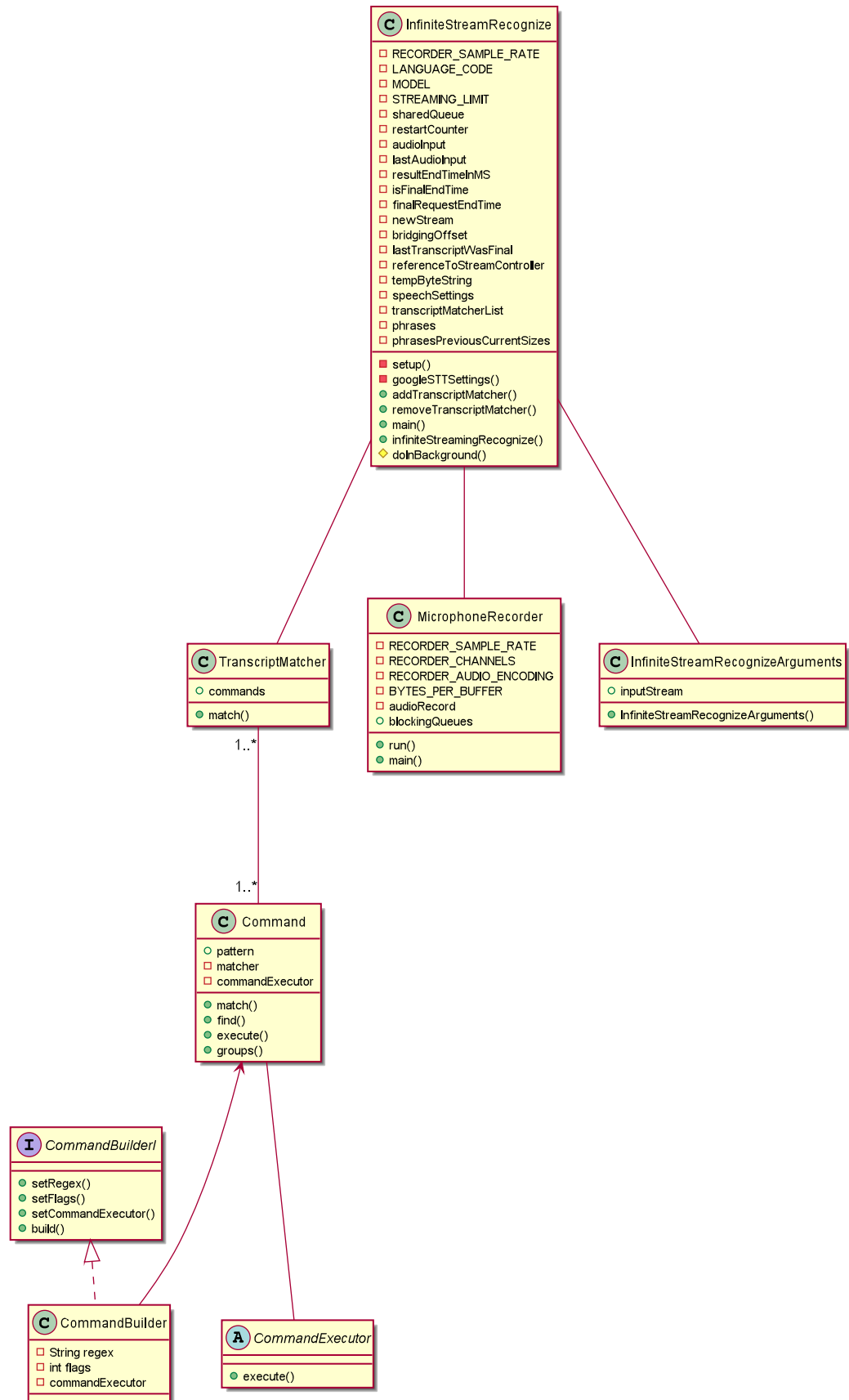
Το IBM Watson Speech to Text API είναι ιδιαίτερα ανθεκτικό στην κατανόηση τα συμφραζόμενα, στηρίζεται στη δημιουργία υποθέσεων και στην αξιολόγηση στη διαμόρφωση της απόκρισης. Είναι επίσης σε θέση να κάνει διάκριση μεταξύ πολλαπλών ομιλητών, γεγονός που το καθιστά κατάλληλο για τις περισσότερες εργασίες μεταγραφής. Μπορείτε ακόμη και να ορίσετε έναν αριθμό φίλτρων, εξαλείφοντας βωμολοχίες, προσθέτοντας εμπιστοσύνη λέξεων και επιλογές μορφοποίησης για εφαρμογές ομιλίας σε κείμενο.

Ένας από τους λόγους για την εντυπωσιακή ακρίβεια των API είναι η δυνατότητα επιλογής μεταξύ διαφορετικών μοντέλων μηχανικής μάθησης, ανάλογα με το τι χρησιμοποιείται η εφαρμογή σας. Αυτό καθιστά επίσης το Google Speech-To-Text μια κατάλληλη λύση για εφαρμογές εκτός από τις σύντομες αναζητήσεις ιστού. Μπορεί επίσης να ρυθμιστεί για ήχο από τηλεφωνικές κλήσεις ή βίντεο. Υπάρχει επίσης μια τέταρτη ρύθμιση, την οποία συνιστά η Google να χρησιμοποιείται ως προεπιλογή. Το API Speech-To-Text διαθέτει επίσης μια εντυπωσιακή ενημέρωση για εκτεταμένες επιλογές στίξης. Αυτό έχει σχεδιαστεί για να κάνει πιο χρήσιμες μεταγραφές, με λιγότερες τρέχουσες προτάσεις ή σφάλματα στίξης. Επίσης το Google cloud Speech-to-Text API αναγνωρίζει πάνω από 120 γλώσσες συμπεριλαμβάνοντας και την Ελληνική γλώσσα που χρησιμοποιείτε για την ανάπτυξη της βιβλιοθήκης.

## **Κεφάλαιο 5**

### **Σχεδίαση Βιβλιοθήκης**





# Γλωσσάρι

Το γλωσσάρι μπορεί να είναι χρήσιμο αν χρησιμοποιείτε πολλά ακρόνυμα και συντομογραφίες. Για παράδειγμα

<https://nordicapis.com/5-best-speech-to-text-apis/>

<https://cloud.google.com/speech-to-text#all-features>

<https://www.ibm.com/cloud/watson-speech-to-text>

<https://azure.microsoft.com/en-us/services/cognitive-services/speech-to-text/>

**TCP** Transmission Control Protocol

