

# Ekstrakcija teksta iz stripa Alan Ford

---

Čagalj, Filipa

Master's thesis / Diplomski rad

2023

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Split, Faculty of Science / Sveučilište u Splitu, Prirodoslovno-matematički fakultet**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:166:493035>

Rights / Prava: [Attribution 4.0 International](#)/[Imenovanje 4.0 međunarodna](#)

Download date / Datum preuzimanja: **2024-07-07**

Repository / Repozitorij:

[Repository of Faculty of Science](#)



SVEUČILIŠTE U SPLITU  
PRIRODOSLOVNO MATEMATIČKI FAKULTET

DIPLOMSKI RAD

**EKSTRAKCIJA TEKSTA IZ STRIPA ALAN  
FORD**

Filipa Čagalj

Split, srpanj 2023.

## Temeljna dokumentacijska kartica

Sveučilište u Splitu  
Prirodoslovno-matematički fakultet  
Odjel za informatiku  
Ruđera Boškovića 33, 21000 Split, Hrvatska

Diplomski rad

### Ekstrakcija teksta iz stripa Alan Ford Filipa Čagalj

U ovom diplomskom radu istražujemo metode i alate za prepoznavanje oblačića i ekstrakciju slika u stripovima, te evaluiramo njihovu točnost i korisnost na primjerima stripa Alan Ford. Posebno se fokusiramo na prilagodbu odabranog alata za analizu Alan Ford stripova. Analiziramo relevantnu literaturu, opisujemo korištene metode i procjenjujemo njihovu učinkovitost. Rad zaključujemo s glavnim nalazima i smjernicama za buduća istraživanja. Ovaj rad pomaže umjetnicima, izdavačima stripova, čitateljima i znanstvenicima u analizi i stvaranju stripova.

**Ključne riječi:** strip, prepoznavanje oblačića, alati, evaluacija, analiza stripova

Rad je pohranjen u knjižnici Prirodoslovno-matematičkog fakulteta, Sveučilišta u Splitu

**Rad sadrži:** 49 stranica, 19 grafičkih prikaza, 2 tablice i 18 literaturnih navoda. Izvornik je na hrvatskom jeziku.

**Mentor:** Izv. prof. dr. sc. Branko Žitko, izvanredni profesor Prirodoslovno-matematičkog fakulteta, Sveučilište u Splitu

**Ocjenjivači:** Izv. prof. dr. sc. Branko Žitko, izvanredni profesor Prirodoslovno-matematičkog fakulteta, Sveučilište u Splitu

Izv. prof. dr. sc. Ani Grubišić, izvanredni profesor Prirodoslovno-matematičkog fakulteta, Sveučilište u Splitu

Lucija Bročić, asistent Prirodoslovno-matematičkog fakulteta, Sveučilište u Splitu

Rad prihvaćen: Srpanj, 2023.

## Basic documentation card

University of Split  
Faculty of Science  
Department of Informatics  
Ruđera Boškovića 33, 21000 Split, Croatia

Diploma Thesis

### Text Extraction from Alan Ford Comic Book

Filipa Čagalj

In this thesis, we investigate methods and tools for speech bubble recognition and image extraction in comics and evaluate their accuracy and usefulness on the examples of Alan Ford comics. In particular, we focus on customizing the selected tool to analyze Alan Ford comics. We analyze the relevant literature, describe the methods used and evaluate their effectiveness. We conclude the paper with the main findings and guidelines for future research. This work helps artists, comics publishers, readers, and scholars analyze and create comics.

**Keywords:** comics, speech balloon recognition, tools, evaluation, comics analysis

Thesis deposited in the library of Faculty of Science, University of Split

**Thesis consists of:** 49 pages, 19 figures, 2 tables and 18 references, original in: Croatian

**Assistant Supervisor:** Branko Žitko, Ph.D. Associate Professor of Faculty of Science, University of Split

**Reviewers:** Branko Žitko, Ph.D. Associate Professor of Faculty of Science, University of Split

Ani Grubišić, Ph.D. Associate Professor of Faculty of Science, University of Split

Lucija Bročić, Assistant of Faculty of Science, University of Split

Thesis accepted: July, 2023.

## Sadržaj

1	UVOD .....	1
2	PREGLED STRUKTURE RADA .....	2
2.1	Definiranje istraživačkog pitanja i ciljevi istraživanja .....	2
2.2	Motivacija za istraživanje.....	3
3	PREGLED ISTRAŽIVANJA .....	5
3.1	Pregled povijesti stripa i uloge oblačića u stripovima.....	5
3.2	Pregled postojećih istraživanja o prepoznavanju oblačića u stripovima.....	6
3.3	Analiza prednosti i nedostataka postojećih metoda za prepoznavanje oblačića u stripovima.....	10
4	OPIS IZVORNOG ALATA „COMIC-BOOK-READER“ .....	12
4.1	Optičko prepoznavanje znakova koristeći Pytesseract.....	13
4.2	Tesseract OCR.....	14
4.3	Korištenje alata.....	16
5	PRILAGODBE ZA STRIP ALAN FORD.....	20
5.1	Izgled stripa Alan Ford.....	20
5.2	Prepoznavanje teksta na hrvatskom jeziku.....	21
5.3	Obrada više stranica .....	23
5.3.1	Format datoteka na ulazu .....	23
5.3.2	Proces obrade na visokom nivou.....	24
5.3.3	Programski kod .....	25
5.3.4	Povratni rezultati .....	29
6	EVALUACIJA PRIMIJENJENIH METODA.....	30
6.1	Opis primjera stripova koji će se evaluirati.....	30
6.2	Metodologija evaluacije .....	31
6.3	Primjena odabranih metoda i alata na stripovima Alan Ford.....	32
6.4	Analiza točnosti i korisnosti odabranih metoda i alata.....	33
6.4.1	Analiza stripa 1.....	33
6.4.2	Analiza stripa 2.....	36
6.4.3	Analiza stripa 3.....	37
6.4.4	Analiza stripa 4.....	38
6.4.5	Analiza stripa 5.....	39
6.4.6	Opća analiza performansi alata .....	41

6.5	Najčešći izazovi i pogreške alata.....	43
7	ZAKLJUČAK .....	44
7.1	Razmatranje mogućih smjerova budućih istraživanja.....	44
	REFERENCE .....	46
	POPIS SLIKA .....	48
	POPIS TABLICA.....	49

# 1 UVOD

Stripovi su umjetnički oblik koji je popularan već stoljećima te predstavljaju izvor zabave za ljude svih dobnih skupina. U suvremenom dobu, stripovi su postali sve dostupniji zahvaljujući digitalizaciji i internetu, što je dodatno pojačalo njihovu popularnost diljem svijeta. Međutim, iako su stripovi iznimno popularni, ekstrakcija i prepoznavanje teksta u stripovima može biti izazovno zbog različitih stilova i fontova teksta, te različitih oblika oblačića u kojima se tekst pojavljuje. Ovome dodatno pridonosi činjenica da riječi u stripovima često nisu gramatički ispravne jer pisci žele tekстом prenijeti emocije, pa koriste onomatopeju i slične gramatičke figure.

Ovo predstavlja izazov za umjetnike, izdavače i čitatelje koji se bave analizom stripova ili stvaranjem novih stripova, što čini prepoznavanje oblačića u stripovima ključnim elementom za daljnji razvoj ove umjetničke forme i njenog utjecaja na modernu kulturu.

## 2 PREGLED STRUKTURE RADA

Ovaj diplomski rad slijedi jasnu strukturu kako bismo uspješno ostvarili naše ciljeve istraživanja o prepoznavanju oblačića u stripovima. U uvodnom poglavlju pregledat ćemo strukturu rada, definirati istraživačko pitanje i postaviti ciljeve istraživanja.

Nakon uvodnog dijela, u drugom dijelu pregledat ćemo relevantnu literaturu na temu diplomskog rada. Pregledat ćemo povijest stripa, ulogu oblačića u stripovima, postojeća istraživanja o prepoznavanju oblačića te analizirati prednosti i nedostatke postojećih metoda i alata za prepoznavanje oblačića. Ovaj pregled literature pomoći će nam da bolje razumijemo kontekst istraživanja i istaknemo važnost razvoja novih tehnika i poboljšanja postojećih alata za prepoznavanje oblačića te samim time ekstrakciju teksta iz stripova.

Nastavno na drugo poglavlje, treći dio rada sadržavat će opis najrelevantnijeg alata kojeg smo odabrali za potrebe izvršavanja prepoznavanja oblačića u stripovima. U ovom dijelu, detaljno ćemo opisati odabrani alat i njegovu implementaciju, za prepoznavanje i analizirati njihove prednosti i nedostatke. Kroz detaljan opis alata čitateljima ćemo pomoći da bolje razumiju zašto smo odabrali taj alat i na koji način smo ga prilagodili da bude primjeren u našem istraživanju.

Četvrti dio rada sadržavat će evaluaciju primijenjenih metoda. U ovom dijelu, primijenit ćemo odabrane alat na primjeru stripova Alan Forda, te ćemo evaluirati točnost i korisnost odabranih metoda i alata za prepoznavanje oblačića. Ova evaluacija omogućit će nam donošenje zaključka o tome koje su metode i alati najprikladniji za prepoznavanje oblačića u stripovima te samu ekstrakciju teksta iz stripova kao glavni cilj prepoznavanja oblačića.

U petom i posljednjem poglavlju rada, iznijet ćemo zaključke o našim glavnim nalazima istraživanja, razmotriti moguće smjerove budućih istraživanja o ovom području te dati općeniti zaključak o istraživanju.

### 2.1 Definiranje istraživačkog pitanja i ciljevi istraživanja

Istraživačko pitanje ovog diplomskog rada postavlja se kako bi se usmjerilo istraživanje prema konkretnim ciljevima i pružilo jasan okvir za analizu. U ovom slučaju, istraživačko pitanje glasi:

Kako prepoznati oblačice u stripovima, te ekstrahirati tekst iz stripova, koristeći odgovarajuće metode i alate, te koje od njih pružaju najbolje rezultate?

S obzirom na postavljeno istraživačko pitanje, ciljevi istraživanja uključuju:

1. Proučavanje problematike prepoznavanja oblačića u stripovima te ekstrakcije teksta, kako bi se steklo razumijevanje izazova koje prepoznavanje oblačića predstavlja u kontekstu analize stripova i stvaranja novih stripova.



2. Pregled postojeće literature na temu prepoznavanja oblačića u stripovima, kako bi se razumjele postojeće metode i alati te njihove prednosti i nedostaci.
3. Identifikacija odgovarajućih metoda i alata za detekciju oblačića i ekstrakciju teksta, temeljena na analizi literature i razumijevanju specifičnih zahtjeva stripova, posebno stripa Alan Ford.
4. Evaluacija odabrane metode i alata na primjerima stripova Alan Forda, kako bi se procijenila njihova točnost i korisnost u praksi te usporedila njihova učinkovitost.
5. Izvlačenje zaključaka iz rezultata evaluacije, kako bi se odredile najpogodnije metode i alati za prepoznavanje oblačića i ekstrakciju teksta u stripovima i ukazalo na moguća područja za daljnji razvoj i unapređenje.

Konačni cilj ovog istraživanja je pružiti smjernice i preporuke za odabir najboljeg alata za prepoznavanje oblačića i ekstrakciju teksta u stripovima, koji bi mogao olakšati analizu i interpretaciju sadržaja.

Kroz ovaj rad, želimo pridonijeti razumijevanju prepoznavanja oblačića i ekstrakcije teksta u stripovima, te istražiti primjenu računalnog vida i strojnog učenja u ovom području..

## **2.2 Motivacija za istraživanje**

Motivacija za istraživanje ekstrakcije teksta iz stripova proizlazi iz nekoliko ključnih razloga koji naglašavaju važnost ovog područja i njegov utjecaj na različite aspekte strip kulture:

1. Pristupačnost sadržaja: Prepoznavanje oblačića te samim time i ekstrakcija teksta olakšava pristup sadržaju stripova za osobe s oštećenim vidom ili one koji koriste čitače zaslona, čime se povećava dostupnost i uključenost ovog umjetničkog oblika.
2. Analiza teksta i sentimenta: Automatska ekstrakcija teksta omogućuje analizu teksta i sentimenta unutar stripova, što može pomoći u istraživanju društvenih, političkih i kulturnih tema koje se obrađuju u ovim djelima.
3. Prevoditeljske usluge: Ekstrakcija teksta olakšava posao prevoditeljima koji rade na prilagodbi stripova za međunarodno tržište, čime se širi doseg ovih djela i povećava njihov kulturni utjecaj.
4. Automatizacija obrade stripova: Ekstrakcija teksta omogućuje automatizaciju obrade stripova, što može ubrzati procese kao što su dizajn, uređivanje i izdavanje, te pomoći u očuvanju naslijeđa stripova digitalizacijom starih i rijetkih izdanja.
5. Razvoj umjetne inteligencije: Napredak u ekstrakciji teksta iz stripova može doprinijeti razvoju umjetne inteligencije i računalnog vida, s obzirom na izazove koji ovaj problem predstavlja u kontekstu vizualnog prepoznavanja i analize teksta.

Ukupno gledano, istraživanje ekstrakcije teksta iz stripova ima širok raspon praktičnih i teorijskih implikacija koje mogu pozitivno utjecati na razvoj strip kulture, pristupačnosti sadržaja i tehnologije umjetne inteligencije. Ova motivacija pokazuje važnost rješavanja problema ekstrakcije teksta i stavlja naš rad u širi kontekst doprinosa znanosti i društvu.

### 3 PREGLED ISTRAŽIVANJA

U ovom dijelu rada, cilj nam je pružiti čitateljima uvid u relevantna istraživanja i literaturu koja se bavi prepoznavanjem oblačića u stripovima te samim time i ekstrakciju teksta iz tih oblačića. Da bismo to postigli, prvo ćemo se osvrnuti na povijest stripa i ulogu oblačića u stripovima, kako bismo bolje razumjeli kontekst i razvoj ove umjetničke forme. Zatim, pregledat ćemo postojeća istraživanja o prepoznavanju oblačića u stripovima, kako bismo saznali koje su metode i tehnike već primijenjene u ovom području te kakvi su rezultati postignuti. Na kraju, analizirat ćemo prednosti i nedostatke postojećih metoda za prepoznavanje oblačića u stripovima, kako bismo identificirali mogućnosti za unaprjeđenje tehnika i razvoj novih pristupa u našem istraživanju.

Ovim pregledom istraživanja želimo pružiti temelj za naš rad i osigurati da naše istraživanje bude utemeljeno na postojećim znanjima i doprinese znanstvenoj zajednici na ovom području. Također, ovaj pregled istraživanja pomoći će nam u prepoznavanju praznina u znanju i potencijalnih smjerova za daljnje istraživanje prepoznavanja oblačića u stripovima.

#### 3.1 Pregled povijesti stripa i uloge oblačića u stripovima

Stripovi su popularni oblik umjetnosti koji je nastao krajem 19. stoljeća u Sjedinjenim Američkim Državama. Originalno su bili objavljivani u novinama i imali su kraće narativne priče koje su se odvijale kroz slike i tekst. U početku su stripovi bili jednostavnog dizajna, često crno-bijeli, s kratkim narativima koji su pratili političke ili društvene teme. Međutim, s vremenom su se razvijali i postajali sve raznolikiji, kako u smislu stilova tako i u smislu sadržaja.

Tijekom 20. stoljeća, stripovi su se razvijali u samostalne publikacije, koje su često bile namijenjene djeci, i postale su vrlo popularne u cijelom svijetu. S pojavom superherojskih stripova, proširili su se na mnoge druge žanrove, uključujući avanturističke, humorističke i znanstveno-fantastične. Stripovi su također postali sredstvo političkog izražavanja, a od 1960-ih godina bili su važan dio kontrakulture.

Jedna od glavnih značajki stripova su oblačići, također poznati kao tekstualni baloni ili govorni baloni, koji se koriste za prikazivanje dijaloga i misli likova. Oblačići su postali prepoznatljiv simbol stripova i igraju ključnu ulogu u prenošenju sadržaja. Tijekom godina su se razvijali u različite oblike i stilove, a njihov dizajn često odražava ton i atmosferu stripa, kao i emocije i osobnosti likova. Oblačići mogu biti različitih oblika i veličina, ovisno o stilu i žanru stripa. Primjerice, stripovi za djecu obično imaju jednostavne oblačiće koji su manji i imaju manje teksta, dok stripovi za odrasle mogu imati oblačiće različitih oblika i veličina, uključujući i tzv. "razbijene oblačiće" koji prikazuju kaotične misli ili govorni kašalj.

Uloga oblačića u stripovima nije samo u prikazivanju dijaloga i misli, već i u povezivanju vizualnih elemenata stripa s tekstualnim sadržajem. Oblačići su važni za razumijevanje narativa i razvoja radnje, jer pomažu čitateljima da prate tok priče i bolje razumiju likove i njihove

interakcije. Također, oblačići mogu igrati ulogu u stvaranju ritma i tempa čitanja stripa, što utječe na doživljaj čitatelja.

Uz to, stripovi su utjecali na mnoge druge aspekte popularne kulture, uključujući filmove, animaciju, videoigre i glazbu. Mnogi popularni filmovi, kao što su "Superman" i "Spider-Man", temelje se na superherojskim stripovima. Animirani serijali poput "Simpsona" i "SpongeBob SquarePants" također su inspirirani stripovima, dok se mnoge videoigre temelje na stripovima.

Stripovi su postali predmetom proučavanja u akademskim krugovima, a mnoga istraživanja se bave analizom žanrova, stilova, narativa i socio-kulturnih implikacija stripova. Osim toga, stripovi su postali popularni i među kolekcionarima, a mnogi rijetki primjerci stripova mogu doseći visoku cijenu na tržištu.

Uz razvoj računalne tehnologije, prepoznavanje oblačića u stripovima postalo je sve važnije za analizu i pristup sadržaju. Različite tehnike i alati razvijeni su za prepoznavanje oblačića u stripovima, a njihova učinkovitost i primjenjivost variraju ovisno o stilu i fontu teksta u stripovima. Analiza povijesti stripova i uloge oblačića u njima pomoći će u razumijevanju razvoja tehnologije za prepoznavanje oblačića i njihovu primjenu u praksi. Iako su se stripovi razvijali kroz različite žanrove, stilove i formate, oblačići su ostali ključni element u stripovima i često predstavljaju jedan od glavnih izazova u automatskom prepoznavanju teksta u stripovima.

Sve u svemu, povijest stripova je bogata i raznolika te je ostavila dubok utjecaj na popularnu kulturu i umjetnost. Razumijevanje povijesti stripova i njihove uloge u društvu ključno je za daljnji razvoj tehnologije za prepoznavanje oblačića u stripovima i pružanje smjernica za daljnja istraživanja u ovom području.

### **3.2 Pregled postojećih istraživanja o prepoznavanju oblačića u stripovima**

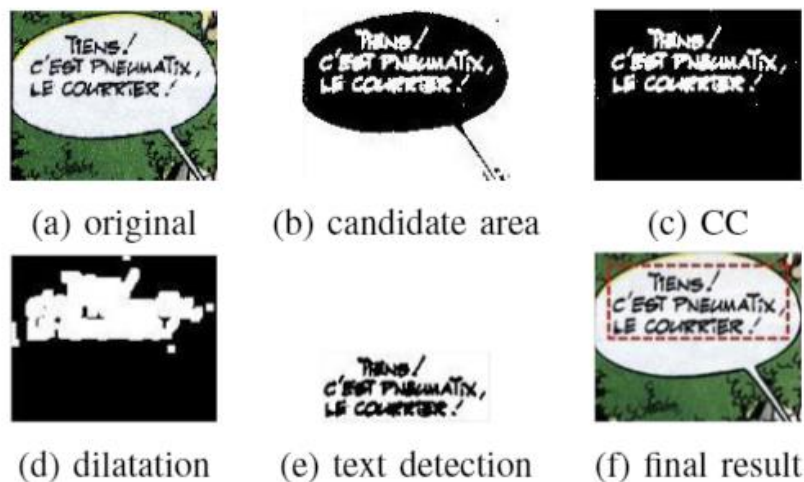
Postoje razne metode i alati koji su se koristili kroz modernu povijest za detekciju oblačića, i potječu iz različitih spektara znanosti, ali su se neki od njih istakli kao bolja i sofisticiranija rješenja.

U novije vrijeme, metode dubokog učenja pokazale su značajan napredak u odnosu na prijašnje pristupe koji se oslanjaju na razne vrste ručno izrađenih značajki za zadatke vizualnog prepoznavanja i lokalizacije. Ipak, zbog ograničene količine označenih skupova podataka i velike raznolikosti stilova crtanja koje koriste crtači stripova, metode temeljene na dubokom učenju još uvijek nisu u potpunosti istražene u kontekstu analize slika stranica stripova. U ovom pregledu postojećih istraživanja o prepoznavanju oblačića u stripovima, prvo ćemo sažeti metode koje se temelje na ručno izrađenim značajkama ili heuristikama za detekciju govornih balona, a zatim ćemo se pozabaviti pristupima koji koriste duboko učenje.

Prvi koji su se okušali u pokušaju prepoznavanja oblačića u stripovima su Arai i Tolle [1]. Oni su predložili metodu segmentacije govornih balona i izdvajanja teksta iz njih temeljenu na analizi povezanih komponenti. Međutim, primjenjivost predložene metode ograničena je samo

na Manga stripove jer se neke heuristike oslanjaju na okomite tekstove. Ako je neka metoda primjenjiva samo na manga stripove, to znači da je dizajnirana ili optimizirana za rad sa specifičnim stilom i karakteristikama manga stripova. Manga stripovi su japanski stripovi koji se razlikuju od zapadnjačkih stripova u nekoliko aspekata, uključujući stil crteža, tematiku, narativnu strukturu i konvencije prikaza. Manga stripovi, primjerice, često koriste vertikalni tekst za dijaloge unutar govornih balona, što nije uobičajeno za zapadnjačke stripove. Također, manga stripovi imaju svojstven stil crteža, s naglaskom na izražajne oči likova, dinamične kutove i detaljne pozadine.

Ho i suradnici [2], koriste sličan pristup, oni su pokušali pronaći tekstualno područje unutar govornih balona primjenjujući prvo identifikaciju bijele boje, zatim kriterije veličine i oblika povezanih komponenti. Ovaj postupak je prikazan na Slika 1 Prepoznavanje oblačića na slici i postupak ekstrakcije teksta, Ho i suradnici , koja ilustrira korake identifikacije ovih elemenata. Zatim se izvodi morfološka dilatacijska operacija kako bi se identificiralo tekstualno područje i konačno otkrili odgovarajući govorni baloni. Ovaj pristup također radi samo s jednostavnim oblicima oblačića.



Slika 1 Prepoznavanje oblačića na slici i postupak ekstrakcije teksta, Ho i suradnici [2]

Rigaud i suradnici [3] razvili su izrazito drugačiji pristup. Predložili su metodu za obradu različitih vrsta balona nepravilnih oblika stvaranjem energetske karte temeljene na domenskom znanju o stripovima u aktivnom konturiranom modelu [4]. U [5], Rigaud i suradnici prvo su koristili adaptivni prag binarizacije, a zatim analizu povezanih komponenti za otkrivanje govornih balona i tekstualnog područja unutar tih govornih balona. Rigaud i suradnici [6] predložili su još jedan učinkovit pristup segmentaciji govornih balona temeljen na analizi odnosa roditelj-dijete ekstrahiranih povezanih komponenti. Također su pokušali otkriti repove govornih balona analizom kontura balona. 2019-te su Rigaud i suradnici [7] pokušali izračunati ocjenu pouzdanosti temeljenu na domenskom znanju (tekst, oblik itd.) stripova za segmentaciju govornih balona. Kriterij za odabir ocjene pouzdanosti za segmentaciju govornih balona može se koristiti za generiranje referentnih slika za algoritme temeljene na strojnom učenju.

Nakon što smo razmotrili metode temeljene na ručno izrađenim značajkama i heuristikama, sada ćemo se usredotočiti na suvremene pristupe koji koriste duboko učenje. Dolaskom dubokog učenja i njegovog utjecaja na područje računalnog vida, istraživači su počeli istraživati kako i na koji način ove napredne tehnike mogu poboljšati prepoznavanje govornih balona u stripovima. U sljedećem dijelu pregleda literature, analizirat ćemo metode temeljene na dubokom učenju te istražiti kako one nadmašuju starije pristupe i otvaraju nove mogućnosti za precizniju analizu stripova.

Nguyen i suradnici [8] predložili su okvir zasnovan na višezadaćnom modelu učenja na temelju Mask R-CNN-a [9] za detekciju i segmentaciju nekoliko elemenata slika stranica stripova istovremeno. Također su pokušali predvidjeti povezane veze između govornih balona i likova iz stripa koristeći isti model. Sličnim pristupom vodili su se i Dubray i suradnici [10], o tome ćemo pisati u nastavku teksta.

Iyyer i suradnici [11] pokušali su modelirati značajan sekvencijalni aspekt stripova, odnosno nužnost čitatelja da premosti prazninu između sličica kako bi razumio tranzicije sličica i koherentno ih integrirao u mentalni model teksta. Konkretno, (a) anotirali su 500 nasumično odabranih stranica kako bi trenirali Faster R-CNN [12] za segmentaciju stranica na otprilike 1,2 milijuna sličica. Nadalje, (b) anotirali su 1.500 sličica za tekstualne okvire kako bi trenirali drugi Faster R-CNN u kombinaciji s unaprijed treniranim VGG kodiranjem značajki iz ImageNeta [13], velika vizualna baza podataka, kako bi segmentirali 2,5 milijuna tekstualnih okvira i (c) izvukli tekst iz tih tekstualnih okvira pomoću Google Cloud Vision OCR-a

Glavni problem jest što su tekstualni elementi anotirani pravokutnim omeđujućim okvirima, bez razlikovanja između opisa i govornih balona. Stoga segmentacije temeljene na ovim anotacijama nedostaju detalja i specifičnosti. Općenito, njihovi rezultati jasno pokazuju da se unaprijed naučene značajke dobivene treniranjem na ImageNetu, mogu vrlo dobro ponovno upotrijebiti za analizu slika dokumenata stripova.

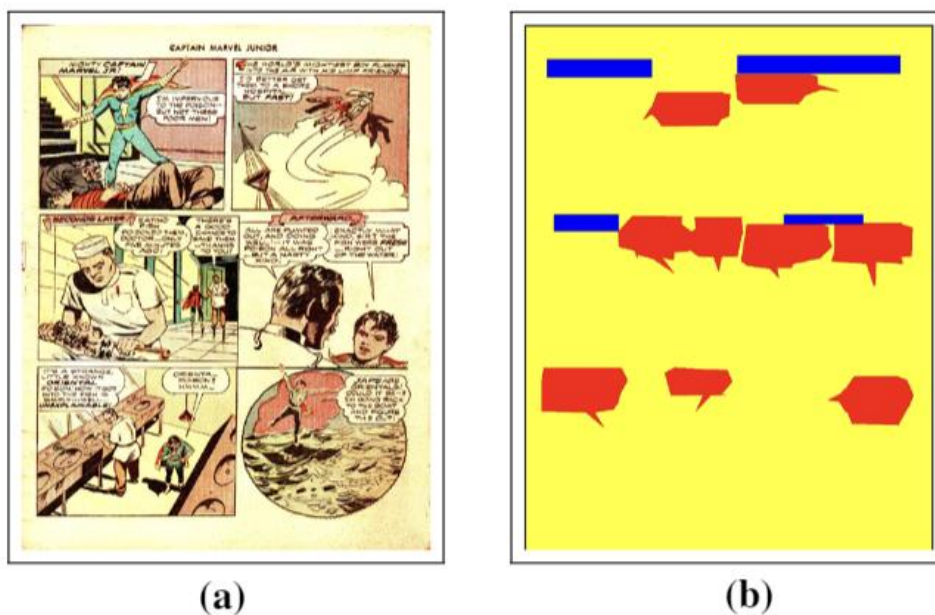
Istraživanja Laubrock i Dubray [13] iz 2018-te također su pokazali da su duboke konvolucijske neuronske mreže pogodne za hvatanje grafičkih aspekata stripova te da se uspješno mogu koristiti za kategorizaciju stila ilustratora, pripremajući teren za vizualnu stilometriju.

Kao što smo već spomenuli, Dubray i Laubrock [10] su 2019. nadogradili svoju metodu, smatrali su da je problem detekcije govornih balona problem semantičke segmentacije te su predložili metodu dubinskog učenja zasnovanu na arhitekturi U-Net [13] i VGG-16 enkoderu [14]. Cilj njihove metode bio je kategorizirati oznaku svakog piksela, odnosno odrediti pripada li govornim balonima ili pozadini. Metoda je testirana na eBDtheque [17], jednoj od najpopularnijih baza podataka različitih slika i stripova. Rezultati sugeriraju da repovi, zakrivljeni uglovi i iluzorne konture ne stvaraju velike poteškoće u detekciji. Model je naučio razlikovati oblačiće od natpisa. Ovo istraživanje je doprinijelo velikim značajem u razvoju automatizacije prepoznavanja oblačića u stripovima, iako je ljudska ruka za provjeru još uvijek potrebna.

Međutim, ova metoda nije uspjela locirati govorne balone na svijetlim područjima zbog pogrešne predikcije ocjene pouzdanosti. Autori su također ukazali na problem nedostatka označenih skupova podataka.

2021. godine, Dutta i suradnici [16] usredotočuju se na rješavanje problema kompleksnih oblika govornih balona i narativnih tekstualnih okvira zbog raznolikosti stilova crtanja.

Autori ovog rada predstavljaju pristup zasnovan na obliku osviještenoj dvostranoj konvolucijskoj neuronskoj mreži za segmentaciju narativnih tekstualnih okvira i govornih balona različitih oblika. U njihovoj dvostranoj arhitekturi, dodatni oblikovni modul obrađuje informacije o rubovima govornih balona i narativnih tekstova u kombinaciji s glavnim modulom. Spajanjem ova dva modula postiže se preciznija segmentacija govornih balona i narativnih tekstualnih okvira. Njihova metoda uspješno locira i govorne balone i narativne tekstualne okvire s značajnom točnošću regije i granice. Važno je napomenuti da njihov predloženi model uspješno detektira preklapajuće govorne balone, kao što je prikazano na Slika 2 Lokalizacija preklapanja govornih oblačića, što je bio velik izazov u prošlosti, ali ih ne može razdvojiti u njihove pojedinačne dijelove.



Slika 2 Lokalizacija preklapanja govornih oblačića

U zaključku ovog pregleda postojećih istraživanja, vidimo kako su se razvile različite metode za prepoznavanje oblačića u stripovima, od onih koje se temelje na ručno izrađenim značajkama do naprednijih pristupa koji koriste duboko učenje. Sada ćemo se prebaciti na analizu prednosti i nedostataka ovih metoda kako bismo bolje razumjeli njihove uspjehe i izazove s kojima se suočavaju.

### **3.3 Analiza prednosti i nedostataka postojećih metoda za prepoznavanje oblačića u stripovima**

U ovoj analizi prednosti i nedostataka postojećih metoda za prepoznavanje oblačića u stripovima, razmotrit ćemo ključne aspekte svake metode te istražiti njihove snage i slabosti.

#### **Metode temeljene na ručno izrađenim značajkama i heuristikama**

Prednosti:

- Jednostavnost pristupa i niži računalni zahtjevi.
- Prilagodljive heuristike koje mogu biti prilagođene specifičnom stilu stripova, kao što su Manga stripovi.

Nedostaci:

- Ograničena primjenjivost na različite stilove stripova.
- Ne mogu se lako generalizirati na složene oblike oblačića ili na različite stilove crteža.
- Ovisnost o domenskom znanju, što može ograničiti skalabilnost i generalizaciju metode.

#### **Metode temeljene na dubokom učenju**

Prednosti:

- Veća točnost i preciznost u detekciji oblačića, čak i kod složenih oblika i stilova crtanja.
- Sposobnost učenja značajki iz podataka, smanjujući potrebu za ručno izrađenim značajkama i heuristikama.
- Mogućnost primjene na različite stilove stripova, bez potrebe za specifičnim prilagodbama.
- Poboljšana detekcija i segmentacija elemenata stripova, kao što su likovi, govorni baloni i narativni tekstovi.

Nedostaci:

- Potreba za velikim označenim skupovima podataka za treniranje, što može biti izazovno i skupo.
- Veći računalni zahtjevi i dulje vrijeme treniranja u usporedbi s metodama temeljenim na ručno izrađenim značajkama.
- Moguća prenaučenosť na specifične stilove stripova ili podatke, što može utjecati na generalizaciju.



- Poteškoće u detekciji i razdvajanju preklapajućih govornih balona, što može utjecati na točnost.

Metode temeljene na dubokom učenju nude značajna poboljšanja u točnosti i preciznosti detekcije oblačića u stripovima, međutim one također dolaze s vlastitim izazovima, kao što su potreba za velikim označenim skupovima podataka i veći računalni zahtjevi. S druge strane, metode temeljene na ručno izrađenim značajkama i heuristikama, iako jednostavnije i manje računalno zahtjevne, imaju ograničeniju primjenjivost i slabiju generalizaciju na različite stilove stripova.

## 4 OPIS IZVORNOG ALATA „COMIC-BOOK-READER“

U ovom radu, kao osnovu za daljnje prilagodbe i istraživanje, koristimo izvorni alat "comic-book-reader" za detekciju oblačića i ekstrakciju teksta iz stripova. Izvorni alat "comic-book-reader" predstavlja rješenje otvorenog koda dostupno na GitHub platformi (<https://github.com/damishshah/comic-book-reader>), koji je namijenjen čitanju stripova u digitalnom formatu. Otvoreni kod se odnosi na softver čiji je izvorni kod dostupan javnosti, omogućujući svima da slobodno pregledavaju, koriste, mijenjaju i distribuiraju softver, čime se potiče suradnja, inovacija i transparentnost u razvoju softvera. Glavni cilj ovog alata jest omogućiti korisnicima detekciju oblačića u stripovima te ekstrakciju teksta iz oblačića.

Alat "comic-book-reader" temelji se na istraživanju objavljenom u radu "Deep Learning for Panel Extraction and Text Recognition in Comic Books" (<https://arxiv.org/abs/1902.08137>), koji smo spomenuli u poglavlju "Pregled postojećih istraživanja o prepoznavanju oblačića u stripovima". U navedenom istraživanju, autori su razvili metode za prepoznavanje oblačića korištenjem konvolucijskih neuronskih mreža (CNN) i segmentacije slike te algoritme za ekstrakciju teksta iz detektiranih oblačića pomoću optičkog prepoznavanja znakova (OCR).

Izvorni alat "comic-book-reader" razvijen je korištenjem tehnologija poput Pythona, Flask-a, Gunicorna, JavaScripta, HTML-a/CSS-a i Nginx-a. Također, za optičko prepoznavanje znakova (OCR) korištena je biblioteka Pytesseract, dok je za obradu slike korištena biblioteka OpenCV. Alat je dizajniran tako da može raditi s različitim stilovima stripova i omogućuje korisnicima da prilagode parametre detekcije i ekstrakcije teksta prema svojim potrebama.

Izvorni alat "comic-book-reader" prvenstveno je namijenjen detekciji i obradi teksta na engleskom jeziku, uključujući autocorrect i prepoznavanje engleskih riječi, stoga će u ovom radu biti potrebno prilagoditi ga kako bi odgovarajuće prepoznavao i obradio tekst na hrvatskom jeziku.

Kroz upotrebu izvornog alata "comic-book-reader", istražujemo mogućnosti i ograničenja metoda temeljenih na dubokom učenju u kontekstu detekcije oblačića i ekstrakcije teksta iz stripova. Ovaj alat omogućuje nam da bolje razumijemo praktične aspekte primjene ovih metoda, kao i njihovu učinkovitost i točnost u različitim uvjetima i situacijama. Kroz proučavanje i poboljšanje ovog alata, nastojimo doprinijeti razvoju tehnologija za prepoznavanje oblačića i ekstrakciju teksta iz stripova, čime se unapređuje pristupačnost, analiza i kulturna razmjena ovog popularnog umjetničkog oblika. Također, kroz prilagodbu ovog alata za prepoznavanje i obradu hrvatskog jezika, želimo proširiti njegovu primjenjivost i učiniti ga korisnim za širi krug korisnika i različite jezične zajednice. Na taj način, rad na ovom alatu pruža nam vrijedna saznanja o potencijalu dubokog učenja u rješavanju problema prepoznavanja oblačića u stripovima, te istovremeno doprinosi širenju znanja i tehnologija u ovom interdisciplinarnom području.

U nastavku ovog rada, detaljno ćemo opisati rad izvornog alata "comic-book-reader", istražujući njegove komponente, funkcije i mehanizme koji omogućuju detekciju oblačića i

ekstrakciju teksta iz stripova. Time želimo pružiti temeljit uvid u način na koji ovaj alat koristi metode dubokog učenja za rješavanje zadatka prepoznavanja oblačića u stripovima te pripremiti čitatelja za razumijevanje naše daljnje prilagodbe i poboljšanja alata.

#### **4.1 Optičko prepoznavanje znakova koristeći Pytesseract**

Optičko prepoznavanje znakova (eng. Optical Character Recognition – OCR) je proces koji transformira dvodimenzionalnu sliku teksta, koja može sadržavati strojno ispisan ili ručno napisan tekst iz svoje slikovne reprezentacije u strojno čitljiv tekst. Drugim riječima, OCR sustavi omogućuju konverziju slika teksta u oblik koji se može urediti i pretraživati. OCR kao proces se općenito sastoji od nekoliko pod procesa kako bi se postigla što veća točnost. Pod procesi su:

- Pred obrada slike
- Lokalizacija teksta
- Segmentacija znakova
- Prepoznavanje znakova
- „Post“ obrada ili krajnja obrada

Već gotovo dva desetljeća, sustavi za optičko prepoznavanje znakova se široko koriste za automatski unos teksta u računalne sustave. Tijekom tog vremena, konvencionalni online OCR sustavi (poput zonalnog OCR-a) nikada nisu prevladali svoju nesposobnost čitanja više od nekoliko vrsta fontova i rasporeda stranica. Kao rezultat toga, konvencionalni OCR nikada nije postigao veliki utjecaj.

Međutim, optičko prepoznavanje znakova doživjelo je značajna unapređenja s razvojem tehnologije i računalnog vida, što omogućuje bolje rezultate i širi spektar primjena. Unatoč izazovima s kojima se konvencionalni OCR sustavi susreću, poput ograničenja u prepoznavanju različitih fontova i rasporeda stranica, novije metode koje koriste duboko učenje i neuronske mreže znatno su poboljšale njihovu učinkovitost i točnost. Integracijom ovih naprednih metoda u OCR sustave, postalo je moguće prepoznati širi raspon pisama, stilova i rasporeda, čime se omogućuje veća automatizacija i digitalizacija teksta u različitim industrijama i sektorima. Budući da OCR tehnologija nastavlja napredovati, očekuje se da će njena primjena postati sve šira i biti još važniji alat za obradu teksta u digitalnom svijetu.

## 4.2 Tesseract OCR

Tesseract je sustav za prepoznavanje teksta otvorenog koda te se može koristiti izravno ili (za programere) putem API-ja za izdvajanje ispisanih tekstova iz slika. Tesseract 4 uključuje novi podsustav neuronske mreže konfiguriran kao prepoznavač tekstnih linija.

Prepoznavanje tekstualnih linija se odnosi na proces identifikacije i izdvajanja teksta iz određenih linija u slikama. Ovaj podsustav koristi mješavinu konvolucijske neuronske mreže (CNN) i mreže dugoročne kratkoročne memorije (eng. Long Short-Term Memory , LSTM). CNN se koristi za prepoznavanje slike koja sadrži pojedini znak, dok se LSTM koristi za rješavanje problema sekvenci znakova, odnosno teksta proizvoljne duljine.

Mreže dugoročne kratkoročne memorije postigle su izvanredne rezultate na polju prepoznavanja rukopisa. U kontekstu prepoznavanja rukom pisanih znakova, važno je napomenuti da se njihov točan položaj i linija na kojoj se temelje (linija osnove) može razlikovati. To je rezultat osobina prirodnog rukopisa, gdje se pojedina slova mogu naslanjati na različite visine, ili se može javiti nagib tijekom pisanja. Stoga su za učinkovito prepoznavanje takvih znakova potrebne složenije metode, poput dvodimenzionalnih rekurentnih mreža.

Kada ljudi pišu rukom, položaj i "linija osnove" pisani [1]h znakova može se mijenjati. To stvara izazov prilikom pokušaja prepoznavanja i interpretacije rukopisnih znakova, jer ne možete pretpostaviti da su sva slova poravnata duž iste linije osnove.

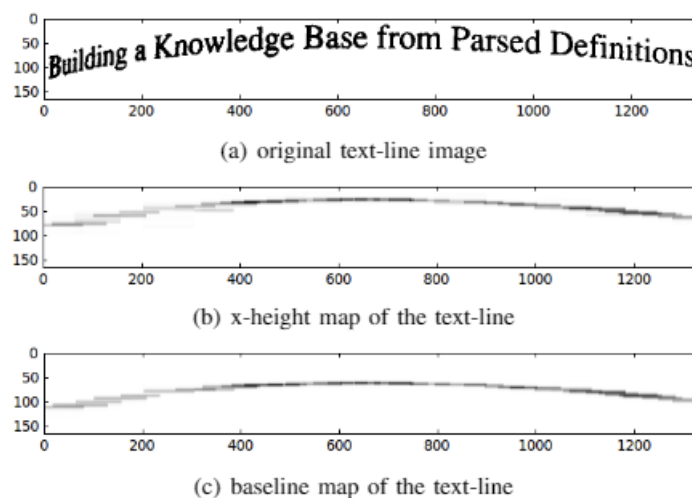
Međutim, u otisnutom tekstu (kao što su knjige, časopisi ili web stranice), slova su obično precizno poravnata duž linije osnove, što pojednostavljuje proces prepoznavanja. Nasuprot tome, za prepoznavanje otisnutog teksta koristi se jednodimenzionalna rekurentna mreža u kombinaciji s novim algoritmom za normalizaciju linije osnove i visine  $x$  gdje visina  $x$  predstavlja informaciju o visini malih slova unutar teksta.



Slika 3 Ulazni i izlazni podaci optičkog prepoznavanja znakova pomoću LSTM-a

Slika 3 Ulazni i izlazni podaci optičkog prepoznavanja znakova pomoću LSTM-a ilustrira ulazne i izlazne podatke OCR-a temeljenog na LSTM-u, demonstrirajući njegove sposobnosti i potencijalne pogreške. U pravilu, OCR sustav koji koristi LSTM metodu učinkovito upravlja situacijama kada se znakovi međusobno dodiruju (kao u primjerima a i b) te s ligaturama (kao što su prikazane u primjerima e i f). No, karakteristike ili stilovi koji se rijetko pojavljuju u skupu za treniranje su skloniji pogrešnom klasificiranju (kao u primjerima c i e), ovaj izazov se može ublažiti korištenjem veće količine podataka za treniranje.

Slika 4 Tekst sa varijabilnom linijom osnove predstavlja proces izdvajanja visine  $x$  i osnovne linije teksta. Na slici (a) možemo vidjeti originalnu sliku teksta. Slika (b) prikazuje vizualizaciju predviđenih lokacija na kojima bi se trebala nalaziti 'x-height' linija, dok slika (c) prikazuje mapu predviđenih lokacija osnovne linije teksta.



Slika 4 Tekst sa varijabilnom linijom osnove

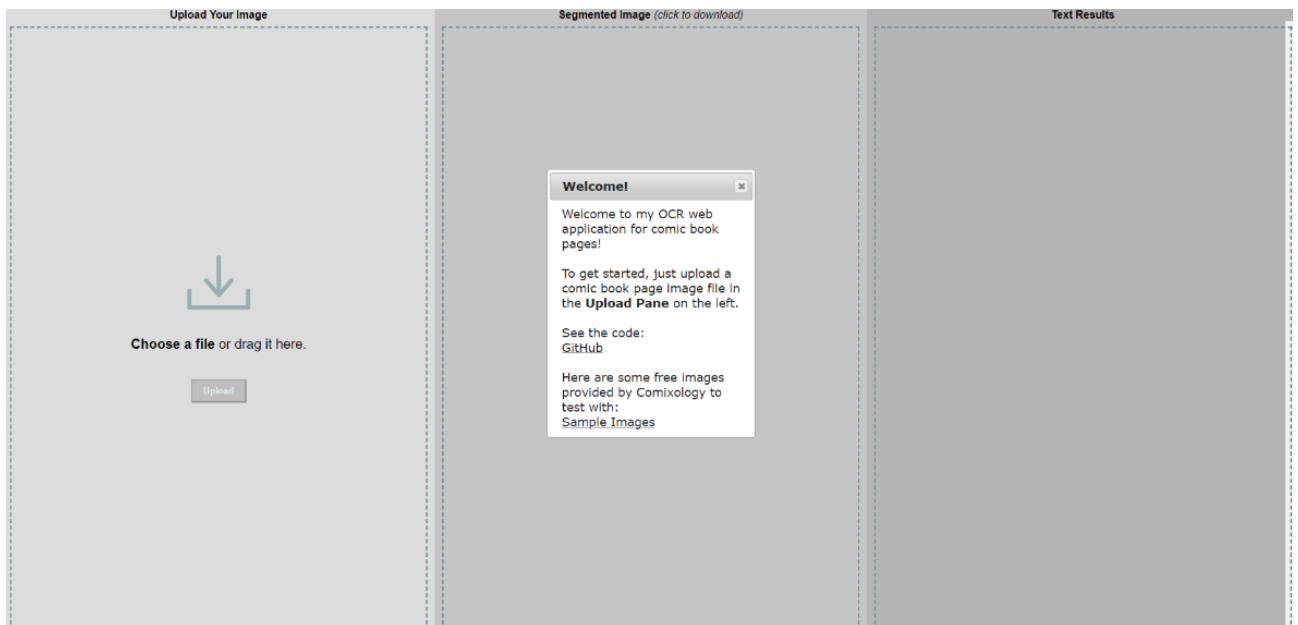
Implementacija LSTM-a u prepoznavanju tekstualnih linija pokazala se osobito korisnom. Zbog svoje sposobnosti učenja dugoročnih ovisnosti, LSTM mreže se izvrsno nose s problemima kontekstualne analize unutar teksta. U praksi, ovo znači da LSTM mreže mogu efikasno uočiti i naučiti složene obrasce unutar linija teksta, uključujući varijacije u stilu pisanja, veličinama slova, te interakcije i odnose između pojedinih slova i riječi. Ova snažna analitička sposobnost LSTM-a olakšava postupak prepoznavanja i omogućava preciznije rezultate, što je ključno za efikasnu obradu i interpretaciju teksta na slikama.

U zaključku, možemo istaknuti da Tesseract, kao jedan od najpopularnijih alata za prepoznavanje optičkih znakova (OCR), koristi napredne LSTM mreže u svojoj implementaciji. Tesseractova upotreba LSTM-a omogućava mu da se nosi s velikom raznolikošću tipova teksta, uključujući tiskani i rukopisni tekst, pružajući visoku razinu preciznosti i robustnosti. Kroz korištenje LSTM-a, Tesseract je u mogućnosti efikasno rukovati izazovima kao što su dodirujući znakovi i ligature, kao i varijabilnost položaja i linija osnove rukom pisanih znakova. Međutim, važno je napomenuti da, kao i kod svakog algoritma strojnog učenja, učinkovitost Tesseracta može biti ograničena kada se susreće s rijetkim znakovima ili stilovima koji nisu dobro zastupljeni u skupu za treniranje. Unatoč tome, ovi se izazovi mogu ublažiti kroz povećanje količine i raznolikosti podataka za treniranje. [1]

### 4.3 Korištenje alata

U ovom dijelu rada prikazat ćemo funkcionalnost i korištenje izvornog alata "comic-book-reader" kroz njegovo korisničko sučelje. Na Slika 5 Izgled korisničkog sučelja alata "comic-book-reader" prikazano je korisničko sučelje, koje je dizajnirano za jednostavno i intuitivno korištenje. Sučelje je podijeljeno u tri dijela: u prvom dijelu nalazi se gumb za učitavanje slike iz stripa, koja se nakon unosa prikazuje u istom odjeljku, u središnjem dijelu sučelja prikazuje se slika s detektiranim oblačićima nakon obrade, dok se u zadnjem dijelu sučelja ispisuje tekst koji je prepoznat unutar oblačića.

Kroz ovaj pregled korisničkog sučelja možemo bolje razumjeti kako alat "comic-book-reader" omogućuje korisnicima da jednostavno učitaju i obrađuju slike iz stripova, te dobiju rezultate detekcije oblačića i ekstrakcije teksta.



Slika 5 Izgled korisničkog sučelja alata "comic-book-reader"

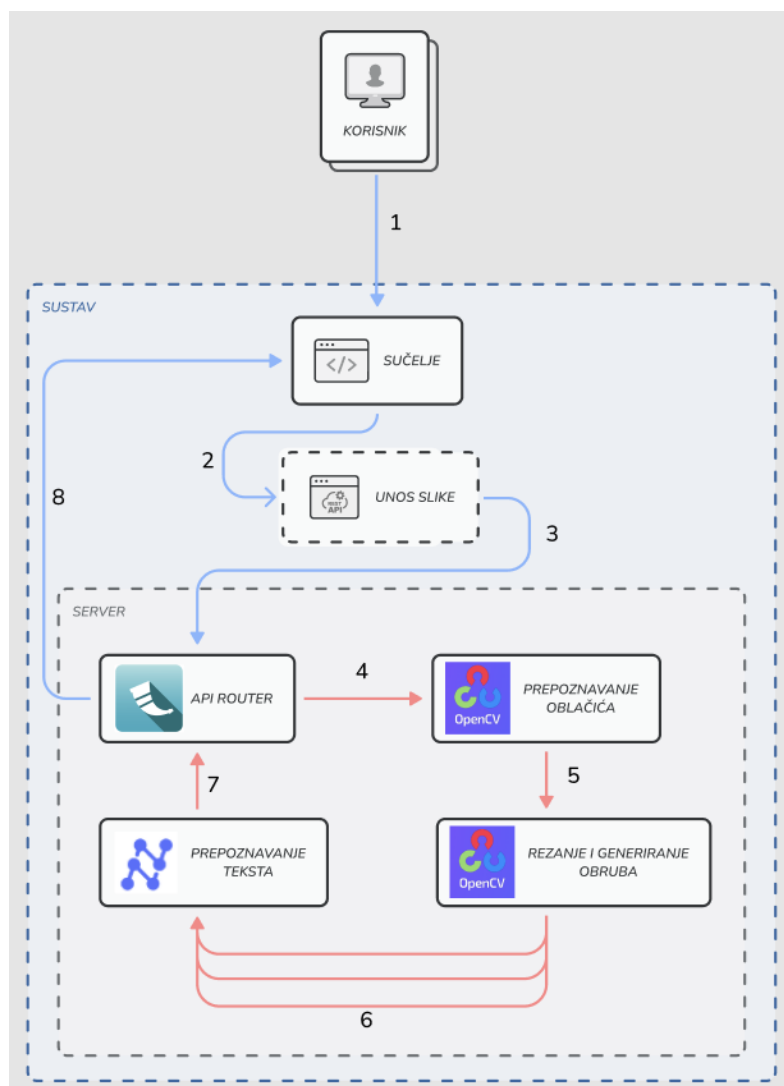
Proces prepoznavanja oblačića u alatu "comic-book-reader" započinje identifikacijom granica oblačića pomoću OpenCV 2 biblioteke za Python.

OpenCV (Open Source Computer Vision) 2 je biblioteka otvorenog koda za obradu slika i računalni vid. Nudi mnoge funkcije i algoritme za detekciju i prepoznavanje oblika, analizu slika, segmentaciju slika, obradu videa i strojno učenje. OpenCV 2 je popularan izbor za programere koji se bave računalnim vidom i analizom slika zbog svoje svestranosti, performansi i podrške za različite platforme i programerske jezike, uključujući Python. U kontekstu našeg projekta, OpenCV 2 se koristi za identifikaciju granica oblačića u stripovima te za izrezivanje tih oblačića kao pripremnu radnju za OCR.

Nakon što se granice oblačića uspješno prepoznaju, algoritam generira obrise i izrezuje svaki pojedini oblačić. Kao rezultat ovog postupka dobivamo niz izrezanih slika oblačića.

Zatim se ovaj niz slika šalje Pytesseract biblioteci, koja provodi optičko prepoznavanje znakova kako bi se detektirao i izdvojio tekst iz svakog oblačića. Na taj način, alat "comic-book-reader" uspješno prepoznaje i ekstrahira tekstualni sadržaj iz slika stripova.

Slika 6 Struktura „comic-book-reader“ sustava donosi grafički prikaz strukture „comic-book-reader“ sustava za obradu slika stripova, čime nastojimo čitateljima pružiti jasniju sliku o načinu rada alata i međusobnoj povezanosti komponenti unutar sustava. Za potrebe boljeg razumijevanja, analizirat ćemo svaku komponentu i njen zadatak u cjelokupnom procesu, prateći brojeve na slici koji se odnose na svaki pojedini korak.



Slika 6 Struktura „comic-book-reader“ sustava

Korak 1 označava ulaznu točku našeg programa, početak puta na kojem korisnik stupa u interakciju sa sustavom.

U Koraku 2, korisnik unosi slike stripova u sustav za obradu. Ovdje je važno naglasiti da prije daljnje obrade, unesene slike prolaze kroz fazu osnovne validacije kako bi se osigurao pravilan format slike.

Kada slike uspješno prođu proces validacije, prelazimo na Korak 3. U ovom koraku, slike se šalju na serversku komponentu koja prihvata taj sadržaj te ga priprema za daljnju obradu. Na temelju formata slike, pozivaju se odgovarajuće funkcije za obradu navedenih podataka.

Korak 4 pokreće prvi korak u transformaciji podataka - prepoznavanje oblačića u stripu. Svaka je slika pažljivo analizirana kako bi se identificirali i označili oblačići, što nam omogućuje daljnju obradu kao što su rezanje i izvlačenje teksta.

Nakon označavanja oblačića, prelazimo na Korak 5 gdje izrezujemo označene oblačiće kako bismo preciznije i brže izvukli sadržaj unutar njih. Skup izrezanih oblačića, kao i cjelovita slika s označenim oblačićima, šalju se dalje na obradu.

Korak 6 obuhvaća slanje dobivenog niza izrezanih oblačića biblioteci Tesseract, koja vrši obradu svakog pojedinog oblačića i vraća rezultirajući tekst.

Nakon završetka obrade, dolazimo do Koraka 7. Ovdje se svi dobiveni podaci šalju natrag na serversku rutu, koja prosljeđuje podatke kao odgovor na korisnikov zahtjev.

Na kraju, u Koraku 8, obrađeni podaci se šalju natrag na korisničko sučelje, dovršavajući ciklus obrade slike stripa, i omogućujući korisniku da vidi finalni rezultat.

Na ovaj način, kroz niz osmišljenih koraka, ovaj alat pruža pouzdano i učinkovito rješenje za obradu i transformaciju slika stripova.

Povratna informacija sa servera sadrži sliku sa iscrtanim oblačićima, te tekst sadržan u slici.

Na Sliku 7 Korištenje korisničkog sučelja alata "comic-book-reader" prikazan je alat nakon što se učita fotografija stripa. U prvom dijelu sučelja vidimo učitanu fotografiju, dok u drugom dijelu vidimo segmentirane oblačiće ocrtane zelenom bojom. U ovom slučaju, na slici možemo uočiti četiri oblačića i jednu pogrešnu detekciju, gdje je glava lika iz stripa prepoznata kao oblačić. Ovaj primjer ilustrira izazove koje smo prethodno spomenuli u radu prilikom detekcije oblačića, gdje je zbog boje i oblika lica lika alatu izgledalo kao oblak. U trećem dijelu sučelja vidimo ispis teksta koji je vrlo precizno očitao. Slika je iz stripa na engleskom jeziku, što je u skladu s našim prethodnim napomenama da je izvorni alat dizajniran za stripove na engleskom jeziku.





Slika 7 Korištenje korisničkog sučelja alata "comic-book-reader"

Kako bismo omogućili korištenje alata u svrhu ovog diplomskog rada, prilagodili smo "comic-book-reader" za detekciju oblačića i ekstrakciju teksta iz stripova na hrvatskom jeziku, uz posebno uzimanje u obzir specifičnosti stripova Alan Ford. U tom smislu, detaljno ćemo opisati arhitekturu i prilagodbe koje smo napravili kako bi alat bio učinkovit u analizi takvih stripova.

## 5 PRILAGODBE ZA STRIP ALAN FORD

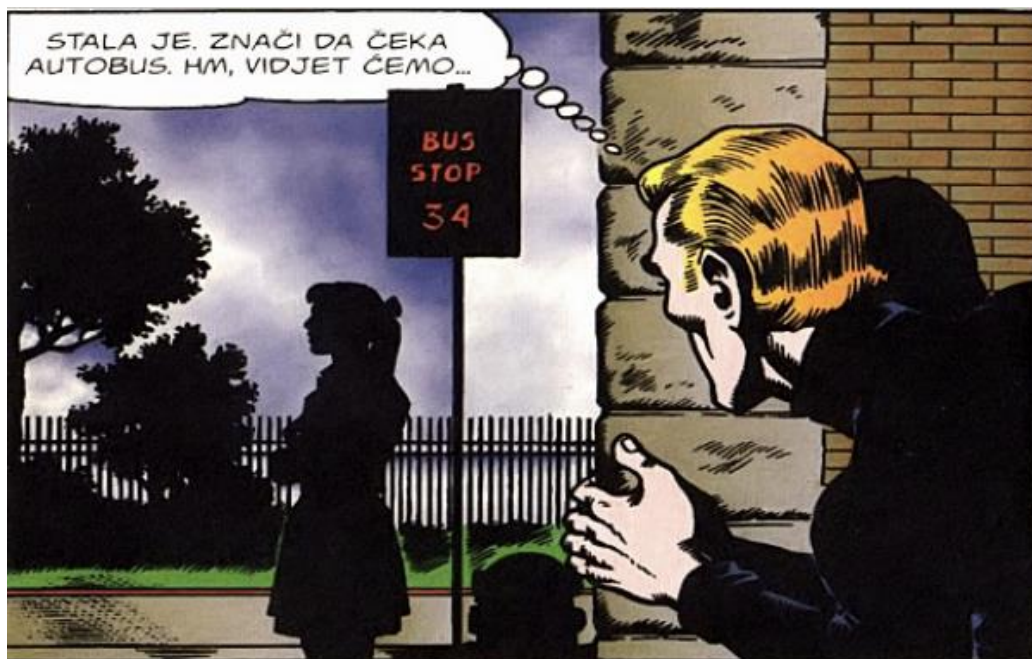
U ovom poglavlju detaljno ćemo razmotriti prilagodbe koje smo napravili za strip Alan Ford. Ispitat ćemo specifične karakteristike stripa, prilagodbu OCR-a za hrvatski jezik i postupak obrade više stranica stripa.

### 5.1 Izgled stripa Alan Ford

Da bismo mogli napraviti prilagodbe za strip Alan Ford, moramo se detaljno upoznati s njegovim izgledom i specifičnostima koje ima. Stripovi Alan Ford konstruirani su na jedinstven način, gdje svaka stranica sadrži dva odvojena kvadrata. Ovi kvadrati, koji svaki za sebe predstavljaju jedinstvenu cjelinu sadržaja i tekstualnih oblačića, stvaraju specifičnu strukturu stripa. Rečeno jednostavnijim jezikom, na svakoj stranici stripa Alan Ford možemo vizualno razaznati dvije distinktno kvadratne priče, koje skupa tvore koherentnu narativnu cjelinu.

Tekst u stripu Alan Ford nalazi se ne samo unutar standardnih oblačića za dijalog, već i unutar narativnih oblačića, koji se često koriste za uvod u priču ili za komentiranje određenih situacija. Pored toga, tekst se može nalaziti i na neobičnim mjestima, kao što su majice likova, reklame ili čak objekti u sceni.

Slika 8 Tekst izvan tekstualnih oblačića: Primjer na znaku autobusne stanice ilustrira jedan takav primjer, gdje se tekst ingeniozno smjestio na znaku autobusne stanice, dodatno obogaćujući dojam i humorni kontekst scene. Sve ove specifičnosti čine ovaj strip izazovnim za analizu, ali i otvaraju mogućnosti za dubinsko razumijevanje konteksta i humora koji je inherentan ovom popularnom stripu.



Slika 8 Tekst izvan tekstualnih oblačića: Primjer na znaku autobusne stanice

Na Slika 9 Višestruki tekstualni oblačići iz Alan Ford stripa: Različitost u veličini, obliku i boji možemo vidjeti bogatu raznolikost oblačića za dijalog koji su često prisutni u stripu Alan Ford. Primijetiti ćemo da se oblačići ne odlikuju univerzalnim oblikom - dok neki imaju glatke, ravne linije, drugi su oštri i zakrivljeni. U stripovima koji koriste boju, određeni oblačići su čak obojani u drugačijim bojama, poput žute, izazov koji ćemo detaljnije obraditi u poglavlju o evaluaciji. Također, važno je napomenuti da oblačići ne sadrže isključivo tekst. Ponekad, kao što je prikazano na slici, oblačići mogu uključivati i dodatne elemente, kao što su glazbene note.



Slika 9 Višestruki tekstualni oblačići iz Alan Ford stripa: Različitost u veličini, obliku i boji

## 5.2 Prepoznavanje teksta na hrvatskom jeziku

Kako bi prilagođeni "comic-book-reader" bio što učinkovitiji u analizi stripova Alan Ford, napravili smo prilagodbe koje su poboljšale preciznost i točnost prepoznavanja teksta. Prilagodba OCR-a za hrvatski jezik i stilizirane fontove bila je ključna jer stilizirani fontovi koji se često koriste u stripovima mogu biti izazovni za OCR sustave, a hrvatski jezik ima svoje specifičnosti u pravopisu i gramatici koje treba uzeti u obzir prilikom prepoznavanja teksta.

Prvi ključni korak koji smo poduzeli bio je proširenje alata Tesseract OCR, kojeg smo detaljnije opisali u poglavlju 3.2.. Tesseract OCR već ima pripremljen skup podataka za hrvatski jezik. U skladu s preporukama, uspješno smo izvršili proces proširenja Tesseract OCR-a koristeći predtrenirani skup podataka specifično za hrvatski jezik. Ovaj proces proširenja omogućio je OCR sustavu da bolje razumije i prepozna hrvatski tekst, uključujući stilizirane fontove koji su česti u stripovima poput Alan Forda.

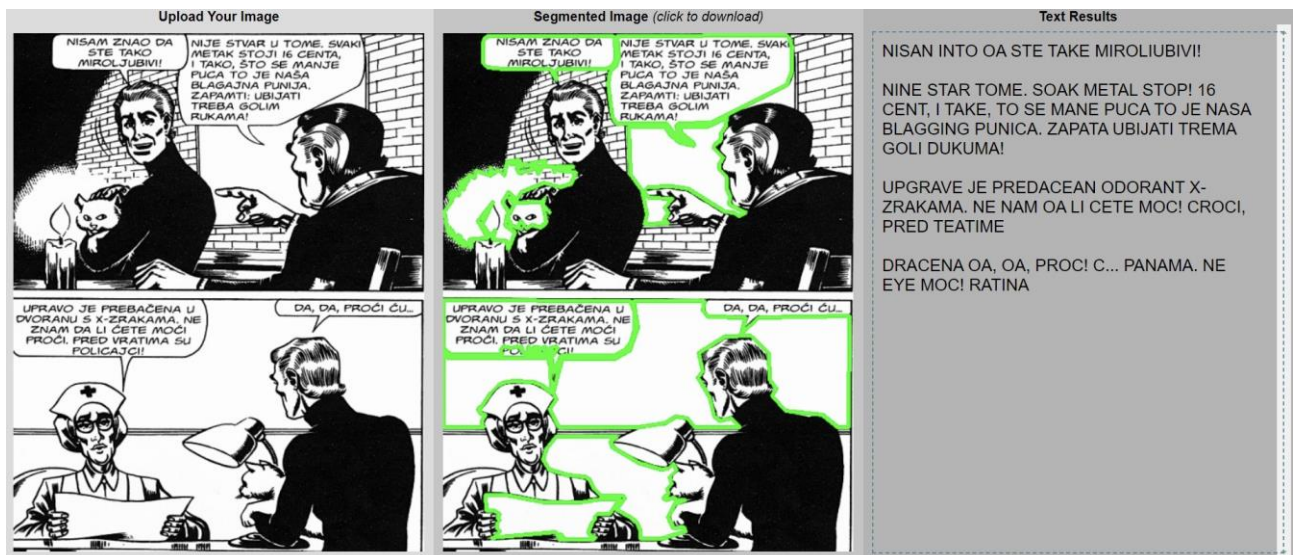
Dodatno smo implementirali pravila pravopisa i gramatike hrvatskog jezika kako bismo značajno poboljšali točnost prepoznavanja teksta, ta pravila zapravo filtriraju krivo prepoznate znakove i nepostojeće veznike i riječi. Naša prilagodba je obuhvatila prepoznavanje specifičnih znakova, interpunkcijskih oblika i naglasaka koji su karakteristični za hrvatski jezik. Posebno smo omogućili OCR sustavu prepoznavanje hrvatskih slova "š, č, ć, ž, đ", koja su učestala u



hrvatskom jeziku. Također, ručno smo dodali prepoznavanje hrvatskih samoglasnika i slova "s" kao samostalnih slova, budući da su oni često korišteni kao veznici u hrvatskom jeziku.

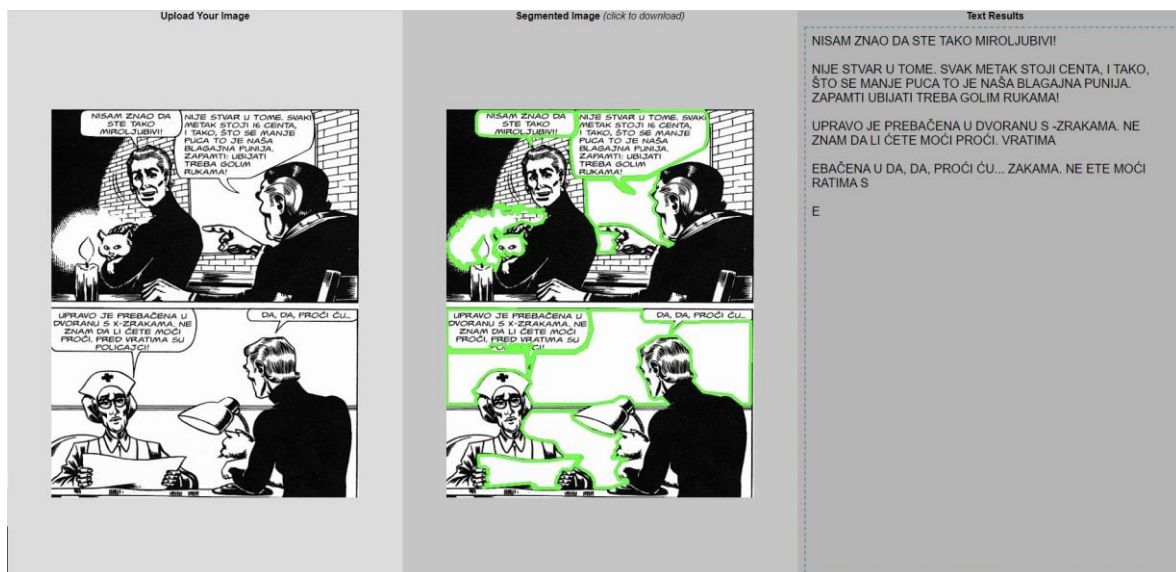
Na sljedećim slikama možemo vidjeti usporedbu između izvornog i prilagođenog alata.

Na Slika 10 Ekstrakcija teksta alatom "comic-book-reader" prikazan je rezultat koji izvorni alat "comic-book-reader" pruža prilikom obrade stranice stripa Alan Ford. Kao što možete vidjeti, iako se neki od tekstualnih elemenata uspješno prepoznaju, postoji niz izazova, posebno u vezi s prepoznavanjem specifičnih hrvatskih znakova.



Slika 10 Ekstrakcija teksta alatom "comic-book-reader"

Nakon što smo implementirali izmjene, rezultati su bili bolji. Slika 11 Ekstrakcija teksta prilagođenim alatom "comic-book-reader" prikazuje kako prilagođeni "comic-book-reader" sada može prepoznati i obraditi hrvatski tekst mnogo točnije i preciznije.



Slika 11 Ekstrakcija teksta prilagođenim alatom "comic-book-reader"

Kombinacija proširenja Tesseract OCR-a modela i dodavanja specifičnih pravila za filtriranje povratnih informacija je rezultirala poboljšanjem preciznosti i točnosti OCR sustava u prepoznavanju teksta u stripovima Alan Ford na hrvatskom jeziku.

### **5.3 Obrada više stranica**

Kao što je već spomenuto, izvorni alat „comic-book-reader“ je usmjeren na obradu pojedinačnih slika. Međutim, kako bismo omogućili korisnicima jednostavnu analizu cijelog stripa umjesto pojedinačnih stranica, temeljito smo prilagodili naš alat. Sada alat prihvaća i obrađuje više stranica stripova, čime korisnicima omogućujemo cjelovitu analizu stripa s lakoćom. Ova prilagodba osigurava da korisnici mogu uživati u jednostavnoj analizi stripova Alan Ford bez potrebe za pojedinačnim pregledavanjem svake stranice zasebno.

No, prva implementacija ovog pristupa donijela je novi izazov. Kada smo izmijenili naš server da prihvaća niz slika umjesto samo jedne, inicijalna obrada cijelog stripa trajala je više od 10 minuta.

Da bismo to poboljšali, uveli smo tehniku paralelnog programiranja, koja omogućuje izvršavanje više procesa istodobno. Konkretno, modificirali smo naš server tako da koristi više "dretvi" (engl. threads), što je rezultiralo značajno bržim vremenom obrade. Dretva predstavlja najmanju sekvencu programskih instrukcija koje operacijski sustav može upravljati. Sustavi sa više dretvi omogućuju paralelno izvršavanje više dretvi unutar jednog procesa, povećavajući tako učinkovitost i performanse.

To je prikazano na Kod 2 Adresa upload - Izmjena servera za obradu više slika u poglavlju 5.3.3, gdje je pojašnjena njegova implementacija.

Nakon implementacije ovih promjena, vrijeme izvršavanja za svaki strip smanjilo se na prosječno 3 minute - što je značajno poboljšanje u odnosu na izvorni kod.

#### **5.3.1 Format datoteka na ulazu**

Stripovi se često spremaju u popularne formate poput CBR (Comic Book RAR) ili PDF (Portable Document Format). CBR format je specifičan datotečni format namijenjen digitalnom spremanju stripova. To je zapravo komprimirani arhivski format koji sadrži više slika, obično u JPEG ili PNG formatu, poredanih po određenom redoslijedu kako bi stvorile pojedinačne stranice stripa. CBR datoteke su zapravo RAR arhivi koji su preimenovani u .cbr ekstenziju kako bi bili kompatibilni s programima za čitanje stripova. Korištenje CBR formata olakšava manipulaciju digitalnim stripovima i omogućuje njihovo čitanje na računalima i drugim elektroničkim uređajima.

Drugi popularan format za spremanje stripova je PDF. PDF je format datoteka koji je razvio Adobe Systems s ciljem zadržavanja izgleda i formata dokumenata bez obzira na hardversku platformu, softversku aplikaciju ili operacijski sustav. PDF datoteke omogućuju jednostavnu

razmjenu dokumenata među različitim računalima i uređajima bez gubitka kvalitete. Osim toga, pružaju mogućnosti enkripcije i digitalnog potpisivanja za zaštitu povjerljivih informacija.

U skladu s tim, naš alat je prilagođen za učitavanje PDF formata umjesto CBR-a, s ciljem veće fleksibilnosti i šire primjene. Razvoj aplikacije za učitavanje PDF formata bio je brži i manje zahtjevan, što rezultira jednostavnijom i učinkovitijom uporabom alata.

### **5.3.2 Proces obrade na visokom nivou**

Proces obrade na visokom nivou uključuje nekoliko ključnih koraka: učitavanje stripa, segmentaciju stranica na panele, izdvajanje oblačića s tekstom unutar panela, prepoznavanje i konverziju teksta te konačno, renderiranje prevedenog stripa za korisnika.

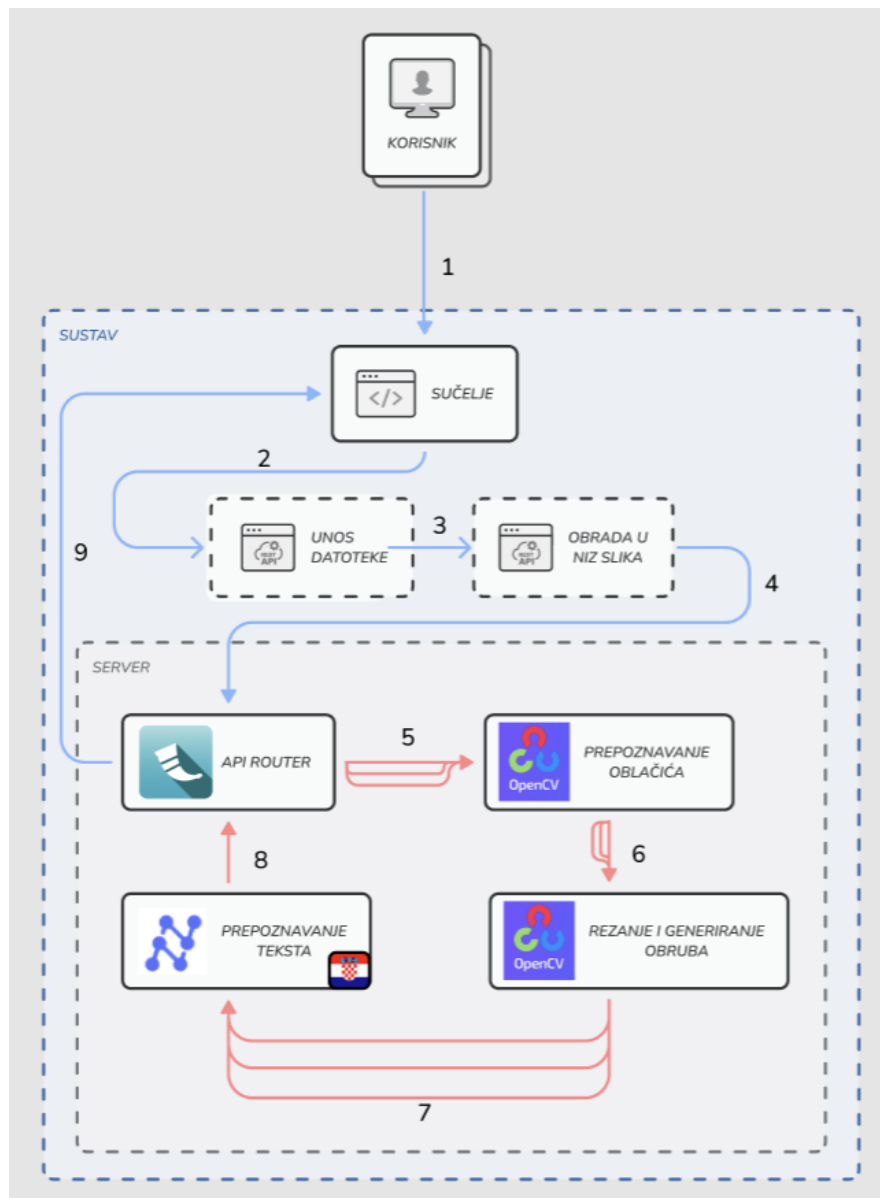
Prateći Slika 12 Proširena struktura prilagođenog „comic-book-reader“ sustava, prilagođeni sustav i proces obrade na visokom nivou sastoje se od 2 glavna modula. Korisničko sučelje, kao prvi modul, omogućuje korisnicima da unesu podatke i odaberu PDF datoteku koju žele obraditi (2). Zatim se ta PDF datoteka pretvara u niz slikovnih datoteka, što olakšava daljnju obradu (3).

Drugi modul predstavlja serversku komponentu koja sadrži sljedeće ključne elemente: osnovni API router, OpenCV biblioteku i Tesseract biblioteku proširenu za hrvatski jezik. Serverska komponenta, preuzima te slikovne datoteke i provodi daljnju obradu. Osnovni API router omogućuje komunikaciju između korisničkog sučelja i procesa obrade na serverskoj strani (4, 9).

U procesu obrade koristi se OpenCV biblioteka koja omogućuje segmentaciju stranica stripa na panele (5), izdvajanje oblačića s tekstom unutar panela te druge manipulacije slikom kako bi se dobili relevantni dijelovi za prepoznavanje teksta.

Tesseract biblioteka, koja je proširena za hrvatski jezik (7), koristi se za prepoznavanje teksta iz slikovnih datoteka. Ona analizira sliku oblačića s tekstom i konvertira ga u tekstualni format.

Navedeni moduli i njihove funkcionalnosti omogućuju prilagođeni sustav obrade podataka koji je optimiziran za hrvatski jezik i specifičnosti stripa. Ovaj proces obrade na visokom nivou omogućuje automatsko prepoznavanje i konverziju teksta te renderiranje prevedenog stripa za korisnika.



Slika 12 Proširena struktura prilagođenog „comic-book-reader“ sustava

### 5.3.3 Programski kod

Kako bismo olakšali rad s podacima, razvili smo intuitivnu frontend (FE) aplikaciju koja omogućuje korisnicima jednostavno učitavanje PDF datoteka. FE aplikacija predstavlja korisničko sučelje koje korisnicima omogućuje interakciju s sustavom putem grafičkog sučelja. Na visokoj razini, FE je prikazan kao prvi modul na slici 9. To obuhvaća sve elemente koje korisnik vidi na zaslonu, kao što su izbornici, tipke, obrasci za unos podataka, tablice i drugo. Za izradu FE aplikacije koristili smo React, popularnu JavaScript biblioteku otvorenog koda koja se koristi za razvoj web aplikacija. Sama FE aplikacija ima jednostavan dizajn i osnovne funkcionalnosti, te ne zahtijeva dodatna objašnjenja vezana uz kod aplikacije.

Nadalje, kako bismo povezali frontend aplikaciju s funkcionalnostima prilagođenog comic-book-readera, koristili smo server zasnovan na Pythonu, koji smo opisali u prethodnom poglavlju. Taj server je također komponiran kao Docker datoteka, što omogućava jednostavno

pokretanje bez obzira na lokalno okruženje. Docker je otvorena platforma koja omogućava pakiranje, distribuciju i pokretanje aplikacija unutar kontejnera. Kontejneri pružaju izolaciju i konzistentno okruženje za izvršavanje aplikacija neovisno o specifičnostima sustava. U nastavku teksta prikazan je Kod 1 Docker kompozicija za pokretanje aplikacije, gdje možete vidjeti da se i ostale biblioteke koje se koriste u alatu također instaliraju i pokreću pomoću Docker-a. Kombinacija FE aplikacije i Python servera omogućuje korisnicima interakciju s comic-book-readerom putem intuitivnog grafičkog sučelja, gdje mogu učitavati PDF datoteke i koristiti različite funkcionalnosti koje su implementirane na serveru.

```
FROM python:3.7

ADD . /comic-book-reader

WORKDIR /comic-book-reader

RUN apt-get update \
    && apt-get install -y python3-enchanted \
    && apt-get install -y tesseract-ocr \
    && apt-get install tesseract-ocr-hrv \
    && apt-get install -y libtesseract-dev \
    && pip install --upgrade pip \
    && pip install flask_cors \
    && pip install --upgrade pip setuptools wheel \
    && pip install -r requirements.txt

EXPOSE 8080

CMD ["gunicorn", "--bind", "0.0.0.0:8080", "--timeout", "60", "--workers=1",
    "--threads=4", "--worker-class=gthread", "app"]
```

*Kod 1 Docker kompozicija*

Kroz ovaj rad, već smo nekoliko puta istaknuli kako je izvorna aplikacija koncipirana da primi jednu sliku u jednom od standardnih formata za pohranjivanje slika.

U sklopu Kod 2 Adresa upload - Izmjena servera za obradu više slika prikazujemo specifični dio naše aplikacije, čija je primarna uloga prijem podataka poslanih iz React klijenta, preciznije, ona obrađuje listu slika. Ovaj serverski segment komunicira koristeći REST API, koristeći se "/upload" krajnjom točkom za prihvatanje podataka.

Također, koristimo **concurrent.futures.ThreadPoolExecutor** za kreiranje skupa radnih threadova, svaki sposoban za nezavisnu obradu slike. Na ovaj način, više slika može se obraditi istodobno, čime se smanjuje ukupno vrijeme obrade.

```
@cbr.route('/upload', methods=['POST'])
def upload_images():
    if request.method == 'POST':
        images = request.files.getlist('file[]')
    if not images:
```



```

        return jsonify({"error": "No images received"}), 400
    results = []
    i=1
    with concurrent.futures.ThreadPoolExecutor(max_workers=5) as executor:
        futures = []
        for image in images:
            future = executor.submit(process_image, image, i)
            futures.append(future)
            i += 1
        for future in concurrent.futures.as_completed(futures):
            result = future.result()
            results.append(result)
    sorted_results = sorted(results, key=lambda x: x['PageNumber'])
    return jsonify(sorted_results)

```

*Kod 2 Adresa upload - Izmjena servera za obradu više slika*

Po uspješnom prijemu podataka, svaka se slika šalje na individualnu obradu. Kompletan postupak obrade slike implementiran je unutar funkcije pod nazivom `process_image`, što je detaljno prikazano u segmentu Kod 3 Procesuiranje pojedine slike.

```

def process_image(file):
    if file and allowed_file(file.filename):
        npimg = numpy.fromstring(file.read(), numpy.uint8)
        img = cv2.imdecode(npimg, cv2.IMREAD_COLOR)
        croppedImageList = segmentPage(img)
        pageText = parseComicSpeechBubbles(croppedImageList)
        return {"pageText": pageText}
    else:
        return {"error": "Invalid file format"}

```

*Kod 3 Procesuiranje pojedine slike*

Funkcija za segmentiranje stranice, ili "segmentPage", je ključni dio koda koji upravlja procesom prepoznavanja i izoliranja oblačića razgovora.

```

def segmentPage(image, shouldShowImage = False):
    contours = findSpeechBubbles(image)
    croppedImageList = cropSpeechBubbles(image, contours)

    cv2.drawContours(image, contours, -1, (0, 0, 0), 2)
    if shouldShowImage:
        cv2.imshow('Speech Bubble Identification', image)
        cv2.waitKey(0)
        cv2.destroyAllWindows()

    return croppedImageList

```

*Kod 4 Funkcija za segmentiranje slike*

Evo kako ta funkcija prikazana na Kod 4 Funkcija za segmentiranje slike djeluje: Funkcija `segmentPage` prima sliku kao ulazni argument, zajedno s opcionalnim parametrom

'shouldShowImage'. Ovaj parametar, ako je postavljen na True, omogućava vizualizaciju slike tijekom procesiranja.

Kao prvi korak, funkcija koristi pomoćnu funkciju 'findSpeechBubbles' koja se koristi za detekciju oblačića na unesenoj slici. Konture ili obrisi oblačića koje ova funkcija detektira zatim se spremaju.

Nakon toga, koristi se funkcija 'cropSpeechBubbles' koja prima originalnu sliku i detektirane konture kao ulazne parametre. Ova funkcija izolira i izrezuje svaki oblačić na slici i vraća listu izrezanih oblačića. Nakon izolacije oblačića, koristi se funkcija 'drawContours' iz biblioteke OpenCV kako bi se konture crtale na originalnoj slici.

Ukoliko je parametar 'shouldShowImage' postavljen na True, slika se prikazuje kroz 'imshow' funkciju OpenCV-a. Ovo omogućava vizualizaciju koraka prepoznavanja oblačića, čime se može provjeriti točnost i efikasnost procesa. Slika ostaje otvorena dok korisnik ne pritisne bilo koju tipku, nakon čega se svi prozori zatvaraju kroz 'destroyAllWindows' funkciju.

Naposljetku, funkcija 'segmentPage' vraća listu izrezanih oblačića kao rezultat. Ova lista zatim se koristi za daljnje procesiranje i ekstrakciju teksta.

Nakon obrade dobivamo niz obrađenih slika koje se zatim šalju na izvlačenje teksta koristeći funkciju prikazanu na Kod 5 Ekstrakcija teksta.

```
def parseComicSpeechBubbles(croppedImageList, shouldShowImage = False):
    scriptList = []

    for croppedImage in croppedImageList:
        # Enlarge cropped image
        croppedImage = cv2.resize(croppedImage, (0,0), fx = 2, fy = 2)
        # # Denoise
        croppedImage = cv2.fastNlMeansDenoisingColored(croppedImage, None, 10,
10, 7, 15)

        if shouldShowImage:
            cv2.imshow('Cropped Speech Bubble', croppedImage)
            cv2.waitKey(0)
            cv2.destroyAllWindows()

        # Pass cropped image to the ocr engine
        script = tesseract(croppedImage)

        # If we don't find any characters, try shrinking the cropped area.
        # This occasionally helps tesseract recognize single word lines, but
increases processing time.
        count = 0
        while (script == '' and count < 3):
            count+=1
            croppedImage = shrinkByPixels(croppedImage, 5)
```

```

script = tesseract(croppedImage)

if script != '' and script not in scriptList:
    scriptList.append(script)

return scriptList

```

*Kod 5 Ekstrakcija teksta*

Sve izmjene izvornog koda koje smo napravili dostupne su na sljedećem GitHub repozitoriju: <https://github.com/cagaljfilipa/comic-book-reader/tree/HRV>. Posjetite ovu poveznicu kako biste pregledali prilagođeni kod i proučili kako smo implementirali promjene za podršku hrvatskog jezika.

Nakon ovih koraka dobivamo sve podatke koji su nam potrebni, te ih šaljemo natrag na korisničko sučelje u JSON formatu.

### 5.3.4 Povratni rezultati

U završnoj fazi aplikacije, server vraća obrađeni tekst klijentu u obliku JSON formata. Taj tekst se potom prikazuje korisniku kao što je prikazano na JSON 1 Povratni rezultati iz aplikacije, omogućavajući mu da ga pročita ili, u našem slučaju, da ga dalje obrađuje. Obradom dobivenih rezultata, odnosno teksta, možemo dalje izračunati metrike kao što su preciznost, opoziv i F1-metrika, kako bismo ocijenili kvalitetu prepoznavanja i ekstrakcije teksta što će biti prikazano u sljedećem poglavlju.

#### **Results:**

```

{"pageText":["U NEKOJ TRGOVINI KERAMIKE I PORCULANA U NJUJORŠKOJ 5. AVENIJI...", "TIITITITIINUKHUHUU LII II CI TIE TII FIT TA 7\''"]}

```

```

{"pageText":["KAKVA IDEJA! POSLATI ME OVAMO, TO ZBILJA MOŽE PASTI NA PAMET SAMO ŠEFU. ALI BIO JE ODLUČAN! NALOZIMA SE NE DISKUTIRA IH SE IZVRŠUJE!", "SE IZVRŠUJE!", "ŠTO OVDJE MOGLO BITI SUMNJIVO OSIM VLASNI KOVA LICA?", "ŠTO OVDJE MOGLO BITI SUMNJIVO OSIM VLASNIKOVA LICA?"]}

```

```

{"pageText":["LOPOVE, VARALICO, UBOJICO, KRADLJIVČE, OTMIČARU I PLJAČKAŠU!", "OPROSTITUTE MI, MR. MINNING, ALI JA NISAM...", "BUDI PAZLJIV,"]}

```

```

{"pageText":["DA NISI OVDJE PO DUŽNOSTI, SA ZADATKOM DA POMOGNEŠ, ODMAH BI TE NAJURIO!", "USUDUJEM PREDLOŽITI, MR. MINNING, ODA MI ZBOG TOG RAZBIJENOG STAKLA ODBIJETE OD PLAĆE.."]}

```

```

{"pageText":["ZNAČI, DOSAD MI DU AKO USPIJEM OOŽIVJETI.. I OVADESET I OVA DANA RADA...", "MA, KAD VEĆ GOVORIŠ, NEMOJ SE PRITOM RAZMAHATI RUKAMA! NAMJERAVAŠ LI OVDJE RADITI DO SMRTI? MTV", "ŽELIO BIH KAD TO BILO KRATKO!"]}

```

```

{"pageText":["GLEDAJ TI TOG TIPČINU GARRETA! UVIJEK KASNI! MORAO JE STIĆI PRIJE 25 MINUTA! ZANIMA ME ŠTO ĆE IZMISLITI? DA SU PUKLI KOTAČI NA PODZEMNOJ ŽELJEZNICI? ŽAO MI JE ŠTO NE OGU VAMA DIJELITI TO MIŠLJENJE, MISS VIOLET!", "NEMOJTE TO UZETI SRCU, MR. FORD. MR. MINNING JE, ZAPRAVO, DOBAR ČOVJEK!", "ŽAO JE ŠTO NE S VAMA DIJELITI TO MIŠLJENJE, MISS VIOLET!"]}

```

*JSON 1 Povratni rezultati iz aplikacije*

## 6 EVALUACIJA PRIMIJENJENIH METODA

U ovom dijelu diplomskog rada, cilj nam je evaluirati prilagođene metode i alat koji se koristi za ekstrakciju teksta iz stripova. Evaluacija metoda znači procijeniti njihovu učinkovitost, točnost i performanse u odnosu na zadani zadatak ili problem. Cilj evaluacije je razumjeti koliko su metode uspješne u rješavanju problema i kako se međusobno uspoređuju.

Evaluacija se često provodi kroz kvantitativne metrike, poput preciznosti, opoziva i F1-mjere, koje omogućuju objektivnu ocjenu metoda. Osim toga, evaluacija može uključivati i kvalitativne aspekte, kao što su jednostavnost implementacije, brzina izvođenja i razumljivost rezultata.

U kontekstu prepoznavanja teksta iz oblačića razgovora, evaluacija metoda podrazumijeva mjerenje koliko dobro algoritam uspijeva detektirati oblačiće i izvući tekst iz njih .

U ovom radu, izračunavamo kvantitativne metrike, preciznost, opoziv i F1-mjera, kako bismo procijenili učinkovitost i točnost naših prilagodbi te odredili njihovu uspješnost u prepoznavanju i ekstrakciji teksta iz oblačića razgovora. Kroz ovu evaluaciju, nadamo se identificirati područja u kojima naše prilagodbe mogu biti poboljšane, kao i potvrditi da su naše prilagodbe učinkovite u postizanju ciljeva koje smo postavili. Rezultati ovog dijela rada mogli bi nam pružiti vrijedne uvide koji će informirati naše buduće radove i istraživanja na ovom području.

### 6.1 Opis primjera stripova koji će se evaluirati

U ovom radu, naš fokus je na evaluaciji primijenjenih metoda na primjeru stripa Alan Ford, poznatog talijanskog serijala. Za potrebe naše analize, usredotočit ćemo se na prvih pet nastavaka serijala, koji se sastoji od sljedećih naslova: "Grupa TNT", "Šuplji zub", "Operacija Frankenstein", "Zimska idila" i "Daj! Daj! Daj!".

Alan Ford je poznat po svom humoru, složenim likovima i vizualno bogatim ilustracijama. Strip kombinira elemente avanture, satire, akcije i drame, čineći ga zanimljivim izvorom za analizu i evaluaciju našeg alata za prepoznavanje teksta.

Baloni s tekстом u stripu dolaze u različitim formatima i veličinama, a često se preklapaju, što predstavlja izazov za alat koji se koristi za ekstrakciju teksta. Osim toga, stripovi često koriste nestandardne fontove i oblike slova, što dodatno komplicira proces ekstrakcije teksta.

Prvih pet nastavaka stripa pružaju raznovrsne situacije, dijaloge i naraciju, što će omogućiti našem alatu da se suoči s različitim izazovima vezanim uz prepoznavanje teksta. Svaki od ovih nastavaka sadrži oko 120 stranica, što nam pruža dovoljno podataka za detaljnu analizu. U daljnjem tijeku našeg istraživanja, posvetit ćemo pažnju pojedinačnoj analizi svakog stripa, nakon čega ćemo na osnovi dobivenih rezultata sintetizirati ukupnu evaluaciju performansi našeg alata.

U sklopu evaluacije, alat će morati prepoznati tekstualni sadržaj slika, kao što su dijalozni likovi, opisi i naracija. Ovaj proces uključuje segmentaciju teksta, prepoznavanje slova i pretvaranje slika u strojno čitljiv format. Evaluacija će se temeljiti na točnosti prepoznavanja teksta, kao i na kvaliteti generiranog sadržaja u usporedbi s izvornim stripom.

Za dodatnu točnost u našoj evaluaciji, provesti ćemo vizualnu, odnosno ručnu provjeru rezultata. Ovaj korak će nam omogućiti da identificiramo specifične probleme u našem postupku ekstrakcije teksta, kao što su pogreške pri prepoznavanju slova ili propušteni dijelovi teksta. Također, to će nam pomoći u procjeni kvalitete generiranog teksta, odnosno koliko dobro prepoznati tekst odražava izvorni sadržaj stripa.

Da bismo bolje razumjeli performanse našeg alata, analizirat ćemo i vrste pogrešaka koje alat čini. Na primjer, možemo istražiti je li alat posebno slab u prepoznavanju određenih vrsta slova ili oblika balona za tekst, ili ima problema s interpretacijom složenijih stranica sa mnoštvom elemenata. Ova vrsta analize može nam pomoći u identifikaciji područja za daljnje poboljšanje.

Na kraju, usporedit ćemo rezultate dobivene od našeg alata s ljudskim tumačenjem stripova. Ova usporedba će nam omogućiti da vidimo koliko dobro alat može reproducirati ljudsku interpretaciju stripa, što je ključni aspekt u ocjenjivanju njegove učinkovitosti.

Uzimajući u obzir sve ove faktore, ovaj dio rada pružit će sveobuhvatnu evaluaciju našeg pristupa ekstrakciji teksta iz stripova, s posebnim naglaskom na strip serijal Alan Ford.

## 6.2 Metodologija evaluacije

1. **Odabir materijala za evaluaciju:** Za temeljitu i robusnu evaluaciju našeg alata, odabrali smo prvih pet nastavaka poznatog stripa Alan Ford. Svaki od tih nastavaka sadrži otprilike 120 stranica, što nam daje značajan uzorak od oko 600 stranica za analizu. Ova količina podataka omogućuje nam da procijenimo performanse alata u širokom rasponu situacija i uvjeta.
2. **Ekstrakcija teksta iz stripova:** Nakon odabira materijala, koristili smo naš alat za izdvajanje i prepoznavanje teksta iz stripova. Alat je dizajniran da identificira oblake teksta, interpretira slova i pretvara vizualni tekst u strojno čitljiv format. Ova faza je ključna za transformaciju grafičkih elemenata stripa u korisne tekstualne podatke.
3. **Ručna usporedba:** Kako bi smo osigurali točnost i pouzdanost rezultata, poduzeli smo korak ručne provjere generiranih tekstova. Svaka stranica stripa je uspoređena s odgovarajućom stranicom generiranog teksta. Tijekom ovog postupka, identificirali smo točno prepoznate riječi (engl. True Positive - TP), netočno prepoznate riječi (engl. False Positive - FP) i propuštene riječi (engl. False Negative - FN).
4. **Izračun metrika:** Na temelju podataka prikupljenih tijekom ručne usporedbe, izračunali smo ključne metrike - preciznost, opoziv i F1 mjeru. Ove metrike služe kao kvantitativna mjera učinkovitosti našeg alata. Preciznost (engl. Precision) je omjer

točno prepoznatih riječi i ukupnog broja prepoznatih riječi. Opoziv (engl. Recall) je omjer točno prepoznatih riječi i ukupnog broja riječi u stripu. F1 mjera je harmonijska sredina preciznosti i opoziva.

- Preciznost =  $TP / (TP + FP)$
- Opoziv =  $TP / (TP + FN)$
- F1 mjera =  $2 * (Preciznost * Opoziv) / (Preciznost + Opoziv)$

5. **Analiza rezultata:** Nakon što smo izračunali kvantitativne metrike, proveli smo dubinsku analizu rezultata. Ova analiza nam omogućava da bolje razumijemo snage i slabosti našeg alata, identificiramo područja za poboljšanje i donosimo informirane odluke o budućim koracima. Kroz ovu analizu, ne samo da procjenjujemo trenutne performanse alata, već i oblikujemo njegov daljnji razvoj i optimizaciju.

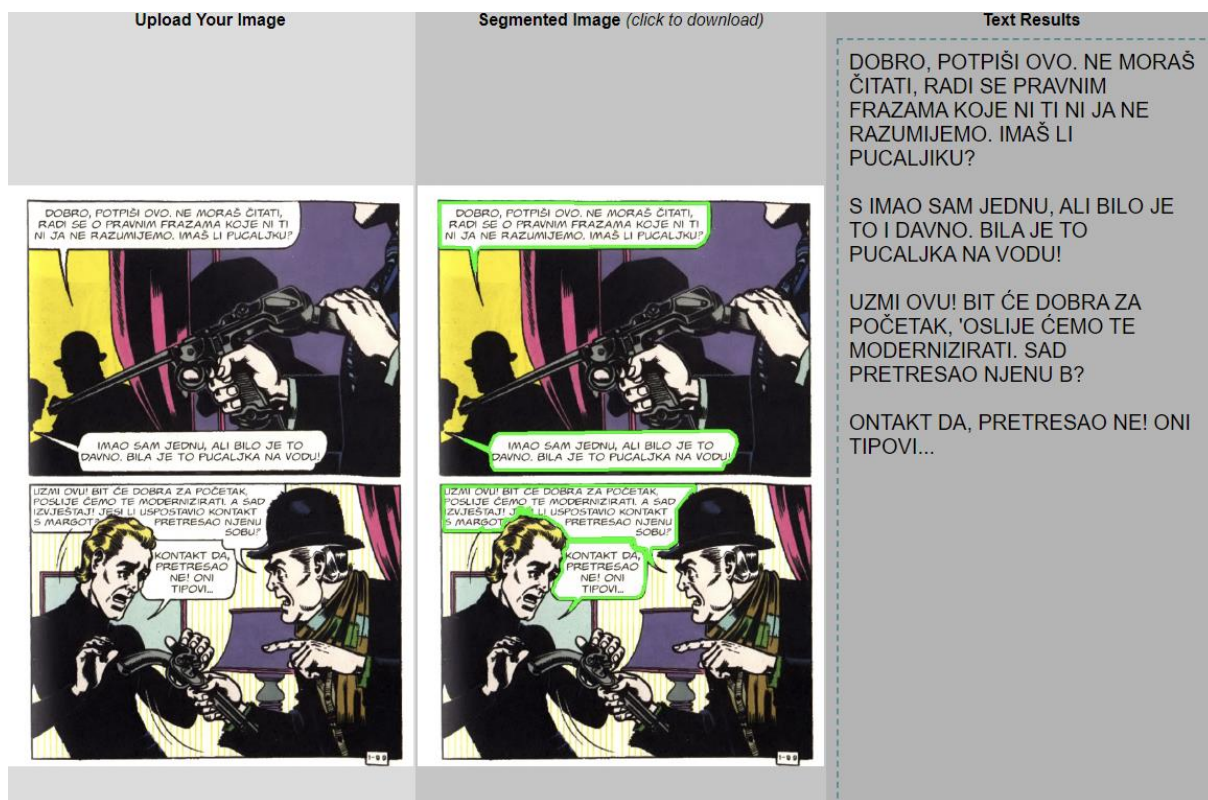
### 6.3 Primjena odabranih metoda i alata na stripovima Alan Ford

Nakon što se stripovi učitaju u aplikaciju i konvertiraju u slike, naš algoritam provodi niz koraka kako bi efikasno obradio balone s tekstom. Ovi koraci uključuju detekciju oblika balona, segmentaciju i OCR čitanje teksta. U nastavku je detaljnije opisan svaki od tih koraka:

1. **Detekcija oblika balona:** Algoritam prvo identificira balone s tekstom na svakoj slici. Detekcija oblika temelji se na kombinaciji različitih značajki, poput boje, kontrasta i geometrijskih oblika. Ovaj korak je ključan za precizno određivanje granica balona, što omogućuje daljnju obradu teksta.
2. **Segmentacija balona:** Nakon što se baloni detektiraju, algoritam provodi segmentaciju kako bi izdvojio pojedinačne tekstualne elemente unutar svakog balona. Segmentacija se odnosi na razdvajanje slova i riječi unutar balona, što olakšava kasnije čitanje i interpretaciju teksta.
3. **OCR čitanje teksta:** Posljednji korak u procesu obrade balona uključuje primjenu optičkog prepoznavanja znakova na segmentirane tekstualne elemente. OCR tehnologija omogućuje prepoznavanje i pretvaranje slova u strojno čitljiv tekst. Rezultat je digitalni tekstualni zapis dijaloga i naracije iz stripa.

Slika 13 Korištenje aplikacije na primjeru jedne stranice Alan Forda ilustrira kako aplikacija prepoznaje i obrađuje balone s tekstom. Baloni su automatski označeni zelenom bojom kako bi se prikazalo područje prepoznavanja, a generirani tekst je prikazan pored njih. Ova vizualna prezentacija omogućuje korisnicima da jednostavno razumiju i prate rad aplikacije te procijene njezinu točnost u prepoznavanju i obradi teksta.

Kroz ovaj proces primjene odabranih metoda i alata na stripovima, dobivamo strojno čitljive tekstualne podatke koji se mogu dalje analizirati i obrađivati.



Slika 13 Korištenje aplikacije na primjeru jedne stranice Alan Forda

## 6.4 Analiza točnosti i korisnosti odabranih metoda i alata

Nakon razmatranja opće strukture procesa i resursa potrebnih za njegovo provođenje, sada ćemo usmjeriti svoju pažnju na detaljno ispitivanje performansi našeg alata kroz konkretne primjere. Svaki od odabranih stripova pružit će jedinstveni uvid u funkcionalnost i efikasnost algoritma.

U sklopu ovog detaljnijeg istraživanja, svaki strip bit će posebno analiziran kako bismo uočili specifične izazove i pogreške koje alat može susresti. Iako su ove prepreke primjenjive na sve stripove, odlučili smo ih pojedinačno prikazati u kontekstu svakog stripa. Time želimo omogućiti čitatelju jasniji i konkretniji uvid u način na koji algoritam djeluje te se prilagođava i reagira na različite scenarije. Sada, krenimo na proučavanje prvog u nizu stripova.

### 6.4.1 Analiza stripa 1

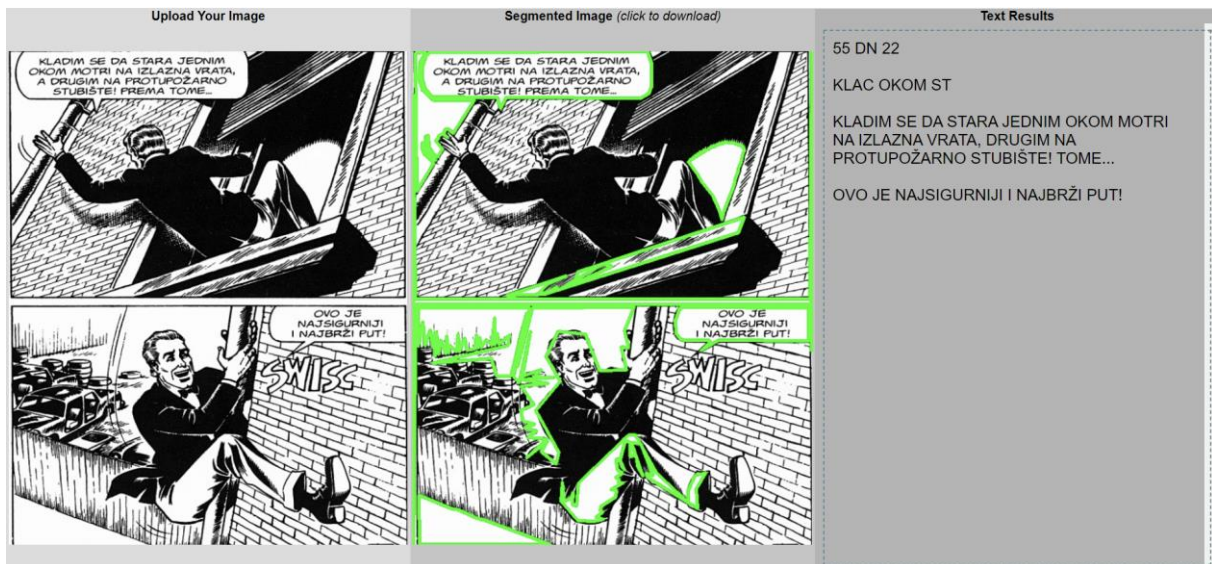
Važno je naglasiti kako se složenost obrade materijala očituje u vremenu potrebnom za završetak procesa segmentacije i ekstrakcije teksta. Našem alatu je trebalo 3.8 minuta da bi obradio sve stranice prvog stripa i vratio rezultate u strojno čitljivom JSON formatu. Iako se to može činiti kao značajan vremenski period, valja uzeti u obzir kompleksnost zadatka koji se obavlja i veliku količinu podataka koja se obrađuje, kako bi se dobilo cjelovito razumijevanje vremenskih zahtjeva procesa.



U kontekstu našeg prvog analiziranog stripa, pod nazivom "Grupa TNT", moramo istaknuti njegovu strukturu. Strip ukupno broji 132 stranice, međutim, stvarni sadržaj stripa nalazi se u rasponu od sedme do 126. stranice, s uvodnim, naslovnim i završnim stranicama koje čine preostali dio.

Strip je u crno-bijelom formatu, konzistentan s prosječnom distribucijom dijaloga, s otprilike tri tekstualna balona po stranici. Dobiveni tekst, kojeg je alat izvukao i konvertirao u JSON format, zauzima 25 stranica i ukupno 3989 riječi te pokazuje opseg i dubinu informacija koje se mogu ekstrahirati iz jednog stripa.

U cilju detaljne ilustracije našeg pristupa, sada ćemo kroz konkretan primjer prvog stripa objasniti postupak analize i izračuna ključnih metrika. Isti postupak bit će primijenjen u svim sljedećim analizama. Za svaki pojedinačni strip, prikazat ćemo ilustracije koje uključuju različite tipove tekstualnih balona, kao i različite razine složenosti.



Slika 14 Korištenje aplikacije na primjeru jedne stranice Alan Forda - 1

Na prethodnoj ilustraciji Slika 14 Korištenje aplikacije na primjeru jedne stranice Alan Forda - demonstriramo kako naš prilagođeni alat interpretira jednu stranicu iz prvog stripa, identificirajući i označavajući tekstualne balone zelenom bojom. Također, važno je naglasiti da u procesu evaluacije metoda i alata za prepoznavanje teksta, potrebno je imati referentni, točni tekst s kojim se rezultati uspoređuju. Ovaj referentni tekst poznat je kao 'ground truth'. U našem slučaju, 'ground truth' predstavlja izvorni tekst iz stripa, s kojim smo ručno uspoređivali riječ po riječ rezultate generirane našim alatom, na prethodnoj ilustraciji vidljiv u lijevom dijelu korisničkog sučelja. Izvorni tekst prvog stripa ima ukupno 3370 riječi.

Uz pomoć 'ground truth' teksta, uspjeli smo precizno odrediti koje je riječi alat točno prepoznao - TP, koje je riječi alat pogrešno prepoznao - FP, te koje riječi alat nije uspio prepoznati - FN. Sve ove informacije zatim koristimo za izračun metrika preciznosti, opoziva i F1.

```
{„pageText“:[„55 DN 22“]}
```



```
{„pageText“:[„KLAC OKOM ST“]}
```

```
{„pageText“:[„KLADIM SE DA STARA JEDNIM OKOM MOTRI NA IZLAZNA VRATA,  
a DRUGIM NA PROTUPOŽARNO STUBIŠTE! Prema TOME...“]}
```

```
{„pageText“:[„OVO JE NAJSIGURNIJI I NAJBRŽI PUT“]}
```

TP- 21 FP – 6 FN - 2

*JSON 2 prikaz rezultata jedne stranice - strip 1*

Da bismo detaljnije prikazali kako smo izračunali ove metrike, odlučili smo vizualno prikazati ranije ilustrirani segment stripa. Na primjeru JSON 2 prikaz rezultata jedne stranice - strip 1 **Error! Reference source not found.**, označili smo TP, FP I FN rezultate, koji predstavljaju ključne faktore u izračunu naših metrika. True Positive, označen zelenom bojom, predstavlja tekst koji je alat točno prepoznao te kao što je vidljivo iz slike na ovom konkretnom primjeru iznosi broj 21.

S druge strane, koncept False Positive odnosi se na riječi koje alat interpretira iz teksta, ali na iskrivljen ili netočan način. Primjerice, u našoj analizi, FP je označen za tekst "55 DN 22" i "KLAC OKOM ST", ilustriran žutom bojom, koji zapravo nije trebao biti prepoznat kao relevantan tekst. Slika iznad ilustrira jedan od ključnih izazova s kojima se naš alat susreće: precizno određivanje granica tekstualnog balona. U stripovima poput Alan Ford, gdje su vizualni elementi bogati i kompleksni, često nailazimo na brojne oblike i konture koje po svojim karakteristikama mogu podsjećati na tekstualne balone. Kao rezultat, naš alat ponekad krivo interpretira te elemente, što dovodi do False Positive rezultata. FN se odnosi na riječi koje alat nije uspio prepoznati. U našem primjeru, to su riječi „a“ i „prema“.

Na osnovu ovog primjera, izračunat ćemo metrike preciznosti, opoziva i F1-metrije kako bismo bolje razumjeli metodologiju izračuna koja se primjenjuje na cijeli skup od prvih pet stripova. Uzimajući u obzir broj točno prepoznatih riječi s gornje slike, dobivamo ukupan broj od 21. Broj pogrešno prepoznatih riječi iznosi 6, dok broj riječi koje nisu uopće prepoznate iznosi 2.

S tim brojevima, možemo izračunati preciznost, opoziv i F1 metriku na sljedeći način:

1. Preciznost =  $TP / (TP + FP) = 21 / (21 + 6) = 0.7778$ , odnosno 77.78%
2. Opoziv =  $TP / (TP + FN) = 21 / (21 + 2) = 0.9130$ , odnosno 91.30%
3. F1 metrika =  $2 * (Preciznost * Opoziv) / (Preciznost + Opoziv) = 2 * (0.7778 * 0.9130) / (0.7778 + 0.9130) = 0.8392$ , odnosno 83.92%

Primjenjujući ove formule, možemo dobiti jasnu sliku o performansama alata na ovom konkretnom primjeru, što nam omogućuje da bolje razumijemo njegovu ukupnu učinkovitost i točnost.

Nakon detaljne analize ovog pojedinačnog primjera, sada prelazimo na sveobuhvatniji pregled performansi našeg alata na cjelokupnom prvom stripu. Temeljem broja točno prepoznatih riječi koji iznosi 2829, broja pogrešno prepoznatih riječi od 619, te broja neprepoznatih riječi koji iznosi 541, alat je za prvi strip postigao sljedeće rezultate:

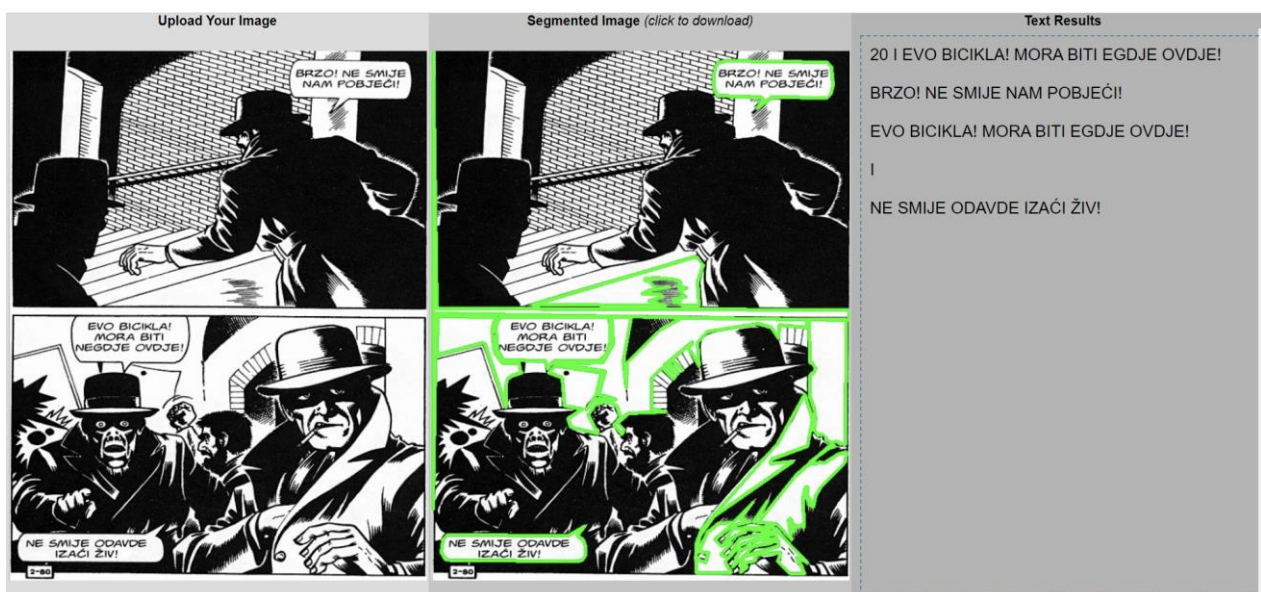
- Preciznost: 82.05%
- Opoziv: 83.95%
- F1 metrika: 82.98%

Ovi rezultati predstavljaju prosječnu učinkovitost alata na svim stranicama prvog stripa, omogućujući dublji uvid u njegove sposobnosti i ograničenja.

### 6.4.2 Analiza stripa 2

Drugom stripu pod nazivom "Šuplji zub". Jednako prvom stripu, i ovaj broji 120 stranica sadržaja. No, ovaj strip donosi nove izazove za naš alat, uključujući brojne vizualne elemente i različite fontove teksta, što je stavilo naš alat na test.

Nakon učitavanja stripa u alat, proces ekstrakcije teksta trajao je nešto duže nego u prvom slučaju, odnosno tri minute i dvije sekunde.



Slika 15 Korištenje aplikacije na primjeru jedne stranice Alan Forda - 2

Na stranici iz drugog stripa Slika 15 Korištenje aplikacije na primjeru jedne stranice Alan Forda - 2, moguće je primijetiti značajne razlike u načinu na koji naš alat određuje granice tekstualnih balona. U gornjem dijelu stranice, alat je pokazao iznimnu preciznost u identifikaciji i označavanju oblačića s tekstom. No, donji dio stranice predstavlja više izazova, gdje alat pokazuje teškoće u preciznom određivanju granica balona.

Posebno je važno naglasiti kako su ove razlike u performansama alata izravno povezane s vizualnim karakteristikama same stranice. Gornji dio stranice obiluje tamnim, crnim tonovima u pozadini, što olakšava prepoznavanje bijelih tekstualnih oblačića. S druge strane, donji dio

stranice sadrži mnogo više bijelih elemenata, stvarajući konfuziju za alat, koji može pogrešno interpretirati ove elemente kao tekstualne balone.

Stoga može se uočiti još jedan problem s kojim se naš alat susreće - ponavljanje teksta. To rezultira time da je broj riječi generiranih od strane naše aplikacije veći od broja riječi u izvornom tekstu. Za strip dva izvorni broj riječi iznosi 3449 dok je broj riječi koji je naša aplikacija prepoznala 4234.

Ovaj problem nastaje uslijed nedostatka jasne strukture balona s tekstom, što rezultira time da alat ne može uvijek točno odrediti poredak čitanja balona. Kada alat naiđe na granice već očitano balona, ponavlja prethodno pročitani tekst.

Ovaj primjer ilustrira koliko je važan kontekst i vizualni okvir za uspješnu identifikaciju i ekstrakciju teksta, te kako promjene u vizualnom okruženju mogu utjecati na performanse alata.

Unatoč tome, broj točno prepoznatih riječi iznosi 3028, broj pogrešno prepoznatih riječi 785 i broj neprepoznatih riječi bile su 421.

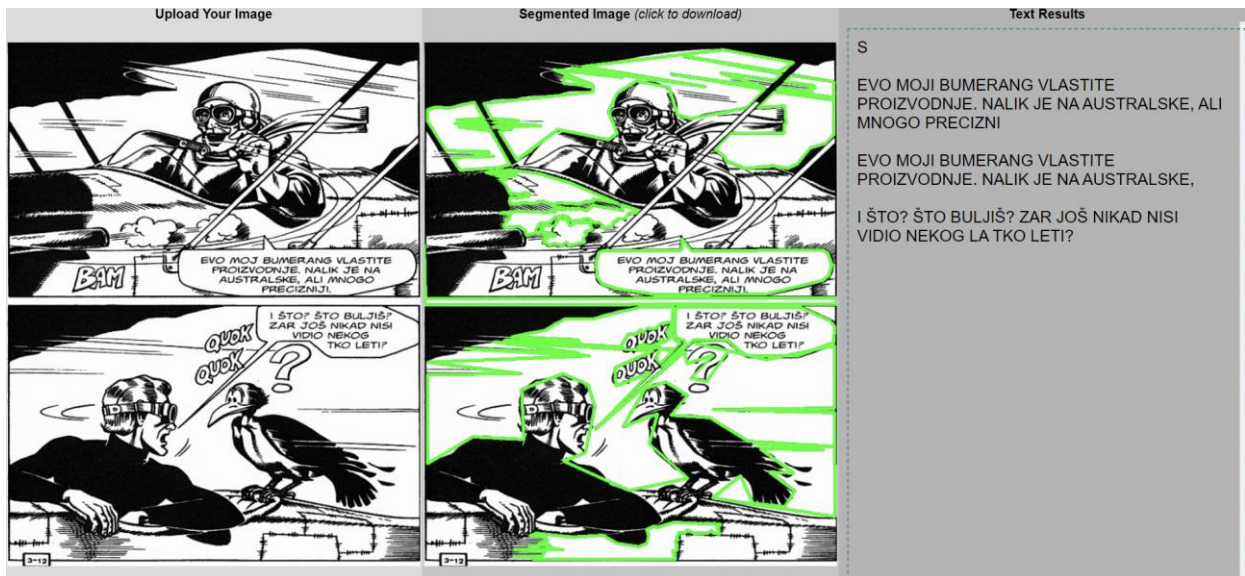
S tim brojevima, izračunali smo preciznost, opoziv i F1 metriku čiji rezultati iznose:

- Preciznost: 79.31%
- Opoziv: 87.79%
- F1 metrika: 83.34%

Sada nastavljamo s analizom trećeg stripa.

### **6.4.3 Analiza stripa 3**

Analiza je nastavljena s trećim stripom, intrigantno nazvanim "Operacija Frankenstein". Ovaj strip, kao i prethodna dva, sastoji se od ukupno 132 stranice, ali njegov suštinski sadržaj smješten je unutar 120 stranica. Kroz ove stranice, naš alat uspješno je ekstrahirao ukupno 5392 riječ, dok izvorni tekst sadrži 4340 riječi, pri čemu je prosječan broj tekstualnih balona po stranici bio 4. Ovaj podatak pokazuje kontinuiranu učinkovitost i preciznost alata, unatoč varijacijama u strukturi i dizajnu različitih stripova.



Slika 16 Korištenje aplikacije na primjeru jedne stranice Alan Forda - 3

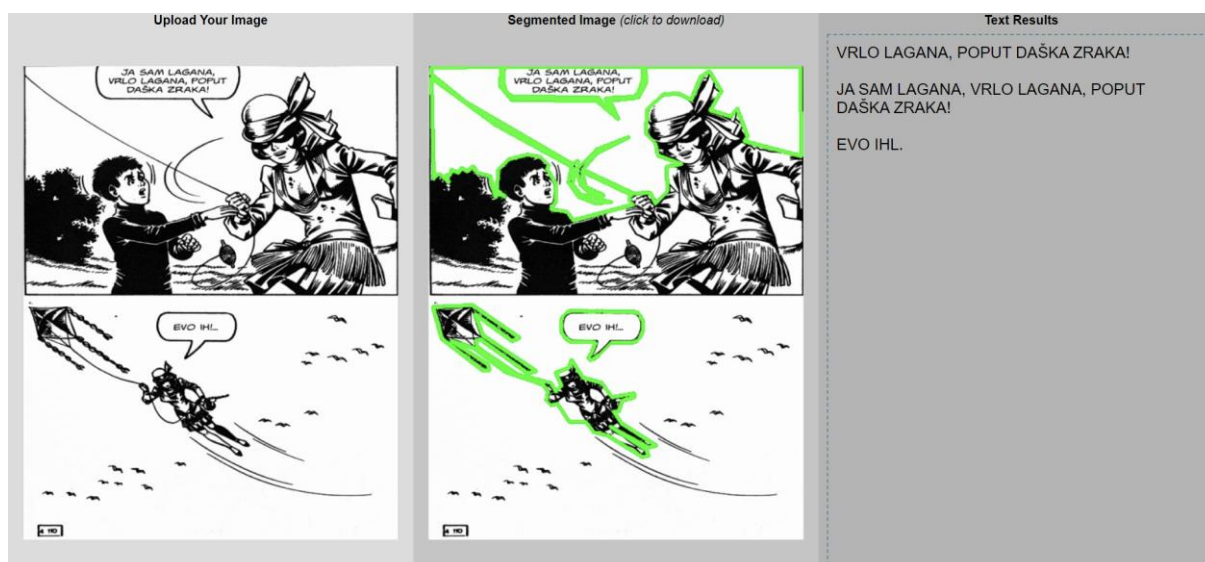
Analiza trećeg stripa, izazvala je izazove za naš alat, posebno na vizualno kompleksnijim stranicama. Jedan od ključnih problema ponovno je bio fenomen ponavljanja teksta, što se pokazalo kao izravna posljedica nesavršene detekcije tekstualnih balona. Na konkretnom primjeru Slika 16 Korištenje aplikacije na primjeru jedne stranice Alan Forda - 3, ilustracija je bila ispunjena čistim bijelim prostorima koje je naš alat pogrešno identificirao kao tekstualne balone. Kao rezultat, OCR algoritam je prolazio kroz sliku više puta iz različitih kutova, uzrokujući neželjeno ponavljanje teksta.

Međutim, važno je primijetiti da kada se izuzme taj ponovljeni tekst, alat je pružio impresivno preciznu ekstrakciju teksta. Također, ovaj strip je ukazao na još jedan specifičan izazov - glasovne efekte. Ti su elementi, poput "BAM" koji ukazuje na udarac, i "QUOK, QUOK" koji simbolizira zvuk ptice, integrirani u slike kako bi pojačali čitateljsko iskustvo. Nažalost, naš alat nije uvijek uspješno prepoznavao ove nekonvencionalne tekstualne elemente. Za strip 3, "Operacija Frankenstein", ostvareni su sljedeći rezultati: preciznost iznosi 79.18%, opoziv je 92.14%, a F1 metrika je postigla vrijednost od 85.17%.

#### 6.4.4 Analiza stripa 4

Prelazimo na analizu četvrtog stripa, poznatog pod nazivom "Kuća duhova". U skladu s dosadašnjim trendom, ovaj strip također sadrži 120 stranica teksta, s prosječnim brojem tekstualnih oblačića, 4 po stranici karakterističnim za Alan Fordove stripove. Tekst ekstrahiran pomoću aplikacije zauzima ukupno 25 stranica teksta, što je u skladu s uobičajenim prosjekom, i sadrži 3953 riječi. Rezultati analize ovog stripa su sljedeći:

Preciznost: 78.79% Opoziv: 86.86% te F1 metrika: 82.63%



Slika 17 Korištenje aplikacije na primjeru jedne stranice Alan Forda - 4

Na ilustrativnom primjeru iz četvrtog stripa prikazanog na Slika 17 Korištenje aplikacije na primjeru jedne stranice Alan Forda - 4, ponovno susrećemo izazov ponavljanja teksta, problema koji smo već identificirali i objasnili u prethodnoj analizi. Ipak, bez obzira na to, tekst je generalno vrlo dobro prepoznat.

Dodatna zanimljivost koju ovaj strip otkriva jest problem prepoznavanja riječi "IH!", gdje je naš alat pogrešno identificirao uskličnik kao slovo "L", rezultirajući s "IHL". Iako je u gornje prepoznatoj rečenici alat precizno detektirao interpunkcijski znak uskličnik, ovaj slučaj ilustrira da alat može naići na teškoće kada se uskličnik nalazi pored skraćenih zamjenica koje alat može teže interpretirati. Ovo ukazuje da, iako alat nije nesposoban za prepoznavanje interpunkcijskih znakova, u određenim okolnostima može doći do određenih poteškoća u njihovom tumačenju.

#### 6.4.5 Analiza stripa 5

Zadnji strip u našem nizu analiza nosi naslov "Daj! Daj! Daj!" i predstavlja značajan odmak od prethodnih stripova zbog jedne ključne razlike - boje. Za razliku od prethodnih crno-bijelih izdanja, ovaj strip smo analizirali u njegovom šarenom izdanju. Ukupan broj stranica ovog stripa je 120, proces ekstrakcije teksta trajao je 3 minuta i 20 sekundi, a ukupan broj izvučenih riječi iznosi 4856.

Valja naglasiti da format stripa igra značajnu ulogu u procesu ekstrakcije teksta. Naime, neki stripovi su u crno-bijelom formatu, dok su drugi u boji. Na temelju naše evaluacije, ustanovili smo da alat efikasnije obrađuje crno-bijele stripove, dok kod stripova u boji može doći do poteškoća u prepoznavanju balona s tekstom. Konkretno, alat ima poteškoća u prepoznavanju balona obojanih u žutu boju, što može dovesti do propusta u ekstrakciji teksta, što je ilustrirano na slici ispod.





Slika 18 Korištenje aplikacije na primjeru jedne stranice Alan Forda - 5

Ove izazove treba uzeti u obzir prilikom evaluacije alata, ali i u budućem razvoju i optimizaciji tehnologija za ekstrakciju teksta iz stripova.

Kako bismo pružili konzistentnu i usporedivu analizu, postupak prikaza analize petog stripa izvest ćemo na sličan način kao i za prvi strip. To će omogućiti jasno uspoređivanje i razumijevanje različitih performansi našeg alata kroz različite uvjete - u ovom slučaju, u odnosu na kolorirani strip nasuprot crno-bijelim stripovima.

```
{„pageText“:[„KAKVA IDEJA! POSLATI ME OVAMO, TO ZBIJLA MOŽE PASTI NA PAMET SAMO ŠEFU. ALI BIO JE ODLUČAN! A o NALOZIMA SE NE DISKUTIRA, već IH SE IZVRŠUJE!“;„ŠTO bi OVDJE MOGLO BITI SUMNJIVO OSIM VLASNIKOVA LIKA?“]}
```

TP- 32 FP-0 FN-4

```
{„pageText“:[„LOPOVE, VARALICO, UBOJICO, KRADLJIVČE, OTMIČARU I PLJAČKAŠU!“;„OPROSTITE MI, MR. MINNING, ALI JA NISAM...“;„BUDI PAŽLJIV“; „Razbila se! Oh, vazda Ming!“]}
```

TP- 16 FP – 0 FN – 5

```
{„pageText“:[„DA NISI OVDJE PO DUŽNOSTI, SA ZADATKOM DA POMOGNEŠ, ODMAH BI TE NAJURIO!“;„USUĐUJEM se PREDLOŽITI, MR. MINNING, ODA da MI ZBOG TOG RAZBIJENOG STAKLA ODBIJETE OD PLAĆE...“]}
```

TP – 25 FP – 1 FN - 2

JSON 3 prikaz rezultata jedne stranice - strip 5

Na gornjim rezultatima JSON 3 prikaz rezultata jedne stranice - strip 5, TP, označen zelenom bojom, predstavlja tekst koji je alat točno prepoznao. S druge strane FP predstavlja riječi koje

je alat očitao iz teksta, ali su neispravno interpretirane, odnosno ekstrahirane su, ali ne pripadaju izvornom tekstu. U našem primjeru, FP je tekst "ODA", označen žutom bojom, koji bi zapravo trebao biti "DA".

Ovdje se suočavamo s dodatnim izazovom u prepoznavanju teksta - identifikacija veznika koji uvodi zavisnu rečenicu. Konkretno, "da" u ovom kontekstu služi kao veznik s ciljem naglašavanja svrhe ili želje koju lik izražava. Naš alat, međutim, susreće poteškoće pri prepoznavanju ovakvih lingvističkih nijansi. FN u našem primjeru su riječi "A o", "već", i "bi". Iz ovog primjera, možemo potvrditi našu prethodnu tvrdnju da alat susreće poteškoće s prepoznavanjem veznika i čestica.

U drugom dijelu teksta na slici, možemo uočiti da alat nije uspio prepoznati mali tekstualni balon u kojem piše "Razbila se! Oh, vazda Ming". Ovaj i slični primjeri se, u kontekstu naše evaluacije, smatraju dodatnim FN primjerima.

Zaključak naše analize posljednjeg, petog stripa "Daj! Daj! Daj!" nudi sljedeće podatke:

- Preciznost: 86.36%
- Opoziv: 92.65%
- F1 metrika: 89.39%

#### **6.4.6 Opća analiza performansi alata**

Nakon što smo demonstrirali proces izračuna preciznosti, opoziva i F1 metrike na svakom od stripova s konkretnim primjerima, sada ćemo se fokusirati na prikazivanje rezultata temeljenih na svih pet stripova iz serijala Alan Ford. Važno je naglasiti da je proces usporedbe između izvornog teksta stripa i generiranog teksta alata obavljen ručno uspoređujući riječ po riječ i s velikom pažnjom kako bi se osigurala što veća točnost rezultata. Ipak, valja uzeti u obzir da postoji mala mogućnost ljudske pogreške u ovom procesu.

Tijekom naše analize, odlučili smo uzeti u obzir ponavljanje teksta, prikazano u prethodnim poglavljima, prilikom izračuna metrike, iako se radi o strukturalnom problemu sa očitavanjima granica tekstualnih oblačića, a ne grešci u prepoznavanju teksta. Također, alatovu nesposobnost da čita tekst iz žutih balona smatrat ćemo greškom. Preciznije, tekst iz žutih balona koji nije prepoznat kategorizirat ćemo kao False Negativ - broj propuštenih riječi. Ovo će nam omogućiti preciznije kvantificiranje učinkovitosti našeg alata.

Po završetku naše analize i detaljne usporedbe svih pet stripova, sada smo u mogućnosti prezentirati kumulativne rezultate. Ove sumirane vrijednosti odražavaju ukupnu uspješnost našeg alata u ekstrakciji teksta iz stripova. U sljedeće dvije tablice Tablica 1 i Tablica 2, detaljno smo prikazali rezultate za svaki strip, kao i ukupne vrijednosti za sve stripove zajedno.

Broj stripa	Broj stranica	Boja	Izvorni tekst (broj riječi)	Ekstrahirani tekst (broj riječi)	True Positive	False Positive	False Negative
1	120	NE	3370	3989	2829	619	541
2	120	NE	3449	4234	3028	785	421
3	120	NE	4340	5392	3999	1052	341
4	120	NE	3204	3953	2783	749	421
5	120	DA	4242	4856	3888	614	354
<b>Ukupno</b>	600	/	18605	22424	16527	3819	2078

Tablica 1 Prikaz rezultata analize po stripovima

Broj stripa	Preciznost	Opoziv	F1 metrika
1	82.05%	83.95%	82.98%
2	79.31%	87.79%	83.34%
3	79.18%	92.14%	85.17%
4	78.79%	86.86%	82.63%
5	86.36%	92.65%	89.39%
<b>Ukupno</b>	81.14%	88.68%	89.39%

Tablica 2 Prikaz rezultata analize po stripovima - 2

Ovi rezultati ukazuju na to da naš alat za prepoznavanje teksta u stripovima postiže dobre performanse u pogledu preciznosti i opoziva, što znači da se većina prepoznatih riječi poklapa s izvornim tekstom. Međutim, preciznost, iako relativno visoka, ukazuje na to da alat ponavlja određeni broj riječi, i također da dosta riječi krivo prepozna, a opoziv ukazuje na to da alat nije uspio prepoznati sve relevantne riječi iz stripa. U konačnici, F1 metrika predstavlja uravnoteženu mjeru između preciznosti i opoziva, što nam omogućuje da bolje razumijemo ukupnu učinkovitost alata u prepoznavanju teksta u stripovima.

S obzirom na ove rezultate, možemo zaključiti da naš alat ima potencijal za daljnje poboljšanje i optimizaciju, posebno u pogledu preciznosti. Detaljna analiza našeg pristupa i metoda koje smo koristili može pružiti uvide u moguće nedostatke i područja za poboljšanje, kako bismo osigurali još bolje performanse u budućim verzijama alata.



## 6.5 Najčešći izazovi i pogreške alata

Unatoč učinkovitosti i visokoj točnosti alata u ekstrakciji teksta iz stripova, tijekom naših analiza identificirali smo nekoliko ponavljajućih izazova i nedostataka koji su utjecali na performanse. U ovom poglavlju sistematizirati ćemo te ključne probleme.

1. **Ponavljanje teksta:** Kao jedan od glavnih izazova ističe se problem ponavljanja teksta. Alat je, zbog svoje metode prepoznavanja tekstualnih oblačića, imao tendenciju više puta prolaziti kroz istu sliku iz različitih smjerova. Kao rezultat, došlo je do ponavljanja ekstrahiranog teksta, posebno kada su slike sadržavale mnogo čistih bijelih elemenata koje je alat pogrešno prepoznao kao tekstualne oblačiće.
2. **Prepoznavanje interpunkcijskih znakova:** Alat je ponekad pokazivao poteškoće u prepoznavanju i interpretaciji interpunkcijskih znakova, posebno kada su se nalazili pored skraćenih zamjenica. Primjerice, uskličnik je alat prepoznao kao slovo L, što je dovelo do pogrešne interpretacije riječi.
3. **Prepoznavanje glasovnih efekata:** Stripovi često koriste glasovne efekte izvan tekstualnih oblačića za pojačavanje dojmljivosti priče. Alat nije uvijek bio sposoban prepoznati ove elemente, što je dovelo do nepotpune ekstrakcije teksta.
4. **Označavanje granica oblačića:** Na nekim primjerima, alat je imao problema s preciznim označavanjem granica tekstualnih oblačića. To se posebno isticalo na slikama s raznolikim vizualnim elementima i različitim nijansama boja.
5. **Varijacije u složenosti slike:** Alat je pokazao varijacije u točnosti na osnovu različitih struktura slike. To je bio izraženije u slučajevima gdje su stripovi bili u boji, što je dodatno kompliciralo proces ekstrakcije teksta.

Iako ove pogreške i izazovi utječu na ukupnu točnost našeg alata, važno je napomenuti da su rezultati i dalje prilično impresivni, uzimajući u obzir kompleksnost zadatka. Izazovi koje smo identificirali pružaju vrijednu priliku za daljnje poboljšanje i optimizaciju alata.

## 7 ZAKLJUČAK

Kroz ovaj diplomski rad, proveli smo dubinsko istraživanje u svrhu razvijanja alata za prepoznavanje i ekstrakciju teksta iz stripova. Naš pristup uključivao je segmentaciju slika, detekciju balona s tekstom i primjenu OCR tehnologije za prepoznavanje teksta. Sustavno smo procijenili performanse našeg alata kroz metodologiju evaluacije koja se temelji na preciznosti, opozivu i F1 metrici.

Ukupna evaluacija rezultata ilustrira da naš alat postiže solidan nivo preciznosti (81.14%), što sugerira da veći dio riječi koje alat prepoznaje precizno odgovara riječima iz izvornog teksta stripa. Unatoč tome, opoziv je malo viši (88.68%), što ukazuje na to da je naš alat propustio identificirati određen broj relevantnih riječi iz stripa. F1 metrika, koja predstavlja harmonijski prosjek preciznosti i opoziva, je 84.7%, ukazujući na relativno uravnotežen odnos između preciznosti i opoziva. Iako su ovi rezultati obećavajući, oni također ističu prostor za daljnja unapređenja našeg alata, posebno u pogledu povećanja preciznosti.

Naše istraživanje otkrilo je i neke izazove s kojima se alat susreće, uključujući poteškoće u prepoznavanju veznika i čestica, problema s ponavljanjem istog teksta te poteškoća u prepoznavanju teksta iz balona obojenih u određene boje.

Kroz ovaj rad, uspješno smo demonstrirali kako se tehnologije poput OCR-a mogu prilagoditi i koristiti za ekstrakciju teksta iz vizualno bogatih medija poput stripova. Unatoč nekim izazovima, alat pokazuje obećavajuće rezultate i otvara mogućnosti za daljnje istraživanje i razvoj u ovom području.

### 7.1 Razmatranje mogućih smjerova budućih istraživanja

U zaključku, iako metode temeljene na dubokom učenju nude značajna poboljšanja u točnosti i preciznosti detekcije oblačića u stripovima, one također dolaze s vlastitim izazovima, kao što su potreba za velikim označenim skupovima podataka i veći računalni zahtjevi.

S druge strane, metode temeljene na ručno izrađenim značajkama i heuristikama, iako jednostavnije i manje računalno zahtjevne, imaju ograničeniju primjenjivost i slabiju generalizaciju na različite stilove stripova. Buduća istraživanja mogla bi se usredotočiti na kombiniranje prednosti oba pristupa kako bi se razvile metode koje su i robusne i skalabilne te mogu se prilagoditi širokom spektru stilova i oblika oblačića u stripovima.

Jedan od mogućih smjerova istraživanja uključuje razvoj hibridnih metoda koje kombiniraju duboko učenje s ručno izrađenim značajkama i heuristikama. Ovakav pristup može pružiti veću fleksibilnost u detekciji oblačića u različitim scenarijima i stilovima stripova. Također, to može pomoći u smanjenju potrebe za velikim označenim skupovima podataka i ubrzavanju procesa treniranja.

Drugi smjer istraživanja može se usredotočiti na poboljšanje metoda dubokog učenja kroz razvoj arhitektura i tehnika koje omogućuju bolju generalizaciju i učinkovitost. Primjerice,

uvođenje tehnika prijenosa učenja ili korištenje generativnih modela može pomoći u smanjenju potrebe za velikim označenim skupovima podataka i poboljšanju detekcije oblačića u različitim stilovima stripova.

U svakom slučaju, daljnja istraživanja u ovom području mogu doprinijeti razvoju sveobuhvatnih metoda za prepoznavanje oblačića u stripovima, olakšavajući analizu i pristup sadržaju stripova za različite primjene, kao što su automatsko prevođenje, obrada prirodnog jezika ili multimedijaska analiza.

## REFERENCE

- [1] Kohei Arai and Herman Tolle. Method for real time text extraction of digital manga comic. *International Journal of Image Processing (IJIP)*, 4(6):669–676, 2011. URL <http://www.cscjournals.org/manuscript/Journals/IJIP/Volume4/Issue6/IJIP-290.pdf>.
- [2] Anh Khoi Ngo Ho, Jean-Christophe Burie, and JeanMarc Ogier. Panel and speech balloon extraction from comic books. In *2012 10th IAPR International Workshop on Document Analysis Systems*, pages 424–428, March 2012. URL [https://www.researchgate.net/publication/237151028\\_Panel\\_and\\_Speech\\_Balloon\\_Extraction\\_from\\_Comic\\_Books](https://www.researchgate.net/publication/237151028_Panel_and_Speech_Balloon_Extraction_from_Comic_Books)
- [3] Rigaud, C., Burie, J.C., Ogier, J.M., Karatzas, D., Van de Weijer, J.: An active contour model for speech balloon detection in comics. In: *2013 12th International Conference on Document Analysis and Recognition*, pp. 1240–1244. IEEE (2013)
- [4] Kass, M., Witkin, A., Terzopoulos, D.: Snakes: active contour models. *Int. J. Comput. Vis.* 1(4), 321–331 (1988)
- [5] Rigaud, C., Burie, J.C., Ogier, J.M.: Text-independent speechballoon segmentation for comics and manga. In: *International Workshop on Graphics Recognition*, pp. 133–147. Springer (2015)
- [6] Rigaud, C., Guérin, C., Karatzas, D., Burie, J.C., Ogier, J.M.: Knowledge-driven understanding of images in comic books. *IJDAR* 18(3), 199–221 (2015)
- [7] Rigaud, C., Nguyen, V., Burie, J.C.: Confidence criterion for speechballoon segmentation. In: *13th IAPR International Workshop on Graphics Recognition* (2019)
- [8] Nguyen, N.V., Rigaud, C., Burie, J.C.: Comic MTL: optimized multi-task learning for comic book image analysis. *Int. J. Doc. Anal. Recognit. IJDAR* 22(3), 265–284 (2019)
- [9] He, K., Gkioxari, G., Dollár, P., Girshick, R.: Mask R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2961–2969 (2017)
- [10] Dubray, D., Laubrock, J.: Deep CNN-based speech balloon detection and segmentation for comic books. *arXiv preprint arXiv:1902.08137* (2019) URL <https://arxiv.org/pdf/1902.08137.pdf>
- [11] Mohit Iyyer, Varun Manjunatha, Anupam Guha, Yogarshi Vyas, Jordan Boyd-Graber, Hal Daumé III, and Larry Davis. The amazing mysteries of the gutter: Drawing inferences between panels in comic book narratives. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.
- [12] Shaoqing Ren, Kaiming He, Ross B. Girshick, and Jian Sun. Faster R-CNN: towards real-time object detection with region proposal networks. *CoRR*, abs/1506.01497, 2015. URL <http://arxiv.org/abs/1506.01497>.

- [13] Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional net-works for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 234–241. Springer (2015)
- [14] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556(2014)
- [15] Jochen Laubrock and David Dubray. CNN-based classification of illustrator style in graphic novels: Which features contribute most? In Kompatsiaris et al. [21], pages 684–695. ISBN 978-3-030-05716-9. URL [https://doi.org/10.1007/978-3-030-05716-9\\_61](https://doi.org/10.1007/978-3-030-05716-9_61).
- [16] Arpita Dutta, Samit Biswas and Amit Kumar Das , CNN-based segmentation of speech balloons and narrative text boxes from comic book page images URL [https://www.researchgate.net/publication/351047919\\_CNN-based\\_segmentation\\_of\\_speech\\_balloons\\_and\\_narrative\\_text\\_boxes\\_from\\_comic\\_book\\_page\\_images](https://www.researchgate.net/publication/351047919_CNN-based_segmentation_of_speech_balloons_and_narrative_text_boxes_from_comic_book_page_images)
- [17] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. 2009 IEEE Conference on Computer Vision and Pattern Recognition, pages 248–255, 2009. URL <https://doi.org/10.1109/CVPR.2009.5206848>.
- [18] Clément Guérin, Christophe Rigaud, Antoine Mercier, Farid Ammar-Boudjelal, Karell Bertet, Alain Bouju, Jean-Christophe Burie, Georges Louis, Jean-Marc Ogier, and Arnaud Revel. eBDtheque: a representative database of comics. In Proceedings of the 12th International Conference on Document Analysis and Recognition (ICDAR), pages 1145–1149, 2013. URL <https://doi.org/10.1109/ICDAR.2013.232>.

## POPIS SLIKA

Slika 1	Prepoznavanje oblačića na slici i postupak ekstrakcije teksta, Ho i suradnici [2].....	7
Slika 2	Lokalizacija preklapanja govornih oblačića.....	9
Slika 3	Ulazni i izlazni podaci optičkog prepoznavanja znakova pomoću LSTM-a .....	14
Slika 4	Tekst sa varijabilnom linijom osnove .....	15
Slika 5	Izgled korisničkog sučelja alata "comic-book-reader" .....	16
Slika 6	Struktura „comic-book-reader“ sustava .....	17
Slika 7	Korištenje korisničkog sučelja alata "comic-book-reader" .....	19
Slika 8	Tekst izvan tekstualnih oblačića: Primjer na znaku autobusne stanice .....	20
Slika 9	Višestruki tekstualni oblačići iz Alan Ford stripa: Različitost u veličini, obliku i boji	21
Slika 10	Ekstrakcija teksta alatom "comic-book-reader" .....	22
Slika 11	Ekstrakcija teksta prilagođenim alatom "comic-book-reader" .....	22
Slika 12	Proširena struktura prilagođenog „comic-book-reader“ sustava.....	25
Slika 13	Korištenje aplikacije na primjeru jedne stranice Alan Forda.....	33
Slika 14	Korištenje aplikacije na primjeru jedne stranice Alan Forda - 1 .....	34
Slika 16	Korištenje aplikacije na primjeru jedne stranice Alan Forda - 2 .....	36
Slika 17	Korištenje aplikacije na primjeru jedne stranice Alan Forda - 3 .....	38
Slika 18	Korištenje aplikacije na primjeru jedne stranice Alan Forda - 4 .....	39
Slika 19	Korištenje aplikacije na primjeru jedne stranice Alan Forda - 5 .....	40

## **POPIS TABLICA**

Tablica 1 Prikaz rezultata analize po stripovima.....	42
Tablica 2 Prikaz rezultata analize po stripovima - 2 .....	42