OCR230
@RARES ANGHEL

Document Release Date
JULY 2025

OCR PENTRU BINE SOCIAL

OCR230
PROJECT DOCUMENTATION

Prepared By:

Rares Mihai Anghel

OCR230

Digitalizarea proceselor birocratice este esențială în era actuală. OCR230 simplifică și automatizează completarea formularelor 230, economisind timp și reducând erorile. Prin tehnologie accesibilă și eficientă, aplicația sprijină ONG-urile să lucreze mai rapid, mai sigur și mai inteligent, pregătindu-le pentru un viitor digital.

Prezentare

Contact:

0733393301

raresanghel2008@gmail.com

Bucharest, Romania

Colegiul National de Informatica Tudor Vianu

CUPRINS

I	Descrierea aplicației	5
II	Autor	6
III	Analiza pieței	7
IV	Începuturile aplicației	8
V	Planificarea dezvoltării	9
VI	Proiectarea arhitecturală	10
VII	Tehnologiile aplicației	11
VIII	Stadiul aplicației	12
IX	Cum funcționează aplicația	13
X	Utilitatea aplicației	14
XI	Rezultate de ieșire ale aplicației	15
XII	Ghidul aplicației	17
XIII	Cerințe și specificații	18
XIV	Project Timeline	19
XV	Design și Arhitectură	20
XVI	Modularitate	21
XVII	Implementare	22
XVIII	Testare și asigurarea calității	23
XIX	Teamwork	24
XX	Vulnerabilități	25
XXI	Obiective	27
XXII	Concuzie	28
XXIII	Recenzii	29
XXIV	Resurse externe	30
XXV	Glosar	31

DESCRIEREA APLICAȚIEI

NUME PROIECT
OCR230

CEO & FONDATOR
RARES MIHAI ANGHEL

START DATE
DECEMBER 2024

EXPECTED END DATE
JUNE 2027

DETALII PROIECT

OCR230 este o aplicație desktop care automatizează extragerea datelor din formularele 230, folosind tehnologie OCR. Procesează fișiere PDF sau imagini, identificând rapid informații precum nume, CNP și adresă.
Aplicația are o interfață simplă, organizează automat fișierele și permite exportul datelor, inclusiv în Excel. Este folosită deja de mai multe ONG-uri și este disponibilă gratuit în Microsoft Store.

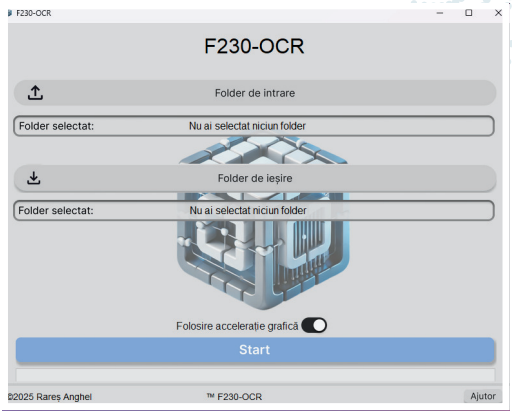


fig. 1 - Fereastra principală a aplicației

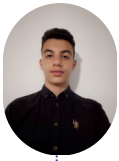
OBIECTIVE

- Implementarea unui sistem OCR pentru reducerea timpului de procesare cu 90% față de cea manuală, cu extragere automată și organizare pe ANAF-uri.
- Dezvoltarea unui sistem de export în Excel cu formatare corectă și organizare automată în foldere, eliminând erorile.
- Asigurarea procesării 100% locale a datelor personale cu validare robustă și gestionare securizată a erorilor, respectând legislația GDPR din România și UE.

SCOPUL PROIECTULUI

Scopul proiectului este de a automatiza și simplifica procesarea formularelor 230, reducând timpul și efortul necesar pentru extragerea și organizarea datelor, oferind astfel o soluție accesibilă și eficientă pentru ONG-uri și utilizatori individuali.

AUTOR



RARES MIHAI ANGHEL

CEO & DEVELOPER

Sunt elev la Colegiul Național de Informatică „Tudor Vianu” din București, o școală cu tradiție în formarea tinerilor pasionați de informatică, matematică și științe exacte. Încă din gimnaziu, am fost atras de tehnologie și rezolvarea problemelor reale prin cod, dar adevărata mea pasiune pentru programare s-a conturat în perioada școlii online, când am început să învăț pe cont propriu și să experimentez intensiv.

Am participat la mai multe olimpiade și concursuri școlare, printre care Olimpiada de Informatică, Tehnologia Informației (TIC) și AcadNet, unde am aprofundat atât conceptele de bază, cât și cele avansate din domeniul IT.

Pe lângă latura academică, sunt pasionat de robotică și am făcut parte din Echipa Națională de Robotică a României – AutoVortex, cu care am participat la competiții internaționale precum FIRST Tech Challenge (FTC) și VEX Robotics, obținând rezultate remarcabile. Această experiență mi-a oferit șansa de a lucra în echipe diverse, de a rezolva probleme complexe sub presiune și de a colabora cu oameni pasionați din toată lumea.

Pasiunea mea pentru dezvoltarea de aplicații utile m-a condus către proiecte independente, cum este și OCR230, dar și către dezvoltarea unor site-uri pentru grădinițe și afaceri locale, unde am învățat să înțeleg cerințele reale ale utilizatorilor și să creez soluții simple și eficiente.

În vara anului 2025, voi avea ocazia să fiu Teacher Assistant la Code Sinaia, o tabără de programare avansată dedicată elevilor pasionați. Mă bucur să pot contribui și eu la formarea altor tineri, oferind din experiența mea celor care vor să învețe mai mult și să-și depășească limitele.

Mă consider o persoană deschisă, curioasă și orientată spre progres. Cred că tehnologia este un instrument esențial pentru schimbare, iar contribuția fiecăruia contează. Îmi doresc ca pe viitor să construiesc aplicații cu impact, să învăț constant și să colaborez cu oameni care cred în inovație și educație.

ANALIZA PIEȚEI

Înainte de a începe dezvoltarea aplicației OCR230, am realizat o analiză detaliată a pieței și a nevoilor reale ale ONG-urilor din România. În țara noastră există peste 130.000 de organizații non-guvernamentale, majoritatea fiind mici sau mijlocii, cu resurse financiare și tehnice limitate. Acestea se confruntă cu o provocare majoră: procesarea manuală a formularelor 230, documente esențiale pentru redirecționarea a 3,5% din impozitul pe venit către cauzele sociale pe care le susțin. Voluntarii și angajații petrec adesea ore întregi, uneori zile, pentru introducerea manuală a datelor în baze de date, un proces obositor, predispus la erori și consumator de timp prețios, ce ar putea fi dedicat direct misiunii organizației.

Pe piața românească nu există în prezent aplicații dedicate care să automatizeze procesarea formularelor 230 pentru ONG-uri. În multe dintre aceste organizații, voluntarii procesează manual fiecare formular, un proces consumator de timp și predispus la erori. OCR230 răspunde acestei nevoi reale, oferind o soluție locală, completă și sigură, care nu necesită încărcarea datelor pe servere externe. Aplicația automatizează întreg fluxul de lucru: de la convertirea documentelor PDF sau imagine în text, extragerea și validarea datelor esențiale, până la organizarea inteligentă a fișierelor în foldere structurate după ani și localități. Fișierele sunt redenumite automat după numele beneficiarului, facilitând astfel arhivarea și accesul rapid ulterior.

Interfața grafică, simplă și intuitivă, permite chiar și utilizatorilor fără cunoștințe tehnice să folosească eficient aplicația, reducând semnificativ timpul și efortul dedicat procesării formularelor 230. Tehnologiile moderne integrate, precum OCR-ul avansat (inclusiv efficient-ocr), convertirea PDF-urilor în imagini și tehnici de paralelizare, asigură o performanță optimă fără costuri suplimentare sau dependențe comerciale.

Astfel, OCR230 nu este doar un instrument tehnic, ci un sprijin real pentru ONG-urile românești, ajutând la reducerea poverii administrative și la creșterea eficienței procesului de finanțare prin redirecționarea impozitului pe venit, contribuind direct la impactul pozitiv în comunitățile pe care le deservesc.

INTRODUCERE

CUM MI-A VENIT IDEA

Ideea dezvoltării aplicației OCR230 a luat naștere dintr-o nevoie reală observată în mediul ONG-urilor mici din România. Mama mea conduce o organizație non-guvernamentală dedicată sprijinirii surorii mele, diagnosticată cu o afecțiune genetică rară. În activitatea sa, a fost nevoită să gestioneze un număr semnificativ de formulare 230, documente prin care contribuabilii pot redirecționa 3,5% din impozitul pe venit către o cauză socială. Procesarea acestor formulare se realiza în mod manual, o activitate repetitivă, consumatoare de timp și predispusă la erori.

Observând dificultățile întâmpinate și impactul lor asupra resurselor limitate ale ONG-urilor, am identificat oportunitatea de a contribui prin competențele mele în domeniul informaticii. Astfel, am început prin dezvoltarea unui script simplu care automatiza extragerea datelor din formularele scanate. Ulterior, pe măsură ce soluția a demonstrat rezultate promițătoare, am extins funcționalitățile, am dezvoltat o interfață grafică intuitivă și am creat o aplicație completă, accesibilă oricărui utilizator, fără necesitatea unor cunoștințe tehnice avansate.

CE ESTE TEHNOLOGIA OCR

OCR (Optical Character Recognition) este o tehnologie ce transformă imagini sau PDF-uri în text digital. De la începuturi simple în anii ’50, a evoluat spre sisteme inteligente bazate pe AI. În OCR230, OCR-ul permite extragerea automată a datelor din formulare scanate.

CUM FUNCȚIONEAZĂ MODELELE OCR ÎN OCR230

- **EasyOCR** folosește o combinație de CNN[1], LSTM[2] și CTC[3] pentru a extrage caractere din imagini. Imaginea este transformată în trăsături vizuale, apoi procesată ca o secvență de text.
- **EffOCR**, dezvoltat recent de Harvard, abordează problema diferit: detectează individual caracterele folosind un model de obiecte și le recunoaște prin comparație vizuală (image retrieval), cu un encoder antrenat contrastiv.

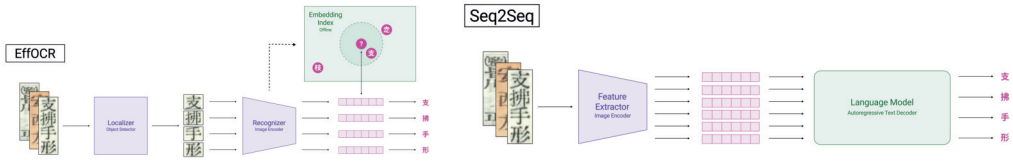


fig. 2 - Comparație între EffOCR si EasyOCR

PLANIFICAREA DEZVOLTĂRII

01

CERCETARE PRELIMINARĂ ȘI PROTOTIPARE

Am explorat soluții OCR și tehnologii conexe (precum conversia PDF), testând librării ca easyocr, efficient-ocr, pdf2image și poppler-utils. În această etapă au fost construite primele prototipuri, necesare pentru a evalua acuratețea și viteza fiecărei soluții.

02

DEZVOLTAREA NUCLEULUI APLICAȚIEI

Am implementat funcționalitatea de bază: citirea documentelor PDF sau imagine, aplicarea OCR-ului și extragerea datelor relevante din formularele 230. S-a pus accent pe identificarea automată a zonelor-cheie din documente și pe extragerea robustă a textului util.

03

PRELUCRAREA ȘI ORGANIZAREA DATELOR

Am dezvoltat module pentru validarea CNP-urilor, filtrarea conținutului inutil și redenumirea automată a fișierelor în funcție de datele extrase. În plus, aplicația organizează documentele în foldere structurate pe ani și localități, optimizând arhivarea și accesul rapid.

04

CREAREA INTERFEȚEI GRAFICE

Designul a fost realizat în Figma, cu accent pe claritate vizuală și accesibilitate. Ulterior, interfața a fost implementată cu biblioteca tkinter, fiind construită special pentru utilizatori fără experiență tehnică.

05

TESTARE FUNCȚIONALĂ ȘI FEEDBACK DE LA ONG-URI

Am testat aplicația pe loturi reale de formulare puse la dispoziție de ONG-uri partenere. Acest lucru a permis identificarea cazurilor-limită, îmbunătățirea stabilității aplicației și adaptarea funcțiilor la cerințe.

06

IMPLEMENTAREA METODELOR DE EXPORT

Aplicația oferă export automat în trei formate esențiale:

- Excel: centralizare structurată, gata de utilizare în raportări.
- CSV: pentru integrare cu baze de date sau sisteme contabile.
- PDF: generare automată de rapoarte vizuale, cu tabele și grafice.

07

STATISTICĂ LIVE ȘI RAPORTARE ÎN TIMP REAL

La finalul procesării, aplicația oferă un tablou de bord interactiv, unde sunt afișate: numărul total de formulare procesate, ratele de validare a CNP-urilor, duplicatele identificate, localitățile predominante etc. Aceste statistici ajută ONG-urile să obțină rapid o imagine de ansamblu asupra datelor colectate.

PROIECTAREA ARHITECTURALĂ

Proiectarea arhitecturală a unei aplicații software este etapa în care se stabilește structura de bază a sistemului, modul în care principalele sale componente sunt organizate, comunică între ele și se raportează la cerințele funcționale și nefuncționale. Pentru OCR230, acest proces a fost esențial în transformarea unei idei tehnice într-o aplicație robustă, scalabilă și ușor de utilizat, în special în contextul sensibil al prelucrării datelor personale.

COMPONENTELE PRINCIPALE

1. Interfața cu utilizatorul (UI)
2. Modulul OCR
3. Conversia fișierelor
4. Prelucrarea datelor
5. Organizarea fișierelor
6. Exportul datelor
7. Statistici live și feedback vizual

STIL ARHITECTURAL ȘI ABORDARE TEHNICĂ

OCR230 adoptă un stil apropiat de modelul client-local + modular monolit, în care aplicația rulează local și înglobează module clar separate. Deși nu este o aplicație distribuită, designul modular permite în viitor migrarea către microservicii sau integrarea cu baze de date externe. Printre principiile arhitecturale respectate se numără:

- **Separarea responsabilităților** – fiecare modul are un rol unic și bine definit.
- **Extensibilitate** – arhitectura modulară permite adăugarea de noi funcții.
- **Portabilitate** – aplicația poate rula pe orice sistem, având totul local.
- **Accesibilitate** – interfața este simplă și pregătită pentru toți utilizatorii.

TEHNOLOGII FOLOSITE ȘI STABILITATEA APLICAȚIEI

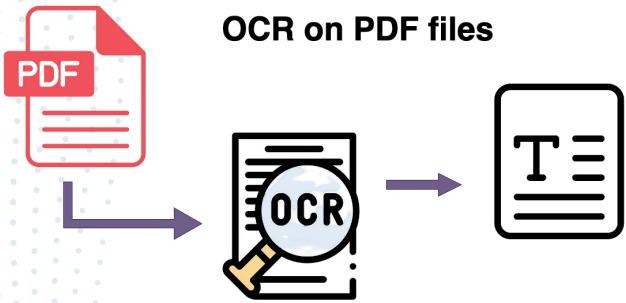
Aplicația **OCR230** este dezvoltată exclusiv cu tehnologii **open-source**, care s-au dovedit a fi **foarte stabile și eficiente** în prelucrarea imaginilor și extragerea textului din **formulare 230**, asigurând o **experiență optimă** chiar și pentru utilizatorii fără cunoștințe tehnice.

Printre tehnologiile cheie se numără **Python** pentru flexibilitate și modularitate, **EasyOCR și EfficientOCR** pentru recunoaștere avansată a textului, **pdf2image** pentru conversia PDF-urilor în imagini, precum și **numpy și Pillow** pentru procesarea imaginilor. Interfața grafică este realizată cu **Tkinter,** folosind elemente **canvas** pentru control vizual detaliat, iar **threading** optimizează procesarea paralelă, crescând viteza și responsivitatea aplicației. De asemenea, **Canva** a fost utilizată pentru crearea unui ghid vizual accesibil.

OCR230 rulează complet local, fără a necesita conexiune la internet sau trimiterea datelor către servere externe, ceea ce **garantează securitatea și confidențialitatea** informațiilor sensibile, precum CNP.

Din punct de vedere al stabilității, aplicația ocupă aproximativ **900 MB**, majoritar datorită librăriilor OCR complexe. Procesarea paralelă prin threading permite menținerea unei interfețe fluide, fără blocaje, chiar și la volume mari de date. **Consumul de memorie este bine gestionat**, cu **eliberarea rapidă a resurselor temporare,** evitând astfel scurgerile de memorie. Spațiul pe disc este utilizat eficient, cu fișiere generate organizat pe localități și ani.

Astfel, **OCR230** oferă o soluție robustă, performantă și sigură, bazată pe tehnologii dovedite, adaptată nevoilor ONG-urilor care gestionează formularele 230.



MATURITATEA ȘI GESTIONAREA PROFESIONALĂ A DEZVOLTĂRII APLICAȚIEI

Aplicația **OCR230** a atins un **nivel solid de maturitate,** fiind **complet funcțională și stabilă**, cu **toate modulele esențiale implementate**: OCR, organizare automată a fișierelor, interfață grafică intuitivă și funcții de securitate locală. Timpul mediu de procesare pentru un formular este de aproximativ **5 secunde**, ceea ce contribuie semnificativ la eficientizarea activității ONG-urilor.

Până în prezent, **peste 3.000 de formulare** au fost procesate prin aplicație de către **6 ONG-uri partenere**, în condiții reale de utilizare. Feedback-ul oferit de aceștia a fost valorificat pentru îmbunătățirea interfeței și funcționalităților. Aplicația este disponibilă **public** în **Microsoft Store**, fiind **ușor de instalat** și folosit fără cunoștințe tehnice avansate.

Dezvoltarea aplicației a fost susținută de un proces bine organizat de versionare prin **Git**, cu repository găzduit pe **GitHub**. Evoluția codului este documentată prin **commit-uri** clare și descriptive. S-au utilizat **branch-uri separate** pentru dezvoltări noi și teste, evitând instabilitatea codului principal. Acest workflow a permis **testarea riguroasă** și **integrarea controlată** a modificărilor.

Prin acest sistem, **OCR230** este nu doar o **aplicație matură** și robustă, ci și un proiect bine gestionat din punct de vedere tehnic, pregătit pentru scalare și extindere viitoare.

CUM FUNCȚIONEAZĂ APLICAȚIA

- **Selectarea folderelor**

Utilizatorul selectează folderul de input (cu formularele F230 – scanate) și un folder de output, unde vor fi salvate fișierele procesate și structurate.

- **Procesarea automată**

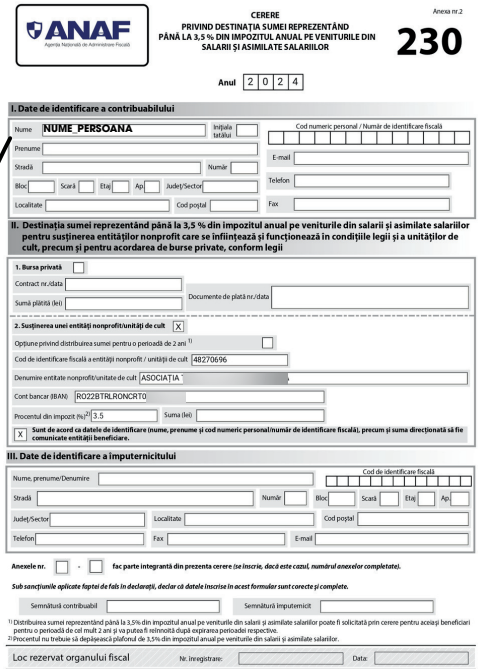
Aplicația parcurge fiecare formular, aplică OCR, extrage datele esențiale (nume, CNP, adresă etc.) și redenumește fișierele în mod inteligent. Acestea sunt apoi organizate automat în subfoldere (ex: pe ani sau localități). Pe parcursul procesării, aplicația afișează în **Dasboardul live** diverse statistici.

- **Exportul datelor**
  - **Excel centralizat** cu toate datele structurate
  - **CSV** pentru compatibilitate și backup
  - **PDF** cu rapoarte și statistici complete
  - **TXT** individual pentru fiecare persoană
- **Completarea borderoului (B230)**

ONG-ul folosește copy-paste pentru a transfera datele din fișierul generat de aplicație în formularul oficial B230, necesar pentru depunerea la ANAF.

fig. 5 - Dashboard **live** cu statistici

UTILITATEA APLICAȚIEI



ANAF
Agenția Națională de Administrare Fiscală

CERERE
PRIVIND DESTINAȚIA SUMEI REPREZENTÂND PÂNĂ LA 3,5 % DIN IMPOZITUL ANUAL PE VENITURILE DIN SALARII ȘI ASIMILATE SALARIILOR

Anexa nr.2
230

Anul 2024

I. Date de identificare a contribuabilului
Nume NUME_PERSOANA Inițiala tatălui
Cod numeric personal / Număr de identificare fiscală
Prenume
E-mail
Stradă Număr
Telefon
Bloc Scară Etaj Ap. Județ/Sector
Localitate Cod poștal Fax

II. Destinația sumei reprezentând până la 3,5 % din impozitul anual pe veniturile din salarii și asimilate salariilor pentru susținerea entităților nonprofit care se înființează și funcționează în condițiile legii și a unităților de cult, precum și pentru acordarea de burse private, conform legii

1. Bursa privată
Contract nr./data
Sumă plătită (lei) Documente de plată nr./data

2. Susținerea unei entități nonprofit/unități de cult X
Opțiune privind distribuirea sumei pentru o perioadă de 2 ani 1)
Cod de identificare fiscală a entității nonprofit / unității de cult 48270696
Denumire entitate nonprofit/unitate de cult ASOCIAȚIA
Cont bancar (IBAN) RO22BTRLRONCRT0
Procentul din impozit (%)2) 3.5 Suma (lei)
X Sunt de acord ca datele de identificare (nume, prenume și cod numeric personal/număr de identificare fiscală), precum și suma direcționată să fie comunicate entității beneficiare.

III. Date de identificare a Imputernicitului
Nume, prenume/Denumire
Cod de identificare fiscală
Stradă Număr Bloc Scară Etaj Ap.
Județ/Sector Localitate Cod poștal
Telefon Fax E-mail

Anexele nr. - fac parte integrantă din prezenta cerere (se înscrie, dacă este cazul, numărul anexelor completate).

Sub sancțiunile aplicate faptei de fals în declarații, declar că datele înscrise în acest formular sunt corecte și complete.

Semnătura contribuabil Semnătura Imputernicit

1) Distribuirea sumei reprezentând până la 3,5% din impozitul anual pe veniturile din salarii și asimilate salariilor poate fi solicitată prin cerere pentru aceiași beneficiari pentru o perioadă de cel mult 2 ani și va putea fi reînnoită după expirarea perioadei respective.
2) Procentul nu trebuie să depășească plafonul de 3,5% din impozitul anual pe veniturile din salarii și asimilate salariilor.

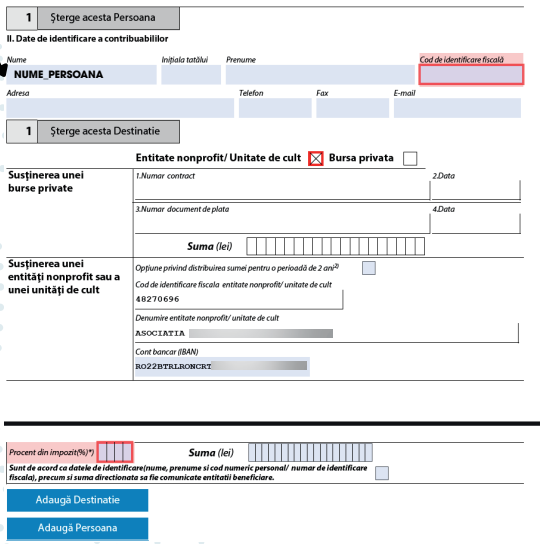
Loc rezervat organului fiscal Nr. înregistrare: Data:

Document care conține date cu caracter personal protejate de prevederile Regulamentului (UE) 2016/679
Cod 14.13.04.13

OCR

fig. 3 - Formular F230[4]

- Aplicația scanează fiecare câmp din formularul F230, completat de persoana fizică.
- Folosind modulul **OCR**, aplicația extrage textul din câmpul dat, pe care apoi îl prelucrează (Ex. se validează, se scriu în excel, csv, txt etc.)
- Apoi, aplicația ia următorul câmp si folosește aceleași iterații.



1 Șterge acesta Persoana

II. Date de identificare a contribuabililor
Nume NUME_PERSOANA Inițiala tatălui Prenume Cod de identificare fiscală
Adresa Telefon Fax E-mail

1 Șterge acesta Destinatie

Entitate nonprofit/ Unitate de cult ☒ Bursa privata ☐

Susținerea unei burse private
1.Numar contract 2.Data
3.Numar document de plata 4.Data
Suma (lei)

Susținerea unei entități nonprofit sau a unei unități de cult
Opțiune privind distribuirea sumei pentru o perioadă de 2 ani2)
Cod de identificare fiscala entitate nonprofit/ unitate de cult
48270696
Denumire entitate nonprofit/ unitate de cult
ASOCIATIA
Cont bancar (IBAN)
RO22BTRLRONCRT

Procent din impozit(%)*) Suma (lei)
Sunt de acord ca datele de identificare(nume, prenume si cod numeric personal/ numar de identificare fiscala), precum si suma directionata sa fie comunicate entitatii beneficiare.

Adaugă Destinatie
Adaugă Persoana

fig. 4 - Borderou B230[5]

- Pentru ca ANAF să vireze fondurile strânse de ONG-uri în conturile bancare, fiecare ONG trebuie să depună **borderoul B230**, în care sunt trecute datele de identificare ale fiecărei persoane care a completat formularul F230.
- Nu numai că persoanele ce completează borderorile sunt ajutate, având posibilitatea să ia datele cu **COPY & PASTE,** dar și persoanele care se ocupă de contabilitatea ONGului au o gestiune mai bună.

EXEMPLE DE OUTPUT

Aplicația generează trei tipuri principale de fișiere ca rezultat al procesării: **PDF, Excel și CSV**, fiecare având un rol distinct. **Fișierul PDF** este un **raport vizual,** profesional, care conține **grafice și tabele sintetizate**, utile pentru prezentări sau arhivare, fiind generat cu ajutorul bibliotecilor **matplotlib și reportlab**. Fișierul **Excel** este structurat tabelar, cu toate **datele extrase** din formulare (nume, CNP, adresă etc.), **formatat clar** pentru a permite corecturi manuale, sortări sau completări ulterioare și este deschis automat la finalul rulării. Fișierul **CSV**, **simplu și compatibil** cu majoritatea platformelor, este destinat integrării rapide cu **baze de date** sau platforme externe, fiind ideal pentru importuri automate în sisteme ANAF sau alte softuri contabile.

Distributia pe ANAF:

ANAF	Numar persoane	Procent
Administratia Judeteana a Finantelor Publice Prahova mic	21	42.0%
Administratia Sectorului 1 a Finantelor Publice	5	10.0%
Unitatea Fiscala Oraseneasca Boldesti Scaeni	4	8.0%
Habud	4	8.0%
Administratia Sectorului 3 a Finantelor Publice	2	4.0%
Administratia Sectorului 6 a Finantelor Publice	2	4.0%
Baraitaru	2	4.0%
Strejnicu	1	2.0%
Administratia Sectorului 4 a Finantelor Publice	1	2.0%
Administratia Sectorului 5 a Finantelor Publice	1	2.0%

Verificare Duplicate:
- CNP-uri duplicate gasite: 0
- Total inregistrari duplicate: 0

Statistici Generale:
- Total inregistrari procesate: 50
- Înregistrari cu CNP valid: 50
- Persoane donatoare pe 2 ani: 27
- Persoane donatoare pe 1 an: 23
- Data generare raport: 30/06/2025 17:31

fig. 6 - Statistici export PDF

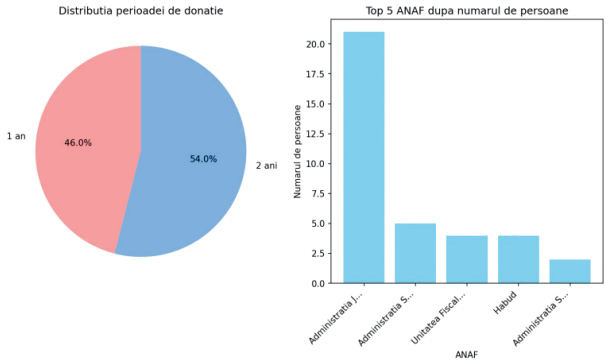


fig. 7 - Grafice export PDF

SUMARUL REZULTATELOR

În plus, pentru a elimina orice barieră tehnică și a face aplicația cât mai prietenoasă, OCR230 este concepută astfel încât să deschidă automat toate fișierele relevante imediat după finalizarea procesării. Mai exact, fișierul Excel generat cu datele extrase, raportul PDF creat și folderul în care au fost salvate toate rezultatele sunt deschise automat, fără ca utilizatorul să fie nevoit să le caute manual în structura de fișiere.
Această funcționalitate este extrem de utilă mai ales pentru utilizatorii fără experiență tehnică, deoarece le permite să acceseze imediat rezultatele muncii lor, fără pași suplimentari sau confuzii. Astfel, procesul este fluid, rapid și intuitiv, contribuind la o experiență pozitivă și eficientă, indiferent de nivelul de familiaritate cu calculatorul.

fig. 9 - Pagina de Rapoarte în aplicație

AJUTOR PENTRU PERSOANELE FĂRĂ CUNOȘTIINȚE TEHNICE

Persoanele fără cunoștințe tehnice se pot confrunta cu dificultăți atunci când utilizează aplicații specializate, mai ales în contexte administrative, cum este procesarea formularelor 230.
Pentru a răspunde acestei nevoi, OCR230 include un mod de ajutor dedicat utilizatorilor non-tehnici, care explică fiecare pas într-un limbaj clar, accesibil.
Ghidul vizual integrat, completat de mesaje explicative în aplicație, permite chiar și voluntarilor fără experiență în domeniul IT să folosească programul cu ușurință și în siguranță.

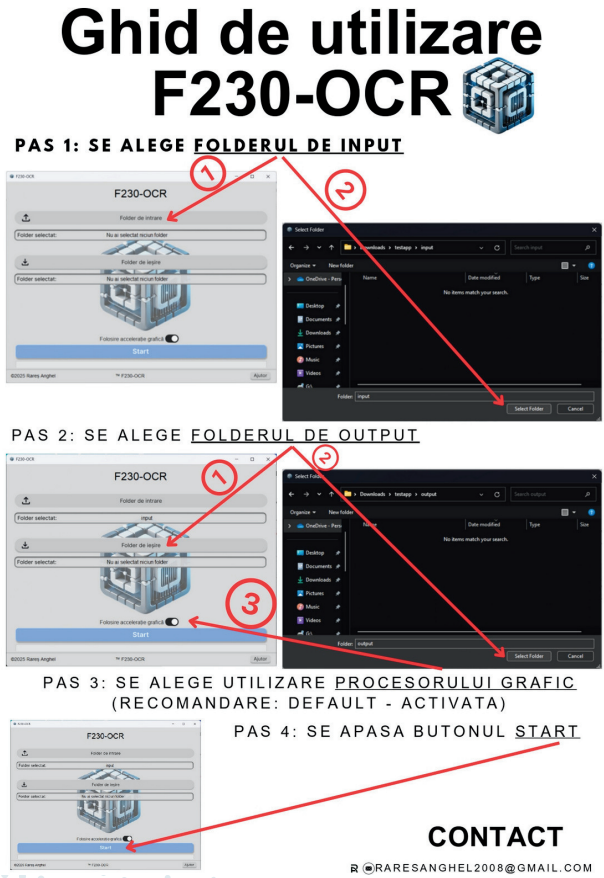
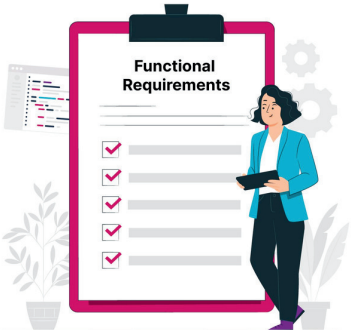


fig. 8 - Ghidul aplicației OCR230

CERINȚE ȘI SPECIFICAȚII

01 CERINȚE FUNCȚIONALE

Aplicația trebuie să permită încărcarea formularelor 230 în format PDF sau imagine, să extragă automat datele relevante prin OCR și să le organizeze în mod structurat, oferind posibilitatea exportului în format text sau Excel.



02 SPECIFICAȚII TEHNICE

Aplicația este dezvoltată în Python, folosind biblioteci precum EasyOCR pentru recunoașterea textului, Tkinter pentru interfața grafică și Pillow, PDF2Image, Numpy pentru procesarea imaginilor. Rulează local, fără conexiune la internet, și este compatibilă cu Windows. Spațiul ocupat este de aproximativ 900 MB, iar timpul mediu de procesare per formular este de 6 secunde.

03 CERINȚE DE INTEGRARE

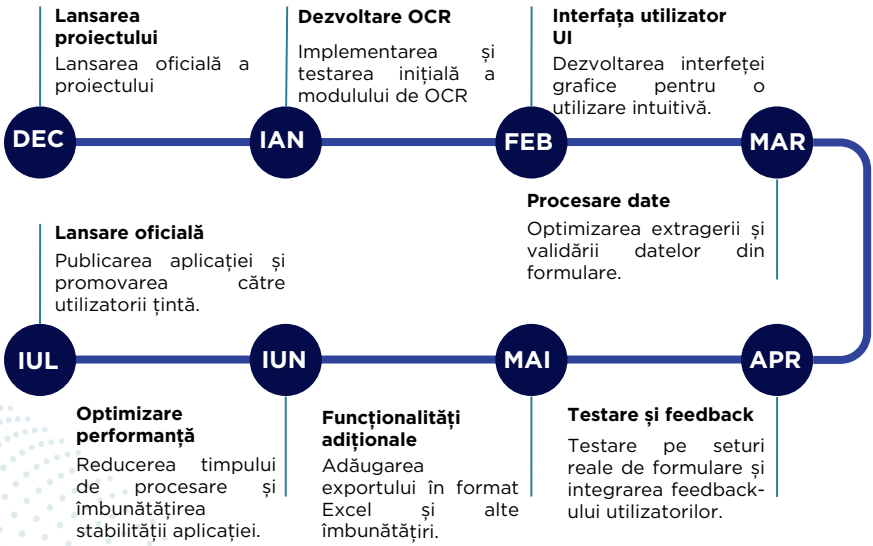
Aplicația nu necesită integrare cu servicii externe și funcționează complet local. Poate fi instalată direct din Microsoft Store pe sisteme Window. Aceaste permite exportul datelor într-un fișier Excel pentru integrare ulterioară cu alte sisteme de gestiune.

PROJECT TIMELINE

IDEEA PROIECTULUI

Ideea cu care am pornit proiectul a fost de a automatiza și simplifica munca laborioasa de procesare a formularelor F230. Astfel, am creeat un soft care permitea acest lucru, care apoi a evoluat într-o aplicație modernă, oferind astfel o soluție accesibilă și eficientă pentru ONG-uri și utilizatori individuali.

REPERE IMPORTANTE & CRONOLOGIE



ALOCAREA RESURSELOR ȘI DEPENDENȚE

Resurse umane

Responsabil dezvoltator principal pentru programare, testare și mentenanță. Colaborare cu ONG-uri pentru feedback și validare.

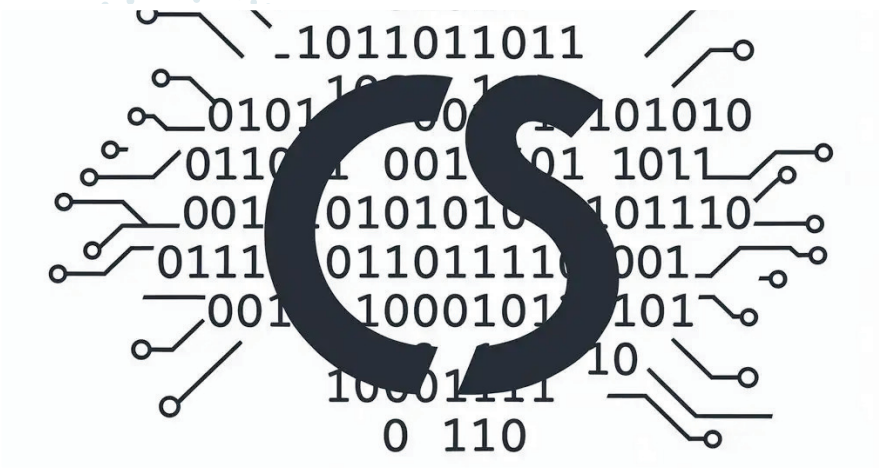
Resurse materiale

Calculator cu performanțe medii, software Python, biblioteci OCR (EasyOCR, EffOCR), mediul de dezvoltare și resurse grafice.

Dependencies

Biblioteci Python (easyocr, numpy, Pillow, pdf2image, tkinter), date oficiale ANAF pentru validare, acces la Microsoft Store pentru distribuție.

DESIGN ȘI ARHITECTURĂ



01

MODULARITATE

Aplicația este structurată în module clar delimitate, fiecare responsabil pentru o funcționalitate specifică (OCR, procesare, interfață). Această separare facilitează mentenanța, testarea și extinderea ulterioară a codului.

02

RESPONSIVITATE ȘI PARALELIZARE

Procesarea formularelor se realizează pe fire de execuție separate, astfel încât interfața rămâne fluidă și utilizatorul beneficiază de o experiență fără blocaje sau întârzieri.

03

SECURITATE ȘI FIABILITATE

Aplicația rulează local, validând input-urile și tratând excepțiile pentru a preveni erorile critice. Acest design asigură protecția datelor și stabilitatea funcționării în condiții variate.

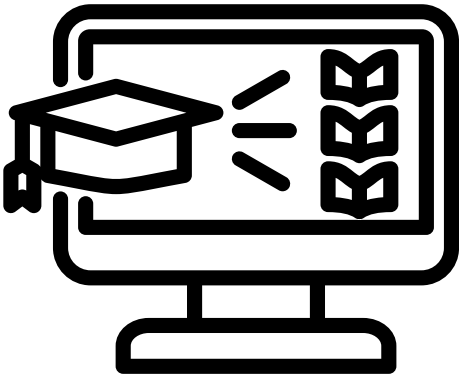
04

ROBUSTEȚE ȘI SECURITATE

Aplicația este construită să gestioneze erorile și intrările invalide în mod eficient, prevenind blocările și protejând datele utilizatorilor prin validări stricte și tratarea excepțiilor.

MODULARITATEA APLICAȚIEI

- Backend & Procesare: Python 3.13+
- OCR Engine: EasyOCR + EffOCR (model custom)
- Interfață grafică: Tkinter cu design modern realizat in FIGMA
- Export date: Excel (openpyxl) + TXT + CSV
- Structură modulară:
  - src/ocr/ – Recunoaștere text și inițializare engine-uri OCR
  - src/processing/ – Extracție și validare date, coordonate zone pentru OCR
  - src/ui/ – Interfață grafică (splash, main_window, rapoarte)
  - src/anaf/ – Module pentru validare județe și sectoare ANAF
  - src/excel/ – Export incremental și gestionare fișier Excel
  - src/utils/, src/names/ – Funcții auxiliare și validare nume
  - Assets/ – Resurse grafice și icons



DETALII DE IMPLEMENTARE

DEZVOLTARE SOFTWARE

Crearea codului modular și integrarea funcționalităților esențiale, precum OCR-ul și interfața grafică, pentru a asigura procesarea rapidă și precisă a formularelor.

ACTIVITĂȚI CHEIE

- Crearea și organizarea codului modular
- Implementarea motorului OCR și procesarea imaginii
- Dezvoltarea interfeței grafice intuitive
- Gestionarea erorilor și validarea datelor
- Integrarea funcției de export Excel
- Testare funcțională și optimizare performanță

MANAGEMENTUL PROIECTULUI

Coordinarea comunicării cu utilizatorii, urmărirea versiunilor în Git și realizarea documentației, pentru a menține proiectul organizat și adaptat cerințelor reale.

ACTIVITĂȚI CHEIE

- Colectarea și integrarea feedback-ului de la utilizatori
- Versionarea codului și gestionarea repository-ului Git
- Planificarea și prioritizarea sarcinilor din roadmap
- Documentarea tehnică și actualizarea ghidurilor de instalare
- Comunicarea și coordonarea cu ONG-urile partenere

PROJECT MANAGEMENT

TESTARE ȘI ASIGURAREA CALITĂȚII

Functional Testing

TESTARE FUNCȚIONALĂ

Această etapă implică testarea aplicației pe o gamă largă de scenarii reale, folosind diverse formulare 230 pentru a verifica dacă toate funcțiile rulează corect și extragerea datelor se face precis și complet. Scopul este de a asigura conformitatea cu cerințele specificate și de a detecta eventualele erori sau comportamente neașteptate înainte de lansare.

TESTARE PERFORMANȚĂ ȘI STABILITATE

Se monitorizează timpul de procesare pentru fiecare formular și consumul de resurse pe diferite configurații hardware, pentru a evalua eficiența și stabilitatea aplicației. Această testare asigură că aplicația rămâne responsivă și funcțională în condiții variate de utilizare, fără să afecteze negativ experiența utilizatorului.

QUALITY ASSURANCE

AUDIT ȘI REVIZUIRE COD

Se realizează revizuiri periodice ale codului sursă, atât manual, cât și cu ajutorul instrumentelor automate, pentru a garanta calitatea, claritatea și mentenabilitatea codului. Aceste audituri ajută la identificarea și corectarea timpurie a problemelor, menținând un standard ridicat de calitate și facilitând extinderile viitoare.

LUCRUL ÎN ECHIPĂ ȘI COLABORAREA CU UTILIZATORII

Deși dezvoltarea aplicației **OCR230** a fost realizată **integral de mine**, succesul proiectului a fost posibil datorită colaborării strânse cu **utilizatori reali și reprezentanți ai mai multor ONG-uri.** Această cooperare activă a avut un rol esențial în modelarea aplicației conform nevoilor practice ale beneficiarilor finali.

Fiind **singurul dezvoltator**, mi-am asumat toate responsabilitățile legate de proiectare, scrierea codului, testarea funcționalităților, gestionarea erorilor, optimizarea performanței, dar și redactarea documentației tehnice și pregătirea materialelor de suport pentru utilizatori. În paralel, am menținut un **contact permanent** cu organizațiile care au testat aplicația în condiții reale, ceea ce mi-a permis să obțin un **flux constant de feedback concret și aplicabil.**

Pentru **comunicare** am folosit emailul și platforme de mesagerie precum **WhatsApp și Facebook** prin care am primit **observații punctuale**, sugestii de îmbunătățiri și raportări de eventuale disfuncționalități. Fiecare rundă de feedback a fost analizată și integrată în dezvoltarea ulterioară a aplicației, contribuind la **rafinarea interfeței, simplificarea fluxurilor de lucru și creșterea stabilității generale.**

În ceea ce privește organizarea muncii, am folosit sistemul Git pentru versionarea codului și urmărirea progresului în timp. Chiar dacă nu am lucrat într-o echipă de dezvoltatori, am aplicat bune practici de gestionare a versiunilor și izolare a **funcționalităților noi** în **branch-uri separate.** Acest mod de lucru a permis menținerea unui **cod stabil și ușor de actualizat.**

Implicarea activă a ONG-urilor în procesul de testare a avut un impact semnificativ asupra calității aplicației. Sesiunile de testare practică au ajutat la identificarea unor cazuri particulare de utilizare, pe care nu le-aș fi anticipat în faza de dezvoltare. Astfel, aplicația a evoluat nu doar din punct de vedere tehnic, ci și funcțional, devenind mai prietenoasă și mai eficientă pentru scopul său: automatizarea procesării formularelor 230.

VULNERABILITĂȚI

Vulnerabilitățile într-o aplicație pot apărea din erori de programare, lipsa validării datelor sau interacțiuni neașteptate cu fișiere. În contextul OCR230, datele personale sensibile și fișierele manipulate necesită o atenție deosebită pentru a preveni scurgeri sau coruperi.

01. IDENTIFICAREA VULNERABILITĂȚILOR

Procesul de identificare implică testarea funcționalităților aplicației pentru a detecta punctele slabe înainte ca acestea să devină probleme reale. Se analizează fluxurile de date, modul de acces la fișiere și comportamentul aplicației în cazuri-limită.

COMPONENTE CHEIE:

01. Testare cu date corupte sau malformate

Aplicația este verificată prin introducerea unor fișiere PDF sau imagini deteriorate, incomplete sau greșit formatate, pentru a evalua cum reacționează la inputuri neprevăzute. Acest tip de testare ajută la descoperirea crash-urilor, a lipsei de validare și a eventualelor blocaje în procesare.

02. Analiza accesului la fișiere

Se monitorizează modul în care aplicația gestionează fișierele din folderul de input și output, inclusiv permisiunile de scriere, citire și redenumire. Astfel, se evită situațiile în care fișiere importante pot fi suprascrise accidental, corupte sau pierdute din cauza accesului incorect.

02. PREVENIREA & REZOLVAREA VULNERABILITĂȚILOR

Prevenirea implică bune practici de programare și verificări constante, iar rezolvarea presupune corectarea rapidă a problemelor identificate și actualizarea aplicației.

COMPONENTE CHEIE:

01. Validare strictă a datelor

Fiecare informație extrasă (nume, CNP, localitate etc.) este verificată înainte de a fi introdusă în fișierele de ieșire. Algoritmii de validare previn erorile logice (de exemplu, CNP-uri incorecte sau câmpuri goale), reducând șansele ca datele greșite să ajungă în rapoartele finale.

02. Rapoarte de eroare clare

Atunci când aplicația întâlnește o problemă (ex. formular imposibil de citit, câmp lipsă, imagine neclară), generează automat un mesaj de eroare clar, atât în consolă, cât și în raportul de procesare. Utilizatorul este informat rapid și poate corecta sau reîncerca procesarea fără intervenții tehnice complexe.

CUM PREVENIM VULNERABILITĂȚILE

OCR230 este o aplicație care procesează formulare fiscale de tip 230, documente ce conțin date personale extrem de sensibile – precum nume complet, adresă, CNP și semnături ale contribuabililor. În acest context, orice breșă de securitate sau vulnerabilitate poate avea consecințe grave, atât pentru organizațiile utilizatoare, cât și pentru persoanele fizice implicate. Tocmai de aceea, OCR230 a fost proiectată având în centru siguranța datelor și reziliența aplicației.

OCR230 este concepută pentru a preveni din start expunerea datelor și apariția erorilor critice. Printre măsurile-cheie implementate:

- Execuție complet locală, fără conexiuni externe
- Aplicația rulează exclusiv pe calculatorul utilizatorului, fără a transmite date către servere externe și fără a necesita conexiune la internet. Toate librăriile și resursele necesare sunt instalate local, reducând semnificativ riscul de interceptare, scurgere de date sau atacuri cibernetice.
- Validare și gestionare robustă a datelor
- Fiecare fișier este verificat înainte de procesare: CNP-urile sunt validate cu algoritmul oficial, iar datele extrase sunt filtrate pentru a exclude caractere sau simboluri eronate. Astfel se asigură coerența și corectitudinea datelor și se minimizează riscul de corupere a fișierelor finale.

OBIECTIVE

OBIECTIVELE PROIECTULUI

Dezvoltarea unei aplicații eficiente și accesibile pentru automatizarea procesării formularelor 230, reducând timpul și efortul manual. Se urmărește o soluție stabilă, cu interfață intuitivă și extragere precisă a datelor, adaptată nevoilor ONG-urilor, cu posibilitatea extinderii prin tehnologii moderne, inclusiv cloud și AI.

OBIECTIVE SPECIFICE

01. OPTIMIZAREA PERFORMANȚEI OCR

Obiectiv : Creșterea vitezei de procesare a formularelor, în special pe sisteme cu hardware redus.

Target : Implementarea unui model de OCR cu o acuratețe mai mare și care ocupă puțină memorie.

Rezultat : Aplicație mai rapidă și mai accesibilă pentru utilizatori cu calculatoare mai puțin performante.

02. INTEGRAREA ÎN CLOUD

Obiectiv : Implementarea suportului pentru procesare și stocare în cloud.

Target : Dezvoltarea unei versiuni care să poată sincroniza datele și formularele printr-o platformă online.

Rezultat : Facilitarea accesului și colaborării între utilizatori, precum și scalabilitate sporită a aplicației.

INDICATORI CHEIE DE PERFORMANȚĂ

INDICATOR CHEIE #01 Procesare în sub 8 secunde per formular

INDICATOR CHEIE #02 Peste 98,9% rată de recunoaștere

INDICATOR CHEIE #03 Implementat în 5 ONG-uri din România

INDICATOR CHEIE #04 Rating pozitiv > 4.5/5

INDICATOR CHEIE #05 Zero incidente de securitate sau confidențialitate

CONCLUZIE ȘI ACTIVITĂȚI VIITOARE

CONCLUZIE

Proiectul OCR230 a demonstrat că automatizarea procesării formularelor 230 este posibilă și eficientă, reducând semnificativ timpul și efortul necesar comparativ cu metoda manuală. Aplicația este stabilă, funcțională și ușor de utilizat, fiind deja implementată cu succes în mai multe ONG-uri. Structura modulară și codul bine organizat permit extinderea și îmbunătățirea continuă, oferind un potențial crescut pentru dezvoltări viitoare.

Realizări:

Aplicația a procesat cu succes **peste 3.000 de formulare,** este **disponibilă în Microsoft Store** și a primit **feedback pozitiv**, demonstrând utilitatea și stabilitatea sa în **condiții reale**.

Impact:

OCR230 a simplificat și accelerat semnificativ procesarea formularelor 230, reducând erorile și timpul de lucru pentru ONG-urile mici și mijlocii.

ACTIVITĂȚI VIITOARE

Optimizare performanță pe hardware redus și integrare cloud	Testarea și ajustarea aplicației pentru a rula eficient pe sisteme cu resurse limitate, plus explorarea integrării în cloud pentru procesare rapidă și scalabilă.
Revizuirea completă a interfeței	Redesign total al UI pentru a îmbunătăți experiența utilizatorului, cu focus pe accesibilitate și intuitivitate.
Dezvoltarea modelelor AI personalizate	Antrenarea și integrarea unor modele OCR bazate pe inteligență artificială pentru o recunoaștere mai precisă a formularelor 230.
Implementarea unui sistem de actualizare automată a aplicației	O funcționalitate care să permită utilizatorilor să primească și să instaleze cele mai noi versiuni ale aplicației direct din interfață, fără descărcări manuale.

RECENZII

În perioada de testare și feedback, aplicația OCR230 a fost utilizată activ de mai multe ONG-uri partenere, care au oferit observații valoroase legate de funcționalitate, interfață și performanță. Pe lângă sugestiile tehnice implementate în versiunile ulterioare, am primit și recenzii din partea utilizatorilor privind experiența generală de lucru cu aplicația.
Mai jos sunt câteva dintre recenziile primite în această etapă:

★ ★ ★ ★ ★

"Aplicația a redus semnificativ timpul necesar procesării formularelor 230 în ONG-ul nostru. În loc să petrecem ore întregi cu verificarea manuală, acum procesăm rapid și eficient toate documentele. Recomand cu încredere!„

Georgiana A., Asociația Totul Pentru Maria Rebeca

★ ★ ★ ★ ★

"Este un instrument esențial pentru ONG-urile mici ca al nostru, care nu dispun de resurse pentru soluții complexe sau costisitoare. O soluție accesibilă și foarte eficientă!„

Voluntar ONG

RESURSE EXTERNE

În dezvoltarea aplicației OCR230 am folosit resurse externe open-source și date oficiale, atent selectate pentru a asigura performanță, stabilitate și securitate.
Biblioteci open-source:

- **EasyOCR** – Componenta principală pentru recunoaștere optică a caracterelor, cu suport bun pentru limba română și rezultate fiabile chiar pe imagini de calitate variabilă.
- **EffOCR** – Utilizat complementar pentru a crește acuratețea în zonele dificile, bazat pe rețele neuronale optimizate.
- **Numpy, pdf2image, Pillow (PIL)** – Pentru procesarea imaginilor, conversia PDF-urilor și manipulări grafice necesare OCR-ului.
- **Tkinter** – Biblioteca grafică folosită pentru realizarea interfeței, datorită integrării simple și personalizabile.
- **Threading** – Permite rularea OCR-ului în paralel, reducând timpul de procesare și menținând interfața responsivă.

Date oficiale:
Datele privind corespondența localităților cu administrațiile fiscale au fost preluate de pe site-ul ANAF:
https://static.anaf.ro/static/10/Anaf/AsistentaContribuabili_r/telefoane_judete/Regiuni.htm
Acestea au fost integrate local pentru validarea automată a formularelor, fără a fi necesară conexiunea la internet.
Toate resursele au fost alese pentru compatibilitate, licențiere permisivă și pentru a menține aplicația complet locală și sigură.

GLOSAR

1. **CNN** - rețea neuronală convoluțională, tip de algoritm de învățare profundă, potrivit în special pentru analiza datelor vizuale, cum ar fi imaginile și videoclipurile.
2. **LSTM** - Long Short-Term Memory, tip de arhitectură de rețea neuronală recurentă (RNN) concepută pentru a aborda problema gradientului de dispariție întâlnită în RNN-urile standard, în special atunci când se lucrează cu secvențe lungi.
3. **CTC** - Connectionist Temporal Classification, un tip de ieșire a unei rețele neuronale utilizat pentru modelarea seriilor temporale care tratează date secvențiale, de exemplu, RNN și LSTM.
4. **Formular F230** - formularul pe care o persoană îl completează pentru a redirecționa 3,5% din impozit unui ONG
5. **Borderou B230** - Borderoul care cuprinde formularele F230, pe care fiecare ONG îl depune la ANAF

OCR230

© Anghel Rares Mihai