



Υπολογιστική Γλωσσολογία

Project report

Κωνσταντίνος Μαλωνάς - 7115112200020
Βασίλειος Κυριακόπουλος - 7115112200017

| | |
|---|-----------|
| Εισαγωγή..... | 3 |
| ΘΕΩΡΗΤΙΚΟ ΠΛΑΙΣΙΟ..... | 4 |
| 2.1 Αντικείμενο της Υπολογιστικής Γλωσσολογίας..... | 4 |
| 2.2 Πεδίο εφαρμογής των Διαλογικών Συστημάτων..... | 6 |
| 2.2.1 Σύνδεση με τη Σημασιολογία / Πραγματολογία..... | 6 |
| 2.3 Σύντομη Περιγραφή του Συστήματος..... | 7 |
| 3. ΠΕΔΙΟ..... | 7 |
| 3.1. ΧΡΗΣΤΕΣ (Χαρακτηριστικά των Χρηστών)..... | 7 |
| 4. ΜΕΘΟΔΟΣ..... | 9 |
| 5. ΑΝΑΛΥΣΗ/ΠΕΡΙΓΡΑΦΗ..... | 9 |
| 5.1 Περιγραφή εντολών..... | 9 |
| 5.2 Περιγραφή προγράμματος..... | 10 |
| 5.2.1 Κλάση GameController..... | 10 |
| 5.2.3 Κλάση Browser..... | 12 |
| 5.2.4 Script grid.py..... | 12 |
| 5.2.5 Script SpeechToText.py..... | 13 |
| 5.3 Οδηγίες εκτέλεσης προγράμματος..... | 14 |
| 5.4 Libraries..... | 16 |
| 6. ΣΥΜΠΕΡΑΣΜΑΤΑ..... | 16 |
| 7. ΕΠΙΛΟΓΟΣ..... | 17 |

Εισαγωγή

Στην εποχή της ραγδαίας τεχνολογικής ανάπτυξης που διανύουμε, η εξέλιξη των τεχνολογικών μέσων συνεχίζει να διαμορφώνει νέους και καινοτόμους τρόπους αλληλεπίδρασης με τους υπολογιστές και το διαδίκτυο. Η αναγνώριση και ο έλεγχος μέσω φωνής είναι ένα από τα βασικά επιστημονικά πεδία πάνω στο οποίο έχουν γίνει πολλές μελέτες και έχουν αναπτυχθεί πληθώρα προϊόντων λογισμικού, προσφέροντας στους χρήστες την δυνατότητα, χωρίς την χρήση χειριών και με έναν εντελώς ενστικτώδη τρόπο πλοήγησης να έχουν πρόσβαση σε διαδικτυακούς πόρους. Η συγκεκριμένη εργασία έχει ως στόχο να αναδείξει την δυναμική του ελέγχου με φωνή περαιτέρω, αναπτύσσοντας μια Εφαρμογή Πλοήγησης και Ελέγχου Ποντικιού με την Φωνή. Η συγκεκριμένη εφαρμογή που αναπτύξαμε αξιοποιεί τη δυνατότητα της αναγνώρισης φωνής, της αυτοματοποίησης της πλοήγησης μεταξύ διαφόρων ιστοσελίδων και των γραφικών διεπαφών χρήστη, επιτρέποντας στους χρήστες να περιηγούνται στον κόσμο του διαδικτύου, να ελέγχουν τον κέρσορα του ποντικιού, να πραγματοποιούν το click και το press button functionality, και να αλληλεπιδρούν με μηχανές αναζήτησης μόνο με τη χρήση φωνητικών εντολών.

Οι διεπαφές ελέγχου με φωνή έχουν αποκτήσει ταχύτατα δημοτικότητα λόγω της ευκολίας και της προσβασιμότητας τους. Το λογισμικό που αναπτύξαμε, συνδυάζει διάφορες τεχνολογίες, όπως βιβλιοθήκες αναγνώρισης φωνής και εργαλεία αυτοματοποίησης ιστοσελίδων, μέσω των οποίων ο χρήστης μπορεί μέσω φωνητικών εντολών να πραγματοποιήσει αναζήτηση μέσα σε κάποια συγκεκριμένη ιστοσελίδα, για να δημιουργήσει μια απρόσκοπτη εμπειρία χρήσης. Ο κύριος στόχος αυτής της εφαρμογής είναι να παρέχει στους χρήστες τη δυνατότητα να εκτελούν καθημερινές εργασίες, όπως αναζητήσεις στο διαδίκτυο, μετακινήσεις του ποντικιού και πληκτρολόγηση κειμένου με φυσικές φωνητικές εντολές.

Η βασική λειτουργία της εφαρμογής περιλαμβάνει την πλοήγηση σε ιστοσελίδες μέσω φωνητικών εντολών, την εκτέλεση αναζητήσεων στη Google, το κλικ σε στοιχεία της ιστοσελίδας και την πληκτρολόγηση κειμένου σε πεδία αναζήτησης, καθώς και την μετακίνηση πίσω στην σελίδα που βρισκόμασταν μετά την επίσκεψη κάποιας σελίδας. Επιπλέον, η εφαρμογή προσθέτει την έννοια ενός διαφανούς πλέγματος που βοηθά τους χρήστες να περιηγούνται με μεγαλύτερη ευκολία στο περιβάλλον τους και προσφέρει επαρκέστερη χωρική αντίληψη κατά την χρήση της φωνητικής διεπαφής. Η λειτουργικότητα αυτή είναι αποτέλεσμα μερικής έμπνευσης από τα περιβάλλοντα όπου χρησιμοποιείται το λειτουργικό σύστημα iOS και όπου εισήχθη αρχικά, για πιο εύκολη περιήγηση των χρηστών.

Στο παρόν project report, θα παρουσιάσουμε τις τεχνικές πτυχές του έργου, αναλύοντας τον σχεδιασμό, την υλοποίηση και την ολοκλήρωση των διαφόρων συστατικών. Θα αναφερθούμε επίσης στις προκλήσεις που αντιμετωπίστηκαν κατά τη διάρκεια της ανάπτυξης και τις λύσεις που εφαρμόστηκαν για την αντιμετώπισή τους. Επιπλέον θα συζητηθούν τα πεδία εφαρμογής του λογισμικού που αναπτύξαμε, συμπεριλαμβανομένης της χρησιμότητάς της για άτομα με κινητικές δυσκολίες ή αναπηρίες που δυσχεραίνουν την παραδοσιακή χρήση ποντικιού και πληκτρολογίου.

Το συγκεκριμένο project θεωρούμε ότι αποτελεί παράδειγμα της δύναμης της διεπιστημονικής συνεργασίας, συνδυάζοντας έννοιες από την αναγνώριση φωνής, την αυτοματοποίηση ιστοσελίδων, τις γραφικές διεπαφές χρήστη και τον σχεδιασμό με εστίαση στον χρήστη. Με τη συνδυασμένη χρήση αυτών των στοιχείων, στοχεύουμε να παρέχουμε στους χρήστες έναν εναλλακτικό και προσβάσιμο τρόπο αλληλεπίδρασης με τους υπολογιστές και το διαδίκτυο. Αυτό το report αποτελεί μια συνοπτική επισκόπηση της Εφαρμογής Πλοήγησης και Ελέγχου Ποντικιού με την Φωνή, αναδεικνύοντας τα χαρακτηριστικά, τη λειτουργικότητα και την πιθανή επίδρασή της στο μέλλον της αλληλεπίδρασης ανθρώπου-υπολογιστή.

ΘΕΩΡΗΤΙΚΟ ΠΛΑΙΣΙΟ

Το συγκεκριμένο project εμπίπτει περισσότερο στο πεδίο των Συστημάτων Διαλόγου. Φυσικά οι εντολές, καθώς και οι απαντήσεις που παίρνει πολλές φορές ο χρήστης για να πει για παράδειγμα κάποιον όρο που επιθυμεί να αναζητήσει στο διαδίκτυο είναι αυστηρά καθορισμένες.

2.1 Αντικείμενο της Υπολογιστικής Γλωσσολογίας

Κατεχοχήν επικοινωνιακή εφαρμογή των Συστημάτων Επεξεργασίας Φυσικής Γλώσσας αποτελούν τα Συστήματα Αλληλεπίδρασης Ανθρώπου-Μηχανής και κατ' επέκταση, τα Διαλογικά Συστήματα. Τα Διαλογικά Συστήματα με επεξεργασία φυσικής γλώσσας επιτρέπουν σε έναν ανθρώπινο Χρήστη να επικοινωνήσει προφορικά με ένα Σύστημα-Μηχανή και μπορούν να χωριστούν σε δύο γενικές κατηγορίες:

(1) τα Διαλογικά Συστήματα για εκτέλεση συγκεκριμένων πράξεων (Task-oriented dialog agents), στα οποία συμπεριλαμβάνονται εφαρμογές Ρομποτικής αλλά και οι ψηφιακοί βοηθοί (digital assistants), όπως οι Siri, Alexa, Google Now, και

(2) τα πιο ομιλητικά Chatbots που στοχεύουν στην προσομοίωση της ανθρώπινης επικοινωνίας

Το δικό μας σύστημα, όπως αναφέρθηκε, ανήκει στην πρώτη κατηγορία, λόγω των αυστηρά καθορισμένων εντολών-απαντήσεων που μπορεί να πάρει ο χρήστης.

Συνήθως, με τον όρο Διαλογικά Συστήματα χαρακτηρίζονται εκείνα τα Συστήματα Επεξεργασίας Φυσικής Γλώσσας (ΣΕΦΓ) τα οποία αφορούν τον προφορικό λόγο, τόσο ως δεδομένα εισόδου στο Σύστημα όσο και ως δεδομένα εξόδου που παράγει το Σύστημα. Τα Διαλογικά Συστήματα διαθέτουν ένα πολύ μικρό ή πολύ περιορισμένο πεδίο γνώσης (για τον κόσμο) σε σχέση με τον άνθρωπο – κάτι που ισχύει γενικότερα για τα ΣΕΦΓ – και δεν μπορούν να επικοινωνήσουν οπουδήποτε και σε κάθε περίπτωση όπως ένας άνθρωπος. Κατά συνέπεια, η χρήση τους, τουλάχιστον μέχρι σήμερα, συνήθως περιορίζεται σε εξειδικευμένα πεδία εφαρμογών. Κάποια βασικά στοιχεία των διαλογικών συστημάτων είναι η κατανόηση φυσικής γλώσσας, η αυτόματη κατανόηση φυσικής γλώσσας με σκοπό για παράδειγμα την μετάφραση της καθώς και η μετατροπή κειμένου σε λόγο.

Οι εισερχόμενες προφορικές ερωτήσεις του Χρήστη επεξεργάζονται με την χρήση Συστήματος Αναγνώρισης Φωνής (Speech Recognition-ASR). Επιπλέον, ανάλογα με τις ανάγκες της εφαρμογής, για την παραγωγή των απαντήσεων του Συστήματος στον Χρήστη, μπορεί να χρησιμοποιηθεί Σύστημα Σύνθεσης Φωνής (Speech Synthesis-TTS) ή ηχογραφημένες προτάσεις από εκπαιδευμένο φυσικό ομιλητή (voice talent).

Επιπλέον, όπως ισχύει για τα Συστήματα Επικοινωνίας Ανθρώπου-Μηχανής, τα Διαλογικά Συστήματα μπορούν να είναι πολυμεσικά (multimodal) και να συνδυάζουν πολλαπλές δυνατότητες εισαγωγής ή παραγωγής δεδομένων, όπως ο συνδυασμός οπτικής εικόνας, χειρονομιών και χρήσης πληκτρολογίου (π.χ στην εφαρμογή που παρουσιάζουμε ο χρήστης μπορεί να πληκτρολογήσει με φωνητικές εντολές).

Για παράδειγμα, στην προφορική έκφραση “Πόσο κάνει αυτό”, ο συνδυασμός δεικτικών χειρονομιών και δεικτικών εκφράσεων αναγνωρίζεται και επεξεργάζεται από το πολυμεσικό Διαλογικό Σύστημα. Ο συνδυασμός πολλαπλών δυνατοτήτων εισαγωγής και παραγωγής δεδομένων σε ένα πολυμεσικό Διαλογικό Σύστημα έχει παρατηρηθεί ότι οδηγεί σε μια σαφή αύξηση της χρησιμότητας και αποτελεσματικότητας του ως προς τον Χρήστη, ιδίως όταν χρησιμοποιείται σε καταστάσεις που χρειάζονται άμεση αντίδραση από τον Χρήστη, όπως Συστήματα που αφορούν θέματα ασφαλείας, αλλά και εφαρμογές με πολλαπλούς Ομιλητές Χρήστες που επικοινωνούν ταυτόχρονα με το Σύστημα και μεταξύ τους.

Παρά ταύτα, πρωταρχικό ρόλο σε όλα τα Διαλογικά Συστήματα παίζει η φωνή η οποία συμβάλλει καθοριστικά στην επικοινωνία Ανθρώπου-Μηχανής (Nass and Brave 2005), ειδικά αν ληφθεί υπόψη ότι η επικοινωνία-“συζήτηση” με ένα Σύστημα διαφέρει σε πάρα πολλά σημεία από την επικοινωνία με έναν Άνθρωπο (Lewis 2009), και επιπροσθέτως, σε αντίθεση με τον Άνθρωπο (Chomsky 1965, 1968, 2001), ο Υπολογιστής δεν διαθέτει εγγενείς μηχανισμούς επεξεργασίας της γλώσσας. Στα περισσότερα Διαλογικά Συστήματα, οι προτάσεις που εκφέρει το Σύστημα είναι σύντομες και σαφείς, προκειμένου να αυξηθεί η προβλεψιμότητα και, κατά συνέπεια, η αναγνωρισιμότητα του περιεχομένου των προτάσεων του Χρήστη. Ταυτόχρονα, επιδιώκεται από τη μία να μην είναι ο διάλογος μακράς διαρκείας και κουραστικός και από την άλλη να χαρακτηρίζεται από φιλικότητα και φυσικότητα.

2.2 Πεδίο εφαρμογής των Διαλογικών Συστημάτων

Ξεκινώντας από ειδικές εφαρμογές, όπως η Ρομποτική, τα Διαλογικά Συστήματα χρησιμοποιούνται ολοένα και περισσότερο σε εξειδικευμένα είδη εφαρμογών και υπηρεσιών που αφορούν ένα ευρύτερο σύνολο Χρηστών, όπως, για παράδειγμα, οι κρατήσεις θέσεων ή/και εισιτηρίων, η παροχή πληροφοριών (για τον καιρό, για τουρίστες ή επισκέπτες, για τα μέσα μεταφοράς κ.α.) ή η παροχή υπηρεσιών στον εμπορικό και επιχειρησιακό τομέα. Σε αυτές τις εφαρμογές για το ευρύ κοινό προστίθενται και οι εφαρμογές στις εργασίες που εκτελούνται στον Υπολογιστή χωρίς την χρήση πληκτρολογίου (για παράδειγμα, σε οχήματα κατά την διάρκεια οδήγησης) ή/ και από απόσταση. Στις δυνατότητες αυτές συμπεριλαμβάνονται και οι ειδικές εφαρμογές για άτομα με αναπηρία ή προβλήματα όρασης και κινητικά προβλήματα. Επιπλέον, αρκετά στοιχεία και πρακτικές των Διαλογικών

Συστημάτων χρησιμοποιούνται και για την κατασκευή διαδραστικού εκπαιδευτικού λογισμικού, ιδίως όταν αφορά εικονικούς ομιλητές και Chatbots.

2.2.1 Σύνδεση με τη Σημασιολογία / Πραγματολογία

Σε όλα τα Διαλογικά Συστήματα πρέπει το Σύστημα να αναγνωρίζει τον προφορικό λόγο και, ταυτόχρονα, να εκφέρει την κατάλληλη αντίδραση ή απάντηση στον Χρήστη. Όσον αφορά τη διαδικασία επιλογής και παραγωγής των κατάλληλων αντιδράσεων από το Σύστημα, η διαδικασία αυτή συνήθως πραγματοποιείται με την εφαρμογή ειδικών κανόνων με την μορφή προγράμματος, τον επονομαζόμενο “Διαχειριστή Διαλόγου” (Dialog Manager, σ.σ. εμείς το έχουμε ονομάσει MouseController). Οι κανόνες αυτοί συνήθως περιέχουν γλωσσολογικές πληροφορίες, κυρίως από την πραγματολογία.

Η διαδικασία της “διαχείρισης διαλόγου” μπορεί να θεωρηθεί ότι παίζει τον ρόλο του “διαιτητή” ή του “τροχονόμου” στη διάδραση-διάλογο που επεξεργάζεται το Σύστημα. Συγκεκριμένα, ελέγχονται οι πράξεις που εκτελούνται στον διάλογο. Οι πράξεις αυτές ονομάζονται διαλογικές πράξεις (dialog acts) και μπορούν να ταυτιστούν με γλωσσικές πράξεις (speech acts). Στα περισσότερα Διαλογικά Συστήματα, για την ερμηνεία του σημασιολογικού περιεχομένου (δηλαδή, των συμφραζομένων και της αναφορικότητας) από τον διαχειριστή διαλόγου (“dialog-manager”) αποθηκεύονται όλες οι προτάσεις ή οι εκφράσεις που έχουν εκφέρει τόσο ο Χρήστης όσο και το Σύστημα για τον συγκεκριμένο διάλογο του Χρήστη με το Σύστημα. Πρόκειται για πρακτική εφαρμογή γλωσσολογικών κανόνων της πραγματολογίας, οι οποίοι πλέον ενισχύονται με νευρωνικά δίκτυα, για την επίτευξη μεγαλύτερης ταχύτητας και ακρίβειας. Κατ’ αυτόν τον τρόπο δεν επιτυγχάνεται μόνο η επίλυση προβλημάτων που συνδέονται με την γλωσσική δομή, όπως η έλλειψη (για παράδειγμα, στην έκφραση “Και για την Δευτέρα;”) ή η αναφορικότητα (για παράδειγμα, στην έκφραση “Επανάλαβε το τελευταίο δρομολόγιο”), αλλά επιτυγχάνεται επίσης και η αναγνώριση αμφισημιών ή παράλογων εκφράσεων που μπορούν να εισάγονται στο Σύστημα ενεργοποιώντας τις κατάλληλες αντιδράσεις εκ μέρους του. Σε αυτού του είδους τις προγραμματισμένες αντιδράσεις ανήκουν και οι επονομαζόμενες “ερωτήσεις επιβεβαίωσης” (verification questions), που μπορούν να διατυπώνονται με άμεσο ή με έμμεσο τρόπο.

Οι σημαντικότερες πληροφορίες που περιλαμβάνονται σε κάθε πρόταση, καθώς και οι σχέσεις μεταξύ των προτάσεων ή/και των διαφορετικών ενοτήτων του διαλόγου μεταξύ τους (συμπεριλαμβανομένης της αναφορικότητας και των προϋποθέσεων για την ορθή ερμηνεία μιας γλωσσικής έκφρασης), καθορίζονται με την βοήθεια της πραγματολογίας

2.3 Σύντομη Περιγραφή του Συστήματος

Η εφαρμογή που κατασκευάσαμε δίνει την δυνατότητα στον χρήστη να ελέγχει το mouse cursor, καθώς επίσης και να πληκτρολογεί, μέσω φωνητικών εντολών. Επιπλέον ο χρήστης μπορεί να κάνει “click” ή “press button” απευθείας λέγοντας “click” ή “press” αντίστοιχα για buttons ακολουθούμενο από το κείμενο του element το οποίο θέλει να κλικάρει ή να πατήσει ο χρήστης. Ο χρήστης λέει κάθε φορά την κατεύθυνση που επιθυμεί να κινηθεί ο cursor και τον αριθμό των pixel που θέλει να τον μετακινήσει π.χ δεξιά 100 (κίνησε το cursor δεξιά 100 pixel). Επίσης μπορεί να πει την εντολή “κλικ” για να πραγματοποιήσει ένα mouse click. Με

την εντολή “type” έχει την δυνατότητα να κάνει χρήση του πληκτρολογίου μέσω φωνής και να πληκτρολογήσει σε κάποιο search bar που πιθανώς να υπάρχει στην σελίδα που έχει επισκεφτεί. Επίσης με την εντολή “νέα αναζήτηση” μπορεί να πραγματοποιήσει μια καινούργια αναζήτηση. Επιπροσθέτως. Με την εντολή “πίσω” μεταφερόμαστε στην προηγούμενη σελίδα από την οποία βρισκόμαστε. Επίσης με την εντολή “open” ή “open grid” εμφανίζεται ένα grid για να βοηθάει τον χρήστη να κινεί τον cursor με μεγαλύτερη ακρίβεια στην οθόνη και με την εντολή “close” το grid εξαφανίζεται. Ο χρήστης θα πρέπει πάντα να κλείνει πρώτα το grid πριν πραγματοποιήσει την εντολή “click”, “type” ή “νέα αναζήτηση”. Τέλος όταν ο χρήστης θέλει να εξέλθει της εφαρμογής λέει απλά “exit”. Η εφαρμογή είναι υλοποιημένη με την προγραμματιστική γλώσσα Python. Για την αναγνώριση της φωνής χρησιμοποιείται η βιβλιοθήκη speech_recognition και η μέθοδος recognize_google.

3. ΠΕΔΙΟ

3.1. ΧΡΗΣΤΕΣ (Χαρακτηριστικά των Χρηστών)

Η εφαρμογή φωνητικού ελέγχου του ποντικιού σχεδιάστηκε για να εξυπηρετήσει μια ποικιλία χρηστών που μπορούν να επωφεληθούν από μια πιο διαδραστική αλληλεπίδραση με τους υπολογιστές τους. Με την χρήση του google api (recognize_google) έχουμε καταφέρει να δημιουργήσουμε μια τέτοια εφαρμογή που μπορεί να βοηθήσει διάφορες ομάδες ατόμων ευπαθείς και μη να χρησιμοποιήσουν τον υπολογιστή τους και να περιηγηθούν στο διαδίκτυο με έναν διαφορετικό τρόπο που στο μέλλον μπορεί να τους φανεί πιο βολικός από τον συμβατικό τρόπο περιήγησης στο διαδίκτυο.

1. Άτομα με Περιορισμένη Κινητικότητα:

Ένας από τους βασικούς ωφελούμενους αυτής της εφαρμογής είναι άτομα με περιορισμένη κινητικότητα, όπως άτομα με αναπηρίες ή κινητικές δυσκολίες. Για αυτούς τους χρήστες, η παραδοσιακή αλληλεπίδραση με το ποντίκι μπορεί να είναι δύσκολη ή ακόμη και αδύνατη. Χρησιμοποιώντας φωνητικές εντολές για τον έλεγχο της κίνησης του ποντικιού και των ενεργειών του, τα συγκεκριμένα άτομα αποκτούν μια νέα ανεξαρτησία όσον αφορά την πλοήγηση στο διαδίκτυο με τους υπολογιστές τους. Αυτή η τεχνολογία τους επιτρέπει να αποκτήσουν πρόσβαση σε πληροφορίες, να περιηγούνται στο διαδίκτυο και να εκτελούν εργασίες που ήταν προηγουμένως δύσκολες ή μη προσβάσιμες.

2. Χρήστες που Πραγματοποιούν Πολλαπλές Εργασίες και Επαγγελματίες:

Η σύγχρονη ζωή απαιτεί τη δυνατότητα πολλαπλών εργασιών, και οι διάφοροι επαγγελματίες που εργάζονται σε δουλειές γραφείου έχουν να ασχοληθούν με ένα μεγάλο εύρος διαφορετικών καθηκόντων. Αυτή η εφαρμογή γίνεται ένα πολύτιμο εργαλείο για πολυάσχολα άτομα που χρειάζονται την γρήγορη εναλλαγή ανάμεσα σε εφαρμογές, να πραγματοποιήσουν αναζητήσεις και να έχουν πρόσβαση σε πληροφορίες χωρίς να απομακρύνουν τα χέρια από το πληκτρολόγιο. Μέσω μιας ομαλής μετάβασης ανάμεσα στη φωνητική είσοδο και τον έλεγχο του ποντικιού, η εφαρμογή βελτιστοποιεί την παραγωγικότητα και την αποδοτικότητα.

3. Χρήστες με προβλήματα όρασης:

Άνθρωποι με προβλήματα όρασης μπορούν επίσης να επωφεληθούν σημαντικά από την συγκεκριμένη εφαρμογή. Ενώ τα προγράμματα ανάγνωσης οθόνης προσφέρουν δυνατότητες μετατροπής κειμένου σε ομιλία, η πλοήγηση σε γραφικά περιβάλλοντα εξακολουθεί να είναι προκλητική. Χρησιμοποιώντας φωνητικές εντολές, τους επιτρέπεται να περιηγηθούν σε ιστοσελίδες, να αλληλεπιδράσουν με οπτικά στοιχεία και να έχουν πρόσβαση σε περιεχόμενο χωρίς να εξαρτώνται αποκλειστικά από λογισμικό ανάγνωσης οθόνης.

4. Ηλικιωμένοι:

Οι ηλικιωμένοι συχνά αντιμετωπίζουν προκλήσεις κατά την προσαρμογή στην τεχνολογία. Αυτή η εφαρμογή προσφέρει μια φιλική προς τον χρήστη προσέγγιση στην περιήγηση στο διαδίκτυο και στις απλές ενέργειες που απαιτούνται σε αυτή, μειώνοντας την πολυπλοκότητα της παραδοσιακής αλληλεπίδρασης με το ποντίκι και το πληκτρολόγιο. Οι φωνητικές εντολές παρέχουν έναν πιο εύκολο και κατανοητό τρόπο για την αλληλεπίδραση με τις συσκευές τους, καθιστώντας τις εργασίες όπως η περιήγηση στον ιστό και η αποστολή email μέσω ηλεκτρονικού ταχυδρομείου πιο επιτεύξιμες και ευχάριστες.

5. Χρήστες παιχνιδιών και Δημιουργοί Περιεχομένου:

Ακόμα και οι κοινότητες των δημιουργών περιεχομένου μπορούν να βρουν χρηστική αξία σε αυτή την εφαρμογή καθώς με την χρήση της μπορούν να διευκολύνει την πλοήγηση και την αλληλεπίδραση με εργαλεία δημιουργίας και πλατφόρμες ανάπτυξης.

Με βάση τις ποικίλες ανάγκες αυτών των χρηστικών προφίλ, η εφαρμογή φωνητικού ελέγχου του ποντικιού αποτελεί ένα ευέλικτο εργαλείο που αυξάνει την προσβασιμότητα, βελτιώνει την παραγωγικότητα και επιτρέπει σε διάφορες ομάδες χρηστών να αξιοποιήσουν πλήρως τη δυναμική της τεχνολογίας.

4. ΜΕΘΟΔΟΣ

Σε αυτή την εργασία για την κατασκευή του λογισμικού μας χρησιμοποιήσαμε έναν συνδυασμό από βιβλιοθήκες της Python. Συγκεκριμένα για την αναγνώριση της χρησιμοποιήσαμε την βιβλιοθήκη **speech_recognition**. Η συγκεκριμένη βιβλιοθήκη μας επέτρεψε να μετατρέψουμε κείμενο σε φωνή ώστε να δοθούν φωνητικές εντολές στον χρήστη καθώς και να γίνει η λήψη φωνής σε μορφή εντολών από τον χρήστη ώστε στην συνέχεια με την χρήση της μεθόδου **recognize_google** της συγκεκριμένης βιβλιοθήκης να μετατρέψουμε τα λεγόμενα του χρήστη σε κείμενο. Για τον έλεγχο του ποντικιού και για την επίτευξη λειτουργιών όπως click, backspace, enter, type, χρησιμοποιήθηκε η βιβλιοθήκη **pyautogui**. Πολύ σημαντικό ρόλο στην εργασία μας έπαιξε η βιβλιοθήκη **selenium** με την οποία μας επιτράπηκε να γίνει αυτόματη αποδοχή cookies αλλά και επιστροφή elements τα οποία ο χρήστης θέλει να κλικάρει (σύνδεσμοι, button elements). Τέλος για τον σχεδιασμό του grid χρησιμοποιήθηκε ένας συνδυασμός βιβλιοθηκών με τον κυρίαρχο ρόλο να τον έχει η βιβλιοθήκη **pygame**.

5. ΑΝΑΛΥΣΗ/ΠΕΡΙΓΡΑΦΗ

5.1 Περιγραφή εντολών

1) **πάνω - κάτω - αριστερά - δεξιά αριθμός π.χ (δεξιά 100)**

Ο mouse cursor θα μετακινηθεί 100 pixel δεξιά

2) **νέα αναζήτηση**

Θα πούμε το καινούργιο search term αφού πρώτα το πρόγραμμα μας πεί "Please say your search term"

3) **click - κλίκ**

Για το mouse click functionality

4) **type - delete**

Όταν το mouse cursor είναι π.χ πάνω από κάποιο search bar και πούμε click και εμφανιστεί το cursor που υποδεικνύει ότι αναμένεται είσοδος από το keyboard, ο χρήστης μπορεί να πεί type και αφού ακουστεί το "Please say what you want to type" ο χρήστης λέει αυτό που θέλει να πληκτρολογήσει και πατιέται αυτοματα **ENTER** αλλιώς μπορεί να πει delete ή διαγραφή για να διαγραφεί οτιδήποτε υπάρχει στο search bar.

5) **Πίσω:**

Για να πάμε στην προηγούμενη σελίδα από την οποία ήμασταν.

6) **click - κείμενο από το link που θέλει να πατήσει ο χρήστης**

Για παράδειγμα ο χρήστης έχει επισκεφτεί μια σελίδα της wikipedia όπου μέσα υπάρχουν διάφορα links. Ο χρήστης λέει click και το κείμενο του link που θέλει να πατήσει και πραγματοποιείται click στο συγκεκριμένο link element.

7) **press - κείμενο από το button που θέλει να πατήσει ο χρήστης**

Αντίστοιχα με την εντολή click (6) ο χρήστης λέει press και το κείμενο που περιέχεται στο button element που επιθυμεί να πατήσει ο χρήστης και πραγματοποιείται το press button functionality.

8) **open grid ή open**

Με την εντολή open grid ή open εμφανίζεται ένα grid στην οθόνη του χρήστη το οποίο του επιτρέπει να μετακινήσει το ποντίκι του με μεγαλύτερη ακρίβεια.

9) **exit**

Για να εξέλθει ο χρήστης από την εφαρμογή.

5.2 Περιγραφή προγράμματος

5.2.1 Κλάση MouseController

Η κλάση MouseController έχει σχεδιαστεί για να ελέγχει τον κέρσορα του ποντικιού και να εκτελεί διάφορες ενέργειες χρησιμοποιώντας φωνητικές εντολές. Επιτρέπει στους χρήστες να μετακινούν τον κέρσορα του ποντικιού, να πραγματοποιούν κλικ, να πληκτρολογούν κείμενο και να αλληλεπιδρούν με στοιχεία στον ιστό χρησιμοποιώντας φωνητικές εντολές.

Παρακάτω εξηγούνται οι μέθοδοι της και οι λειτουργίες τους:

__init__(self)

- Ο κατασκευαστής αρχικοποιεί μια νέα οντότητα της κλάσης MouseController.
- Δημιουργεί επίσης ένα αντικείμενο Browser από το module br (script Browser.py) για αλληλεπίδραση με ιστοσελίδες. Κυρίως για την αποδοχή cookies κατά το πρώτο google search του χρήστη αλλά και για την επιστροφή link και button elements.
- Οι μεταβλητές num_of_search και search_bar_text αρχικοποιούνται για να παρακολουθούν τον αριθμό των αναζητήσεων και το κείμενο στη γραμμή αναζήτησης.
- Η μεταβλητή text_query είναι η μεταβλητή στην οποία αποθηκεύεται το κείμενο το οποίο έγινε extract από την φωνή του χρήστη
- Με την χρήση της βιβλιοθήκης pyautogui μετακινούμε το mouse cursor, αρχικά στο κέντρο της οθόνης.

type_in_search_bar(self)

- Αυτή η μέθοδος χρησιμοποιείται για να πληκτρολογήσει κείμενο σε μια γραμμή αναζήτησης.
- Περιμένει να λάβει μια φωνητική εντολή από το χρήστη, συγκεκριμένα την εντολή "type."
- Όταν αναγνωριστεί η εντολή "type," παραπέμπει τον χρήστη να πει τι επιθυμεί να πληκτρολογήσει και περιμένει να ακούσει τον χρήστη, στην συνέχεια μετατρέπει την φωνή του χρήστη σε κείμενο και το πληκτρολογεί στο αντίστοιχο search bar.
- Εάν ο χρήστης αναφέρει έναν όρο αναζήτησης, τότε πληκτρολογεί αυτόν τον όρο στη γραμμή αναζήτησης και πατά "Enter" για να ξεκινήσει την αναζήτηση.
- Αν ο χρήστης πει "delete," καθαρίζει το κείμενο στη γραμμή αναζήτησης.

start(self)

- Η μέθοδος start αποτελεί τον κύριο βρόγχο της κλάσης MouseController.
- Συνεχώς περιμένει να λάβει φωνητικές εντολές από τον χρήστη και πράττει αναλόγως εκτελώντας τις αντίστοιχες ενέργειες.
- Μπορεί να εκτελέσει διάφορες ενέργειες με βάση τις αναγνωρισμένες φωνητικές εντολές, όπως έναρξη αναζήτησης στο Google, κλικ σε στοιχεία, πληκτρολόγηση στη γραμμή αναζήτησης, κλικ σε στοιχεία με την χρήση selenium χωρίς την χρήση ποντικιού, έναρξη νέας αναζήτησης, επιστροφή στην προηγούμενη σελίδα από αυτή στην οποία βρισκόμαστε, άνοιγμα και κλείσιμο του grid για ακριβέστερη μετακίνηση του ποντικιού, μετακίνηση του ποντικιού πάνω-κάτω-αριστερά-δεξιά.

parse_number(self, num_str: str)

- Αυτή η μέθοδος χρησιμοποιείται για την ανάλυση αριθμητικών τιμών από τις φωνητικές εντολές του χρήστη. Συγκεκριμένα για την μετατροπή των αριθμητικών τιμών που είναι σε μορφή κειμένου σε ακέραιους αριθμούς.
- Χειρίζεται περιπτώσεις όπου ο χρήστης μπορεί να πει, για παράδειγμα, "χίλια" και το μετατρέπει στον ακέραιο αριθμό 1000.
- Βεβαιώνεται ότι οι αριθμητικές τιμές αναγνωρίζονται σωστά και μετατρέπονται σε ακέραιους αριθμούς.

move_mouse(self, direction, num)

- Η μέθοδος move_mouse χρησιμοποιείται για να μετακινήσει τον κέρσορα του ποντικιού με βάση τις φωνητικές εντολές του χρήστη.
- Παίρνει ως παραμέτρους την κατεύθυνση και τον αριθμό των pixels για τη μετακίνηση.
- Ανάλογα με την καθορισμένη κατεύθυνση (πάνω, κάτω, αριστερά, δεξιά), υπολογίζει τη νέα θέση του κέρσορα και τον μετακινεί ανάλογα.
- Εάν ο κέρσορας φτάσει στα όρια της οθόνης, προκαλεί μια ενέργεια κύλισης για να συνεχίσει τη μετακίνηση.

Συνολικά, η κλάση MouseController παρέχει μια διεπαφή που ελέγχεται μέσω φωνητικών εντολών για την εκτέλεση διαφόρων ενεργειών που σχετίζονται με την αλληλεπίδραση με τον υπολογιστή και τον ιστό. Αυτό επιτρέπει στους χρήστες να εκτελούν διάφορες εργασίες, όπως αναζητήσεις στον ιστό, πλοήγηση σε ιστοσελίδες, πληκτρολόγηση κ.α χρησιμοποιώντας απλά τη φωνή τους.

5.2.3 Κλάση Browser

Η κλάση Browser αντιπροσωπεύει τον web browser που χρησιμοποιείται για την αυτοματοποιημένη αλληλεπίδραση με ιστοσελίδες και την εκτέλεση αναζητήσεων στο Google. Εδώ είναι η περιγραφή των μεθόδων και των λειτουργιών της κλάσης:

__init__(self)

- Αυτή η μέθοδος είναι ο κατασκευαστής της κλάσης και καλείται όταν δημιουργείται ένα αντικείμενο της Browser.
- Στον κατασκευαστή, δημιουργείται ένα νέο παράθυρο περιήγησης Google Chrome με τη χρήση της βιβλιοθήκης Selenium και του WebDriver Manager για την αυτόματη διαχείριση του Chrome Driver.
- Το παράθυρο του browser μεγιστοποιείται στην οθόνη του υπολογιστή.

perform_google_search(self, term)

- Αυτή η μέθοδος πραγματοποιεί μια αναζήτηση στο Google με βάση την παρεχόμενη λέξη-κλειδί term.
- Ανοίγει μια νέα καρτέλα του προγράμματος περιήγησης και προβάλλει τα αποτελέσματα αναζήτησης για το συγκεκριμένο term.

accept_google_cookies(self)

- Αυτή η μέθοδος χρησιμοποιείται για την αυτόματη αποδοχή των cookies που παρουσιάζονται στην αρχική σελίδα του Google.
- Περιμένει για ένα σύντομο χρονικό διάστημα (0,5 δευτερόλεπτα) και στη συνέχεια εντοπίζει το κουμπί "Αποδοχή" (Accept) και το πατά, γεγονός που επιτρέπει τη συνέχιση της αναζήτησης.

visit_page(self, url)

- Με αυτή την μέθοδο επισκεπτόμαστε την σελίδα που μας υποδεικνύεται από την παράμετρο url

5.2.4 Script grid.py

Το script grid.py δημιουργεί ένα παράθυρο προβολής πλέγματος στην οθόνη υπολογιστή.

Αρχικοποίηση

- Το script αρχίζει με την εισαγωγή των απαραίτητων βιβλιοθηκών και την αρχικοποίηση του Pygame.
- Υπολογίζει το πλάτος και το ύψος της οθόνης υπολογιστή με βάση τις παραμέτρους της οθόνης.
- Υπολογίζει το μέγεθος του πλέγματος (GRID_SIZE) ανάλογα με τις διαστάσεις της οθόνης. Μπορεί να προσαρμοστεί αλλάζοντας τον διαιρέτη στην γραμμή υπολογισμού.
- Ορίζει το χρώμα των γραμμών του πλέγματος (GRID_COLOR). Το χρώμα του πλέγματος είναι σχεδόν διάφανο ώστε ο χρήστης να μην δυσκολεύεται να διακρίνει τα στοιχεία της σελίδας στην οποία περιηγείται

Μέθοδος draw_grid(screen)

- Αυτή η μέθοδος δέχεται ένα παράθυρο Pygame ως όρισμα.
- Σχεδιάζει το πλέγμα στην οθόνη με χρήση γραμμών, με το χρώμα που καθορίστηκε στην αρχικοποίηση (GRID_COLOR).
- Οι γραμμές του πλέγματος σχεδιάζονται κάθε GRID_SIZE pixel.

Μέθοδος grid()

- Αυτή η μέθοδος είναι υπεύθυνη για την εμφάνιση του πλέγματος στην οθόνη.
- Δημιουργεί ένα παράθυρο Pygame με τις διαστάσεις της οθόνης και αφαιρεί την περιγραφή του παραθύρου για να φαίνεται ως παράθυρο χωρίς πλαίσιο (no-frame).
- Ορίζει τη διαφάνεια του παραθύρου (μόνο στα Windows) για να είναι διαφανές.
- Στη συνέχεια, ανεβάζει το πλέγμα πάνω στο παράθυρο και ενημερώνει την οθόνη.
- Η ενημέρωση της οθόνης πραγματοποιείται με συγκεκριμένο ρυθμό καρέ (frame rate) το οποίο περιορίζεται σε 60 FPS.

Μέθοδος close()

- Αυτή η μέθοδος κλείνει το παράθυρο Pygame και τερματίζει το πρόγραμμα.

Το συγκεκριμένο script είναι χρήσιμο για τη δημιουργία του πλέγματος πάνω από την οθόνη του υπολογιστή, το οποίο μπορεί να χρησιμοποιηθεί για διάφορους σκοπούς, όπως τον προσδιορισμό συντεταγμένων σε εφαρμογές που απαιτούν τον χειρισμό του ποντικιού, για παιχνίδια κ.λπ.

5.2.5 Script SpeechToText.py

Το script SpeechToText.py περιλαμβάνει λειτουργίες για την αναγνώριση φωνητικών εντολών και την εκτέλεση αναζητήσεων στο Google με βάση αυτές τις εντολές.

Μέθοδος speak(message)

- Αυτή η συνάρτηση χρησιμοποιείται για να εκφωνήσει ένα μήνυμα χρησιμοποιώντας τη φωνή του υπολογιστή.
- Χρησιμοποιεί τη βιβλιοθήκη pyttsx3 για τη σύνθεση και αναπαραγωγή του μηνύματος με φωνητική έκφραση.

Μέθοδος obtain_audio_from_mic(message)

- Αυτή η συνάρτηση χρησιμοποιείται για να λάβει ήχο από το μικρόφωνο του υπολογιστή.
- Χρησιμοποιεί τη βιβλιοθήκη speech_recognition για να ανιχνεύσει και να καταγράψει τον ήχο.

Μέθοδος convert_speech_to_text(audio)

- Αυτή η συνάρτηση δέχεται ένα αντικείμενο ήχου (ήχο από το μικρόφωνο) ως είσοδο.
- Χρησιμοποιεί τη βιβλιοθήκη speech_recognition για να μετατρέψει την ηχητική είσοδο σε κείμενο, χρησιμοποιώντας την υπηρεσία αναγνώρισης φωνής της Google.
- Επιστρέφει το αποτέλεσμα της αναγνώρισης ως κείμενο.

Μέθοδος perform_google_search(query)

- Αυτή η συνάρτηση δέχεται ένα κείμενο ως ερώτημα για αναζήτηση στο Google.
- Δημιουργεί το URL για την αναζήτηση στο Google βάσει του ερωτήματος.
- Ανοίγει ένα νέο παράθυρο του προγράμματος περιήγησης που είναι ήδη εγκατεστημένο στον υπολογιστή για να εκτελέσει την αναζήτηση στο Google.

Αυτές οι συναρτήσεις επιτρέπουν στο την αλληλεπίδραση με τον χρήστη μέσω φωνητικών εντολών, την καταγραφή του ήχου από το μικρόφωνο, την μετατροπή της φωνής σε κείμενο και την αναζήτηση στο Google με βάση το κείμενο που αναγνωρίστηκε.

5.2.6 Script WORDS.py

Στο συγκεκριμένο script έχουμε αναθέσει σε μεταβλητές όλες τις πιθανές εντολές που μπορεί να εκφωνήσει ο χρήστης για την χρήση της εφαρμογής. Κατασκευάσαμε το συγκεκριμένο script ώστε στο μέλλον σε περίπτωση που επιθυμούμε να αλλάξουμε μια εντολή απλά να εκχωρήσουμε άλλη τιμή στην αντίστοιχη μεταβλητή στο script WORDS.py.

5.3 Οδηγίες εκτέλεσης προγράμματος

Αρχικά κάνουμε execute το main.py. Το πρόγραμμα μας παραπέμπει να εκφωνήσουμε τον όρο, τον οποίο θέλουμε να υποβάλουμε υπό αναζήτηση στο google. Αφού ο χρήστης πει τι επιθυμεί να αναζητήσει, η φωνή του χρήστη μετατρέπεται σε text και πραγματοποιείται η αναζήτηση στο google. Σε περίπτωση που δεν γίνει σωστά η λήψη της φωνής ή σε περίπτωση που ο χρήστης δεν πει τίποτα το πρόγραμμα περιμένει την είσοδο από τον χρήστη. Μπορείτε επίσης να παρακολουθήσετε το βίντεο στον παρακάτω σύνδεσμο για μια πιο ενδελεχή κατανόηση του συστήματος μας

[VIDEO SYSTEM PRESENTATION](#)

Πάνω-Κάτω-Δεξιά-Αριστερά

Αφού ο χρήστης έχει πει τον όρο που επιθυμεί να υποβάλει υπο αναζήτηση το πρόγραμμα αναμένει τις επόμενες εντολές. Για παράδειγμα ο χρήστης μπορεί να πει Δεξιά 100 και να κινήσει το ποντίκι δεξιά 100 pixel. Ο χρήστης μπορεί να συνεχίσει να κουνάει το ποντίκι με την εκάστοτε κατάλληλη εντολή ώστε να τοποθετήσει το ποντίκι πάνω από το στοιχείο που επιθυμεί. Σε περίπτωση που αναφέρει αριθμό pixel έξω από τα όρια της οθόνης πραγματοποιείται scroll down ή scroll up.

Open grid - Open - Close grid - Close

Υπάρχει η δυνατότητα ο χρήστης να δώσει τις παραπάνω εντολές στο σύστημα ώστε να εμφανιστεί ένα διάφανο πλέγμα στην οθόνη με σκοπό να μετακινεί το ποντίκι με μεγαλύτερη ακρίβεια. Να σημειωθεί πως έχουμε προσαρμόσει τα κουτάκια του grid σύμφωνα με την δικιά μας οθόνη όπου και εξετάσαμε το σύστημα η οποία είναι 17 ιντσών και τα προαναφερθέντα κουτάκια-κελιά είναι 54X54 pixel, σε κάποια μεγαλύτερη ή μικρότερη οθόνη η μεταβλητή που καθορίζει το μήκος και το πλάτος των κελιών, πιθανόν να πρέπει να γίνει re-assign με διαφορετική τιμή ώστε να φαίνονται τα στοιχεία της σελίδας με μεγαλύτερη ακρίβεια.

Νέα αναζήτηση

Με την εντολή Νέα Αναζήτηση ο χρήστης μπορεί σε οποιοδήποτε σημείο βρίσκεται κατά την διάρκεια της περιήγησης του, να εκφωνήσει έναν νέο όρο προς αναζήτηση. Προσθέσαμε την συγκεκριμένη εντολή με γνώμονα όχι μόνο το να υπάρχει η δυνατότητα της αναζήτησης με κάποιον καινούργιο όρο, αλλά και για τον λόγο ότι το google api δεν αναγνωρίζει πάντα σωστά την φωνή του χρήστη, επιστρέφοντας έτσι λανθασμένα αποτελέσματα. Με αυτή την εντολή, σε περίπτωση που εκληφθεί λάθος η φωνή του χρήστη από το σύστημα, ο χρήστης θα μπορεί να υποβάλει ξανά τον όρο που επιθυμεί για αναζήτηση.

Type - Delete

Με την εντολή `type` ο χρήστης μπορεί να εκφωνήσει τον όρο που επιθυμεί να πληκτρολογήσει και αυτός ο όρος να εμφανιστεί στο αντίστοιχο `search bar` και να πατηθεί αυτόματα `enter`. Ο χρήστης αρχικά πρέπει να τοποθετήσει το `mouse cursor` πάνω από το `search bar`, να πει την εντολή `click` και μετέπειτα να πει `type`. Αφού πει `type`, το σύστημα τον προτρέπει να πει τον όρο που επιθυμεί να πληκτρολογήσει (`Please say what you want to type`). Στην συνέχεια, αφού ο χρήστης πει τον προς πληκτρολόγηση όρο, το αντίστοιχο `text` εμφανίζεται στο `search bar` και πατιέται αυτοματα `enter`.

Με αντίστοιχο τρόπο χρησιμοποιείται η εντολή `delete`. Σε περίπτωση που υπάρχει ήδη κάποιο κείμενο στο `search bar` και ο χρήστης επιθυμεί να το σβήσει, πρέπει να τοποθετήσει το `mouse cursor` πάνω από το εν λόγω `search bar`, να πει `type` και στην συνέχεια αφού το σύστημα τον προτρέψει να εκφωνήσει τον όρο που επιθυμεί να πληκτρολογήσει να πει `delete`.

Click

Με την εντολή `click` πραγματοποιείται ένα `mouse click`. Ο χρήστης τοποθετεί το `mouse cursor` πάνω από το αντίστοιχο `element` που επιθυμεί να κλικάρει, λέει `click` και κλικάρεται το αντίστοιχο στοιχείο.

Click (text του element που θέλουμε να κλικάρουμε)

Με την χρήση της βιβλιοθήκης `selenium` μπορούμε να κάνουμε `retrieve elements` από το κείμενο που έχουν ως `value` και να τα κλικάρουμε, χωρίς να χρειαστεί να τοποθετήσουμε το ποντίκι πάνω από τα αντίστοιχα στοιχεία στην οθόνη μας. Ο χρήστης απλώς λέει `click` ακολουθούμενο από το κείμενο που βλέπει στην οθόνη και το αντίστοιχο στοιχείο κλικάρεται.

Press (text του button element που θέλουμε να κλικάρουμε)

Αντίστοιχα με την προηγούμενη εντολή, χρησιμοποιούμε την εντολή `press` για να κλικάρουμε `button elements`. Για παράδειγμα αν έχουμε ένα `button` που περιέχει το κείμενο `Accept Cookies` ο χρήστης μπορεί να πει `press Accept Cookies` και να πατηθεί το αντίστοιχο `button`.

Πίσω

Με την εντολή `πίσω` ο χρήστης μπορεί να μεταφερθεί στην προηγούμενη σελίδα από την οποία βρίσκεται.

Exit

Με την εντολή `exit` ο χρήστης μπορεί να κλείσει το πρόγραμμα

5.4 Libraries

Για την εκτέλεση του συγκεκριμένου project απαιτείται Python 3. Επίσης πρέπει να γίνουν `pip install` τα `requirements.txt`

6. ΣΥΜΠΕΡΑΣΜΑΤΑ

Επιτυχής Πλοήγηση στο Διαδίκτυο μέσω Φωνητικών εντολών :

Το project παρουσίασε με επιτυχία την υλοποίηση ενός συστήματος φωνητικού ελέγχου για την αυτοματοποίηση της πλοήγησης στο διαδίκτυο. Οι χρήστες μπορούν να αλληλεπιδρούν με τις σελίδες του ιστού, να πραγματοποιούν αναζητήσεις, να κάνουν κλικ σε συνδέσμους και να πλοηγούνται σε σελίδες χρησιμοποιώντας φωνητικές εντολές, καθιστώντας το πιο βολικό μέσο πλοήγησης στον ιστό.

Βελτίωση της Προσβασιμότητας:

Τα συστήματα φωνητικού ελέγχου βελτιώνουν την προσβασιμότητα για άτομα με σωματικές αναπηρίες ή για όσους προτιμούν την χρήση υπολογιστών χωρίς τη χρήση χεριών. Αυτή η τεχνολογία έχει το δυναμικό να καταστήσει την πλοήγηση στον ιστό και την αλληλεπίδραση με τον υπολογιστή πιο προσιτή.

Συνδυασμός Τεχνολογιών-Τεχνικών:

Το έργο ολοκληρώθηκε με επιτυχία συνδυάζοντας αρκετές τεχνολογίες, συμπεριλαμβανομένης της αναγνώρισης φωνής, του web scraping και της αυτοματοποίησης browser, για τη δημιουργία ενός συνολικού συστήματος. Αυτή η ενσωμάτωση απαιτούσε προσεκτική εξέταση των βιβλιοθηκών λογισμικού και της συμβατότητας τους.

Χρήση Εργαλείων Ανοιχτού Κώδικα:

Το έργο χρησιμοποίησε εργαλεία και βιβλιοθήκες ανοιχτού κώδικα, όπως το Selenium, το PyAutoGUI και η βιβλιοθήκη SpeechRecognition, τα οποία είναι εύκολα διαθέσιμα για τους προγραμματιστές. Αυτό όχι μόνο μείωσε το κόστος ανάπτυξης αλλά και προάγει τη συνεργασία και την περαιτέρω ανάπτυξη στην κοινότητα του ανοιχτού κώδικα.

Προκλήσεις και Περιορισμοί:

Παρά την επίτευξη των βασικών στόχων, το έργο αντιμετώπισε και προκλήσεις και περιορισμούς. Η ακρίβεια της αναγνώρισης φωνής μπορεί να επηρεάζεται από τον θόρυβο του περιβάλλοντος και τις προφορικές διαφορές, ενώ ενδέχεται να υπάρχουν δυσκολίες στην αναγνώριση φωνητικών εντολών σε γλώσσες διαφορετικές από τα Αγγλικά, όπως στην περίπτωση μας τα Ελληνικά. Επιπλέον, η χρηστικότητα του συστήματος θα μπορούσε να βελτιωθεί με επιπρόσθετες φωνητικές εντολές και χειρισμό σφαλμάτων.

Μελλοντική Ανάπτυξη:

Υπάρχει σημαντικό δυναμικό για τη μελλοντική ανάπτυξη και βελτίωση του συστήματος. Αυτό περιλαμβάνει τη βελτίωση των ικανοτήτων επεξεργασίας φυσικής γλώσσας του συστήματος για την κατανόηση πιο πολύπλοκων ερωτημάτων, τη διεύρυνση του φάσματος των υποστηριζόμενων διαδικτυακών υπηρεσιών, και την ενίσχυση της ακρίβειας της φωνητικής αναγνώρισης.

Ευελιξία και Επεκτασιμότητα:

Το έργο έχει σχεδιαστεί με ευελιξία και επεκτασιμότητα, επιτρέποντας την προσθήκη νέων λειτουργιών και επεκτάσεων μεταξύ γλωσσών και υπηρεσιών.

7. ΕΠΙΛΟΓΟΣ

Το έργο αυτό αντιπροσωπεύει μια σημαντική προσπάθεια για την ενσωμάτωση της φωνητικής αλληλεπίδρασης στην καθημερινή χρήση του internet και την προώθηση της προσβασιμότητας σε όλους τους χρήστες. Η επιτυχής υλοποίηση αυτού του συστήματος φωνητικού ελέγχου ανοίγει νέους ορίζοντες για την αλληλεπίδραση των ανθρώπων με την τεχνολογία και τον κυβερνοχώρο.

Με τις προκλήσεις που αντιμετωπίσαμε και τους περιορισμούς που ανακαλύψαμε, κατανοήσαμε τη σημασία της συνεχούς βελτίωσης και αναβάθμισης αυτής της τεχνολογίας. Η ανάπτυξη του συστήματος συνεχίζεται, με στόχο την ενίσχυση της ακρίβειας της φωνητικής αναγνώρισης, τη διεύρυνση των διαθέσιμων λειτουργιών, και τη βελτίωση της χρηστικότητας.

Το έργο επίσης υπογραμμίζει τη σημασία της χρήσης ανοιχτού κώδικα και της συνεργασίας στην κοινότητα του προγραμματισμού. Μέσω της ανάπτυξης και της μοιρασιάς ανοιχτού κώδικα λογισμικού, η τεχνολογία γίνεται προσιτή σε περισσότερους και επιτρέπει την ταχύτερη εξέλιξη των εφαρμογών.

Το έργο αυτό αντικατοπτρίζει τη δέσμευσή μας στη δημιουργία τεχνολογίας που ωφελεί την κοινότητα και προωθεί την προσβασιμότητα και την ευκολία χρήσης. Επίσης, ενισχύει την πίστη μας στη δυνατότητα της τεχνολογίας να δημιουργεί καινοτόμες λύσεις για τις καθημερινές μας ανάγκες.