



ΔΙΕΘΝΕΣ  
ΠΑΝΕΠΙΣΤΗΜΙΟ  
ΕΛΛΑΔΟΣ

ΔΙΕΘΝΕΣ ΠΑΝΕΠΙΣΤΗΜΙΟ ΤΗΣ ΕΛΛΑΔΟΣ  
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ  
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ,  
ΥΠΟΛΟΓΙΣΤΩΝ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

**Διερευνητική μελέτη εφαρμογής φίλτρων εικόνας σε  
μηχανές OCR και έλεγχος απόδοσης» (Exploratory study  
and performance control of image filters application in OCR  
engines.**

**Πτυχιακή Εργασία του  
Γεωργαλή Ανδρέα (4428)**

Επιβλέπων: Δρ. Τσιμπίρης Αλκιβιάδης,  
Επίκουρος Καθηγητής

**ΣΕΡΡΕΣ, ΙΟΥΝΙΟΣ 2021**

**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, Υπολογιστών και Τηλεπικοινωνιών του Διεθνούς Πανεπιστημίου της Ελλάδας.

## Ευχαριστίες

Στο σημείο αυτό, θα ήθελα να εκφράσω τις θερμές ευχαριστίες μου σε όλους όσους με βοήθησαν και συνέβαλλαν στην ολοκλήρωση της εργασίας αυτής. Πρώτον από όλους ευχαριστώ τον επιβλέποντα καθηγητή μου κ. Τσιμπίρη για την πολύ καλή και ευχάριστη συνεργασία που είχαμε και για τις γνώσεις που με τη βοήθειά του αποκόμισα όλο αυτό το διάστημα. Συνεχίζοντας, θα ήθελα να πω ένα μεγάλο ευχαριστώ στην οικογένειά μου, για την πολύτιμη στήριξη της όλα αυτά τα χρόνια, συμβάλλοντας καθοριστικά στην εξέλιξή μου. Τέλος, ευχαριστώ όλους τους φίλους και τα κοντινά μου πρόσωπα για την υπομονή, τη βοήθεια και τη στήριξη που μου πρόσφεραν καθ'όλη τη διάρκεια των σπουδών μου.

## **Περίληψη**

Η παρούσα πτυχιακή εργασία ασχολείται με την μελέτη απόδοσης διαφόρων φίλτρων επεξεργασίας εικόνας και η χρήση αυτών σε μηχανές οπτικής αναγνώρισης χαρακτήρων(OCR). Στα πρώτα κεφάλαια αναφέρονται βασικές έννοιες και χρήσεις της επεξεργασίας ψηφιακής εικόνας και της τεχνολογίας οπτικής αναγνώρισης χαρακτήρων. Έπειτα από την ιστορική αναδρομή και της περαιτέρω ανάλυσης της βιβλιοθήκης tesseract-ocr, στην οποία και βασίζεται η παρούσα εργασία, παρουσιάζεται μια πλήρη περιγραφή της εφαρμογής και τον τρόπο χρήσης της για την μελέτη των φίλτρων που αναφέρθηκαν, όπως επίσης και μελλοντικές βελτιστοποιήσεις και επέκτασης αυτής της εφαρμογής.

## **Abstract**

The present thesis concerns the study of the performance of various image processing filters and their use in optical character recognition(OCR) machines. The first chapters touched upon basic concepts and uses of digital images processing and optical character recognition technology. After the historical review and further analysis of the tesseract-ocr library, on which the present work is base, a complete description of the application and how to use it for the study of the mentioned filters is presented, as well as future optimizations and extension thereof.

## Περιεχόμενα

Κεφάλαιο 1 .....	6
Κεφάλαιο 2 .....	7
Κεφάλαιο 3 .....	17
Κεφάλαιο 4 .....	39
Κεφάλαιο 5 .....	47
Κεφάλαιο 6 .....	66
Βιβλιογραφία .....	67

### **1. Εισαγωγή**

Ο απώτερος σκοπός της παρούσας πτυχιακής εργασίας είναι η κατανόηση της τεχνολογίας οπτικής αναγνώρισης χαρακτήρων (OCR) και η δημιουργία μιας διαδικτυακής εφαρμογής που αφορά την μελέτη απόδοσης αυτής της τεχνολογίας έπειτα από επεξεργασία εικόνας μετά την εφαρμογή διαφόρων φίλτρων.

Η συνεχής ανάγκη εύρεσης και επεξεργασίας πληροφοριών, καθιστά αναγκαία μια εφαρμογή μετατροπής έντυπου εγγράφου σε επεξεργάσιμο ψηφιακό κείμενο. Η δυσκολία παρουσιάζεται στην εκπαίδευση απαιτούμενων βιβλιοθηκών και η παραμετροποίησή τους, ώστε να έχουμε το επιθυμητό αποτέλεσμα.

Έχει αναπτυχθεί πλήθος παρόμοιων εφαρμογών είτε σε μορφή desktop είτε διαδικτυακής μορφής, πολλές από αυτές παρουσιάζουν σφάλματα στην ελληνική γλώσσα, καθώς και είναι η δυσκολότερη, λόγω του ότι δεν ακολουθεί το λατινικό αλφάβητο όπως οι περισσότερες γλώσσες και οι χαρακτήρες εμπεριέχουν τόνους. Άλλες πάλι λειτουργούν άψογα, δίνοντας αποτελέσματα υψηλών ποσοστών επιτυχίας με πολύ μικρά έως και μηδαμινά σφάλματα, όμως το κόστος τους είναι αρκετά μεγάλο με αποτέλεσμα να αποτρέψει πολλούς χρήστες.

Οι θεωρητικές έννοιες για την κατανόηση λειτουργίας της εφαρμογής, αναπτύσσονται με στόχο την πρόκληση ενδιαφέροντος του αναγνώστη για το γνωστικό αντικείμενο της εργασίας, και την εξοικείωσή του με την τεχνολογία OCR.

### **2. Οπτική Αναγνώριση Χαρακτήρων (OCR)**

Η Οπτική Αναγνώριση Χαρακτήρων (Optical Character Recognition) είναι η μηχανική ή ηλεκτρονική μετατροπή σαρωμένων εικόνων χειρόγραφου ή τυποποιημένου κειμένου σε κωδικοποιημένη μορφή χαρακτήρων κειμένου ηλεκτρονικού υπολογιστή.

Χρησιμοποιείται ως ένας τρόπος εξαγωγής δεδομένων από διάφορες μορφές εντύπων κειμένων, όπως έγγραφα, γράμματα αλληλογραφίας, αποδείξεις πωλήσεων, βιβλία, πινακίδες αυτοκινήτων κ.τ.λ. Είναι μια μέθοδος ψηφιοποίησης (Digitization) εντύπων κειμένων έτσι ώστε να μπορούν να προσπελαστούν ηλεκτρονικά, να αποθηκευτούν, να είναι προσβάσιμα από τον παγκόσμιο ιστό και να μπορούν να επεξεργαστούν περαιτέρω για διαδικασίες όπως είναι η αυτόματη μετάφραση (machine translation), η εξόρυξη δεδομένων (data mining) ή η μετατροπή κειμένου σε φωνή (text-to-speech) .

Η οπτική αναγνώριση χαρακτήρων αποτελεί ένα ευρύ πεδίο έρευνας για τους τομείς της αναγνώρισης μοτίβων (pattern recognition), της τεχνητής νοημοσύνης (artificial intelligence) και της μηχανικής όρασης (computer vision).

Οι πρώιμες εφαρμογές του OCR βασίζονταν στην απομνημόνευση ενός-ενός των γραμμάτων από φωτογραφίες, των οποίων και έψαχναν μεμονωμένα μέσα στα προς έλεγχο κείμενα. Με τον καιρό αναπτύχθηκαν εξυπνότερα συστήματα τα οποία εξέταζαν τις λέξεις σαν σύνολα και τα οποία προχωράν σε «συλλογιστικά άλματα» προσπαθώντας να λαμβάνουν υπ' όψιν και το νόημα της λέξης στόχου στην προσπάθειά τους να κάνουν τη σωστή αναγνώριση (Με την χρήση λεξικού, μπορεί να μαντέψουν σωστά την λέξη ακόμα και εάν δεν την έχουν αναγνωρίσει σωστά). Μερικά συστήματα είναι ικανά ακόμη και αναπαράγουν σε κάποιο βαθμό τη διαμόρφωση της πρωτότυπης σελίδας (Δηλαδή τις εικόνες, τις στήλες ή άλλο γραφιστικό και στιλιστικό περιεχόμενο).

Το OCR όμως δεν μπορεί να σταθεί σαν αυτόνομη λύση στην επικοινωνία ανθρώπου-μηχανής. Το κύριο πρόβλημα με το OCR είναι ότι παράγει μόνο μη δομημένους χαρακτήρες. Αυτό απαιτεί τον συνδυασμό άλλων τεχνολογιών μηχανικής μάθησης στο

OCR. Με αυτόν τον τρόπο, οι χρήστες μπορούν να προσεγγίσουν δομημένα δεδομένα από τα έγγραφά τους.



## 2.1 Ιστορία

Η οπτική αναγνώριση χαρακτήρων μπορεί να καταταχθεί ως επέκταση της τηλεγραφίας και στις προσπάθειες ανάπτυξης συσκευών ανάγνωσης για τους τυφλούς.

Το 1914 ο Emanuel Goldberg ανέπτυξε μια μηχανή η οποία διάβαζε χαρακτήρες και τους μετέτρεπε σε τυποποιημένο τηλεγραφικό κώδικα (Morse). Την ίδια περίοδο, ο Edmund Fournier d' Albe (Edmund Edward Fournier) ανέπτυξε το Optophone (Optophone), έναν φορητό σαρωτή εικόνας ο οποίος όταν κινούταν κατά μήκος μια εκτυπωμένης γραμμής κειμένου παρήγαγε ηχητικούς τόνους που αντιστοιχούσαν στους χαρακτήρες που συναντούσε. Πρόκειται για μια από τις πρώτες γνωστές εφαρμογές της ηχητικοποίησης

Ο Goldberg συνέχισε με την ανάπτυξη τεχνολογίας οπτικής αναγνώρισης χαρακτήρων για εισαγωγή δεδομένων. Αργότερα, πρότεινε μια μέθοδο φωτογράφισης δεδομένων τα οποία μετά, με τη χρήση φωτοκύτταρων, αντιστοιχίζονταν σε κάποια μοτίβα αναγνώρισης. Το 1929 ο Gustav Tauschek (Gustav Tauschek) αποδείχθηκε ότι είχε παρόμοιες ιδέες και εξασφάλισε μια πατέντα οπτικής αναγνώρισης χαρακτήρων στη Γερμανία. Ο Tauschek ήταν ένας ιδιοφυής αυτοδίδακτος βιεννέζος μηχανικός, με περισσότερες από 200 πατέντες στον τομέα της πληροφορικής στο ενεργητικό του, ο οποίος εργαζόταν για την IBM και ο οποίος, εκτός από την πρώτη συσκευή OCR, εφηύρε επίσης πολλές συσκευές και συστήματα για τα μηχανήματα με διάτρητες κάρτες, καθώς και τη μνήμη μαγνητικού τυμπάνου. Το 1935 ο Tauschek εξασφάλισε αμερικάνικη πατέντα OCR για τη μεθόδό του.

Το ίδιο και ο Paul W. Handel, ο οποίος εξασφάλισε κι αυτός μια πατέντα ερασιτεχνίας που αφορούσε μια τεχνική αντιστοίχισης προτύπων στις Ηνωμένες Πολιτείες το 1933.

Το 1949, ερευνητές της RCA (Radio Corporation of America) εργάστηκαν για πρώτη φορά πάνω σε μια πρώιμη μορφή ενός συστήματος οπτικής αναγνώρισης χαρακτήρων με τη βοήθεια ηλεκτρονικού υπολογιστή για λογαριασμό της Γενικής Γραμματείας Απόστρατων των Ηνωμένων Πολιτειών Αμερικής, το οποίο όμως αντί να μετατρέπει το κείμενο σε κείμενο υπολογιστή είχε ως στόχο να εκφωνεί τα γράμματα. Ήταν δηλαδή μία πρώιμη εφαρμογή text-to-speech, αλλά αποδείχθηκε πολύ ακριβή και ριζοσπαστική για την εποχή της ώστε να καθιερωθεί.

Το 1950 ο David H. Shepard (David H. Shepard), ένας εκ των κορυφαίων κρυπταναλυτών (κρυπτανάλυση) της Υπηρεσίας Ασφαλείας του Αμερικανικού Στρατού, κατασκεύασε μία μηχανή ονόματι «Gismo» (Gismo) η οποία μπορούσε να αναγνωρίζει 23 γράμματα του αλφαβήτου, όπως παράγονται από μια τυπική γραφομηχανή, να καταλάβει κώδικα μορς και μουσικά σύμβολα και να εκφωνήσει και να αντιγράψει εκτυπωμένες σελίδες. Έλαβε πατέντα για την κατασκευή του το 1953. Ο Shepard προχώρησε ιδρύοντας την εταιρία IMR(Intelligent Machines Research Corporation) η οποία σε λίγο καιρό ανέπτυξε τα πρώτα συστήματα οπτικής αναγνώρισης χαρακτήρων για εμπορική χρήση.

Το 1955, το πρώτο εμπορικό σύστημα οπτικής αναγνώρισης χαρακτήρων εγκαταστάθηκε στα γραφεία του περιοδικού Reader's Digest και χρησιμοποιήθηκε για να εισάγει αναφορές πωλήσεων σε έναν ηλεκτρονικό υπολογιστή. Μετέτρεπε τις δακτυλογραφημένες αναφορές σε διάτρητες κάρτες τις οποίες αργότερα το προσωπικό του τμήματος συνδρομών έβαζε σ' έναν υπολογιστή που εκτελούσε την επεξεργασία των δεδομένων που αφορούσαν τις 15-20 εκατομμύρια αποστολές εντύπων που εκτελούσαν κάθε χρόνο. Το δεύτερο σύστημα πουλήθηκε στην Standard Oil και χρησιμοποιήθηκε για την ανάγνωση αποδείξεων συναλλαγών πιστωτικών καρτών. Άλλα συστήματα που δημιουργήθηκαν και πουλήθηκαν από την IMR περιελάμβαναν έναν αναγνώστη τιμολογίων για την τηλεφωνική εταιρία Bell του Ohio κι έναν σαρωτή σελίδων για την αμερικάνικη πολεμική αεροπορία ο οποίος χρησιμοποιήθηκε για την τηλεγράφιση δακτυλογραφημένων μηνυμάτων. Τα επόμενα χρόνια ο Shepard πούλησε της πατέντες του σε διάφορες εταιρίες, μεταξύ των οποίων και στην IBM.

Γύρω στο 1965, το Reader's Digest και η RCA συνεργάστηκαν για την κατασκευή ενός αναγνώστη εγγράφων που είχε ως στόχο την ψηφιοποίηση των σειριακών αριθμών κουπονιών του περιοδικού που επιστράφηκαν από διαφημίσεις. Τα γράμματα που χρησιμοποιούταν στα έγγραφα αυτά είχαν εκτυπωθεί από έναν εκτυπωτή τυμπάνου της ίδιας εταιρίας με μια γραμματοσειρά ειδικά σχεδιασμένη για τους σκοπούς της οπτικής αναγνώρισης. Μάλιστα ο αναγνώστης αυτός ήταν απ' ευθείας συνδεδεμένος με έναν υπολογιστή RCA 301, έναν από τους πρώτους ηλεκτρονικούς υπολογιστές τεχνολογίας ημιαγωγών (transistor technology). Ένας άλλος ειδικός αναγνώστης που επίσης αναπτύχθηκε σε συνεργασία, τοποθετήθηκε στα γραφεία της αεροπορικής εταιρίας TWA και χρησιμοποιήθηκε για την επεξεργασία του αποθέματος των καρτών επιβίβασης. Ο αναγνώστης αυτός διάβαζε έγγραφα με το ρυθμό των 1500 σελίδων το

λεπτό απορρίπτοντας τις σελίδες που δεν μπορούσε να διαβάσει πλήρως. Το προϊόν αυτό μπήκε τελικά στην κανονική γραμμή παραγωγής της RCA όπου και προωθήθηκε εμπορικά τα επόμενα χρόνια για διάφορες εργασίες αντίστοιχου τύπου.

Το αμερικάνικο ταχυδρομείο χρησιμοποιούσε μηχανές οπτικής αναγνώρισης χαρακτήρων ήδη από το 1965 βασιζόμενο σε τεχνολογία του μεγάλου επιστήμονα κι εφευρέτη Jacob Rabinow. Η πρώτη χρήση τεχνολογίας οπτικής αναγνώρισης χαρακτήρων στην Ευρώπη έγινε από το βρετανικό ταχυδρομείο. Το 1965 ξεκίνησε το σχεδιασμό ενός ολόκληρου τραπεζικού συστήματος, το «National Giro» το οποίο θα χρησιμοποιούσε στον πυρήνα των συναλλαγών του τεχνολογία οπτικής αναγνώρισης χαρακτήρων, μια ιδέα ρηξικέλευθη η οποία άλλαξε από τότε τη μορφή των τραπεζικών συναλλαγών στη Βρετανία.

Το καναδικό ταχυδρομείο χρησιμοποιούσε τεχνολογία αυτόματης αναγνώρισης χαρακτήρων από το 1971. Το σύστημα διάβαζε το όνομα και τη 14 από 65 διεύθυνση του παραλήπτη και εκτύπωνε πάνω στο φάκελο έναν γραμμικό κωδικό (bar code) με οδηγίες δρομολόγησης ανάλογα με τον ταχυδρομικό κώδικα του προορισμού. Για να μην συγχέονται οι γραμμικοί κωδικοί με σύμβολα που μπορεί να σχεδιάσει ο αποστολέας οπουδήποτε πάνω στο φάκελο η εκτύπωση αυτή γινόταν με ειδικό πορτοκαλί μελάνι το οποίο είχε πολύ υψηλά ανακλαστικά χαρακτηριστικά κάτω από υπεριώδες φως. Από αυτό το στάδιο και πέρα ήταν εύκολο τα γράμματα να ταξινομηθούν και να δρομολογηθούν μηχανικά από κάποιο σύστημα ελέγχου γραμμικών κωδικών.

## **2.2 Σημασία για τους τυφλούς**

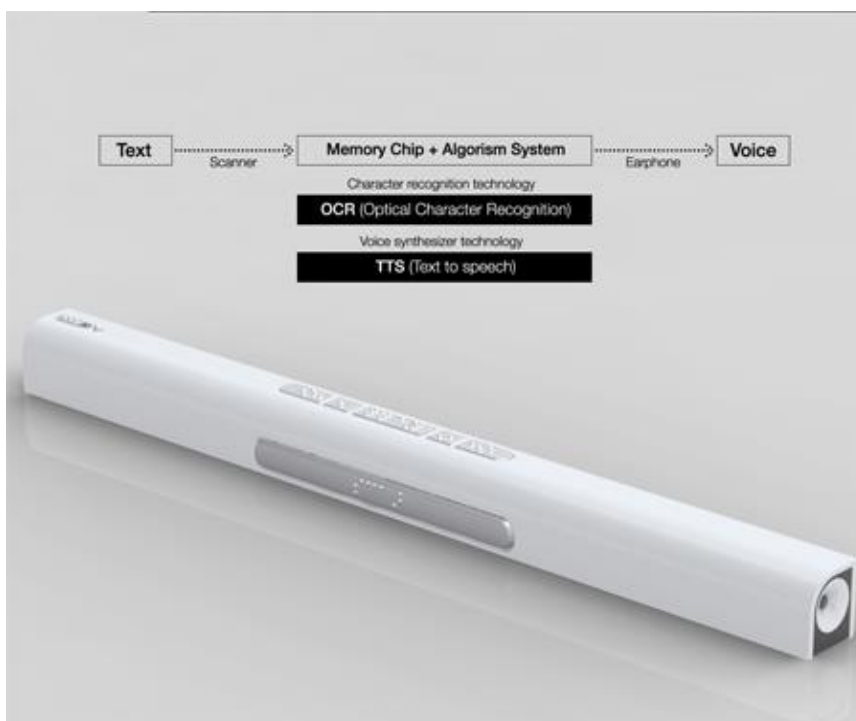
Το 1974 ο Ray Kurzweil (Ray Kurzweil) ίδρυσε την εταιρία Kurzweil Computer Products Inc και συνέχισε την ανάπτυξη προϊόντων οπτικής αναγνώρισης χαρακτήρων τα οποία λειτουργούσαν ανεξαρτήτως γραμματοσειράς. Αποφάσισε πως το καλύτερο προϊόν αυτής της τεχνολογίας θα ήταν αυτό που θα βοηθούσε τους τυφλούς ανθρώπους

να διαβάζουν. Δηλαδή ένα προϊόν με το οποίο ένας ηλεκτρονικός υπολογιστής θα μπορούσε να εκφωνεί κείμενα στους τυφλούς.

Για να υλοποιηθεί μια τέτοια συσκευή ήταν απαραίτητη η ανακάλυψη δύο ακόμα τεχνολογιών. Του επίπεδου σαρωτή CCD και του ηλεκτρονικού συνθέτη φωνής. Στις 13 Ιανουαρίου 1976, το τελειωμένο προϊόν αποκαλύφθηκε στο κοινό με μεγάλη δημοσιότητα κατά τη διάρκεια μιας κοινής συνέντευξης τύπου του Kurzweil και της αμερικανικής ένωσης τυφλών.

Το 1978 ξεκίνησε ανοικτά η διάθεση του προγράμματος στην αγορά. Ένας από τους πρώτους αγοραστές του προγράμματος ήταν η εταιρία LexisNexis η οποία το αγόρασε για να ανεβάσει επίσημα έγγραφα και αποκόμματα εφημερίδων στη εκκολαπτόμενη τότε online βάση δεδομένων τους.

Το 1980, ο Kurzweil πούλησε την εταιρία του στη Xerox την οποία ενδιέφερε έντονα εκείνη την περίοδο η περαιτέρω εμπορευματοποίηση της μετατροπής εντύπων σε αρχεία ηλεκτρονικού υπολογιστή. Τελικά η Xerox απεμπόλησε τα δικαιώματα της μεταβιβάζοντας τα στην εταιρία Scansoft, η οποία στη συνέχεια συγχωνεύτηκε με την εταιρία Nuance Communications όπου και παραμένουν μαζί μέχρι και σήμερα



Εικόνα 1 Voice Stick: Φορητή συσκευή σάρωσης κειμένου για άτομα με προβλήματα όρασης (πρόσβαση:20-06-2021)

## 2.3 Χρήση σε εφαρμογές

Η τεχνολογία οπτικής αναγνώρισης χαρακτήρων ενσωματώνεται σε προϊόντα με διάφορες μορφές.

**Desktop εφαρμογές :** Τα συστήματα οπτικής αναγνώρισης χαρακτήρων είναι αναλυτικές μηχανές τεχνητής νοημοσύνης (artificial intelligence) που εξετάζουν σειρές χαρακτήρων αντί για ολόκληρες λέξεις ή φράσεις. Στηρίζονται στην ανάλυση του σχήματος των χαρακτήρων και βασίζονται σε λεξικά για να κάνουν τις καλύτερες δυνατές προβλέψεις. Είναι οι πρώτες εφαρμογές που κυκλοφόρησαν ευρέως και χρησιμοποιούνται εδώ και πολλά χρόνια για επαγγελματικές και καθημερινές εργασίες. Αρκετές εφαρμογές τέτοιου είδους συμπεριλαμβάνονται συχνά στο λογισμικό εγκατάστασης συσκευών σάρωσης, εκτύπωσης ή συνδυασμού.

**Online εφαρμογές :** Με την εξέλιξη της τεχνολογίας και η εμφάνιση του παγκόσμιου ιστού, η χρήση της οπτικής αναγνώρισης χαρακτήρων μεταφέρθηκε από έναν υπολογιστή σε μια πληθώρα από πλατφόρμες, το διαδικτυακό νέφος (cloud computing) και οι φορητές συσκευές (mobile devices). Η διαδικτυακή οπτική αναγνώριση χαρακτήρων (Web OCR, Online OCR, Web Based OCR) έχει γίνει η νέα τάση του χώρου για την κάλυψη των αναγκών μεγαλύτερων ομάδων χρηστών. Από το 2000 και μετά ένας μεγάλος αριθμός νέων εταιριών μπήκε στο παιχνίδι της προσφοράς διαδικτυακών εφαρμογών οπτικής αναγνώρισης χαρακτήρων, με πολλές από αυτές να τις προσφέρουν ακόμα και δωρεάν στους τελικούς χρήστες, αντλώντας τα έσοδα τους από διαφημίσεις.

**Εφαρμογές προσαρμοσμένες στην περίπτωση:** Με την ραγδαία ανάπτυξη του πεδίου, προέκυψε η ανάγκη για συστήματα περισσότερο εξειδικευμένα στις ανάγκες κάποιων περιστάσεων. Για παράδειγμα, φωτογραφίες φθαρμένες, τσαλακωμένες, λερωμένες, χαμηλής ανάλυσης, θολές, κείμενα με ορολογίες που δεν συναντώνται στο τυπικό λεξικό, ειδικές γραμματοσειρές, όλα αυτά είναι συνθήκες που συναντώνται σε πραγματικά περιβάλλοντα εργασίας στην καθημερινότητα και καθιστούν πολύ δύσκολη τη σωστή λειτουργία τυποποιημένων προγραμμάτων γενικής χρήσης.

Έτσι δημιουργήθηκαν διάφορες εφαρμογές πιο εξελιγμένες για περιπτώσεις εισόδου όπως: τιμολόγια (invoice OCR), επαγγελματικές κάρτες (business-card OCR),

στιγμιότυπα οθόνης (screenshot OCR), ταυτότητες (ID card OCR), διπλώματα οδήγησης (driver license OCR) κ.τ.λ.

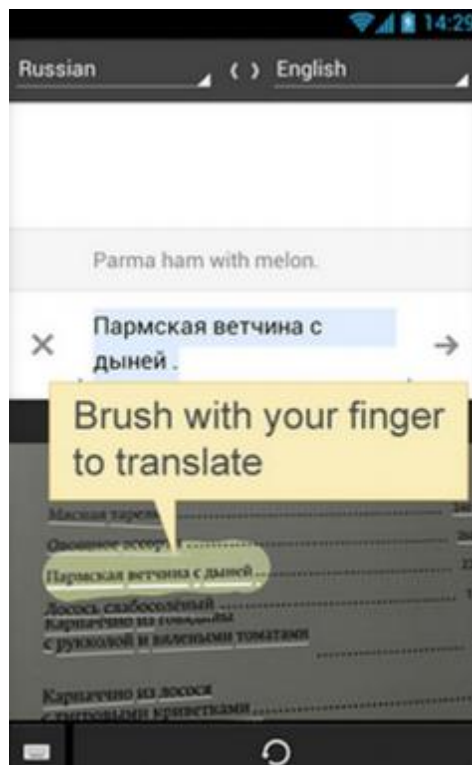
## 2.4 Χρήση στην καθημερινή ζωή

Το OCR έχει αναπτυχθεί εκτεταμένα για πάνω από 60 χρόνια, και έχει εφαρμοστεί σε διάφορους τομείς. Το OCR βοήθησε τον άνθρωπο να μειώσει ένα τεράστιο ποσό της χειρωνακτικής εργασίας όπως η πληκτρολόγηση.

Τα τελευταία λίγα χρόνια με την άνθηση των υπηρεσιών cloud (νέφος), τις έξυπνες κινητές συσκευές και την ευρεία αναγνώριση των OCR εργαλείων, η οπτική αναγνώριση εικόνων δεν παίζει μόνο το ρόλο αποδέσμευσης των ανθρώπων από κόπο και εργατοώρες, αλλά παρέχει καλύτερη εμπειρία στον χρήστη. Τώρα το OCR έχει με τον τρόπο του διεισδύσει στη καθημερινή μας ζωή.

Μερικά παραδείγματα δείχνονται ως εξής.

Google Translate είσοδος (input) από κάμερα. Το 2012, μια λειτουργία OCR προστέθηκε στην υπηρεσία Google Translate, όπως φαίνεται στην εικόνα 2, έτσι ώστε ο χρήστης να μπορεί να μεταφράζει το κείμενο χρησιμοποιώντας μόνο τον φακό της κάμερας του έξυπνου κινητού του.



Εικόνα 2 Google Translate για Android συσκευές πρόσθεσε OCR δυνατότητες (πρόσβαση: 20-06-2021)

## Αυτόματη αναγνώριση πινακίδων κυκλοφορίας (ALPR)

Το OCR βοηθάει τους αστυνομικούς για χρόνια.



*Εικόνα 3 ALRP μηχανήμα σε αστυνομικό όχημα.*

Ο σαρωτής παραπάνω είναι ένα τοποθετημένο σύστημα ALPR σε ένα όχημα, όπως φαίνεται στην εικόνα 3.

Βοηθά τους αξιωματικούς ειδοποιώντας τους όταν εντοπίσει όχημα ή οχήματα που αναζητούνται για οποιοδήποτε λόγο στην εκάστοτε περιοχή.

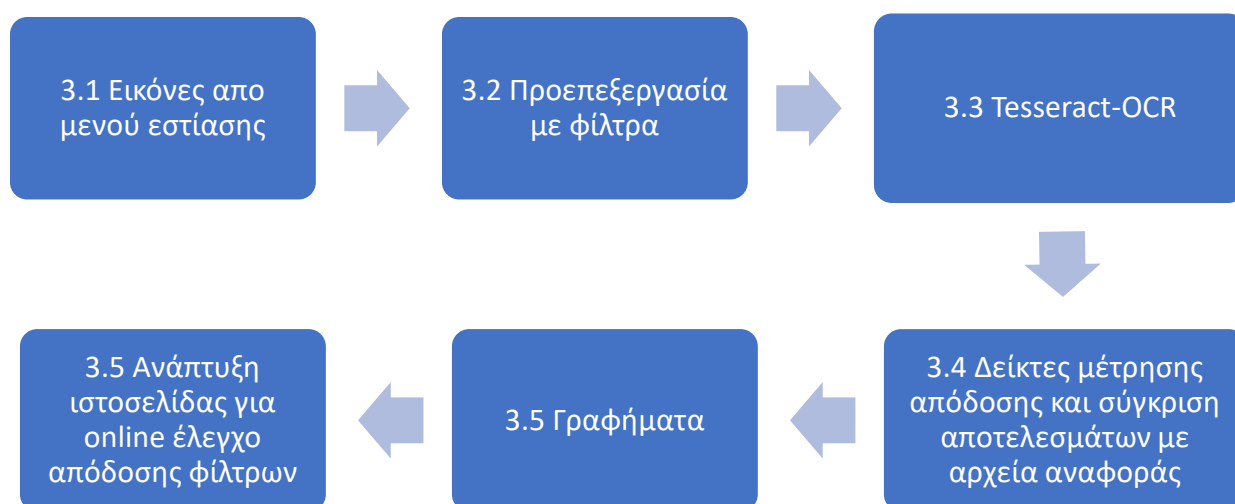


## Κεφάλαιο 3

### 3. Μεθοδολογία

Στην παρούσα πτυχιακή εργασία μελετάμε την απόδοση του Tesseract-OCR έπειτα από την εφαρμογή φίλτρων σε εικόνες και πιο συγκεκριμένα σε καταλόγους εστίασης. Έπειτα της εύρεσης απόδοσης, μελετάται η σύγκριση μεταξύ εικόνων και φίλτρων και παρουσιάζονται διαγράμματα καθώς και αποτελέσματα όπως βέλτιστη και χειρότερη απόδοση.

Η μεθοδολογία η οποία ακολουθήθηκε είναι η εξής:



Εικόνα 4 Flowchart μεθοδολογίας

Εργάστηκα ως ακολούθως, σύμφωνα με το παραπάνω διάγραμμα,

3.1) Αρχικά απαραίτητη ήταν η επιλογή του dataset δηλαδή οι εικόνες υπό εξέταση και επεξεργασίας καθώς και ένα αρχείο κειμένου στο οποίο είναι καταγεγραμμένο οτιδήποτε αναφέρεται στην εικόνα.

Premium		Grand	
Ανάμειξη διαφόρων ποικιλιών καφέ. Πλούσια γεύση & ουδέτερο άρωμα.		Ποικιλία καφέ Arabica 100%. Καθαρή γεύση καφέ & έντονο άρωμα.	
Espresso (ristretto, stretto, lungo, americano)	1,10€ / 1,40€	Espresso (ristretto, stretto, lungo, americano)	1,30€ / 1,60€
Freddo espresso	1,40€	Freddo espresso	1,60€
Cappuccino	1,40€ / 1,60€	Cappuccino	1,60€ / 1,90€
Freddo Cappuccino	1,60€	Freddo Cappuccino	1,90€
Latte cappuccino	1,50€	Macchiato	1,40€ / 1,70€
Macchiato	1,20€ / 1,50€	Mochaccino	2,50€
*ειρόπι καραμέλλα, φουντούκι, σοκολάτα +0,20€		**κρέμα +0,40€	
Καφές φίλτρου	1,50€	Φραπέ	1,20€
Ελληνικός καφές	1,20€ / 1,50€	Φυσικός χυμός πορτοκάλι	2,20€
Ελληνικός καφές *Μαστίχα Χίου	1,20€ / 1,50€	Αναψυκτικά	1,20€
Nescafe	1,20€	Νερό	0,40€

\*Για κάθε τηλεφωνική παραγγελία καφέ - δώρο 1 μπουκάλι νερό 500ml

Εικόνα 5 Παράδειγμα εικόνας dataset

3.2) Έπειτα ακολουθεί η προ-επεξεργασία δηλαδή η εφαρμογή των φίλτρων η οποία θα αναλυθεί περαιτέρω σε επόμενο κεφάλαιο.

3.3) Στη συνέχεια εισάγεται η προεπεξεργασμένη εικόνα στην Tesseract-OCR μηχανή.

3.4) Τα αποτελέσματα της μηχανής αυτής καταμετρούνται και συγκρίνονται με το αρχείο κειμένου που δόθηκε και υπολογίζονται κάποιοι δείκτες μέτρησης απόδοσης.

3.5) Έχοντας αυτά τα αποτελέσματα, είναι πλέον εφικτή η δημιουργία διαγραμμάτων για την παρουσίασή τους.

3.6) Τέλος, εμφανίζονται όλα τα διαγράμματα σε μία ειδικά σχεδιασμένη ιστοσελίδα, δίνοντας την δυνατότητα χρήσης της εφαρμογής από τον παγκόσμιο ιστό.

### 3.2 Προ-επεξεργασία

Το πιο σημαντικό στάδιο στην καταμέτρηση και βελτίωση της απόδοσης μιας μηχανής OCR είναι η διαδικασία της προεπεξεργασίας (image post-processing).

Στην εφαρμογή αυτή επιλέχθηκαν τα εξής 11 φίλτρα για την ολοκλήρωση της προαναφερθείσας διαδικασίας.

Συνομογραφία για διευκόλυνση στην αναφορά	Όνομα φίλτρου
Φ1	CLAHE (clahe)
Φ2	Thresholding OTSU (Otsu's thresholding)
Φ3	Opening (opening)
Φ4	Closing (closing)
Φ5	Gradient (gradient)
Φ6	Gaussian Blur (Gaussian blur)
Φ7	Median Blur (median blur)
Φ8	Equalize (equalize)
Φ9	Dilation (dilation)
Φ10	Erosion (erosion)
Φ11	Conv 2D Kernel (kernel)

## 3.2.1 Ανάλυση φίλτρων

### 3.2.1.1 Clahe( $\Phi 1$ )

Το φίλτρο clahe (Προσαρμοστική εξισορρόπηση ιστογράμματος περιορισμένης αντίθεσης) είναι μια παραλλαγή της προσαρμοστικής ισοστάθμισης ιστογράμματος (Adaptive histogram equalization - AHE), η οποία φροντίζει για την υπερ-ενίσχυση της αντίθεσης.

Το Clahe λειτουργεί σε μικρές περιοχές της εικόνας, που ονομάζονται πλακίδια, αντί για ολόκληρη την εικόνα. Στη συνέχεια, τα γειτονικά πλακίδια συνδυάζονται με διγραμμική παρεμβολή για την αφαίρεση των τεχνητών ορίων. Αυτός ο αλγόριθμος μπορεί να εφαρμοστεί για τη βελτίωση της αντίθεσης των εικόνων. Μπορούμε επίσης να εφαρμόσουμε τον CLAHE σε έγχρωμες εικόνες, όπου συνήθως εφαρμόζεται στο κανάλι φωτεινότητας και τα αποτελέσματα μετά την εξίσωση μόνο του καναλιού φωτεινότητας μιας εικόνας HSV είναι πολύ καλύτερα από την εξίσωση όλων των καναλιών της εικόνας BGR.



*Εικόνα 6 Γκρι εικόνα χωρίς φίλτρο (πρόσβαση 20-06-2021)*



*Εικόνα 7 Εικόνα με εφαρμογή φίλτρου Clahe(πρόσβαση 20-06-2021))*

### 3.2.1.2 Thresholding(Φ2)

Το thresholding (κατωφλίωση) είναι η απλούστερη μέθοδος τμηματοποίησης εικόνων. Από μια γκρι εικόνα, η κατωφλίωση μπορεί να χρησιμοποιηθεί για τη δημιουργία δυαδικών εικόνων.

Αυτή η μέθοδος αντικαθιστά κάθε εικονοστοιχείο από μια εικόνα με ένα μαύρο εικονοστοιχείο εάν η ένταση της εικόνας είναι μικρότερη από κάποια σταθερή σταθερά ή με ένα λευκό εικονοστοιχείο εάν η ένταση της εικόνας είναι μεγαλύτερη από την εν λόγω σταθερά. Στην εικόνα 8, αυτό έχει ως αποτέλεσμα το σκούρο δέντρο να γίνεται εντελώς μαύρο και το λευκό χιόνι να γίνεται εντελώς λευκό.



Εικόνα 8 Εικόνα μετά την εφαρμογή φίλτρου thresholding



Εικόνα 9 Εικόνα χωρίς φίλτρο

### 3.2.1.3 Opening(Φ3) και Closing(Φ4)

Το opening είναι μια μορφολογία όπου στα μαθηματικά είναι η διαστολή της διάβρωσης ενός συνόλου A από ένα δομικό στοιχείο B:

$$A \circ B = (A \ominus B) \oplus B$$

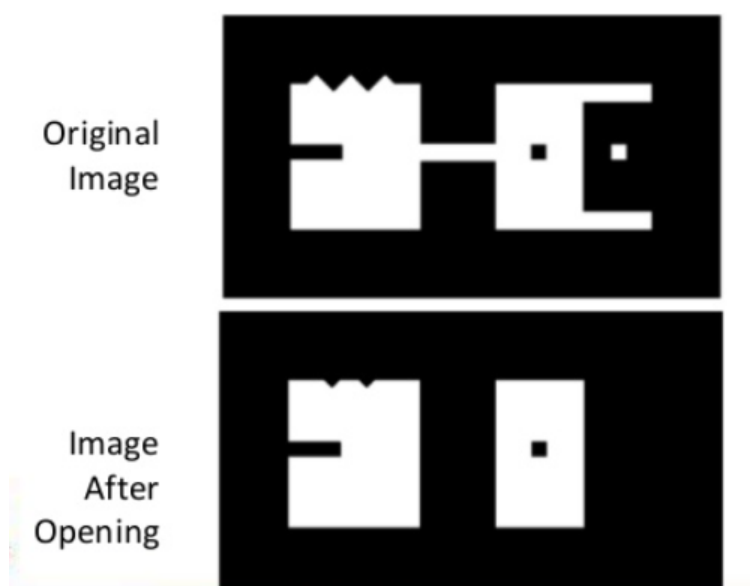
Το closing ενός συνόλου (δυαδική εικόνα) A από ένα δομικό στοιχείο B είναι η διάβρωση της διαστολής αυτού του συνόλου:

$$A \bullet B = (A \oplus B) \ominus B$$

Όπου το  $\ominus$  και το  $\oplus$  υποδηλώνουν τη διάβρωση και τη διαστολή αντίστοιχα.

Μαζί με το **closing**, το **opening** χρησιμεύει στο computer vision και την επεξεργασία εικόνας ως βασικό εργαλείο της μορφολογικής αφαίρεσης θορύβου.

Το **opening** αφαιρεί μικρά αντικείμενα από το προσκήνιο (που συνήθως λαμβάνονται ως τα φωτεινά εικονοστοιχεία) μιας εικόνας, τοποθετώντας τα στο φόντο, ενώ το **closing** αφαιρεί μικρές τρύπες στο προσκήνιο, μετατρέποντας μικρές νησίδες του φόντου σε προσκήνιο. Αυτές οι τεχνικές μπορούν επίσης να χρησιμοποιηθούν για την εύρεση συγκεκριμένων σχημάτων σε μια εικόνα.



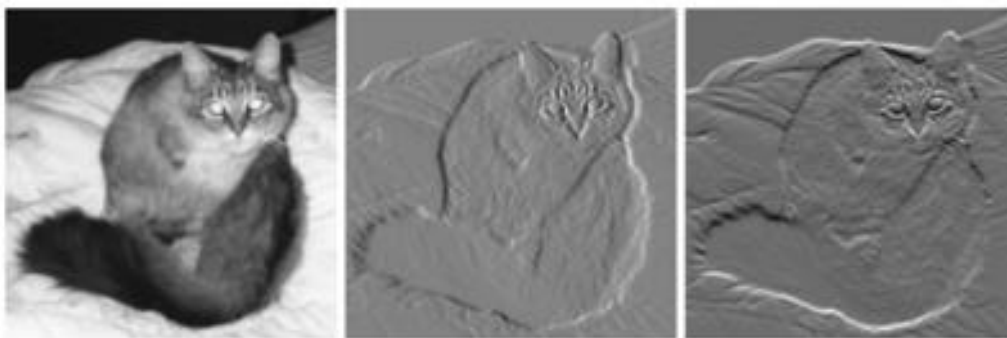
Εικόνα 10 Παράδειγμα φίλτρου opening. (πρόσβαση 20-06-2021)

### 3.2.1.4 Gradient(Φ5)

Μια κλίση εικόνας(Gradient) είναι μια αλλαγή κατεύθυνσης στην ένταση ή το χρώμα σε μια εικόνα. Η κλίση της εικόνας είναι ένα από τα θεμελιώδη δομικά στοιχεία της επεξεργασίας εικόνας.

Κάθε εικονοστοιχείο μιας εικόνας κλίσης μετρά την αλλαγή της έντασης του ίδιου σημείου της αρχικής εικόνας, προς μια δεδομένη κατεύθυνση. Για να ληφθεί το πλήρες εύρος της κατεύθυνσης, υπολογίζονται εικόνες κλίσης στις διευθύνσεις  $x$  και  $y$ .

Μια από τις πιο συνηθισμένες χρήσεις είναι η ανίχνευση ακμών. Μετά τον υπολογισμό των εικόνων κλίσης, τα εικονοστοιχεία με μεγάλες τιμές κλίσης γίνονται πιθανά εικονοστοιχεία ακμών. Τα εικονοστοιχεία με τις μεγαλύτερες τιμές κλίσης στην κατεύθυνση της κλίσης γίνονται εικονοστοιχεία ακμών και οι ακμές μπορούν να ανιχνευθούν στην κατεύθυνση κάθετα προς την κατεύθυνση της κλίσης.



Εικόνα 11 Παράδειγμα φίλτρου Gradient (πρόσβαση 20-06-2021)

Αριστερά, η πρωτότυπη γκρι εικόνα μιας γάτας.

Στο κέντρο, μια εικόνα κλίσης στην κατεύθυνση  $x$  που μετρά την οριζόντια μεταβολή της έντασης. Δεξιά, μια εικόνα κλίσης στην κατεύθυνση  $y$  που μετρά την κατακόρυφη μεταβολή της έντασης. Τα γκριζα εικονοστοιχεία έχουν μικρή κλίση- τα μαύρα ή λευκά εικονοστοιχεία έχουν μεγάλη κλίση.

### 3.2.1.5 Gaussian Blur(Φ6)

Στην επεξεργασία εικόνας, η γκαουσιανή θόλωση(Gaussian Blur) (επίσης γνωστή ως γκαουσιανή εξομάλυνση, Gaussian smoothing) είναι το αποτέλεσμα της θόλωσης μιας εικόνας με μια γκαουσιανή συνάρτηση (που πήρε το όνομά της από τον μαθηματικό και επιστήμονα Carl Friedrich Gauss).

Είναι ένα ευρέως χρησιμοποιούμενο εφέ στο λογισμικό γραφικών, συνήθως για τη μείωση του θορύβου της εικόνας και τη μείωση της λεπτομέρειας. Το οπτικό αποτέλεσμα αυτής της τεχνικής θόλωσης είναι ένα ομαλό θόλωμα που μοιάζει με αυτό της θέασης της εικόνας μέσα από μια ημιδιαφανή οθόνη, σαφώς διαφορετικό από το εφέ bokeh που παράγεται από έναν φακό εκτός εστίασης ή τη σκιά ενός αντικειμένου υπό συνήθη φωτισμό.

Η γκαουσιανή εξομάλυνση χρησιμοποιείται επίσης ως στάδιο προεπεξεργασίας σε αλγόριθμους υπολογιστικής όρασης προκειμένου να βελτιωθούν οι δομές της εικόνας σε διαφορετικές κλίμακες



*Εικόνα 12 Αριστερά η πρωτότυπη εικόνα, Δεξιά η εικόνα μετά την εφαρμογή του φίλτρου Gauss Blur.(πρόσβαση 20-06-2021)*



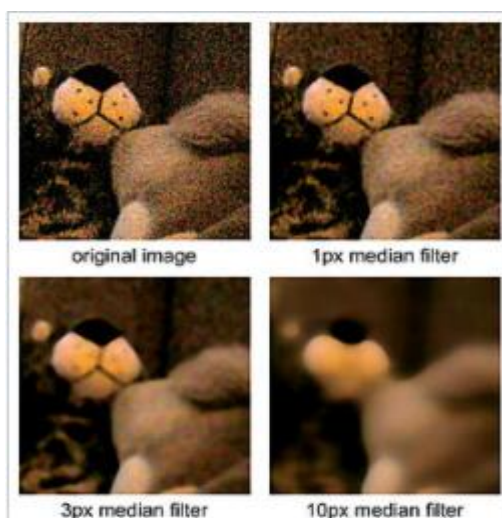
### 3.2.1.6 Median Blur(Φ7)

Το διάμεσο(median) φίλτρο είναι μια μη γραμμική τεχνική ψηφιακού φιλτραρίσματος, που χρησιμοποιείται συχνά για την απομάκρυνση του θορύβου από μια εικόνα ή ένα σήμα. Η εν λόγω μείωση του θορύβου αποτελεί τυπικό βήμα προεπεξεργασίας για τη βελτίωση των αποτελεσμάτων μεταγενέστερης επεξεργασίας (για παράδειγμα, ανίχνευση ακμών σε μια εικόνα).

Το φιλτράρισμα με διάμεσο χρησιμοποιείται πολύ ευρέως στην ψηφιακή επεξεργασία εικόνας, επειδή, υπό ορισμένες συνθήκες, διατηρεί τις ακμές ενώ αφαιρεί το θόρυβο (αλλά βλέπε τη συζήτηση παρακάτω), έχοντας επίσης εφαρμογές στην επεξεργασία σήματος.

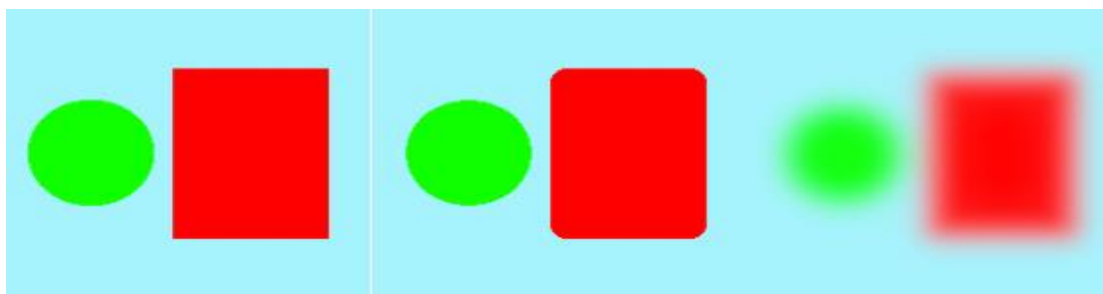
Η βασική ιδέα του φίλτρου είναι να διατρέξει το σήμα είσοδο προς είσοδο, αντικαθιστώντας κάθε είσοδο με τη διάμεσο των γειτονικών εισόδων. Το μοτίβο των γειτονικών σημείων ονομάζεται "παράθυρο", το οποίο ολισθαίνει, είσοδο προς είσοδο, σε ολόκληρο το σήμα.

Για μονοδιάστατα σήματα, το πιο προφανές παράθυρο είναι μόνο οι πρώτες λίγες προηγούμενες και επόμενες καταχωρήσεις, ενώ για δισδιάστατα (ή υψηλότερης διάστασης) δεδομένα το παράθυρο πρέπει να περιλαμβάνει όλες τις καταχωρήσεις εντός μιας δεδομένης ακτίνας ή ελλειψοειδούς περιοχής (δηλαδή το φίλτρο διάμεσης δεν είναι διαχωρίσιμο φίλτρο).



*Εικόνα 6 διάμεσων φίλτρων διαφορετικής ακτίνας που εφαρμόζονται στην ίδια θορυβώδη φωτογραφία. (πρόσβαση 20-06-2021)*

## Διαφορά Gauss και Median Blur



Εικόνα 14 Σύγκριση Gaussian και Median Blur (πρόσβαση 20-06-2021)

Ενώ το Gaussian blur υπολογίζει τη μέση τιμή των γειτονικών εικονοστοιχείων, το Median blur υπολογίζει τη διάμεσο

### 3.2.1.7 Equalize(Φ8)

Η ισοστάθμιση ιστογράμματος είναι μια μέθοδος στην επεξεργασία εικόνας για την προσαρμογή της αντίθεσης με χρήση του ιστογράμματος της εικόνας.

Η μέθοδος αυτή συνήθως αυξάνει τη συνολική αντίθεση πολλών εικόνων, ειδικά όταν τα χρήσιμα δεδομένα της εικόνας αντιπροσωπεύονται από κοντινές τιμές αντίθεσης. Μέσω αυτής της προσαρμογής, οι εντάσεις μπορούν να καταταξιολογηθούν καλύτερα στο ιστογράμμα. Αυτό επιτρέπει στις περιοχές με χαμηλότερη τοπική αντίθεση να αποκτήσουν υψηλότερη αντίθεση. Η εξίσωση ιστογράμματος το επιτυγχάνει αυτό με την αποτελεσματική εξάπλωση των πιο συχνών τιμών έντασης.

Η μέθοδος είναι χρήσιμη σε εικόνες με φόντο και προσκήνιο που είναι και τα δύο φωτεινά ή και τα δύο σκοτεινά. Ειδικότερα, η μέθοδος μπορεί να οδηγήσει σε καλύτερη προβολή της δομής των οστών σε εικόνες ακτίνων X και σε καλύτερη λεπτομέρεια σε φωτογραφίες που είναι υπερεκτεθειμένες ή υποεκτεθειμένες

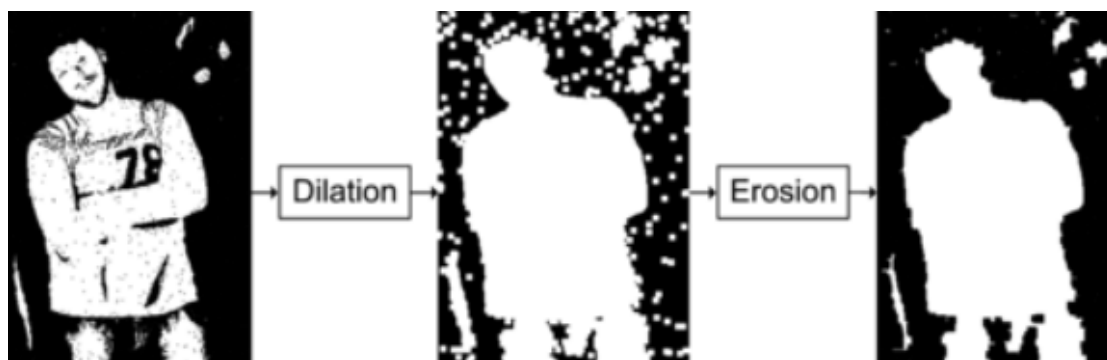


Εικόνα 157 Αριστερά πρωτότυπη εικόνα, δεξιά με το φίλτρο Equalize (πρόσβαση 20-06-2021)

### 3.2.1.8 Dilation( $\Phi_9$ ) και Erosion( $\Phi_{10}$ )

Η διαστολή(Dilation) και η διάβρωση(Erosion) είναι δύο θεμελιώδεις μορφολογικές πράξεις. Η διαστολή προσθέτει εικονοστοιχεία στα όρια των αντικειμένων μιας εικόνας, ενώ η διάβρωση αφαιρεί εικονοστοιχεία στα όρια των αντικειμένων.

Ο αριθμός των εικονοστοιχείων που προστίθενται ή αφαιρούνται από τα αντικείμενα μιας εικόνας εξαρτάται από το μέγεθος και το σχήμα του δομικού στοιχείου που χρησιμοποιείται για την επεξεργασία της εικόνας.



Εικόνα 16 Διαφορές φίλτρων Dilation και Erosion. (πρόσβαση 20-06-2021)

### 3.2.1.9 2D Convolution Kernel(Φ11)

Η συνέλιξη(Convolution) που περιλαμβάνει μονοδιάστατα σήματα αναφέρεται ως συνέλιξη 1D ή απλά συνέλιξη. Διαφορετικά, εάν η συνέλιξη εκτελείται μεταξύ δύο σημάτων που εκτείνονται κατά μήκος δύο αμοιβαία κάθετων διαστάσεων (δηλαδή εάν τα σήματα είναι διδιάστατα στη φύση τους), τότε θα αναφέρεται ως συνέλιξη 2D(Convolution 2D).

Η συνέλιξη είναι η διαδικασία πρόσθεσης κάθε στοιχείου της εικόνας στους τοπικούς του γείτονες, σταθμισμένους με τον πυρήνα. Αυτό σχετίζεται με μια μορφή μαθηματικής συνέλιξης. Η πράξη του πίνακα που εκτελείται - η συνέλιξη - δεν είναι ο παραδοσιακός πολλαπλασιασμός πινάκων, παρά το γεγονός ότι συμβολίζεται ομοίως με  $*$ .



*Εικόνα 17 Conv2dKernel φίλτρο  
(πρόσβαση 20-06-2021)*

### 3.3 Υπάρχουσες βιβλιοθήκες OCR

Κατά την έρευνα εύρεσης κατάλληλης μηχανής OCR για τις ανάγκες της εφαρμογής, μελετήθηκαν κάποιες ήδη υπάρχουσες βιβλιοθήκες και επιλέχθηκε μία από αυτές: Μελετήθηκαν οι εξής βιβλιοθήκες:

#### **ABBYY – OCR**

Η ABBYY (abbyy) είναι μια πολύ γνωστή εταιρεία που παρέχει υπηρεσίες αναγνώρισης εγγράφων, συλλογή δεδομένων και γλωσσολογικών λογισμικών. Χρησιμοποιώντας το SDK τους, οι προγραμματιστές μπορούν να αναπτύξουν εφαρμογές που βασίζονται στο OCR για κινητά τηλέφωνα.

Παρέχουν τα SDK για ακόλουθες κινητές πλατφόρμες:

- Android
- Windows Mobile
- Mobile Linux (Tizen)
- iPhone

Η μηχανή αυτή παίρνει ως είσοδο όλους τους γνωστούς τύπους αρχείων εικόνας όπως: .png, .gif, .jpg, κ.λπ., και μετατρέπει το κείμενο της εικόνας σε επεξεργάσιμο και αναζητήσιμο. Μπορούμε στη συνέχεια να χρησιμοποιήσουμε το εξαγόμενο κείμενο για περαιτέρω επεξεργασία και ανάπτυξη πολλών εφαρμογών

#### **Tesseract- OCR**

Το Tesseract-OCR (tesseract) είναι μια μηχανή ανοιχτού κώδικα γραμμένη σε C / C ++ και αναπτύχθηκε στα εργαστήρια της Hewlett-Packard (HP) μεταξύ 1985 και 1996. Ήταν μία από τις κορυφαία 3 μηχανές OCR στη δοκιμή ακρίβειας στο UNLV 1995. Το 2005, κυκλοφόρησε ως μηχανή ελεύθερου λογισμικού.

Από το 2007, η Google έχει λάβει την εποπτεία της μηχανής αναγνώρισης χαρακτήρων για την περαιτέρω ανάπτυξη και τη συντήρησή της. Παίρνει γκρι ή έγχρωμη εικόνα ως είσοδο και δίνει έξοδο σε μορφή κειμένου. Στην αρχή, μόνο .tiff τύπος εικόνων υποστηριζόταν αλλά τώρα υποστηρίζει επίσης άλλους τύπους εικόνων, όπως .png, .jpg,

κλπ. Μπορεί να διαβάσει τα δεδομένα σε οποιαδήποτε γλώσσα από την εικόνα όπως τα αγγλικά, σουηδικά, δανικά, κλπ και προγραμματιστές μπορούν να “εκπαιδεύσουν” τη δική τους γλώσσα, εάν η υποστήριξη για μια συγκεκριμένη γλώσσα δεν είναι διαθέσιμη.

Οι προγραμματιστές της Google δοκιμάζουν την λειτουργία του στο λειτουργικό σύστημα Ubuntu και Windows αλλά λειτουργεί επίσης και σε άλλες πλατφόρμες Linux και Mac, κλπ. Μπορούμε επίσης να χρησιμοποιήσουμε αυτή τη βιβλιοθήκη σε κινητές πλατφόρμες όπως το Android και iPhone κ.λπ.

## **Aspire-OCR**

Η Aspire-OCR (asprise ocr) είναι μια μηχανή οπτικής αναγνώρισης που παρέχει APIs για την γλώσσα προγραμματισμού Java, C, C ++, Delphi, CSharp.Net, Visual Basic.Net και Visual C ++. Net. Περιλαμβάνει λειτουργικότητα για Barcode ανάγνωση και για την ανάπτυξη εφαρμογών που εκμεταλλεύονται την μηχανή Aspire-OCR απαιτείται το δικό τους SDK έχοντας υποστήριξη για τις ακόλουθες πλατφόρμες.

- Windows
- Mac
- Linux
- Solaris

Παίρνει σχεδόν κάθε μορφή - τύπο εικόνας, όπως .png, .jpg, .bmp, κλπ σαν είσοδο και επιστρέφει ένα κείμενο εικόνας σε μορφή συμβολοσειράς.

## **Online OCR API**

Τα διαδικτυακά OCR είναι υπηρεσία που παρέχει APIs για OCR, δίνοντας την δυνατότητα να αναπτυχθούν διάφορα είδη εφαρμογών για έξυπνα κινητά τηλέφωνα ή και web-based εφαρμογές.. Υποστηρίζει .png, .jpg, .gif και .tiff τύπους εικόνων και μπορεί να διαβάσει το κείμενο από διάφορες γλώσσες όπως αγγλικά, γερμανικά, γαλλικά.

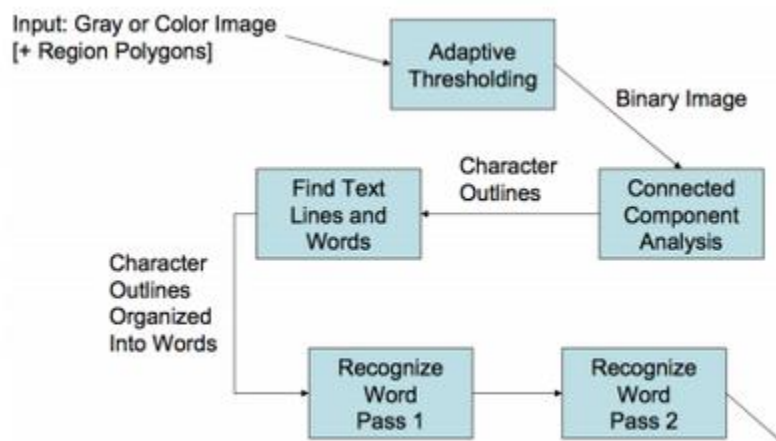
### 3.3.1 Κριτήρια επιλογής OCR

Υπάρχουν πολλοί πάροχοι μηχανών / υπηρεσίες cloud (νέφος) OCR, όπως είδαμε παραπάνω, προσφέροντας εύκολες προσεγγίσεις για την ανάπτυξη της εφαρμογής OCR σε πολλούς διαφορετικούς τομείς. Ωστόσο, οι περισσότερες από τις μηχανές OCR ή πάροχοι υπηρεσιών είναι εμπορικές και κλειστού κώδικα.

Για την παρούσα εργασία επιλέχθηκε η μηχανή αναγνώρισης χαρακτήρων tesseract-OCR. Διότι είναι ελεύθερο λογισμικό και δίνει την επιλογή στον προγραμματιστή της εφαρμογής να προσθέσει τη δική του γλώσσα σαν λεξικό.

#### 3.3.1.1 Tesseract OCR αρχιτεκτονική

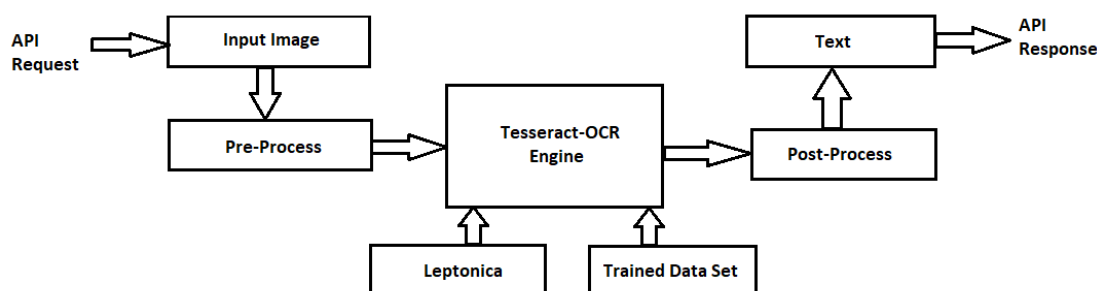
Η διαδικασία OCR του Tesseract ακολουθεί μια σταδιακή μεθοδολογία, όπως φαίνεται την Εικόνα 18.



Εικόνα 18 Tesseract OCR Αρχιτεκτονική από Ray (2007) (πρόσβαση: 20-06-2021)

Το Tesseract υποθέτει ότι η εικόνα εισόδου είναι μια πολυγωνική δυαδική περιοχή (polygonal binary area). Τα περιγράμματα του στοιχείου συγκεντρώνονται σε σταγόνες (Blops). Οι σταγόνες (Blops) τοποθετούνται σε γραμμές κειμένου. (Στην οπτική αναγνώριση εικόνας οι σταγόνες (Blops) αναφέρονται ως εκείνες οι περιοχές στην ψηφιακή εικόνα όπου εντοπίζονται να είναι διαφορετικές από τις γύρω περιοχές

σε σχέση με το χρώμα ή τη φωτεινότητα). Το Tesseract ανιχνεύει τις περιοχές κειμένου από τον αλγόριθμο εύρεσης γραμμής.



### OCR Process Flow

*Εικόνα 19 Διαδικασία ροής OCR (πρόσβαση 20-06-2021)*

### **Εύρεση γραμμής και λέξης**

Το Tesseract αρχικά φιλτράρει έξω σταγόνες που είναι μικρότερες από κάποιο κλάσμα του διάμεσου ύψους, το οποίο θα μπορούσε ενδεχομένως να είναι σημεία στίξης ή θόρυβος. Στη συνέχεια το Tesseract επεξεργάζεται τις φιλτραρισμένες σταγόνες σε οριζόντια θέση και τις εκχωρεί σε μια γραμμή κειμένου. Αφού οι σταγόνες έχουν εκχωρηθεί, οι φιλτραρισμένες σταγόνες θα πρέπει να τοποθετηθούν και πάλι στην κατάλληλη θέση. Μόλις οι γραμμές κειμένου έχουν εντοπιστεί, οι βασικές γραμμές τοποθετούνται με μεγαλύτερη ακρίβεια, χρησιμοποιώντας μια τετραγωνική spline. Οι γραμμές βάσης τοποθετούνται με τον διαχωρισμό των σταγόνων σε ομάδες με λογική συνεχή μετατόπιση για την αρχική ευθεία γραμμή βάσης. Η τετραγωνική spline κρατάει το αποτέλεσμα του υπολογισμού σε ένα σταθερό επίπεδο. Ωστόσο, οι πολλαπλές αυλακώσεις μπορούν να οδηγήσουν σε ασυνέχειες με αυτό να είναι ένα μειονέκτημα. Όπως φαίνεται στην Εικόνα 10, το Tesseract χρησιμοποιεί τη γραμμή βάσης (πράσινη), τη γραμμή κάτω αρχής (μπλε), τη μέση γραμμή (ροζ) και τη γραμμή ανάβασης (κυανό) για να εντοπίσει τη γραμμή του κειμένου.

*Εικόνα 20 Ένα παράδειγμα όπου η βασική γραμμή καμπυλώνει από Ray (πρόσβαση: 20-06-2021)*



Αφού η περιοχή κειμένου έχει ανιχνευθεί, το Tesseract πραγματοποιεί τμηματοποίηση της λέξης σε κομμάτια από χαρακτήρες αναλύοντας αν είναι ισοδύναμου πλάτους, όπως φαίνεται στην Εικόνα 20.



Εικόνα 21 Μια τεμαχισμένη λέξη σταθερού πλάτους από Ray(πρόσβαση 20-06-2021)

Όταν το κείμενο με σταθερό πλάτους βρεθεί, η διαδικασία απομόνωσης της λέξης σε γράμματα σταματάει και ξεκινάει η διαδικασία της αναγνώρισης των λέξεων

### Αναγνώριση λέξης

Το Tesseract βελτιώνει το μη ικανοποιητικό αποτέλεσμα τεμαχίζοντας τις σταγόνες(blobs). Προσπαθεί να βρει σημεία υποψήφια κοπής από κοίλες κορυφές μιας πολυγωνικής προσέγγισης του περιγράμματος. Τουλάχιστον τρία ζεύγη σημείων κοπής διασφαλίζουν ότι ένας ξεχωριστός χαρακτήρας είναι από το σύνολο ASCII . Η Εικόνα 7 δείχνει πώς το Tesseract αναγνωρίζει τον χαρακτήρα όταν ο χαρακτήρας «r» αγγίζει το χαρακτήρα «m».



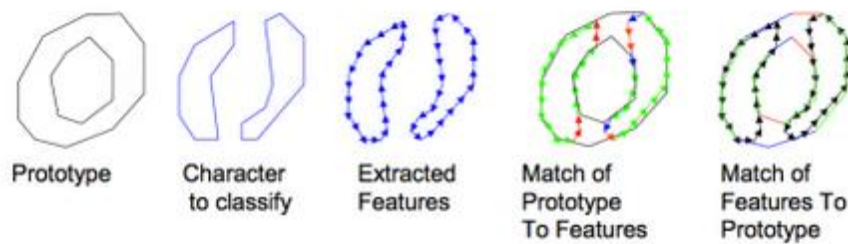
Εικόνα 22 Σημεία υποψήφια κοπής από Ray(πρόσβαση 20-06-2021)

Αν τα σημεία κοπής έχουν όλα χρησιμοποιηθεί και το αποτέλεσμα εξακολουθεί να μην είναι ικανοποιητικό, η διαδικασία της ανάλυσης θα πρέπει να αναληφθεί από τον associator. Ο associator πραγματοποιεί την πρώτη καλύτερη αναζήτηση και κάνει το γράφημα σε υποψήφιους χαρακτήρες. Η Εικόνα 23 δείχνει ένα παράδειγμα του “σπασμένου” χαρακτήρα.



Εικόνα 23 Μια “σπασμένη” λέξη έτοιμη για να αναγνωριστεί Ray(πρόσβαση 20-06-2021)

Τα χαρακτηριστικά του “σπασμένου” χαρακτήρα εξάγονται από το τεμαχισμό του περιγράμματος από το στατικό ταξινομητή.



Εικόνα 24 Χαρακτηριστικά γνωρίσματα και αντιστοίχιση ταίριασμα Ray (πρόσβαση 20-06-2021)

### Δεδομένα εκπαίδευσης (Training Data)

Για να βελτιώσουν την ακρίβεια και την επέκταση του περιορισμού της γλώσσας, η μηχανή Tesseract έχει σχεδιαστεί για να είναι πλήρως εκπαιδευσιμη. Μετά την τμηματοποίηση του χαρακτήρα, μια διαδικασία δύο φάσεων πραγματοποιείται. Στην πρώτη φάση, γίνεται μια προσπάθεια να αναγνωριστεί κάθε μια λέξη τη φορά . Ο ικανοποιητικός χαρακτήρας περνά στον δυναμικό ταξινομητή ως δεδομένα εκπαίδευσης(training data). Στην δεύτερη φάση, οι λέξεις που δεν είχαν αναγνωριστεί ικανοποιητικά αναγνωρίζονται και πάλι. Οι χρήστες μπορούν να εκπαιδεύσουν το Tesseract βελτιώνοντας την ακρίβεια της αναγνώρισης για τις δικές τους περιπτώσεις.

Για την παρούσα εργασία επειδή οι χαρακτήρες είναι τυπωμένοι και όχι χειρόγραφοι δεν χρειάστηκε να γίνει εκπαίδευση των δεδομένων περαιτέρω από την ήδη υπάρχουσα και πραγματοποιημένη από την Google.

### 3.4 Μέτρηση ποιότητας OCR

Η έξοδος των μηχανών OCR συχνά περιέχει αρκετά λάθη, όπως ορθογραφικά λάθη ή ψευδείς χαρακτήρες. Προκειμένου να ληφθεί ένα μέτρο ανεξάρτητο από το μέγεθος του κειμένου, ο αριθμός των λαθών κανονικοποιείται συνήθως στο μήκος του αναμενόμενου περιεχομένου (το κείμενο της αλήθειας εδάφους). Το πηλίκο μεταξύ του αριθμού των λαθών και του μήκους του κειμένου είναι γνωστό ως ποσοστό σφάλματος.

Αυτό το ποσοστό σφάλματος συνήθως μετριέται σε δύο διαφορετικά επίπεδα:

- Ποσοστό σφάλματος χαρακτήρα (Character error rate , **CER**)
- Ποσοστό σφάλματος λέξης (Word error rate , **WER**)

Αξίζει να σημειωθεί ότι τα ποσοστά σφάλματος **λέξεων** είναι συνήθως υψηλότερα από τα ποσοστά σφάλματος **χαρακτήρων**. Για παράδειγμα, ένας σχετικά μικρός αριθμός λανθασμένων χαρακτήρων, όπως ένα ποσοστό σφάλματος 10%, σημαίνει ότι περίπου το ήμισυ των λέξεων με έξι γράμματα θα περιέχει τουλάχιστον ένα λάθος χαρακτήρα (για να είμαστε πιο ακριβείς, περίπου το 47% των λέξεων 6 γραμμάτων θα είναι λάθος με την υπόθεση ότι η κατανομή σφαλμάτων μεταξύ χαρακτήρων είναι ομοιογενής).

Ωστόσο, μπορεί να αναμένεται καλή συσχέτιση μεταξύ των δύο μετρήσεων (εκτός, ίσως, για πολύ υψηλά ποσοστά σφάλματος) όταν και τα δύο μέτρα χρησιμοποιούνται για τη σύγκριση διαφορετικών μηχανών OCR.

Τα πειράματα δείχνουν ότι οι άνθρωποι είναι σχετικά ανεκτικοί σε σφάλματα με την ακόλουθη έννοια: μια ακρίβεια λέξεων κάτω από το 85% οδηγεί σε χαμηλότερη παραγωγικότητα εάν εφαρμοστεί χειροκίνητη διόρθωση στην έξοδο σε σύγκριση με εκείνη της πληκτρολόγησης εξ ολοκλήρου από το μηδέν. Για παράδειγμα, υποστηρίχθηκε ότι η **καλή ακρίβεια** OCR σημαίνει περίπου **98-99%** ακριβή, ενώ η ακρίβεια κάτω από το **90%** είναι **κακή**.

Φυσικά, τα ποσοστά σφάλματος σχετίζονται με την αναμενόμενη έξοδο και, επομένως, η εκτίμησή τους απαιτεί κείμενο αναφοράς. Αυτή η αναφορά είναι συνήθως ένα ψηφιακό έγγραφο που έχει παραχθεί με το χέρι για να αντιπροσωπεύει, όσο το δυνατόν ακριβέστερα, το περιεχόμενο κειμένου της αρχικής πηγής.

Τέτοιες μεταγραφές συχνά ονομάζονται **αλήθεια εδάφους, ground text**, (ένας όρος δανεισμένος από χαρτογραφία), καθώς αντιστοιχούν την πηγή στο πραγματικό περιεχόμενο κειμένου. Παρόλο που το πραγματικό περιεχόμενο φαίνεται να είναι μια σαφής έννοια, η δημιουργία του πρέπει συχνά να αντιμετωπίζει κάποιες λεπτές αποφάσεις. Για παράδειγμα, οι σύνδεσμοι πρέπει να αντιπροσωπεύονται ως ένας μοναδικός χαρακτήρας Unicode ή ως δύο ανεξάρτητοι χαρακτήρες;

Τα θολά και παραμορφωμένα γράμματα θα μεταγραφούν ως αυτά που προσποιούνται ότι είναι ή μοιάζουν πραγματικά; Φυσικά, δεν υπάρχει συγκεκριμένη απάντηση σε τέτοιου είδους ερωτήσεις, καθώς μόνο η εικόνα προέλευσης περιέχει όλες τις πληροφορίες και μια σχολαστική δημιουργία ground text μπορεί να απαιτήσει μια προσπάθεια για να αξίζει.

### 3.4.1 Υπολογισμός Ποσοστών Σφαλμάτων

Όσο αυξάνεται ο αριθμός των σφαλμάτων τόσο δυσκολότερος γίνεται ο υπολογισμός αυτών, για παράδειγμα θα μπορούσε κανείς να υποστηρίξει ότι το ποσοστό σφάλματος χαρακτήρων μεταξύ "ernest" (αναφορά) και "nester" (έξοδος) είναι 6 αφού κανένας χαρακτήρας στην αναφορά δεν εμφανίζεται στο "nester" στην απαιτούμενη θέση.

Ωστόσο, δύο πανομοιότυπες ακολουθίες "nest" μπορούν να ευθυγραμμιστούν και στις δύο λέξεις. Μια ακολουθία μιας λέξης είναι οποιαδήποτε συμβολοσειρά χαρακτήρων που μπορεί να ληφθεί απλώς αφαιρώντας μερικούς από τους πρωτότυπους (δεν επιτρέπεται η ταξινόμηση αφού η σειρά των γραμμάτων είναι απαραίτητη για τη σωστή ορθογραφία). Για παράδειγμα, το "nest" είναι το αποτέλεσμα όταν αφαιρούνται οι δύο αρχικοί χαρακτήρες στο "ernest" και το "rnst" είναι η ακολουθία που περιέχει μόνο σύμφωνα με την αρχική λέξη.

Είναι σαφές ότι υπάρχει μεγαλύτερη ομοιότητα μεταξύ αυτού του ζεύγους λέξεων από αυτό που χαρακτηρίζεται από το 100% CER: πράγματι αρκεί η αφαίρεση 2 επιπλέον χαρακτήρες ("er") στην αρχή της πρώτης λέξης και στη συνέχεια η εισαγωγή των δύο τελικών χαρακτήρων ("er") για την λήψη της δεύτερης λέξης. Με αυτές τις 4 λειτουργίες, το CER(Computing error rate) γίνεται περίπου 67%. Αντίθετα, η ευθυγράμμιση και των δύο "er" ακολουθιών θα απαιτούσε 8 λειτουργίες και, ως εκ τούτου, θα οδηγούσε σε CER 133%.

Επομένως, το CER υπολογίζεται με τον ελάχιστο αριθμό λειτουργιών που απαιτούνται για τη μετατροπή του κειμένου αναφοράς στην έξοδο (ένας αριθμός που είναι γνωστός ως απόσταση **Levenshtein**. Ουσιαστικά, όσο μεγαλύτερος είναι ο αριθμός, τόσο πιο διαφορετικά και τα δύο κείμενα είναι). Ωστόσο, η εύρεση του βέλτιστου συνόλου βασικών λειτουργιών δεν είναι ασήμαντη, τουλάχιστον για μεγάλες και θορυβώδεις εξόδους.

Ορισμός του Word Error Rate (WER) :

$$WER = (i_w + s_w + d_w) / n_w$$

Όπου:  $n_w$  ο αριθμός των λέξεων στο κείμενο αναφοράς,

$s_w$  ο αριθμός των λέξεων που αντικαθίστανται,

$d_w$  ο αριθμός των λέξεων που διαγράφηκαν,

$I_w$  ο αριθμός των λέξεων που εισάχθηκαν.

Παρόμοια ορίζεται και το Character Error Rate (CER) :

$$CER = (i + s + d) / n$$

Αλλά χρησιμοποιώντας τον συνολικό αριθμό χαρακτήρων (  $n$  ) και τον ελάχιστο αριθμό λέξεων εισαγωγής (  $i$  ), διαγραφής (  $d$  ), αντικατάστασης (  $d$  ).

## **Κανονικοποίηση**

Δεδομένου ότι ο αριθμός των σφαλμάτων μπορεί να είναι μεγαλύτερος από το μήκος του κειμένου αναφοράς και να οδηγεί σε rate μεγαλύτερο από 100% (για παράδειγμα, ένα έγγραφο αλήθειας εδάφους με 100 χαρακτήρες και μια έξοδο OCR που περιέχει 120 λάθος χαρακτήρες δίνουν ένα σφάλμα 120% και αρνητική ακρίβεια), μερικές φορές ο αριθμός των λαθών διαιρείται με το άθροισμα (  $i + s + d + c$  ) του αριθμού των λειτουργιών επεξεργασίας (  $i + s + d$  ) και του αριθμού  $c$  των σωστών συμβόλων, που είναι πάντα μεγαλύτερο από τον αριθμητή.

## White spaces

Τα κενά σε ένα κείμενο παίζουν σημαντικό ρόλο καθώς διαχωρίζουν τις λέξεις μεταξύ τους και γι'αυτό πρέπει να λαμβάνεται υπόψη κατά τον υπολογισμό του ποσοστού σφάλματος. Για την κατανόηση του κειμένου δεν είναι τόσο σχετικό το μήκος του κενού, επομένως τα συνεχόμενα κενά θεωρούνται συχνά ισοδύναμα. Έτσι, το CER μεταξύ “were wolf” ( με δύο κενά μεταξύ τους) με του “werewolf” είναι 0.125.

Αξίζει ωστόσο να σημειωθεί ότι τα κενά δεν είναι χαρακτήρας όπως οι άλλοι: όταν η ευθυγράμμιση μεταξύ κειμένων επιτρέπει την αντικατάσταση του κενού με έναν εκτυπώσιμο χαρακτήρα, οδηγεί μερικές φορές σε ανεπιθύμητα αποτελέσματα. Για παράδειγμα, κατά την σύγκριση του “bad man” και του “batman” η διαγραφή του d και η αντικατάσταση του κενού με το t έχει ίδιο κόστος Levenshtein(δύο πράξεις) από την διαγραφή του κενού συν την αντικατάσταση του d με το t (επίσης δύο πράξεις). Ωστόσο η δεύτερη επιλογή είναι πιο φυσική. Αυτή η συμπεριφορά μπορεί να ελαχιστοποιηθεί εάν δεν επιτρέπεται η αντικατάσταση του κενού με χαρακτήρες.

## Κεφάλαιο 4

### 4.1 Γλώσσα προγραμματισμού Python 3.7.6

Για την ανάπτυξη της εφαρμογής χρησιμοποιήθηκε γλώσσα προγραμματισμού Python διότι είναι ευρέως διαδεδομένη καθώς και έχει πολλές χρήσιμες βιβλιοθήκες ανοικτού κώδικα.

Η Python (Python) είναι μια διερμηνευμένη γλώσσα προγραμματισμού υψηλού επιπέδου γενικού σκοπού. Η σχεδιαστική φιλοσοφία της Python δίνει έμφαση στην αναγνωσιμότητα του κώδικα με την αξιοσημείωτη χρήση σημαντικών εσοχών. Οι γλωσσικές της δομές καθώς και η αντικειμενοστραφής προσέγγισή της έχουν ως στόχο να βοηθήσουν τους προγραμματιστές να γράψουν σαφή, λογικό κώδικα για μικρά και μεγάλα έργα.

#### 4.1.1 Βιβλιοθήκες

Η κύρια βιβλιοθήκη που χρησιμοποιήθηκε είναι η pytesseract (pyTesseract) αφού είναι ο πυρήνας της εφαρμογής.

#### Python-Tesseract

Το **python-tesseract** είναι ένα εργαλείο οπτικής αναγνώρισης χαρακτήρων (OCR) για την python. Δηλαδή, θα αναγνωρίζει και θα "διαβάζει" το κείμενο που είναι ενσωματωμένο σε εικόνες.

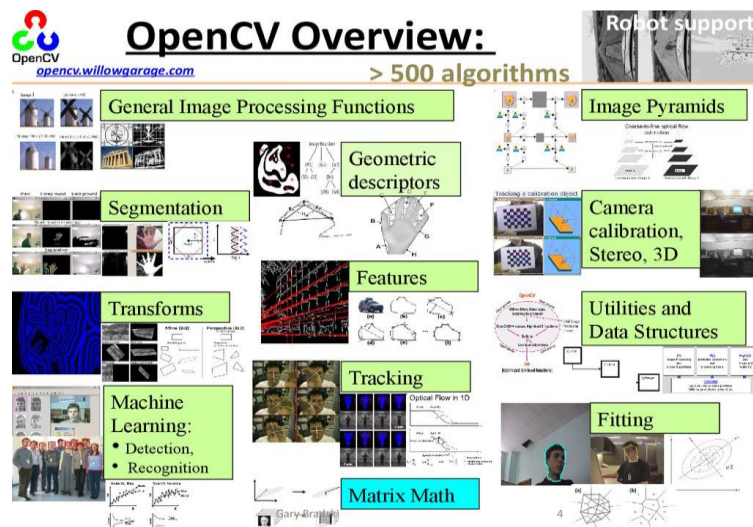
Το python-tesseract είναι ένα περιτύλιγμα για την Tesseract-OCR Engine της Google. Είναι επίσης χρήσιμο ως αυτόνομο σενάριο κλήσης του tesseract, καθώς μπορεί να διαβάσει όλους τους τύπους εικόνων που υποστηρίζονται από τις βιβλιοθήκες απεικόνισης Pillow και Leptonica, συμπεριλαμβανομένων των jpeg, png, gif, bmp, tiff και άλλων. Επιπλέον, αν χρησιμοποιηθεί ως δέσμη ενεργειών, το Python-tesseract θα εκτυπώσει το αναγνωρισμένο κείμενο αντί να το γράψει σε αρχείο.

## OpenCV

Η OpenCV (Open Source Computer Vision) βιβλιοθήκη είναι μια βιβλιοθήκη η οποία χρησιμοποιείται σε εργασίες που απαιτούν αποτελεσματική και σε πραγματικό χρόνο επεξεργασία εικόνας, αποτελεί συναρτήσεις για υλοποίηση μηχανικής όρασης σε πραγματικό χρόνο.

Η OpenCV διατίθεται στο κοινό υπό την άδεια BSD, είναι ελεύθερο λογισμικό τόσο για ακαδημαϊκή όσο και για εμπορική χρήση. Έχει υλοποιηθεί με την χρήση των γλωσσών προγραμματισμού όπως C ++, C, και Python και σύντομα θα αποκτήσει και Java διεπαφή δίνοντας την δυνατότητα στους χρήστες να την τρέχουν στο αγαπημένο τους λογισμικό, Windows, Linux, Android ή ακόμα και macOS.

Η βιβλιοθήκη έχει περισσότερους από 2500 βελτιστοποιημένους αλγόριθμους. Χρησιμοποιείται ευρέως σε όλο τον κόσμο με πάνω από 2.5 εκατομμύρια κατεβάσματα και περισσότερους από 40 χιλιάδες ανθρώπους και ομάδες χρηστών. Χρησιμοποιείται σήμερα από τις διαδραστικές τέχνες, μέχρι και την προηγμένη ρομποτική.



Εικόνα 24 Επισκόπηση των λειτουργιών της OpenCV(20-06-2021)



## **Flask**

Η βιβλιοθήκη αυτή χρησιμοποιήθηκε για το σερβίρισμα της εφαρμογής στον ιστό. Δηλαδή μέσω αυτής μπορεί η εφαρμογή να χρησιμοποιηθεί μέσω browser σε html σελίδες καθώς είναι μορφοποιημένες με CSS.

Το Flask (Flask) είναι web framework γραμμένο σε Python. Ταξινομείται ως microframework επειδή δεν απαιτεί συγκεκριμένα εργαλεία ή βιβλιοθήκες. Δεν έχει επίπεδο αφαίρεσης βάσεων δεδομένων, επικύρωση φόρμας ή άλλα στοιχεία όπου προϋπάρχουσες βιβλιοθήκες τρίτων παρέχουν κοινές λειτουργίες. Ωστόσο, το Flask υποστηρίζει επεκτάσεις που μπορούν να προσθέσουν χαρακτηριστικά εφαρμογών σαν να είχαν υλοποιηθεί στο ίδιο το Flask. Επεκτάσεις υπάρχουν για χαρτογράφους αντικειμένου-σχεσιακής απεικόνισης, επικύρωση φόρμας, χειρισμό μεταφόρτωσης, διάφορες ανοικτές τεχνολογίες αυθεντικοποίησης και διάφορα κοινά εργαλεία που σχετίζονται με το πλαίσιο.

Υποστηρίζει Python έκδοση 3.6 και νεότερες.

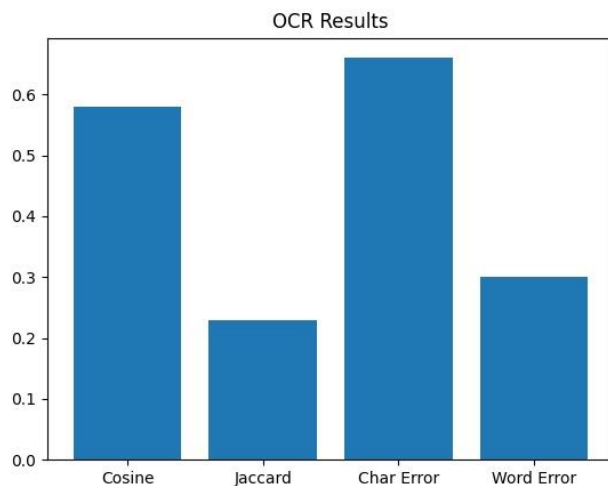
## **Imageio**

Το Imageio (Imageio) είναι μια βιβλιοθήκη Python που παρέχει μια εύκολη διεπαφή για την ανάγνωση και εγγραφή ενός ευρέος φάσματος δεδομένων εικόνας, συμπεριλαμβανομένων κινούμενων εικόνων, ογκομετρικών δεδομένων και επιστημονικών μορφών. Είναι cross-platform, τρέχει σε Python 3.5+ .

## **Matplotlib**

Η Matplotlib (Matplotlib) είναι μια βιβλιοθήκη γραφικών παραστάσεων για τη γλώσσα προγραμματισμού Python και την αριθμητική μαθηματική επέκταση NumPy (NumPy). Παρέχει ένα αντικειμενοστραφές API (API) για την ενσωμάτωση γραφικών παραστάσεων σε εφαρμογές που χρησιμοποιούν εργαλείοθκες GUI γενικού σκοπού.

Χρησιμοποιήθηκε για την δημιουργία γραφικών παραστάσεων και την αναπαράσταση της απόδοσης του OCR.



Εικόνα 24 Διάγραμμα της Matplotlib

## Jiwer

Αυτή η βιβλιοθήκη (Jiwer) υπολογίζει την ελάχιστη απόσταση επεξεργασίας μεταξύ της πρότασης αλήθειας και της πρότασης υπόθεσης. Η απόσταση ελάχιστης επεξεργασίας υπολογίζεται με τη χρήση της μονάδας python C python-Levenshtein. Γενικότερη χρήση της για τον υπολογισμό μετρικών αποστάσεων.

## Numpy

Η NumPy (NumPy) είναι μια βιβλιοθήκη για τη γλώσσα προγραμματισμού Python, η οποία προσθέτει υποστήριξη για μεγάλους, πολυδιάστατους πίνακες και πίνακες, μαζί με μια μεγάλη συλλογή μαθηματικών συναρτήσεων υψηλού επιπέδου για τη λειτουργία αυτών των πινάκων. Ο πρόγονος της NumPy, η Numeric, δημιουργήθηκε αρχικά από τον Jim Hugunin με τη συμβολή πολλών άλλων προγραμματιστών. Το 2005, ο Travis Oliphant δημιούργησε το NumPy ενσωματώνοντας χαρακτηριστικά του ανταγωνιστικού Numarray στο Numeric, με εκτεταμένες τροποποιήσεις. Το NumPy είναι λογισμικό ανοικτού κώδικα και έχει πολλούς συνεισφέροντες.

## 4.2 Χρήση Flask με python

Κομμάτι της παρούσας πτυχιακής εργασίας ήταν η χρήση της εφαρμογής που αναπτύχθηκε σε διαδικτυακά περιβάλλοντα απο φυλλομετρητές (Browsers) δίνοντας την δυνατότητα της πρόσβασης μέσω του παγκόσμιου ιστού.

Η χρήση της HTML (HyperText Markup Language) και CSS (Cascading Style Sheets) ήταν απαραίτητη για την παραπάνω επίτευξη αφού η βιβλιοθήκη flask είναι υπεύθυνη για την σύνδεση της εφαρμογής, γραμμένης σε python, με την ιστοσελίδα χρησιμοποιώντας την html και css.

Η Flask χρησιμοποιεί έναν ιδιότροπο σχεδιασμό του project ο οποίος απαιτεί μία πολύ συγκεκριμένη διάταξη στους φακέλους που περιέχουν οτιδήποτε αρχεία χρειαστούν.

Η παρακάτω εικόνα περιγράφει αυτή την διάταξη

```
project directory/  
  templates/  
    home2.html  
    results.html  
    multipleFilters.html  
    results_many_filters.html  
    results_many_imgs.html  
  static/  
    bestPerfImg/  
    emptyPlot/  
    imagesGen/  
    ocr_text/  
    plots_gen/  
    worstPerfImg/  
    main.css  
  uploads/  
  ocr_main.py  
  ocr_All_filters_600DPI_Tesseract.py  
  serOcrEval.py  
  serPPF.py  
  pytesseract.py
```

Εικόνα 25 Απαραίτητη μορφή διάταξης για την ολοκλήρωση της πτυχιακής εργασίας

Η παρούσα πτυχιακή χωρίζεται σε 3 υποφακέλους: **'templates'**, **'static'**, **'uploads'** καθώς και το root directory έχοντας το ocr\_main.py με τον βασικό κώδικα και καρδιά του προγράμματος ,τα 3 .py αρχεία (ocr\_All\_filters\_600DPI\_Tesseract.py, serOcrEval.py, serPPF.py) με κώδικα απαιτούμενο για την χρήση του προγράμματος και το pytesseract.py την μηχανή OCR που χρησιμοποιείται για όλη την αναμενόμενη λειτουργία.

Μερική Ανάλυση:

- Ο **'templates'** φάκελος εμπεριέχει όλα τα απαραίτητα html αρχεία, τα οποία είναι τα:
  - **home2.html**

Η βασικότερη σελίδα που φορτώνεται πρώτη κατά την εκτέλεση του προγράμματος.  
Επιτρέπει ανέβασμα πολλαπλών εικονών και .txt αρχείων και επιλογή ενός φίλτρου.
  - **Results.html**

Σελίδα αποτελεσμάτων μιας μόνο εικόνας και φίλτρου της home2.html προβάλλοντας την αρχική εικόνα, την εικόνα αφού εφαρμοστεί το επιλεγμένο φίλτρο εμφανίζεται με περίγραμμα κάθε γράμματος που εντοπίστηκε από την μηχανή, το κείμενο που εισάχθηκε (ground text), το κείμενο-αποτέλεσμα της μηχανής OCR και τέλος το διάγραμμα απόδοσης έχοντας τις μετρικές τιμες cosine, jaccard, character error και word error.
  - **multipleFilters.html**

Σελίδα με δυνατότητα επιλογής πολλαπλών φίλτρων σε μία μόνο εικόνα.

- **Results\_many\_filters.html**

Τα αποτελέσματα της παραπάνω σελίδας εμφανίζοντας την αρχικής εικόνα και ένα διάγραμμα με την απόδοση όλων των επιλεγμένων φίλτρων.

- **Results\_many\_imgs.html**

Σελίδα με τα αποτελέσματα έπειτα από ανέβασμα πολλαπλών εικόνων. Εμφανίζονται τέσσερα διαγράμματα με την απόδοση κάθε εικόνας μετά την εφαρμογή του επιλεγμένου φίλτρου καθώς και η εικόνα με την καλύτερη και την χειρότερη απόδοση.

Συνέχεια ανάλυσης:

- Ο **'static'** φάκελος περιέχει όλα τα στατικά αρχεία που δημιουργούνται από το πρόγραμμα, δηλαδή όλα αυτά που θα εμφανιστούν από τα `.html` αρχεία.

Υποφάκελοι:

- **bestPerfImg**: Εδώ αποθηκεύονται η εικόνα με την καλύτερη απόδοση
- **emptyPlot**: Υπάρχει ένα default άδειο διάγραμμα για την περίπτωση που η εικόνα είναι μη αναγνωρίσιμη από την μηχανή.
- **imagesGen**: Φάκελος με όλες τις επεξεργασμένες εικόνες έπειτα από εφαρμογή του επιθυμητού φίλτρου
- **ocr\_text**: Στον φάκελο αυτόν αποθηκεύονται τα αποτελέσματα του OCR σε μορφή κειμένου `.txt`
- **plots\_gen**: Γίνεται η αποθήκευση όλων των παραγόμενων διαγραμμάτων
- **worfPerfImg**: Εδώ αποθηκεύονται η εικόνα με την χειρότερη απόδοση

- `main.css`: Είναι το βασικό αρχείο `.css` για την σωστή εμφάνιση των αποτελεσμάτων σε σελίδες που χρησιμοποιούν Bootstrap 4.
  
- Στον **'uploads'** φάκελο αποθηκεύονται όλες οι εικόνες που θα ανεβάσει ο χρήστης.

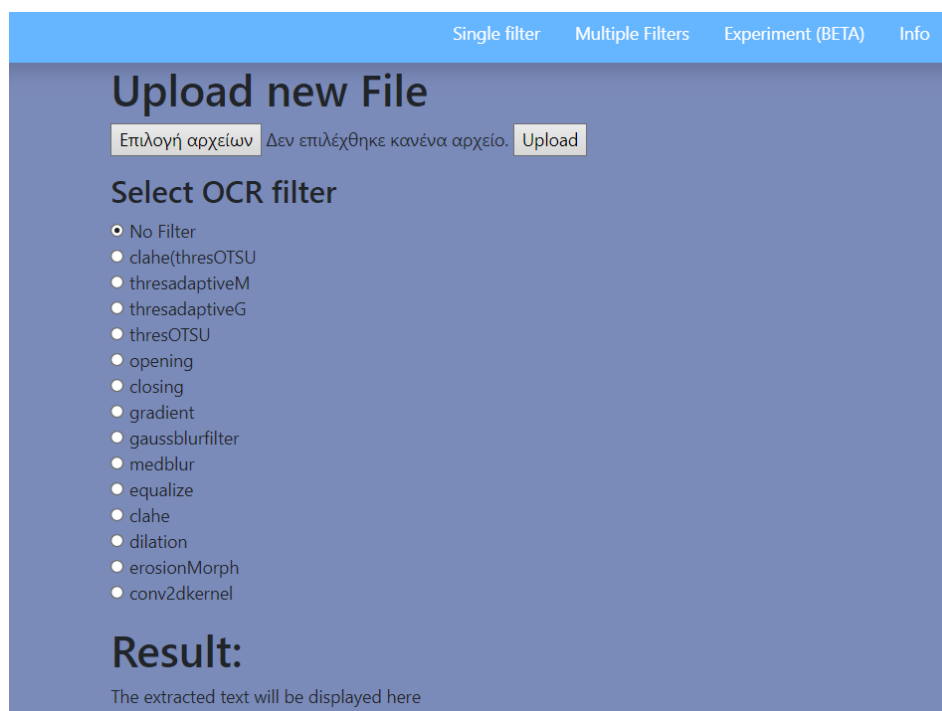
Επιπρόσθετα εδώ θα αποθηκευτούν προσωρινές εικόνες που είναι απαραίτητες για την σωστή λειτουργία του προγράμματος που στο τέλος διαγράφονται τελείως.

## Κεφάλαιο 5

### 5. Αποτελέσματα

Παρακάτω θα αναπτυχθεί ο τρόπος χρήσης της εφαρμογής μέσω της ιστοσελίδας καθώς και η ανάλυση κάποιων αποτελεσμάτων από εικόνες σε γραφήματα.

Κατά την αρχική επίσκεψη της ιστοσελίδας, ο χρήστης έχει την επιλογή χρήσης ενός φίλτρου ή πολλών φίλτρων. Έπειτα ζητείται από τον χρήστη το ανέβασμα μιας ή πολλαπλών εικόνων και του ground text αρχείου και η επιλογή ή μη φίλτρου.



The screenshot shows a web application interface with a blue header. The header contains navigation links: "Single filter", "Multiple Filters", "Experiment (BETA)", and "Info". Below the header, the main content area is titled "Upload new File". It features a file selection area with a button labeled "Επιλογή αρχείων" (File selection) and a message "Δεν επιλέχθηκε κανένα αρχείο." (No file selected). To the right of the message is an "Upload" button. Below this is a section titled "Select OCR filter" with a list of radio buttons for various filters: "No Filter", "clah(thresOTSU)", "thresadaptiveM", "thresadaptiveG", "thresOTSU", "opening", "closing", "gradient", "gaussblurfilter", "medblur", "equalize", "clah", "dilation", "erosionMorph", and "conv2dkernel". At the bottom, there is a "Result:" section with the text "The extracted text will be displayed here".

Εικόνα 26 home2.html, η πρώτη επίσκεψη στην ιστοσελίδα

Για την παρουσίαση χρησιμοποιήθηκε η παρακάτω εικόνα και κείμενο αληθείας:

**ΑΝΑΨΥΚΤΙΚΑ**

Coca cola	.....	1.00
Fanta	.....	1.00
Sprite	.....	1.00
Σόδα	.....	1.00
Νερό φιάλη	.....	0.50

**ΠΟΤΑ**

Vergina	.....	330ml	1.00
Ρετσίνα Μαλαματίνα (μικρή)	.....	1.80	
Ouzo	.....	5.00	
Τσίπουρο	.....	6.00	

**Καθημερινά κοντά σας  
Δευτέρα έως Σάββατο  
από τις 12:00 π.μ. Εως 16:00 μ.μ.  
Κυριακές και αργίες κλειστά**

**DELIVERY  
25410 27198**

**Θερμοπυλών 19 ΞΑΝΘΗ**

Οι τιμές μπορούν να τροποποιηθούν χωρίς προειδοποίηση

**αλάτι και πιπέρι**  
σπιτικό μαγειρεμένο φαγητό

Ρωτήστε μας για το ημερήσιο μενού στο  
**25410 27198**  
Θερμοπυλών 19 ΞΑΝΘΗ

**www.alatikaipiperixanthi.gr**

Στο Αλάτι και πιπέρι Θα βρείτε μια μεγάλη ποικιλία από φαγητά καθώς και πιάτα ημέρας.  
Όλα τα φαγητά μας παρασκευάζονται από φρέσκα και αγνά υλικά.  
Μαγειρεμένα με μεράκι και αγάπη σαν το σπίτι σας.

Εικόνα 27 Εικόνα μελέτης

Αναψυκτικά  
Coca cola 1.00  
Fanta 1.00  
Sprite .00  
Σόδα 1.00  
Νερό φιάλη 0.50  
Ποτά  
Vergina 330ml 1.00  
Ρετσίνα Μαλαματίνα (μικρή) 1.80  
Ούζο 5.00  
Τσίπουρο 6.00

Καθημερινά κοντά σας  
Δευτέρα έως Σάββατο από τις 12:00 π.μ. Εως 16:00μ.μ.  
Κυριακές και αργίες κλειστά

DELIVERY  
2541027198

Θερμοπυλών 19 ΞΑΝΘΗ  
Οι τιμές μπορούν να τροποποιηθούν χωρίς προειδοποίηση

αλάτι και πιπέρι  
σπιτικό μαγειρεμένο φαγητό

Ρωτήστε μας για το ημερήσιο μενού στο  
25410 27198  
Θερμοπυλών 19 ΞΑΝΘΗ

**www.alatikaipiperixanthi.gr**

Στο Αλάτι και πιπέρι θα βρείτε μια μεγάλη ποικιλία από φαγητά καθώς και πιάτα ημέρας.  
Όλα τα φαγητά μας παρασκευάζονται από φρέσκα και αγνά υλικά.  
Μαγειρεμένα με μεράκι και αγάπη σαν το σπίτι σας

Εικόνα 28 Ground text(κείμενο αληθείας)






Έπειτα αριστερά αναφέρεται το ground text όπως ανέβηκε από τον χρήστη, δεξιά το κείμενο όπως ακριβώς το αναγνώρισε το tesseract-ocr και τέλος παρουσιάζονται τα αποτελέσματα των μετρικών σε διάγραμμα με την μορφή των μπάρων (Bars plots).

Παρατηρώντας τα αποτελέσματα είναι φανερό από το ‘Char Error’ του γραφήματος ότι κατά περίπου 76% οι χαρακτήρες του κειμένου αληθείας είναι ίδιοι με το κείμενο που αναγνώρισε η μηχανή καθώς μόλις 44% (Word Error) οι λέξεις είναι ίδιες μεταξύ ground text και output text.


### Φ1 Clahe:

## Successfully processed

**Original Image :**



**Image after filter with detected words (Consider hard refreshing the page, for opera ctr+click on refresh button) :**



Εικόνα 31 Παράδειγμα με Clahe

Original Text :

Αναψυκτικά Coca cola 1.00 Fanta 1.00 Sprite .00 Σόδα 1.00 Νερό φιάλη 0.50 Ποτά Vergina 330ml 1.00 Ρετσίνα Μαλαματίνα (μικρή) 1.80 Ούζο 5.00 Τσίπουρο 6.00 Καθημερινά κοντά σας Δευτέρα έως Σάββατο από τις 12:00 π.μ. Έως 16:00μ.μ. Κυριακές και αργίες κλειστά DELIVERY 2541027198 Θερμοπυλών 19 ΞΑΝΘΗ Οι τιμές μπορούν να τροποποιηθούν χωρίς προειδοποίηση αλάτι και πιπέρι σπυτικό μαγειρεμένο φαγητό Ρωτήστε μας για το ημερήσιο μενού στο 25410 27198 Θερμοπυλών 19 ΞΑΝΘΗ www.alatikairiperixanthi.gr Στο Αλάτι και πιπέρι θα βρείτε μια μεγάλη ποικιλία από φαγητά καθώς και πιάτα ημέρας. Όλα τα φαγητά μας παρασκευάζονται από φρέσκα και αγνά υλικά. Μαγειρεμένα με μεράκι και αγάπη σαν το σπίτι σας

Plot (Consider hard refreshing the page, for opera ctr+click on refresh button) :

Metric	Value
Cosine	0.58
Jaccard	0.21
Char Error	0.72
Word Error	0.36

**Result:**

The extracted text from the image is:

Νερό φιάλη .....0.50 VOI GINA wees 330ml 1.00 Ρετσίνα Μαλαματίνα (Μικρή; ..... 1,80 QUZO . 5 00 Τσίπουρο .....οοοοοοοωνωνωνενινωνων. 6,00 Ka@ηρερίνα Κοvрта oas Δευτέρα εως Σάββατο οαππό tus 12:00 π.ρ. Ews 16:00 μ-μ- Κυριακές Καί αργίts KNÉ101a ELIVERY 95410 27198 Οερρονίον 19 SANOH O:1 τιμές μπορούν να τροποποιηθούν χωρίς προειδοποίηση ONITIKO ραγείρερο ναvnto Ρωτήστε μας VIA TO ημερήσιο μενού ΟΤΟ 25410 27198 ΟερρονναΑων 19 SANOH www.alatikairiperixanthi.gr Στο Αλάτι και πιπέρι Οα βρείτε μια WEYGAN ποικιλία από φαγητά καθώς και πιάτα ημέρας. Όλα Τα φαγητά μας παρασκευάζονται από φρέσκα Καί αγνά υλικά. Μαγειρεμένα με μεράκι ΚΑΙ αγάπη ΟΑV TO σπίτι σας. mecca ra hics 2541078890 ↗

Εικόνα 32 Αποτελέσματα με Clahe

Με το ίδιο μοτίβο εμφανίζονται για κάθε φίλτρο τα παραπάνω αποτελέσματα.

Εφαρμόζοντας το Φ1 Clahe φίλτρο, παρατηρείται ελαφρός μείωση της απόδοσης του tesseract-ocr περίπου κατά 4% αφού πλέον έχουμε 72% Char Error και μόλις 36% Word Error.

Με τα αποτελέσματα αυτά αμέσως γίνεται κατανοητό ότι η εφαρμογή φίλτρων δεν αποδίδει πάντα σε καλύτερη απόδοση της μηχανής Tesseract-OCR αφού η αρχική εικόνα είναι ήδη αρκετά ‘καθαρή’ έχοντας ένα μονόχρωμο background το οποίο βοηθάει κατά πολύ στην καλή απόδοση.



### Φ3 Opening:

## Successfully processed

Original Image :

The image shows two versions of a Greek restaurant menu. The left version is the original image, and the right version is the image after a filter, showing detected words in red and some text in green boxes.

**Original Image:**

- ΑΝΑΨΥΚΤΙΚΑ**
  - Coca cola ..... 1,00
  - Fanta ..... 1,00
  - Sprite ..... 1,00
  - Σόδα ..... 1,00
  - Νερό φιάλη ..... 0,50
- ΠΟΤΑ**
  - Vergina ..... 330ml 1,00
  - Ρετσίνα Μαλαματίνα (μικρή) ..... 1,80
  - Ούζο ..... 5,00
  - Τσίπουρο ..... 6,00
- Καθημερινά κοντά σας Δευτέρα έως Σάββατο από τις 12:00 π.μ. Εως 16:00 μ.μ. Κυριακές και αργίες κλειστά**
- DELIVERY 25410 27198**
- Θερμοπυλών 19 ΞΑΝΘΗ**
- www.alatikaipiperixanthi.gr

**Image after filter with detected words:**

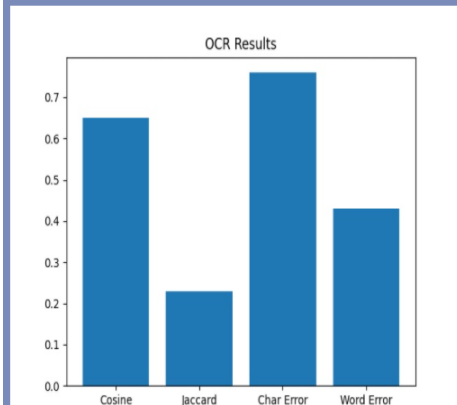
- ΑΝΑΨΥΚΤΙΚΑ**
  - Coca cola ..... 1,00
  - Fanta ..... 1,00
  - Sprite ..... 1,00
  - Σόδα ..... 1,00
  - Νερό φιάλη ..... 0,50
- ΠΟΤΑ**
  - Vergina ..... 330ml 1,00
  - Ρετσίνα Μαλαματίνα (μικρή) ..... 1,80
  - Ούζο ..... 5,00
  - Τσίπουρο ..... 6,00
- Καθημερινά κοντά σας Δευτέρα έως Σάββατο από τις 12:00 π.μ. Εως 16:00 μ.μ. Κυριακές και αργίες κλειστά**
- DELIVERY 25410 27198**
- Θερμοπυλών 19 ΞΑΝΘΗ**
- www.alatikaipiperixanthi.gr

Εικόνα 35 Παράδειγμα Opening

Original Text :

Αναψυκτικά Coca cola 1.00 Fanta 1.00 Sprite .00 Σόδα 1.00 Νερό φιάλη 0.50 Ποτά Vergina 330ml 1.00 Ρετσίνα Μαλαματίνα (μικρή) 1.80 Ούζο 5.00 Τσίπουρο 6.00 Καθημερινά κοντά σας Δευτέρα έως Σάββατο από τις 12:00 π.μ. Εως 16:00μ.μ. Κυριακές και αργίες κλειστά DELIVERY 2541027198 Θερμοπυλών 19 ΞΑΝΘΗ Οι τιμές μπορούν να τροποποιηθούν χωρίς προειδοποίηση αλάτι και πιπερί σπιτικό μαγειρεμένο φαγητό Ρωτήστε μας για το ημερήσιο μενού στο 25410 27198 Θερμοπυλών 19 ΞΑΝΘΗ www.alatikaipiperixanthi.gr Στο Αλάτι και πιπερί θα βρείτε μια μεγάλη ποικιλία από φαγητά καθώς και πιάτα ημέρας. Όλα τα φαγητά μας παρασκευάζονται από φρέσκα και αγνά υλικά. Μαγειρεμένα με μεράκι και αγάπη σαν το σπίτι σας

Plot (Consider hard refreshing the page, for opera ctr+click on refresh button) :



## Result:

The extracted text from the image is:

ΑΝΑΨΥΚΤΙΚΑ Coca cola ..... 1.00 FANTG uu... 1.00 Sprite wc... 1.00 ΥΟΟG .ο 1.00 Νερό φιάλη .....0.50 ΨΘΓΟΩΠ .....οοοοοοονοωννωεωεωοι 330ml 1.00 Ρετσίνα Μαλαματίνα (Μκρή; ..... 1,80 ΟυΖΟ οieccccccsescsseeeceeeessesseeeateeeeees 5,00 Τσίπουρο .....οοοοοοονωνωνεννυννου. 6,00 Ka@ηρηpiva Κλονρα οας Δευτέρς εws Σάββατο ano tug 12:00 n.p. Eos 16:00 np. Κυριακές Καί αργίης ΚΝÉ101ta ELIVERY 254 10 27195 Οερροννίον 19 SANOH O:1 τιμές μπορούν να τροποποιηθούν χωρίς προειδοποίηση ΟΝΙΤΙΚΟ ηγειρεμένο ραητο Ρωτήστε μας VIA TO ημερήσιο μενού ΟΤΟ 25410 27198 ΟερρονναAwν 19 SANOH www.alatikaipiperixanthi.gr Στο Αλάτι και πιπερί Οα βρείτε WIA WEVGAN ποικιλία από φαγητά καθώς και πιάτα ημέρας. Όλα Τα φαγητά μας παρασκευάζονται από φρέσκα Καί αγνά υλικά. Μαγειρεμένα με μεράκι ΚΑΙ αγάπη ΟΑV TO σπίτι σας. mecca gQgraphics 2541078890 ↗

Εικόνα 36 Αποτελέσματα Opening

Παρατηρούνται πάλι σχεδόν ίδια αποτελέσματα παρά μόνο μια μικρή μείωση στην απόδοση λέξης καθώς το Word Error είναι 43%.

#### Φ4 Closing:

### Successfully processed

Original Image :

Image after filter with detected words (Consider hard refreshing the page, for opera ctr+click on refresh button) :

Εικόνα 37 Παράδειγμα Closing

Original Text :

Αναψυκτικά Coca cola 1.00 Fanta 1.00 Sprite .00 Σόδα 1.00 Νερό φιάλη 0.50 Ποτά Vergina 330ml 1.00 Ρεσίνα Μαλαματίνα (μικρή) 1.80 Ούζο 5.00 Τσίπουρο 6.00 Καθημερινά κοντά σας Δευτέρα έως Σάββατο από τις 12:00 π.μ. Εως 16:00μ.μ. Κυριακές και αργίες κλειστά DELIVERY 2541027198 Θερμοπυλών 19 ΞΑΝΘΗ Οι τιμές μπορούν να τροποποιηθούν χωρίς προειδοποίηση αλάτι και πιπέρι σπιτικό μαγειρεμένο φαγητό Ρωτήστε μας για το ημερήσιο μενού στο 25410 27198 Θερμοπυλών 19 ΞΑΝΘΗ [www.alatikaiperixanthi.gr](http://www.alatikaiperixanthi.gr) Στο Αλάτι και πιπέρι θα βρείτε μια μεγάλη ποικιλία από φαγητά καθώς και πιάτα ημέρας. Όλα τα φαγητά μας παρασκευάζονται από φρέσκα και αγνά υλικά. Μαγειρεμένα με μεράκι και αγάπη σαν το σπίτι σας.

Plot (Consider hard refreshing the page, for opera ctr+click on refresh button) :

Result:

The extracted text from the image is:

```

Coca cola ..... 1.00 FANG oo... eee 1.00 SPVite oc... 1.00 LOO oooo 1.00
Νερό @IdAn .....0.50 ΨΦΓΟΩΠ .....ooooooooooooooooooooo, 330ml
1.00 Ρεσίνα Μαλαματίνα (Μικρή); ..... 1.80 ΟΟΥΖΟ
oooooooooecetetseseeeeeseteteneeeeeees 5,00 Τσίπουρο
.....ooooooooooeeeneeweneewenevoni 6,00 Ka@ηπερίνα Κοvr1a oas
Δευτέρx ews Σάθθατο ano tug 12:00 n.p. Ews 16:00 nn. Κυριακές Kai
aryits πδεεστά ELIVERY 95410 27198 Οερρονίον 19 SANΘΗ Ο:1 τιμές
μπορούν να τροποποιηθούν χωρίς προειδοποίηση ΟΝΙΤΙΚΟ
μηγειρεμένο φαγητο Ρωτήστε μας VIA ΤΟ ημερήσιο μενού ΟΤΟ 25410
27198 ΟερρονναΑwn 19 SANΘΗ www.alatikaiperixanthi.gr 2ΤΟ
Αλάτι και πιπέρι Οa βρείτε WIA WEGAN ποικιλία από φαγητά
καθώς και πιάτα ημέρας. Όλα Τα φαγητά μας παρασκευάζονται από
φρέσκα ΚΑΙ αγνά υλικά. Μαγειρεμένα με μεράκι ΚΑΙ αγάπη σαν ΤΟ
σπίτι σας. meca graphics 2541078890 ^

```

Εικόνα 38 Αποτελέσματα Closing

Δεν παρατηρούνται διαφορές.

## Φ5 Gradient:

### Successfully processed

Original Image :


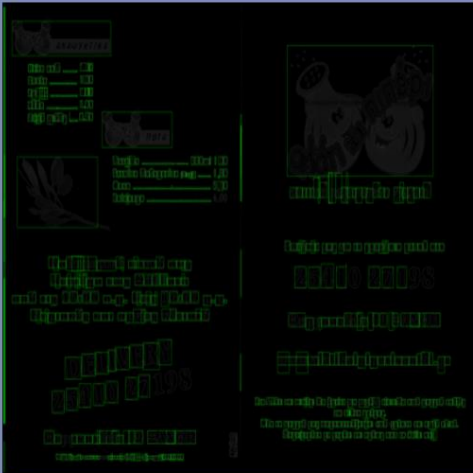


Image after filter with detected words (Consider hard refreshing the page, for opera ctr+click on refresh button) :

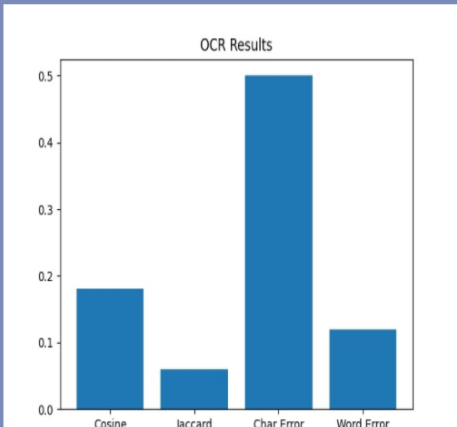


Εικόνα 39 Παράδειγμα Gradient

Original Text :

Αναψυκτικά Coca cola 1.00 Fanta 1.00 Sprite .00 Σόδα 1.00 Νερό φιάλη 0.50 Ποτά Vergina 330ml 1.00 Ρετσίνα Μαλαματίνα (μικρή) 1.80 Ούζο 5.00 Τσίπουρο 6.00 Καθημερινά κοντά σας Δευτέρα έως Σάββατο από τις 12:00 π.μ. Εως 16:00 μ.μ. Κυριακές και αργίες κλειστά DELIVERY 2541027198 Θερμοπυλών 19 ΞΑΝΘΗ Οι τιμές μπορούν να τροποποιηθούν χωρίς προειδοποίηση αλάτι και πιπέρι σπιτικό μαγειρεμένο φαγητό Ρωτήστε μας για το ημερήσιο μενού στο 25410 27198 Θερμοπυλών 19 ΞΑΝΘΗ www.alatikaiperixanthi.gr Στο Αλάτι και πιπέρι θα βρείτε μια μεγάλη ποικιλία από φαγητά καθώς και πιάτα ημέρας. Όλα τα φαγητά μας παρασκευάζονται από φρέσκα και αγνά υλικά. Μαγειρεύονται με μεράκι και αγάπη σαν το σπίτι σας

Plot (Consider hard refreshing the page, for opera ctr+click on refresh button) :



Metric	Value
Cosine	~0.18
Jaccard	~0.06
Char Error	~0.50
Word Error	~0.12

### Result:

The extracted text from the image is:

Coca Cola ..... 1-00 I. SPITE ..... 1 00 ρο | 00 Νερό φιάλη .....0.50 VEOTGING ..... 330M 1.00 Ρετσίνα ΜαΑαυαρίνα (Μxpr) ..... | 80 OQUZO cscscoserscnsceereesscevececeerensensesesee DQ ΤΟΙΠΟΥΡΟ ..... © OQ | ) Πη pee UCU ECON λς NEWEST νς όση( ONO π ως TZ TO ORE peel COC 11(δ9(0)9) τογας INUPIMRES RA ADVISES RASIOVH (ε ρο Ο ο ANG (θ)εήτι ε SHO ρο υναναμτ ρο πο Πο ΠΟ ο ν ΡΙG ρ ΟΕΙΟΠΟ ΠΟ = ι ς--- CHICO [IAYSIPSVEVO CHADTO PWINOTE UAC VIC TO NUEONOIO UEVOU COTO ZD4UO ZIUVMΣ ep ΓΟ ' Ξ ANG) alatiikaiperixantini. or Στο AAGTI KCI THHTEOI Od βρείτε ia EVGA TIOIKIAG απιπτό φαψηγτά καθώς Κ ππιάτα ημέρας. Όλα ΤΑ φαψηγτά μας ππαρασκευάζονται από φρέσκα ΚΟΝ αφνά υλικά. ΜaveίρεVeva UE μεράκι ΚΟΝ CYaTT Cav TO στίτι Cas. ⤴

Εικόνα 40 Αποτελέσματα Gradient

Μετά την εφαρμογή του Φ4 Gradient παρατηρείται σημαντική πτώση της απόδοσης αφού η επεξεργασμένη εικόνα όπως φαίνεται στην Εικόνα 39 έχει μεταεπεξεργαστεί σε σημείο που είναι σχεδόν αδύνατη η αναγνώριση των χαρακτήρων της.







## Φ8 Equalize:

### Successfully processed

Original Image :

Image after filter with detected words (Consider hard refreshing the page, for opera ctr+click on refresh button) :

Εικόνα 45 Παράδειγμα Equalize

Original Text :

Αναψυκτικά Coca cola 1.00 Fanta 1.00 Sprite .00 Σόδα 1.00 Νερό φιάλη 0.50 Ποτά Vergina 330ml 1.00 Ρετσίνα Μαλαματίνα (μικρή) 1.80 Ούζο 5.00 Τσίπουρο 6.00 Καθημερινά κοντά σας Δευτέρα έως Σάββατο από τις 12:00 π.μ. Εως 16:00μ.μ. Κυριακές και αργίες κλειστά DELIVERY 2541027198 Θερμοπυλών 19 ΞΑΝΘΗ Οι τιμές μπορούν να τροποποιηθούν χωρίς προειδοποίηση αλάτι και πιπέρι σπιτικό μαγειρεμένο φαγητό Ρωτήστε μας για το ημερήσιο μενού στο 25410 27198 Θερμοπυλών 19 ΞΑΝΘΗ [www.alatikaipiperixanthi.gr](http://www.alatikaipiperixanthi.gr) Στο Αλάτι και πιπέρι θα βρείτε μια μεγάλη ποικιλία από φαγητά καθώς και πιάτα ημέρας. Όλα τα φαγητά μας παρασκευάζονται από φρέσκα και αγνά υλικά. Μαγειρεμένα με μεράκι και αγάπη σαν το σπίτι σας

Plot (Consider hard refreshing the page, for opera ctr+click on refresh button) :

Metric	Value
Cosine	0.65
Jaccard	0.22
Char Error	0.75
Word Error	0.38

Result:

The extracted text from the image is:

Coca cola ..... 1.00 Ραηία..... 100 . Sprite ..... 1.00 . Σόδα.....- 1.00 Νερό φιάλη .....0.50 Ρετσίνα Μαλαματίνα (Μικρή) οση 1,80 Ουζο wessesssssssstssseessssssseesssneeess we 0,00 'σίπουρο ..... sseessssasserseeseere 0,00 Ka@Oηπερίνα κοντά cas ' Δευτέρα εως Σάββατο από ug 12:00 π.μ. Eos 16:00 BB. Κυριακές και αργίες ελειστά ειNEBY 2ξ410 21 198 Θερμοιυλων 19 ΞΑΝΘΗ Οι.....κ..... ΨοΓηΠΔ .....σοουμου. ' 330 1.00 : σπιτικό μομιμμο παγοrs Ρωτίοte Wag για το ημερήσιο μενού στο 254 10 27198 Θερμοπυλών 19 ΕΑΝΘΗ | τ. ' πιωΜΜμω rix anθigr S10 Αλάτι και πιπέρι Οα βρείτε rία peγαAn ποικιλία από φαγητά καθώς και πιάτα ημέρας. Όλα τα φαγητα μας παρασκευάζονται από φρέσκα και αγνά υλικά. Μαγειρεμένα με μεράκι...και αγάπη σαν το σπίτι σας. ♣

Εικόνα 46 Αποτελέσματα Equalize





# Φ11 Conv 2D Kernel:

**Successfully processed**

Original Image :






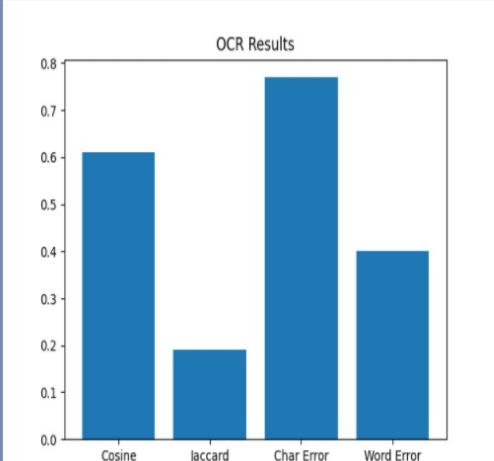
Image after filter with detected words (Consider hard refreshing the page, for opera ctr+click on refresh button) :

Εικόνα 51 Παράδειγμα Conv2DKernel

**Original Text :**

Αναψυκτικά Coca cola 1.00 Fanta 1.00 Sprite .00 Σόδα 1.00 Νερό φιάλη 0.50 Ποτά Vergina 330ml 1.00 Ρετινία Μαλαματίνα (μικρή) 1.80 Ούζο 5.00 Τσίπουρο 6.00 Καθημερινά κοντά σας Δευτέρα έως Σάββατο από τις 12:00 π.μ. Εως 16:00μ.μ. Κυριακές και αργίες κλειστά DELIVERY 2541027198 Θερμοπυλών 19 ΞΑΝΘΗ Οι τιμές μπορούν να τροποποιηθούν χωρίς προειδοποίηση αλάτι και πιπέρι σπιτικό μαγειρεμένο φαγητό Ρωτήστε μας για το ημερήσιο μενού στο 25410 27198 Θερμοπυλών 19 ΞΑΝΘΗ www.alatikaipiperixanthi.gr Στο Αλάτι και πιπέρι θα βρείτε μια μεγάλη ποικιλία από φαγητά καθώς και πιάτα ημέρας. Όλα τα φαγητά μας παρασκευάζονται από φρέσκα και αγνά υλικά. Μαγειρεμένα με μεράκι και αγάπη σαν το σπίτι σας

Plot (Consider hard refreshing the page, for opera ctr+click on refresh button) :



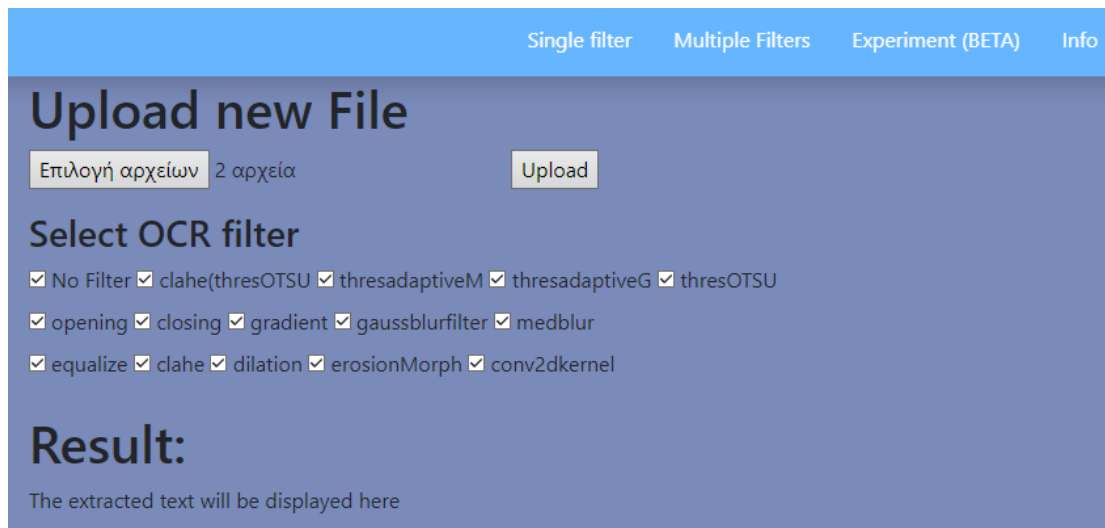
**Result:**

The extracted text from the image is:

tC (ee Coca cola ..... 1.00 Ραηῖα..... 1.00 Sprite  
 οσωωωωωωωωωωω 1.00 Σόδα..... 100 . - Νερό φιάλη .....0.50 ' c, ο  
 - υ ' VOT GINA .....οοοονωωωωωωωωω 330ml 1.00 Ρετινία  
 Μαλαματίνα (μίχρη) ..... | 80 QUZO  
 οίcccccceesssessessesseseeseeeteeaaeeeees 5 00 ' 4 ' Ισίπουρο  
 .....οοοοοωνωνωνεννεννεωο, 6,00 ON1tIKO payeipepevo  
 QMAYnto "en Ρωτήστε μας VIA TO ημερήσιο μενού OTO K@ηperίva  
 πικκουτά oas Δευτέρς ews Σάθθατο 2 54 ] Oo y J y | ] 98 apo ts 12:00  
 n.p. Ews 16:00 μ-μ- | Κυριακές kai arγiys KAN£101a Θερμοπυδλών 19  
 ΞΑΝΘΗ | DELIVERY www.alatikaipiperixanthi.gr 2 54 1 0 y2 Στο AAGTI  
 Kai πιπέρι Oa βρείτε μια WEYGAN ποικιλία από φαγητά καθώς | και  
 πιάτα ημέρας. Oda Τα φαγητά yas παρασκευάζονται από φρέσκα Kai  
 αγνά υλικά. Μαγειρεμένα με μεράκι KAI αγάπη Can TO σπίτι σας.  
 Θερμοπυδλών 19 SANOH O:1 τιμές μπορούν να τροποποιηθούν  
 χωρίς προειδοποίηση ARGS 2541070690 ⬆

Εικόνα 52 Αποτελέσματα Conv2DKernel

Στην συνέχεια, επιλέγοντας ‘Multiple Filters’, χρησιμοποιώντας την ίδια εικόνα και ελέγχοντας όλα τα φίλτρα συγκρίνοντάς τα, παίρνουμε τα εξής αποτελέσματα.



Εικόνα 53 Επιλογή πολλαπλών φίλτρων.



Εικόνα 54 Αποτελέσματα πολλαπλών φίλτρων

Παρατηρώντας τα αποτελέσματα της σύγκρισης όλων των διαθέσιμων φίλτρων Φ1-Φ11, διαπιστώνεται ότι την συνολική καλύτερη απόδοση είχε το Φ2 Thresholding OTSU για την συγκεκριμένη εικόνα, αφού είχε τα υψηλότερα ποσοστά της τάξης των 76% για την απόδοση ανά χαρακτήρα(char error) και 47% για ανά λέξη(word error).

Στην συνέχεια παρουσιάζονται αποτελέσματα για σύγκριση πολλαπλών εικόνων με εφαρμογή ενός φίλτρου.

Χρησιμοποιήθηκαν οι 4 εικόνες από τον ίδιο κατάλογο εστίασης.



Εικόνα 55 Παράδειγμα πολλαπλών εικόνων 1



Εικόνα 56 Παράδειγμα πολλαπλών εικόνων 2



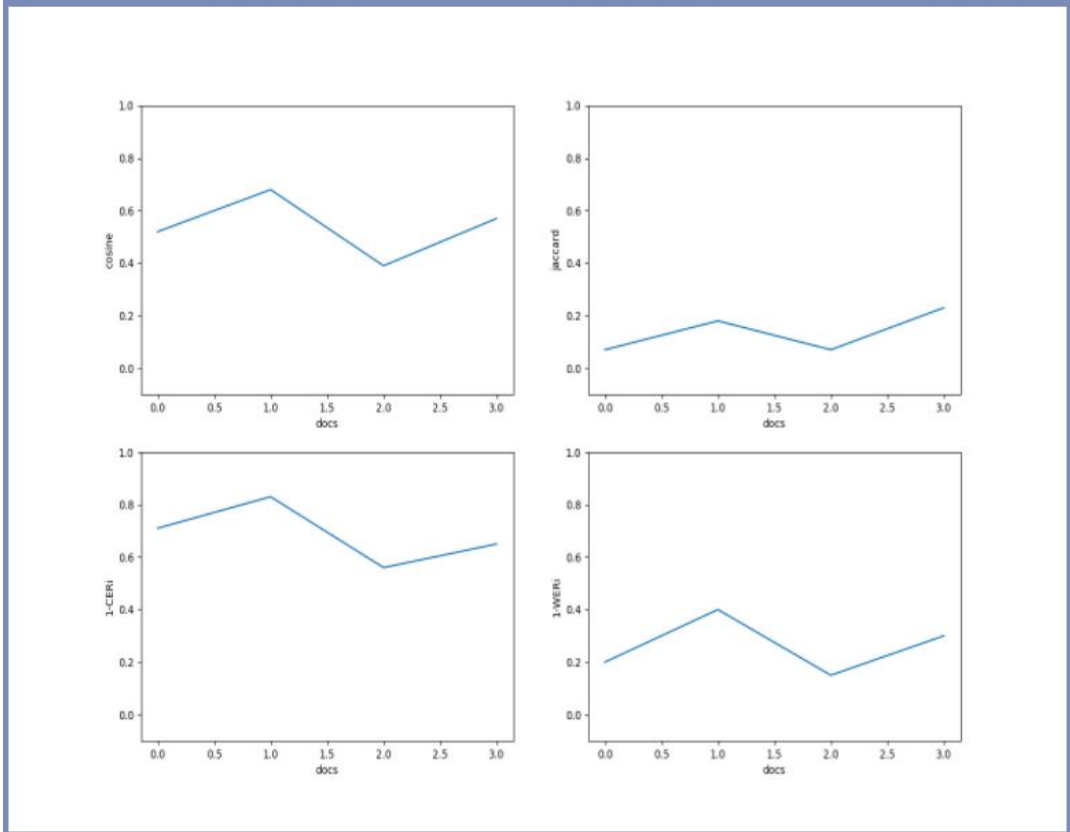
Εικόνα 57 Παράδειγμα πολλαπλών εικόνων 3



Εικόνα 58 Παράδειγμα πολλαπλών εικόνων 4

# Successfully processed

Plot (Consider hard refreshing the page, for opera ctr+click on refresh button) :

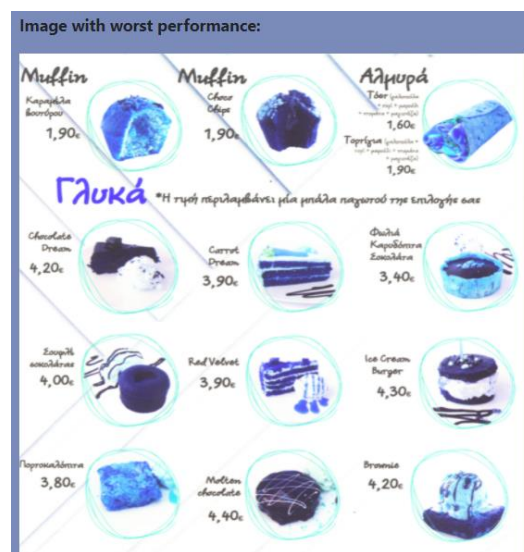


Εικόνα 59 Αποτελέσματα πολλαπλών εικόνων

Φανερή είναι η απόδοση των τεσσάρων εικόνων όπως φαίνεται από τα γραφήματα, αλλά για την καλύτερη εμπειρία του χρήστη, εμφανίζονται παρακάτω οι καλύτερη και η χειρότερη απόδοση μεταξύ των συγκρινόμενων εικόνων.



Εικόνα 60 Εικόνα με την καλύτερη απόδοση



Εικόνα 61 Εικόνα με την χειρότερη απόδοση





### **6. Συμπεράσματα και Μελλοντικές επεκτάσεις**

Όπως πλέον έχει γίνει φανερό, η χρήση της μηχανής Tesseract-OCR είναι εύκολη στην χρήση που μπορεί ο καθένας να χρησιμοποιήσει για την δημιουργία χρήσιμων εφαρμογών για την διευκόλυνση της ζωής του ανθρώπου. Τα ποσοστά επιτυχίας αναγνώρισης κειμένου των διαθέσιμων μηχανών OCR ακόμα έχουν περιθώρια βελτίωσης αλλά ήδη βρίσκονται σε ικανοποιητικά επίπεδα ώστε να έχουν πρακτική χρήση στην καθημερινότητα.

Η παρούσα εργασία έχει τις βάσεις για την μετέπειτα εξέλιξή της η οποία θα μπορούσε να υπολογίσει και να συγκρίνει τις εικόνες μετά την εφαρμογή σε αυτών πολλαπλά ταυτόχρονα φίλτρα. Δηλαδή να γίνει συνδυασμός φίλτρων σε μία εικόνα και σαν αποτέλεσμα να εμφανίζει τον καλύτερο αυτόν συνδυασμό.

Επίσης σημαντική είναι μία επέκταση όπως αυτή της σύγκρισης μεταξύ διαφορετικών μηχανών OCR (Google vision vs Tesseract-OCR vs ABYY vs Asprise).

Επιπρόσθετα, με την πλέον καθημερινή χρήση των smartphone, μια ακόμη επέκταση είναι η δημιουργία εφαρμογής android για την ευκολότερη και γρηγορότερη χρήση του προγράμματος.

## Βιβλιογραφία

- abbyy*. (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/ABBYY\\_FineReader API](https://en.wikipedia.org/wiki/ABBYY_FineReader_API). (χ.χ.). Ανάκτηση από <https://en.wikipedia.org/wiki/API>
- artificial intelligence*. (χ.χ.). Ανάκτηση από <https://www.ibm.com/cloud/learn/what-is-artificial-intelligence>
- asprise ocr*. (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Asprise\\_OCR](https://en.wikipedia.org/wiki/Asprise_OCR)
- Bars plots*. (χ.χ.). Ανάκτηση από [https://www.tutorialspoint.com/matplotlib/matplotlib\\_bar\\_plot.htm](https://www.tutorialspoint.com/matplotlib/matplotlib_bar_plot.htm)
- Blops*. (χ.χ.). Ανάκτηση από [https://tesseract-ocr.repairfaq.org/tess\\_glossary.html](https://tesseract-ocr.repairfaq.org/tess_glossary.html)
- Browsers*. (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Web\\_browser](https://en.wikipedia.org/wiki/Web_browser)
- Cascading Style Sheets*. (χ.χ.). Ανάκτηση από <https://el.wikipedia.org/wiki/CSS>
- clahe*. (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Adaptive\\_histogram\\_equalization](https://en.wikipedia.org/wiki/Adaptive_histogram_equalization)
- closing*. (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Closing\\_\(morphology\)](https://en.wikipedia.org/wiki/Closing_(morphology))
- computer vision*. (χ.χ.). Ανάκτηση από [https://www.sas.com/el\\_gr/insights/analytics/computer-vision.html](https://www.sas.com/el_gr/insights/analytics/computer-vision.html)
- data mining*. (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Data\\_mining](https://en.wikipedia.org/wiki/Data_mining)
- David H. Shepard*. (χ.χ.). Ανάκτηση από <https://www.nytimes.com/2007/12/11/us/11shepard.html>
- Digitization*. (χ.χ.). Ανάκτηση από <https://en.wikipedia.org/wiki/Digitization>
- dilation*. (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Dilation\\_\(morphology\)](https://en.wikipedia.org/wiki/Dilation_(morphology))
- Edmund Edward Fournier*. (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Edmund\\_Edward\\_Fournier\\_d%27Albe](https://en.wikipedia.org/wiki/Edmund_Edward_Fournier_d%27Albe)
- equalize*. (χ.χ.). Ανάκτηση από <https://towardsdatascience.com/histogram-equalization-5d1013626e64>
- erosion*. (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Erosion\\_\(morphology\)](https://en.wikipedia.org/wiki/Erosion_(morphology))
- Flask*. (χ.χ.). Ανάκτηση από <https://flask.palletsprojects.com/en/2.0.x/>
- Gaussian blur*. (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Gaussian\\_blur](https://en.wikipedia.org/wiki/Gaussian_blur)
- Gismo*. (χ.χ.). Ανάκτηση από <https://history-computer.com/gismo-of-david-shepard/>
- Goldberg, E. (2021). Ανάκτηση από <https://history-computer.com/emanuel-goldberg/>
- gradient*. (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Image\\_gradient](https://en.wikipedia.org/wiki/Image_gradient)
- Grayscale*. (χ.χ.). Ανάκτηση από <https://en.wikipedia.org/wiki/Grayscale>

*Gustav Tauschek.* (χ.χ.). Ανάκτηση από <https://www.ithistory.org/honor-roll/mr-gustav-tauschk>

*HP.* (χ.χ.). Ανάκτηση από <https://en.wikipedia.org/wiki/Hewlett-Packard>

*HyperText Markup Language.* (χ.χ.). Ανάκτηση από <https://el.wikipedia.org/wiki/HTML>

*image post-processing.* (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Image\\_editing](https://en.wikipedia.org/wiki/Image_editing)

*Imageio.* (χ.χ.). Ανάκτηση από <https://pypi.org/project/imageio/>

*Jiwer.* (χ.χ.). Ανάκτηση από <https://pypi.org/project/jiwer/>

*kernel.* (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Kernel\\_\(image\\_processing\)](https://en.wikipedia.org/wiki/Kernel_(image_processing))

*machine translation.* (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Machine\\_translation](https://en.wikipedia.org/wiki/Machine_translation)

*Matplotlib.* (χ.χ.). Ανάκτηση από <https://matplotlib.org>

*median blur.* (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Median\\_filter](https://en.wikipedia.org/wiki/Median_filter)

*NumPy.* (χ.χ.). Ανάκτηση από <https://numpy.org>

*Open Source Computer Vision.* (χ.χ.). Ανάκτηση από <https://en.wikipedia.org/wiki/OpenCV>

*opening.* (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Opening\\_\(morphology\)](https://en.wikipedia.org/wiki/Opening_(morphology))

*Optical Character Recognition.* (χ.χ.). Ανάκτηση από [https://el.wikipedia.org/wiki/Οπτική\\_Αναγνώριση\\_Χαρακτήρων](https://el.wikipedia.org/wiki/Οπτική_Αναγνώριση_Χαρακτήρων)

*Optophone.* (χ.χ.). Ανάκτηση από <https://en.wikipedia.org/wiki/Optophone>

*Otsu's thresholding.* (χ.χ.). Ανάκτηση από <https://hbyacademic.medium.com/otsu-thresholding-4337710dc519>

*pattern recognition.* (χ.χ.). Ανάκτηση από <https://serokell.io/blog/pattern-recognition>

*pyTesseract.* (χ.χ.). Ανάκτηση από <https://pypi.org/project/pytesseract/>

*Python.* (χ.χ.). *Python.* Ανάκτηση από <https://el.wikipedia.org/wiki/Python>

*Ray Kurzweil.* (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Ray\\_Kurzweil](https://en.wikipedia.org/wiki/Ray_Kurzweil)

*tesseract.* (χ.χ.). Ανάκτηση από [https://en.wikipedia.org/wiki/Tesseract\\_\(software\)](https://en.wikipedia.org/wiki/Tesseract_(software))

*text-to-speech.* (χ.χ.). *Speech synthesis.* Ανάκτηση από [https://en.wikipedia.org/wiki/Speech\\_synthesis](https://en.wikipedia.org/wiki/Speech_synthesis)

*κρυπτανάλυση.* (χ.χ.). Ανάκτηση από <https://en.wikipedia.org/wiki/Cryptanalysis>

*νέφος.* (χ.χ.). Ανάκτηση από [https://el.wikipedia.org/wiki/Υπολογιστικό\\_νέφος](https://el.wikipedia.org/wiki/Υπολογιστικό_νέφος)

