

SVEUČILIŠTE U SPLITU
FAKULTET ELEKTROTEHNIKE, STROJARSTVA I
BRODOGRADNJE

DOKTORSKI STUDIJ ELEKTROTEHNIKE I INFORMACIJSKE
TEHNOLOGIJE

KVALIFIKACIJSKI DOKTORSKI ISPIT

AUTOMATSKO PREPOZNAVANJE
VIŠEJEZIČNIH DEKLARACIJA NA
AMBALAŽI PRETPAKIRANIH
PREHRAMBENIH PROIZVODA

Dario Vranješ

Split, kolovoz 2026.

SADRŽAJ

1. Uvod	1
2. Ambalaža i označavanje pretpakiranih prehrambenih proizvoda	3
2.1. Pojam, uloga i vrste ambalaže.	3
2.2. Obvezne informacije na deklaracijama pretpakirane hrane	5
2.3. Višejezične ambalaže i izazovi zajedničkog tržišta	7
3. Zakonodavni okvir i usklađenost po državama	8
3.1. Regulatorna Europske unije	8
3.1.1. Označavanje hrane i aktualni razvoj propisa.	8
3.1.2. Ambalaža i ambalažni otpad.	8
3.1.3. Jezični zahtjevi i višejezične deklaracije u EU	9
3.2. Nacionalni propisi odabranih država	10
3.2.1. Hrvatska.	10
3.2.2. Odabrane države članice EU.	10
3.2.3. Regija zapadnog Balkana.	13
3.2.4. Ostale države izvan Europske unije	13
3.3. Problemi i pridržavanje propisa u pojedinim državama	15
3.3.1. Neoznačeni alergeni kao vodeći problem.	15
3.3.2. Inspekcijski nalazi i sankcije.	16
3.3.3. Točnost nutritivnih deklaracija i čitljivost	16
3.3.4. E-trgovina i prijevare s označavanjem.	16
4. Tehnologije automatskog prepoznavanja teksta na ambalaži	18
4.1. Optičko prepoznavanje znakova (OCR)	18
4.1.1. Razvoj pristupa: od klasičnog OCR-a do transformera.	18
4.1.2. Deklaracija kao problem prepoznavanja teksta u sceni	19
4.1.3. Usporedba otvorenih OCR biblioteka	19

4.2. OCR na klijentskoj strani (on-device)	20
4.2.1. Ugrađena mobilna rješenja.	20
4.2.2. Usporedba obrade na uređaju i u oblaku	21
4.2.3. Multimodalni jezični modeli kao alternativa	22
4.2.4. Cjevovod obrade deklaracije	22
4.3. Automatsko prepoznavanje jezika i obrada višejezičnog teksta	23
4.3.1. Identifikacija jezika i pisma	23
4.3.2. Segmentacija višejezičnih blokova.	24
4.3.3. Povezivanje i normalizacija višejezičnih podataka.	24
4.3.4. Prepoznavanje imenovanih entiteta u domeni hrane.	25
4.4. Postojeća rješenja i srodne aplikacije	25
4.4.1. Open Food Facts i Yuka.	25
4.4.2. Ostale aplikacije i infrastruktura	26
4.4.3. Analiza praznina (GAP analiza)	26
4.5. Metode za kontinuirano poboljšanje i analizu prikupljenih podataka	26
4.5.1. Detekcija anomalija i provjera konzistentnosti strukturiranih podataka	28
4.5.2. Aktivno učenje i učenje iz ljudskih korekcija	28
4.5.3. Učenje na uređaju i federativno učenje	29
4.5.4. Dovršavanje grafa znanja i predviđanje veza	29
5. Aplikacija ScanEat i prijedlog nadogradnje	31
5.1. Postojeća arhitektura i funkcionalnosti	31
5.2. Automatsko OCR skeniranje na klijentskoj strani	32
5.3. Prepoznavanje višejezičnih ambalaža	35
5.4. Metodologija vrednovanja i očekivani rezultati	35
5.5. Smjerovi doktorskog istraživanja.	37
5.6. Istraživački potencijal i očekivani doprinosi daljnjeg doktorskog istraživanja	39
6. Zaključak	40
LITERATURA	42

POPIS OZNAKA I KRATICA	58
SAŽETAK	59

1. UVOD

Ambalaža pretpakiranih prehrambenih proizvoda danas je znatno više od fizičke zaštite sadržaja: ona je primarni komunikacijski kanal između proizvođača i potrošača te nositelj zakonom propisanih informacija o hrani. Deklaracija na ambalaži potrošaču mora pružiti točne i čitljive podatke o sastavu, alergenima, hranjivim vrijednostima, roku trajanja i podrijetlu proizvoda, a regulatornim tijelima služi kao temelj nadzora tržišta. U Europskoj uniji ovo područje uređuje Uredba (EU) br. 1169/2011 o informiranju potrošača o hrani [1], dok samu ambalažu kao proizvod i otpad od 2025. godine uređuje nova Uredba (EU) 2025/40 o ambalaži i ambalažnom otpadu [2].

Unatoč detaljnoj regulativi, u praksi postoje značajni problemi: deklaracije su često ispisane slovima na granici propisane čitljivosti, informacije na višejezičnim ambalažama znaju biti nepotpune ili pogrešno prevedene, a istraživanja i sustavi brzog uzbunjivanja pokazuju da je neoznačeni alergen jedan od najčešćih uzroka opoziva prehrambenih proizvoda s tržišta [3]. Istodobno, zahtjevi pojedinih država razlikuju se i unutar EU-a (jezik deklaracije, nacionalne oznake za odvajanje ambalažnog otpada), a pogotovo izvan njega, što proizvođače navodi na uporabu zajedničkih višejezičnih etiketa za više tržišta odjednom. Takve etikete, tipične za regiju jugoistočne Europe, dodatno smanjuju čitljivost i otežavaju potrošaču pronalaženje informacija na vlastitom jeziku.

Potrošaču koji želi donositi informirane odluke pri kupnji pretpakiranih prehrambenih proizvoda na raspolaganju su mobilne aplikacije za skeniranje proizvoda. Postojeća rješenja, poput aplikacija Yuka i Open Food Facts [4], u pravilu se oslanjaju na skeniranje crtičnog koda i kolaborativne baze podataka, pa za proizvode kojih nema u bazi podataka, što je čest slučaj kod regionalnih i višejezičnih proizvoda, zahtijevaju spor i pogreškama sklon ručni unos podataka. U sklopu prethodnog istraživanja autora razvijena je aplikacija ScanEat [5] za analizu pretpakiranih prehrambenih proizvoda, koja u postojećoj inačici počiva na poluautomatskom unosu podataka s deklaracije. Aplikacija pritom nije istraživački prototip, nego produkcijski sustav s više od 150.000 korisnika i više od 2 milijuna skeniranja, iz čijega je dosadašnjeg rada nastao korpus s ručno provjerenim strukturiranim podacima za preko 50.000 proizvoda (v. potpoglavlja 5.1 i 5.4).

Cilj je ovog rada dvojak. Prvo, dati sustavan pregled područja: funkcija i vrsta ambalaže pretpakirane hrane, zakonodavnog okvira označavanja i ambalaže u EU i izvan nje, jezičnih zahtjeva pojedinih država te dokumentiranih problema u pridržavanju propisa. Drugo, na temelju pregleda tehnologija optičkog prepoznavanja znakova (engl. Optical Character Recognition, OCR) i metoda obrade prirodnog jezika, predložiti nadogradnju aplikacije ScanEat automatskim OCR skeniranjem deklaracija na klijentskoj strani (izravno na mobilnom uređaju) i automatskim prepoznavanjem višejezičnih ambalaža, čime bi se poluautomatski unos zamijenio automatizira-

nim, brzim i privatnosti sklonim postupkom.

Rad je strukturiran na sljedeći način. U 2. poglavlju obrađuju se pojam, funkcije i označavanje ambalaže pretpakiranih prehrambenih proizvoda. U 3. poglavlju daje se pregled zakonodavnog okvira u EU i odabranim državama izvan EU-a, jezičnih zahtjeva te analiza problema pridržavanja propisa po državama. U 4. poglavlju pregledno se obrađuju tehnologije automatskog prepoznavanja teksta na ambalaži s naglaskom na izvođenje na klijentskoj strani i višejezičnost. U 5. poglavlju opisuje se aplikacija ScanEat, predlaže arhitektura nadogradnje i metodologija vrednovanja. Rad završava zaključkom i smjernicama za daljnje istraživanje.

2. AMBALAŽA I OZNAČAVANJE PRETPAKIRANIH PREHRAMBENIH PROIZVODA

2.1. Pojam, uloga i vrste ambalaže

Pravna definicija ambalaže u Europskoj uniji potječe iz Direktive 94/62/EZ o ambalaži i ambalažnom otpadu, koja ambalažu dijeli u tri kategorije: prodajnu ili primarnu ambalažu, koja čini prodajnu jedinicu ponuđenu krajnjem korisniku na mjestu kupnje; skupnu ili sekundarnu ambalažu, koja grupira više prodajnih jedinica; te transportnu ili tercijarnu ambalažu, koja olakšava rukovanje i prijevoz [6]. S njom je usko povezan pojam pretpakirane hrane iz Uredbe (EU) br. 1169/2011: to je hrana stavljena u ambalažu prije prodaje na način da se sadržaj ne može promijeniti bez otvaranja ili promjene ambalaže [1].

U literaturi se funkcije ambalaže prehrambenih proizvoda uobičajeno dijele na zaštitnu (očuvanje proizvoda od mehaničkih, kemijskih i mikrobioloških utjecaja te produljenje roka trajanja), logističku (rukovanje, skladištenje i transport), prodajnu odnosno marketinšku (privlačenje pozornosti i razlikovanje od konkurencije) te informativnu funkciju. Upravo je informativna funkcija u središtu ovog rada: ambalaža je fizički nositelj deklaracije, zakonom propisanog skupa informacija o hrani, i time primarno sučelje između proizvođača, potrošača i nadzornih tijela.

Ambalaža mora zadovoljiti i bitne zahtjeve glede sastava i proizvodnje: obujam i masa moraju biti ograničeni na najmanju mjeru koja osigurava sigurnost i higijenu proizvoda, sadržaj opasnih tvari mora biti minimiziran, a zbroj koncentracija olova, kadmija, žive i šesterovalentnog kroma ne smije prijeći 100 ppm masenog udjela [6]. Ambalaža koja dolazi u neposredan dodir s hranom podliježe i Uredbi (EZ) br. 1935/2004 o materijalima i predmetima koji dolaze u dodir s hranom, prema kojoj materijali ne smiju otpuštati tvari u količinama opasnim za zdravlje, a predmeti namijenjeni dodiru s hranom označavaju se riječima “za kontakt s hranom” ili simbolom čaše i vilice [7].

Prema materijalu izrade ambalaža pretpakirane hrane najčešće je plastična (PET, HDPE, PP, višeslojni laminati), papirno-kartonska, staklena, metalna (aluminij, bijeli lim) ili kompozitna (npr. višeslojna kartonska ambalaža za napitke). Za identifikaciju materijala Odlukom Komisije 97/129/EZ uspostavljen je sustav brojčanih i alfanumeričkih kodova, čija je uporaba na razini EU dobrovoljna [8]. Odabrani kodovi prikazani su u tablici 2.1.

Tablica 2.1. Odabrane oznake ambalažnih materijala prema Odluci Komisije 97/129/EZ [8]

Materijal	Kratica	Brojčani kod
PET (polietilen-tereftalat)	PET	1
HDPE (polietilen visoke gustoće)	HDPE	2
PVC (polivinil-klorid)	PVC	3
LDPE (polietilen niske gustoće)	LDPE	4
Polipropilen	PP	5
Polistiren	PS	6
Valoviti karton	PAP	20
Nevaloviti karton	PAP	21
Papir	PAP	22
Čelik	FE	40
Aluminij	ALU	41
Bezbojno staklo	GL	70
Kompozit papir/plastika/aluminij	C/PAP	84

Osim oznaka materijala, na ambalaži se pojavljuje niz dodatnih simbola različitog pravnog statusa: Möbiusova petlja (ISO 7000, simbol 1135) kao samodeklarirana tvrdnja recikabilnosti prema normi EN ISO 14021 [9], Zelena točka kao financijska oznaka sudjelovanja proizvođača u sustavu proširene odgovornosti proizvođača (a ne tvrdnja o recikabilnosti) [10], Tidyman kao apel protiv bacanja smeća, nacionalne oznake za razvrstavanje otpada (npr. francuski Triman) te metrološki znak “e” kojim proizvođač jamči sukladnost nazivne količine punjenja prema Direktivi 76/211/EEZ [11]. Ova heterogenost simbola, o kojoj će detaljnije biti riječi u 3. poglavlju, otežava snalaženje potrošača i motivira harmonizaciju označavanja koju donosi nova Uredba (EU) 2025/40 [2].

Suvremeni trendovi u ambalaži hrane dodatno usložnjavaju sliku: višeslojni laminati zamjenjuju se monomaterijalnim strukturama radi lakšeg recikliranja [12], za pojedine se formate zahtijeva industrijska kompostabilnost prema normi EN 13432 [13], a razvijaju se i aktivna ambalaža (koja namjerno otpušta ili apsorbira tvari radi produljenja roka trajanja) i inteligentna ambalaža (koja prati stanje hrane, npr. vremensko-temperaturnim indikatorima), uređene Uredbom (EZ) br. 450/2009 [14]. Pametne etikete s QR kodovima, RFID/NFC oznakama i indikatorima svježine približavaju ambalažu digitalnim sustavima poput aplikacije koja je predmet ovog rada.

2.2. Obvezne informacije na deklaracijama pretpakirane hrane

Sadržaj deklaracije pretpakirane hrane u EU iscrpno uređuje Uredba (EU) br. 1169/2011 o informiranju potrošača o hrani (engl. Food Information to Consumers, FIC), koja se primjenjuje od 13. prosinca 2014., a obvezu nutritivne deklaracije uvela je od 13. prosinca 2016. [1, 15]. Prema članku 9. te uredbe obvezni su sljedeći podaci: naziv hrane; popis sastojaka; tvari koje izazivaju alergije ili intolerancije; količina određenih sastojaka (QUID); neto količina; datum minimalne trajnosti (“najbolje upotrijebiti do”) ili datum uporabe (“upotrijebiti do”); posebni uvjeti čuvanja i/ili uporabe; ime i adresa odgovornog subjekta; zemlja ili mjesto podrijetla kada je to propisano; upute za uporabu; stvarna alkoholna jakost za pića s više od 1,2% vol. alkohola; te nutritivna deklaracija [1]. Primjer deklaracije pretpakirane hrane prikazan je na slici 2.1.

Za čitljivost deklaracije, a time i za njezino strojno očitavanje, ključan je članak 13.: obvezni podaci tiskaju se slovima čija je x-visina (visina malog slova “x”) najmanje 1,2 mm, odnosno najmanje 0,9 mm na ambalaži čija je najveća površina manja od 80 cm² [1]. Ambalaža s najvećom površinom manjom od 25 cm² izuzeta je od nutritivne deklaracije, a na ambalaži manjoj od 10 cm² obvezni su samo naziv, alergeni, neto količina i datum trajnosti. Naziv hrane, neto količina i alkoholna jakost moraju se nalaziti u istom vidnom polju. Ovi zakonski minimumi izravno definiraju donju granicu veličine teksta koju sustav automatskog prepoznavanja deklaracija mora pouzdano čitati: slova x-visine 0,9–1,2 mm na fotografiji snimljenoj mobilnim uređajem tipično zauzimaju svega desetak do dvadesetak piksela, što je na granici mogućnosti suvremenih OCR sustava (o čemu detaljnije u 4. poglavlju).

Posebna se pozornost posvećuje alergenima. Prilog II. Uredbe navodi 14 kategorija tvari koje izazivaju alergije ili intolerancije: žitarice koje sadrže gluten, rakove, jaja, ribu, kikiriki, soju, mlijeko (uključujući laktozu), orašaste plodove, celer, gorušicu, sjemenke sezama, sumporov dioksid i sulfite u koncentraciji većoj od 10 mg/kg ili 10 mg/L, lupinu i mekušce [1]. Alergeni se u popisu sastojaka moraju istaknuti slogom koji ih jasno razlikuje od ostatka popisa, primjerice podebljanim slovima. Za sustav automatske analize deklaracija to je dvostruko važno: isticanje je vizualni signal koji se OCR-om može iskoristiti za pouzdanije izdvajanje alergena, a istodobno je upravo prisutnost i ispravno isticanje alergena jedan od zahtjeva čiju bi sukladnost aplikacija mogla provjeravati.

Nutritivna deklaracija prema člancima 29. do 35. obvezno sadrži energetska vrijednost (u kJ i kcal) te količine masti, zasićenih masnih kiselina, ugljikohidrata, šećera, bjelancevina i soli, tim redoslijedom, izraženo na 100 g ili 100 mL proizvoda, uz dopušteno dodatno izražavanje po porciji [1]. Prikaz je u pravilu tabličan, što nutritivnu tablicu čini strukturno najpredvidljivijim dijelom deklaracije i zahvalnom metom automatske ekstrakcije podataka. Članak 35. dopušta dodatne oblike izražavanja i prikazivanja, što je pravni temelj dobrovoljnih shema označavanja na prednjoj strani ambalaže (engl. front-of-pack, FOP) poput Nutri-Scorea [15].

Navođenje zemlje podrijetla obvezno je kada bi izostavljanje moglo obmanuti potrošača te za svježe, rashlađeno i smrznuto svinjsko, ovčje i kozje meso te meso peradi; Provedbena uredba Komisije (EU) 2018/775 dodatno propisuje da se, kada je podrijetlo hrane istaknuto, a glavni sastojak potječe iz drugog područja, mora navesti i podrijetlo glavnog sastojka, u istom vidnom polju i slovima x-visine od najmanje 75% x-visine oznake podrijetla [16]. Uz deklaraciju u užem smislu, na ambalaži se nalazi i oznaka serije odnosno lota s prefiksom “L” prema Direktivi 2011/91/EU, ključna za sljedivost i opozive proizvoda [17], te prehrambene i zdravstvene tvrdnje koje smiju biti samo one dopuštene prema Uredbi (EZ) br. 1924/2006 [18].

Cream banana

HR Pjenasti proizvod okusa banane s kakaovim preljevom (22 %)

Sastojci: šećer, kakaov preljev 22% [šećer, potpuno hidrogenizirana palmirina mast, kakao u prahu smanjene masti 12%, emulgatori: lecitini (iz SOJE), E476, E492; banana u prahu 0,01%, aroma], glukozni sirup, voda, tvar za želiranje: E406; BJELANJAK JAJE u prahu, kiselina: limunska kiselina, aroma banane, konzervansi: E202; bojilo: E100. Može sadržavati tragove pšeničnog glutena. Kakaov preljev sadrži 5% bezmasne suhe tvari kakaovih dijelova i 25% ukupne masti. Čuvati na suhom i hladnom mjestu. Zemlja podrijetla: Srbija; Zemlja podrijetla šećera: Srbija.

SLO Penasti izdelek z okusom banane in kakaovim prelivom (22 %)

Sestavine: sladkor, kakaov preliv 22 % [sladkor, popolnoma hidrogenizirana palmova maščoba, kakav v prahu z manj maščobe 12 %, emulgatorji: lecitini (iz SOJE), E476, E492; banana v prahu 0,01 %, aroma], glukozni sirup, voda, snov za želiranje: E406; JAJČNI BELJAK v prahu, kislina: limonska kislina aroma banane, konzervansi: E202; barvilo: E100. Lahko vsebuje sledi pšeničnega glutena. Kakavov preliv vsebuje 5 % nemastne suhe snovi kakaovih delov in 25 % skupne maščobe. Hraniti na suhem in hladnem mestu. Država porekla: Srbija; država porekla sladkorja: Srbija.

Uvoznik za HR/SLO: Hanael food d.o.o., Radnički dol 12, 10000 Zagreb, Hrvatska – www.hanaelfood.hr

Neto količina/ Neto količina: **510 g**

21 PAP

Prosječna hranjiva vrijednost na 100 g/ Povprečna hranilna vrednost na 100 g

	1810 kJ/ 431 kcal	22 %
Energija/ Energija		
Masti/ Maščobe	12 g	17 %
od kojih su zasićene masne kiseline/ od tega nasičene maščobne kisline	11 g	55 %
Ugljikohidrati/ Ogljikovi hidrati	79 g	30 %
od kojih šećeri/ od tega sladkorji	41 g	46 %
Bjelančevine/ Beljakovine	2,0 g	4 %
Sol/ Sol	0,3 g	5 %

*PU – Preporučeni unos za prosječnu odraslu osobu/ *PV – priporočen vnos za povprečno odraslo osebo (8 400 kJ/ 2 000 kcal)

Obvezni podaci prema Uredbi (EU) br. 1169/2011, čl. 9. st. 1.

- 1 Naziv hrane (čl. 9. st. 1. t. a) – trgovački i zakonski naziv
- 2 Popis sastojaka (čl. 9. st. 1. t. b) – uvod „Sastojci:”, padajući redoslijed po masi
- 3 Alergeni istaknuti slogom (čl. 9. t. c, čl. 21., Prilog II.) – SOJA, JAJE, gluten (tragovi)
- 4 Količina sastojka – QUID (čl. 9. t. d, čl. 22.) – „kakaov preljev 22 %”
- 5 Neto količina (čl. 9. t. e, čl. 23.) – 510 g
- 6 Datum minimalne trajnosti (čl. 9. t. f, čl. 24.) – „Najbolje upotrijebiti do”
- 7 Uvjeti čuvanja (čl. 9. t. g) – „Čuvati na suhom i hladnom mjestu”
- 8 Naziv i adresa subjekta u poslovanju s hranom (čl. 9. t. h) – uvoznik
- 9 Zemlja podrijetla (čl. 9. t. i, čl. 26.) – uključujući podrijetlo glavnog sastojka
- 10 Nutritivna deklaracija (čl. 9. t. l, čl. 30.–34.) – na 100 g, + referentni unos (PU)
- 11 Nije obvezno po Uredbi: oznake ambalaže/recikliranja (21 PAP)

Slika 2.1. Primjer deklaracije s označenim obveznim elementima

2.3. Višejezične ambalaže i izazovi zajedničkog tržišta

Članak 15. Uredbe (EU) br. 1169/2011 propisuje da se obvezne informacije o hrani navode na jeziku lako razumljivom potrošačima država članica u kojima se hrana stavlja na tržište, pri čemu države članice mogu propisati uporabu jednog ili više svojih službenih jezika, a višejezično navođenje izričito je dopušteno [1]. Hrvatska je tu mogućnost iskoristila Zakonom o informiranju potrošača o hrani, prema kojemu informacije o hrani na tržištu Republike Hrvatske moraju biti na hrvatskom jeziku i latiničnom pismu, uz dopušteno istodobno navođenje na drugim jezicima [19].

Ekonomska logika zajedničkog tržišta i malih nacionalnih tržišta navodi proizvođače da jednu te istu ambalažu opremaju deklaracijama za više država odjednom. Ta je praksa posebno izražena u jugoistočnoj Europi, gdje se na istoj etiketi rutinski pojavljuju blokovi na hrvatskom, slovenskom, srpskom, bosanskom, crnogorskom, makedonskom i albanskom jeziku, često uz kombinaciju latinice i ćirilice. Slične zajedničke etikete poznaju i nordijske zemlje (finski, švedski, norveški, danski) te zemlje Beneluksa.

Višejezične deklaracije imaju izravne posljedice za potrošača i za automatsku obradu. Prvo, gomilanje jezika na ograničenoj površini ambalaže tjera proizvođače na najmanju dopuštenu veličinu slova, čime deklaracija postaje teško čitljiva i ljudskom oku i OCR sustavu. Drugo, informacije među jezičnim blokovima nisu uvijek istovjetne: prijevodi popisa sastojaka i alergena znaju biti nepotpuni ili pogrešni, a naknadno lijepljene naljepnice s lokalnim prijevodom često prekrivaju izvorne podatke. Treće, strojna obrada takvih deklaracija zahtijeva ne samo prepoznavanje teksta nego i identifikaciju jezika pojedinih blokova, njihovu segmentaciju i međusobno povezivanje istovjetnih informacija, što su problemi kojima se bavi 4. poglavlje. Konačno, višejezičnost otvara i regulatorno pitanje: je li na konkretnom tržištu prisutan blok na propisanom jeziku sa svim obveznim podacima, što je upravo vrsta provjere koju predložena nadogradnja aplikacije ScanEat može automatizirati.

3. ZAKONODAVNI OKVIR I USKLAĐENOST PO DRŽAVAMA

3.1. Regulativa Europske unije

3.1.1. Označavanje hrane i aktualni razvoj propisa

Okosnicu označavanja hrane u EU čini Uredba (EU) br. 1169/2011, čije su materijalne odredbe obrađene u potpoglavlju 2.2. Uz nju djeluje niz povezanih propisa: Uredba (EZ) br. 1924/2006 o prehrambenim i zdravstvenim tvrdnjama, prema kojoj su na deklaraciji dopuštene samo propisane prehrambene tvrdnje i odobrene zdravstvene tvrdnje [18]; Direktiva 2011/91/EU o oznaci serije (lota) [17]; metrološke direktive 76/211/EEZ (sustav prosječnog punjenja pretpakovina i znak “e”) [11] i 2007/45/EZ, koja je ukinula obvezne nizove nazivnih količina za većinu hrane, zadržavši ih za vino i jaka alkoholna pića, čime je pravno omogućena i pojava smanjivanja pakiranja uz zadržanu cijenu, takozvane “shrinkflacije” [20].

Najavljena sveobuhvatna revizija Uredbe 1169/2011, koja je u sklopu strategije “Od polja do stola” trebala donijeti usklađeno obvezno nutritivno označavanje na prednjoj strani ambalaže, nutritivne profile, prošireno označavanje podrijetla i reviziju označavanja rokova trajanja, do sredine 2026. godine nije rezultirala zakonodavnim prijedlogom. Stavovi država članica ostali su polarizirani, osobito oko sheme Nutri-Score [21, 22]. Nutri-Score tako ostaje dobrovoljan u sedam europskih zemalja, uz različite inačice algoritma po državama. Komisija procjenjuje da je do 10% od 88 milijuna tona godišnje bačene hrane u EU povezano s pogrešnim tumačenjem oznaka datuma (“najbolje upotrijebiti do” naspram “upotrijebiti do”) [23]. U međuvremenu su donesene sektorske promjene: za vina od prosinca 2023. vrijedi obveza nutritivne deklaracije i popisa sastojaka uz mogućnost elektroničke e-etikete putem QR koda [24, 25], a od 14. lipnja 2026. mješavine meda moraju navesti sve zemlje podrijetla s postotnim udjelima [26]. E-etiketa za vino prva je masovna primjena QR koda kao nositelja obveznih informacija o hrani u EU i važan presedan za digitalne sustave poput ScanEata.

3.1.2. Ambalaža i ambalažni otpad

Ambalažu kao proizvod i otpad tri je desetljeća uređivala Direktiva 94/62/EZ s bitnim zahtjevima (minimizacija, ograničenje teških metala, obnovljivost) i ciljevima recikliranja koji su izmjenama iz 2018. podignuti na 65% do kraja 2025. i 70% do kraja 2030. godine [6]. Direktivu od 12. kolovoza 2026. zamjenjuje izravno primjenjiva Uredba (EU) 2025/40 o ambalaži i ambalažnom otpadu (PPWR), koja je stupila na snagu 11. veljače 2025. [2]. Njezine su ključne odredbe: sva

ambalaža mora biti reciklabilna, s razredima uspješnosti A, B i C (od 2030. dopušten najmanje razred C, od 2038. samo A ili B); minimalni reciklirani udio u plastičnoj ambalaži od 2030. (10–35% ovisno o vrsti, do 50–65% 2040.); ograničenje omjera praznog prostora na 50% za skupnu, transportnu i e-trgovinsku ambalažu; zabrane pojedinih formata od 2030. (npr. jednokratna plastična ambalaža za svježe voće i povrće lakše od 1,5 kg, hotelske minijature); granice za PFAS spojeve u ambalaži u dodiru s hranom već od kolovoza 2026.; ciljevi ponovne uporabe; te obvezni sustavi povratne naknade kojima se do 2029. mora odvojeno skupljati 90% jednokratnih plastičnih boca i metalnih limenki za napitke [2, 27].

Za temu ovog rada posebno je važno harmonizirano označavanje: od 12. kolovoza 2028. (odnosno 24 mjeseca nakon donošenja provedbenih akata) sva ambalaža mora nositi usklađenu piktogramsku oznaku sastava materijala radi razvrstavanja otpada, a višekratna ambalaža od veljače 2029. i QR kod za praćenje obrtaja; nakon toga države članice više ne smiju zahtijevati vlastite nacionalne oznake [2]. Time se rješava dosadašnja fragmentacija tržišta, ali se i količina obveznih grafičkih elemenata na ambalaži dodatno povećava, što označavanje čini još gušćim, a automatsku strojnu interpretaciju ambalaže još korisnijom. Valja napomenuti da provedbeni akti o oznakama u trenutku pisanja rada kasne za zakonskim rokom [28].

3.1.3. Jezični zahtjevi i višejezične deklaracije u EU

Kako je uvedeno u potpoglavlju 2.3, članak 15. Uredbe (EU) br. 1169/2011 zahtijeva jezik lako razumljiv potrošačima države članice, dopušta državama propisivanje vlastitih službenih jezika i izričito dopušta višejezično navođenje [1]. Komisija ne objavljuje središnji popis država koje su tu mogućnost iskoristile, ali od 2021. održava alat FLIS (Food Labelling Information System) s pregledom obveznih oznaka po državama [29]. Iz primarnih nacionalnih propisa proizlazi sljedeća slika: Hrvatska traži hrvatski jezik i latinično pismo [19]; Francuska francuski (čl. R412-7 Code de la consommation, u tradiciji zakona Toubon) [30]; Njemačka njemački (uredba LMIDV) [31]; Belgija jezik ili jezike jezične regije u kojoj se proizvod prodaje, pa nacionalna distribucija praktično znači nizozemski, francuski i njemački [32]; Luksemburg prihvaća bilo koji od tri jezika (francuski, njemački ili luksemburški) [29]; Finska u dvojezičnim općinama traži finski i švedski [33]; Irska prihvaća engleski [34]; Španjolska traži najmanje kastiljski [35]; Italija talijanski, uz kazne za jezične povrede od 3.000 do 24.000 EUR [36].

U praksi višejezičnih etiketa zanimljiv je nordijski “Scandimix”: kombiniranje danskoga, norveškoga i švedskoga u jedan tekst dopušteno je jer se ti jezici razlikuju tek neznatno u pravopisu, dok Finska takvo miješanje ne prihvaća [37]. Upravo Scandimix ilustrira i rizik “lažnih prijatelja” među srodnim jezicima (npr. švedski “glass” znači sladoled, a norveški “sopp” gljive naspram švedskoga “soppa”-juha), zbog čega struka preporučuje da se alergeni i propisana terminologija nikada ne spajaju među jezicima [37] – nalaz izravno prenosiv na južnoslavenske jezike. Za naknadno lijepljene naljepnice s prijevodom vrijedi ograničenje iz čl. 13. Uredbe: obvezne

informacije ne smiju biti skrivene ili prekinute drugim materijalom, a Komisija je pojasnila da su višeslojne (peel-off) etikete dopuštene samo ako ključni podaci ostaju na gornjem sloju [1, 38].

Industrijske smjernice o čitljivosti konstatiraju temeljni sukob: količina obveznih informacija raste, a površina ambalaže ostaje ista, pa se višejezične deklaracije tiskaju na donjoj granici dopuštene veličine slova [39]. Kvantitativna ilustracija iz jurisdikcije s istovjetnim pragom od 1,2 mm: u studiji čitljivosti u Hong Kongu oko 90% od 55 ispitanih uzoraka imalo je najmanji font ispod propisanog minimuma, s najmanjim izmjerenim slovima od svega 0,5 mm [40]. Za EU takva sustavna studija veze broja jezika i veličine fonta ne postoji, što je jedna od istraživačkih praznina koje ovaj rad identificira.

3.2. Nacionalni propisi odabranih država

3.2.1. Hrvatska

U Hrvatskoj su za temu rada mjerodavna tri sloja propisa. Prvi je označavanje hrane: Zakon o informiranju potrošača o hrani (NN 56/13, 14/14, 56/16, 32/19) provodi Uredbu (EU) br. 1169/2011, propisuje hrvatski jezik i latinično pismo te prekršajne odredbe za nesukladne deklaracije [19]. Drugi je sloj ambalaža i otpad: Zakon o gospodarenju otpadom (NN 84/21) [41] i Pravilnik o ambalaži i otpadnoj ambalaži, plastičnim proizvodima za jednokratnu uporabu i ribolovnom alatu koji sadržava plastiku (NN 137/23), koji propisuje obvezni GTIN kod na primarnoj ambalaži pića i novu oznaku sustava povratne naknade, dok je identifikacija materijala prema Odluci 97/129/EZ dobrovoljna [42]. Treći je sloj sustav povratne naknade, operativan od 2006. godine: obuhvaća jednokratnu ambalažu pića od PET-a, stakla i metala obujma 0,20–3 L, s naknadom koja od 1. siječnja 2025. iznosi 0,10 EUR, a od 2027. proširuje se na višeslojnu kartonsku ambalažu, ostale plastike i ambalažu manju od 0,2 L [43, 44]. Sa stopom povrata ambalaže u sustavu naknade od oko 80%, ali ukupnom stopom recikliranja ambalažnog otpada od 52% (2022.), Hrvatska je prema izvješću Europske komisije među državama u riziku od promašaja cilja od 65% za 2025. godinu [45, 46].

3.2.2. Odabrane države članice EU

Prije harmonizacije koju donosi PPWR, označavanje ambalaže za razvrstavanje otpada razvijalo se nacionalno i međusobno nekompatibilno. Francuska od 2022. zahtijeva znak Triman s uputama za razvrstavanje (Info-tri) na gotovo svim proizvodima za kućanstva [47]; Komisija je zbog te obveze pokrenula postupak povrede prava EU i 2025. uputila Francusku Sudu Europske unije, ocjenjujući da nacionalna oznaka ometa slobodno kretanje robe [48]. Italija od 2023. obvezuje na alfanumeričke kodove materijala prema Odluci 97/129/EZ na svoj ambalaži te upute za odlaganje

na potrošačkoj ambalaži, uz izričito dopuštene digitalne kanale (QR kodovi) [49]. Njemačka nema obveznu potrošačku oznaku, ali zahtijeva registraciju svih proizvođača u registar LUCID i sudjelovanje u dualnim sustavima, uz najstariji veliki sustav povratne naknade za jednokratnu ambalažu (0,25 EUR) [50]. Španjolska od 2025. obvezuje na oznaku frakcije otpada za kućansku ambalažu i zabranjuje paušalne tvrdnje o ekološkoj prihvatljivosti, a od 2023. primjenjuje i porez od 0,45 EUR po kilogramu neregiciklirane plastike u jednokratnoj ambalaži [51]. Ova mozaičnost (četiri velika tržišta, četiri različita režima) ilustrira regulatorni teret za proizvođače koji posluju prekogranično te vrijednost strojno čitljive, po državama parametrizirane baze pravila kakvu predviđa modul sukladnosti u 5. poglavlju. Usporedno s oznakama ambalaže, po državama se razlikuju i jezični zahtjevi za deklaracije: pregled propisanih jezika za države obrađene u potpoglavlju 3.1.3 i u nastavku ovog potpoglavlja daje tablica 3.1.

Tablica 3.1. Jezični zahtjevi za deklaracije pretpakirane hrane u odabranim državama [1, 19, 30, 31, 32, 33, 52, 53, 54, 55, 56, 57, 58, 59]

Država/regija	Propisani jezik deklaracije	Napomena
Hrvatska	hrvatski, latinično pismo	dopušteno istodobno i na drugim jezicima
Francuska	francuski	čl. R412-7 Code de la consommation
Njemačka	njemački	uredba LMIDV
Belgija	jezici jezične regije (NL/FR/DE)	prema regiji prodaje
Luksemburg	francuski, njemački ili luksemburški	dovoljan jedan od tri
Finska	finski i švedski	u dvojezičnim općinama
Irska	engleski	irski dopušten dodatno
Španjolska	kastiljski	iznimke za regionalne proizvode
Italija	talijanski	kazne 3.000–24.000 EUR
Srbija	srpski	pismo nije propisano (ćirilica ili latinica)
BiH	jedan od službenih jezika	latinica ili ćirilica
Crna Gora	crnogorski	prema Zakonu o zaštiti potrošača
Sjeverna Makedonija	makedonski (ćirilica) i albanski	dvojezičnost od 2022.
Albanija	albanski	dopuštena prelijepljena etiketa uvoznika
SAD	engleski	uz strani jezik svi obvezni podaci i na tom jeziku
Kanada	engleski i francuski	obvezna dvojezičnost
Švicarska	najmanje jedan od DE/FR/IT	u praksi trojezične deklaracije
Japan	japanski	obvezno
Kina	pojednostavljeni kineski	GB 7718
Zaljevske države (GCC)	arapski	drugi jezik dopušten uz arapski
EAEU (Rusija i dr.)	ruski + nacionalni jezici	TR CU 022/2011

3.2.3. Regija zapadnog Balkana

Za regionalno tržište na kojem ScanEat primarno djeluje ključno je da sve države zapadnog Balkana imaju propise sadržajno preslikane s Uredbe (EU) br. 1169/2011, ali s fiksiranim nacionalnim jezikom i lokalnim specifičnostima. Srbija *Pravilnikom o deklarisanju, označavanju i reklamiranju hrane* (Sl. glasnik RS 19/2017 s izmjenama) traži srpski jezik, pri čemu pismo nije propisano pa su prihvatljive i ćirilica i latinica; višejezičnost je dopuštena, a nutritivna deklaracija obvezna je od 2018. [52, 60]. Bosna i Hercegovina *Pravilnikom o pružanju informacija potrošačima o hrani* (Sl. glasnik BiH 68/13) traži jedan od službenih jezika, latinicom ili ćirilicom, uz izričitu zabranu prekrivanja deklaracije naljepnicama [53]. Crna Gora je *Uredbom o informisanju potrošača o hrani* (Sl. list CG 22/18) prenijela Uredbu 1169/2011, a jezična obveza proizlazi iz Zakona o zaštiti potrošača; kazne dosežu 20.000 EUR [61]. Sjeverna Makedonija uz makedonski na ćirilici od studenoga 2022. Zakonom o zaštiti potrošača dodatno traži i jezik kojim govori najmanje 20% građana (albanski), pa regionalna etiketa za to tržište mora sadržavati dva dodatna jezična bloka [54, 62]. Albanija traži albanski jezik na izvornoj ili prelijepljenoj etiketi uvoznika te, kao specifičan administrativni teret bez pandana u EU, obveznu prijavu etikete svakog proizvoda nadležnoj agenciji [63].

Upravo ta kombinacija, sadržajno gotovo istovjetni zahtjevi, ali različiti obvezni jezici, dva pisma i lokalne iznimke, objašnjava raširenost zajedničkih višejezičnih etiketa u regiji i čini je idealnim poligonom za automatsku analizu višejezičnih deklaracija. Valja istaknuti da raširenost regionalne višejezične etikete u literaturi nije kvantificirana nijednom objavljenom studijom, što je izravna prilika za empirijski doprinos budućeg istraživanja.

3.2.4. Ostale države izvan Europske unije

Globalnu referencu označavanja pretpakirane hrane čini Codex Alimentarius (FAO/WHO) s općim standardom CXS 1-1985, revidiranim 2018. (oznake datuma) i 2024. (alergeni: sezam postaje globalno obavezan, soja se premješta na nacionalne liste, žitarice s glutenom navode se poimence) [64, 65]; standardi Codexa služe i kao mjerilo u trgovinskim sporovima u okviru WTO-a [66].

U Sjedinjenim Američkim Državama hranu označava okvir FDA-e: panel Nutrition Facts reformiran je 2016. (obvezni dodani šećeri, realistične veličine porcija) [67], alergeni se deklariraju prema zakonu FALCPA iz 2004. (osam alergena) dopunjenom zakonom FASTER iz 2021. koji je dodao sezam kao deveti [68]. Sve obvezne informacije moraju biti na engleskom; pojavi li se na etiketi bilo koja tvrdnja na stranom jeziku, svi obvezni podaci moraju se ponoviti i na tom jeziku [55]. Genetski modificirana hrana deklarira se prema standardu NBFDS iz 2018., koji među prvima dopušta objavu putem QR koda [69], a 2025. FDA je predložila i obveznu

interpretativnu oznaku “Nutrition Info” na prednjoj strani ambalaže [70]. Strukturno, američka se nutritivna deklaracija iskazuje po porciji uz postotke dnevnih vrijednosti, za razliku od europske na 100 g/mL – razlika koju svaki sustav automatske usporedbe proizvoda mora normalizirati.

Ujedinjeno Kraljevstvo nakon Brexita zadržalo je Uredbu 1169/2011 kao asimilirano domaće pravo (uz EU verziju u Sjevernoj Irskoj prema Windsorskom okviru) [71], ali je dodalo vlastite slojeve: od listopada 2021. tzv. Natasha's Law obvezuje na puni popis sastojaka s istaknutim alergenima i na hrani zapakiranoj za izravnu prodaju [72], od 2024. obvezna je adresa subjekta u UK, a za robu koja ide u Sjevernu Irsku koristi se oznaka “Not for EU” [73]. Dobrovoljna shema “semafora” na prednjoj strani ambalaže ostaje široko prihvaćena u maloprodaji.

Kanada zahtijeva potpunu dvojezičnost (engleski i francuski) svih obveznih podataka, nutritivnu tablicu po porciji te, kao prva država skupine G7, od 1. siječnja 2026. obvezni crno-bijeli simbol povećala “High in / Élevé en” za proizvode s visokim udjelom zasićenih masti, šećera ili natrija [74, 75]. Švicarska autonomno preslikava pravo EU: deklaracija mora biti na najmanje jednom od tri službena jezika (njemački, francuski, talijanski), pa engleski nije dovoljan [56]. Australija i Novi Zeland dijele Food Standards Code s nutritivnim panelom po porciji i na 100 g te dobrovoljnom shemom Health Star Rating, za koju je početkom 2026. pokrenut postupak mandatiranja [76]. Japan zahtijeva deklaraciju na japanskom s vlastitom listom od osam obveznih alergena koja uključuje i heljdu [77]. Kina je 2025. objavila nove standarde GB 7718-2025 i GB 28050-2025 (obvezni od 2027.) koji proširuju nutritivnu deklaraciju, prvi put čine deklaraciju alergena obveznom i uređuju digitalne etikete putem QR kodova [57]. Zaljevske države standardom GSO 9:2022 zahtijevaju arapski jezik [58], a Euroazijska ekonomska unija tehničkim propisom TR CU 022/2011 ruski i nacionalne jezike, uz ćirilčno pismo [59]. Konačno, Čile je 2016. uveo obvezne crne osmerokutne oznake upozorenja “ALTO EN” i time pokrenuo val obveznog upozoravajućeg označavanja u Latinskoj Americi [78].

Za automatsku analizu deklaracija iz ove usporedbe (tablica 3.2) proizlaze četiri sistemske razlike koje sustav mora modelirati po jurisdikcijama: format nutritivne deklaracije (na 100 g naspram po porciji; sol naspram natrija; kJ i kcal naspram samo kcal), različiti popisi alergena (14 u EU, 9 u SAD-u, 11 u Kanadi, 8 u Japanu), jezici i pisma (latinica, ćirilica, arapsko pismo, kineski znakovi) te međusobno neprevodive sheme označavanja na prednjoj strani ambalaže (Nutri-Score, semafor, Health Star Rating, kanadski i čileanski simboli upozorenja).

Tablica 3.2. Usporedba režima označavanja pretpakirane hrane: EU i odabrane države izvan EU-a [1, 57, 64, 67, 68, 74]

Obilježje	EU	SAD	Kanada	Japan	Kina
Nutritivna deklaracija	na 100 g/mL, kJ i kcal	po porciji, kcal, %DV	po porciji, %DV	po porciji ili 100 g	na 100 g, NRV
Sol/natrij	sol	natrij	natrij	ekvivalent soli	natrij
Broj obveznih alergena	14	9	11	8	8 (od 2027.)
Jezik	jezik države članice	engleski	engleski i francuski	japanski	pojednostavljeni kineski
FOP shema	dobrovoljno (Nutri-Score i dr.)	predloženo (Nutrition Info)	obvezno od 2026.	dobrovoljno	upozorenja od 2027.

3.3. Problemi i pridržavanje propisa u pojedinim državama

3.3.1. Neoznačeni alergeni kao vodeći problem

Najozbiljniji dokumentirani problem deklaracija jesu neoznačeni alergeni. U europskom Sustavu brzog uzbunjivanja za hranu i hranu za životinje (RASFF) 2023. godine zabilježeno je ukupno 4.695 obavijesti, od čega 217 u kategoriji alergena, s mlijekom kao najčešće navedenim alergenom; u mreži administrativne pomoći i suradnje “pogrešno označavanje ili tvrdnje” bilo je druga najčešća kategorija nesukladnosti s 32% svih obavijesti [79]. Godine 2024. ukupan broj obavijesti u mreži porastao je na 9.460, od čega 5.250 u RASFF-u [80]. Britanska studija opoziva prehrambenih proizvoda u razdoblju od 2016. do 2021. pokazala je da je čak 597 od 1.036 opoziva (57,6%) bilo uzrokovano alergenima, najčešće mlijekom, a glavni je uzrok bilo jednostavno izostavljanje alergena s popisa sastojaka (oko 40% slučajeva) [81]. Irska agencija za sigurnost hrane u 2024. izdala je 126 javnih upozorenja, od čega 58 zbog alergena, što je porast od 77% u odnosu na prethodnu godinu [82].

Poseban rizik predstavljaju uvozni proizvodi s neprevedenim ili nepotpunim deklaracijama: studija 768 uvoznih proizvoda iz 12 azijskih zemalja kupljenih online u Ujedinjenom Kraljevstvu našla je preventivne oznake alergena na samo 22,5% proizvoda (s čak 24 različite formulacije), 36% nedosljednosti između etikete i online opisa, a laboratorijska analiza poduzorka pokazala je nenamjernu prisutnost alergena u 31,2% proizvoda [83]. U britanskom nadzoru maloprodaje 2025./26. od 45 uzoraka pomodne “Dubai” čokolade samo je jedan prošao sve provjere, njih 18 imalo je nedeklarirani kikiriki ili sezam, a 42 druge pogreške označavanja, uključujući nedostatak alergena na engleskom jeziku [84]. Ovi nalazi izravno motiviraju funk-

cionalnost provjere konzistentnosti višejezičnih blokova u predloženoj nadogradnji ScanEata: upravo prijevodni slojevi deklaracije statistički su najslabija karika.

3.3.2. Inspekcijski nalazi i sankcije

U Hrvatskoj je sanitarna inspekcija Državnog inspektorata u 2023. godini u području hrane provela 7.392 nadzora i utvrdila nesukladnosti u 33% nadziranih objekata, uz 1.558 upravnih rješenja; u kontroli voća i povrća pri uvozu od 84 utvrđene nepravilnosti njih 71 odnosilo se na nedostatke u deklaraciji, a 9 na potpuni izostanak deklaracije [85]. Prekršajne odredbe Zakona o informiranju potrošača o hrani stupnjuju kazne prema težini povrede: za lakše povrede (npr. pogrešan redoslijed sastojaka) pravnoj osobi prijeti kazna 10.000–30.000 kuna (oko 1.300–4.000 EUR), za nedostajuće obvezne oznake ili nepravilno isticanje alergena 30.000–70.000 kuna, a za najteže povrede kao što su nekorištenje hrvatskoga jezika i latiničnog pisma ili potpuno nenavođenje alergena, 70.000–100.000 kuna (oko 9.300–13.300 EUR) [19]. Usporedno, Italija za jezične povrede propisuje 3.000–24.000 EUR [36], Crna Gora do 20.000 EUR [61], Sjeverna Makedonija 500–10.000 EUR za trgovca [62], a Francuska uz prekršajne kazne za označavanje poznaje i kazneno djelo prijevare potrošača s kaznama do 750.000 EUR kada proizvod ugrožava zdravlje [30].

3.3.3. Točnost nutritivnih deklaracija i čitljivost

I formalno sukladna deklaracija može biti netočna. Komisijin vodič o tolerancijama dopušta odstupanje deklariranih nutritivnih vrijednosti od analitički utvrđenih od $\pm 20\%$ za većinu makronutrijenata u srednjim rasponima [86]. Empirijske provjere pokazuju da se ni te velikodušne tolerancije ne poštuju uvijek: portugalska studija 209 gotovih jela našla je 27% proizvoda izvan tolerancije za zasićene masti, a samo polovica proizvoda s višim udjelom soli bila je unutar tolerancije [87]; kanadska agencija CFIA izvijestila je o 16,7% proizvoda izvan granice od $\pm 20\%$ [86]. Za aplikacije koje računaju nutritivne pokazatelje iz deklaracija to znači da je i savršena ekstrakcija podataka tek onoliko točna koliko i sama deklaracija. Sustav bi trebao ugraditi svijest o tolerancijama i mogućnost označavanja sumnjivih vrijednosti. Uz točnost, trajni je problem čitljivost: potrošačke organizacije dokumentiraju da točnost odgovora potrošača na pitanja o nutritivnim podacima varira između 19 i 51% ovisno o državi i formatu deklaracije [40].

3.3.4. E-trgovina i prijevare s označavanjem

Prodaja hrane na daljinu dodatni je sloj problema: prema čl. 14. Uredbe 1169/2011 sve obvezne informacije osim datuma trajnosti moraju biti dostupne prije zaključenja kupnje [88], no koordini-

rana kontrola Komisije nad online ponudom hrane u 25 država članica pronašla je 779 nesukladnih ponuda na oko 1.100 pregledanih web-mjesta [89]. Konačno, označavanje je i poprište namjernih prijevara: skandal s konjskim mesom 2013. (4,66% pozitivnih DNK testova u koordiniranom EU testiranju) paradigmatički je slučaj lažnog deklariranja sastava [90]; u akciji “From the Hives” čak 46% uzoraka uvoznog meda bilo je sumnjivo na patvorenje šećernim sirupima [91]; a mreža EU-a za prijevare s hranom u 2023. zabilježila je 758 sumnji, s mesom, medom i maslinovim uljem na vrhu liste i zavaravajućim označavanjem kao tipičnim mehanizmom [92]. Ti primjeri pokazuju da deklaracija nije samo informacijski nego i forenzički dokument te da alati koji deklaracije masovno digitaliziraju i uspoređuju (poput ScanEata) mogu doprinijeti i otkrivanju sustavnih nepravilnosti na tržištu.

4. TEHNOLOGIJE AUTOMATSKOG PREPOZNAVANJA TEKSTA NA AMBALAŽI

4.1. Optičko prepoznavanje znakova (OCR)

4.1.1. Razvoj pristupa: od klasičnog OCR-a do transformera

Optičko prepoznavanje znakova (engl. Optical Character Recognition, OCR) tehnologija je pretvorbe slike teksta u strojno čitljiv tekst. Klasični predstavnik je otvoreni sustav Tesseract, koji od inačice 4 koristi neuronski LSTM mehanizam za prepoznavanje redaka teksta i službeno podržava više od 100 jezika i 35 pisama [93]. Na čistim skeniranim dokumentima Tesseract postiže točnost znakova veću od 95%, no na fotografijama iz stvarnog svijeta (tzv. scene text) točnost znatno fluktuirala: male promjene osvjetljenja i kuta snimanja uzrokuju velike pogreške [94].

Suvremeni sustavi temeljeni na dubokom učenju problem u pravilu dijele u dva stupnja: detekciju teksta (pronalaženje regija sa slovima) i prepoznavanje teksta (pretvorbu regije u niz znakova). Među detektorima su se istaknuli EAST, koji na skupu ICDAR2015 postiže F-mjeru 0,782 uz 13,2 slike u sekundi, ali podržava samo četverokutne regije pa se ruši na zakrivljenom tekstu [95]; CRAFT, koji praćenjem afiniteta među znakovima dobro prati zakrivljene retke (F-mjera 0,869 na ICDAR2015, 0,836 na skupu Total-Text) [96]; te DBNet s diferencijabilnom binarizacijom, danas standardni detektor u više biblioteka (F-mjera 0,873 uz ResNet-50, odnosno 0,823 uz 48 slika u sekundi s lakšom okosnicom) [97]. Kod prepoznavanja, klasičan je model CRNN s CTC dekodiranjem sa svega 8,3 milijuna parametara [98]; pozornosni (attention) dekoderi donose oko 1,1% veću točnost uz osjetno veću latenciju [99], a model SVTR postiže visoku točnost jedinstvenim vizualnim modelom bez rekurencije. Njegova inačica SVTR-T sa 6 milijuna parametara i 4,5 ms po slici temelj je mobilnih sustava PP-OCR [100].

Treću generaciju čine transformerski modeli koji spajaju cijeli cjevovod: TrOCR kombinira vizualni transformerski enkoder i autoregresivni dekoder (62 do 558 milijuna parametara), s vrhunskim rezultatima na rukopisu i računima, ali je prevelik za mobilne uređaje bez agresivne kompresije [101]; Donut je predstavnik pristupa bez eksplicitnog OCR-a (engl. OCR-free) koji izravno iz slike generira strukturirani izlaz, čime eliminira propagaciju OCR pogrešaka, no kao generativni model može “halucinirati” sadržaj [102].

4.1.2. Deklaracija kao problem prepoznavanja teksta u sceni

Za ovaj je rad ključno da fotografija deklaracije na ambalaži nije uredan dokument, nego težak scene-text problem: površine su zakrivljene (boce, konzerve, vrećice), materijali reflektiraju bljeskalicu, tekst je gust, višejezičan i tiskan fontovima od 6 do 8 tipografskih točaka, što na fotografiji daje visinu znaka od svega 10 do 20 piksela, a to je ispod pretpostavki mnogih modela prepoznavanja [103]. Empirijske studije to potvrđuju: na skupu od 1.628 stvarnih fotografija ambalaže svi ispitani otvoreni OCR sustavi praktički padaju. Najbolja stopa pogreške znakova (CER) iznosila je čak 0,912 (Tesseract), dok su PaddleOCR, TrOCR i EasyOCR bili i lošiji [103]. Slično, višejezični benchmark HalalBench s 1.043 slike deklaracija na 14 jezika izvještava najbolju end-to-end F1 mjeru ekstrakcije sastojaka od svega 0,193 (docTR), uz ML Kit 0,180 i EasyOCR 0,167 [104]. Iz toga slijede dva zaključka: sirovi OCR bez predobrade i postprocesiranja nije dovoljan za pouzdanu analizu deklaracija, a namjenskih javnih skupova podataka za OCR na ambalaži gotovo nema, što je istraživačka praznina relevantna za buduće doktorsko istraživanje.

4.1.3. Usporedba otvorenih OCR biblioteka

U tablici 4.1 uspoređene su najvažnije otvorene OCR biblioteke prema kriterijima bitnima za mobilnu višejezičnu primjenu. Tesseract je najmanji CPU sustav s najširoom jezičnom pokrivenošću, ali s najslabijim rezultatima na fotografijama iz ruke [93, 94]. EasyOCR podržava više od 80 jezika, uključujući ćirilicu i arapsko pismo, no memorijski je zahtjevan i projekt stagnira [105]. PaddleOCR s generacijom PP-OCRv5 donosi oko 13 postotnih bodova bolju end-to-end točnost od prethodnika i pokriva čak 106 jezika modelima grupiranim po pismima (latinica, ćirilica, devanagari, arapsko pismo i dr.), a mobilne inačice detektora i recognizera zajedno zauzimaju oko 20 MB [106, 107, 108]. Biblioteka docTR nudi vrlo male modele (recognizer od 2,1 milijun parametara), ali su objavljeni modeli fokusirani na latinicu [109]; Surya postiže visoku točnost na 90+ jezika, no zahtijeva GPU i ima komercijalno ograničenu licencu [110]; MMOCR je vrijedan istraživački okvir koji se više aktivno ne održava [111].

Tablica 4.1. Usporedba otvorenih OCR biblioteka (stanje: srpanj 2026.) [93, 105, 106, 109, 110]

Biblioteka	Broj jezika	Podrška ćirilici	Veličina modela	Prikladnost za mobilne uređaje
Tesseract 5	100+	da	~10 MB + jezične datoteke	ograničena (CPU, spor na scene-textu)
EasyOCR	80+	da	~500 MB tipično	slaba (velik memorijski otisak)
PaddleOCR (PP-OCRv5)	106	da (~30 jezika)	~20 MB (mobilna inačica)	dobra (namjenski mobilni modeli)
docTR	pretežno latinica	ne (objavljeni modeli)	od 2,1 mil. parametara	dobra uz ONNX, samo latinica
Surya	90+	da	650 mil. parametara (VLM)	nije primjenjiv (zahtijeva GPU)

4.2. OCR na klijentskoj strani (on-device)

4.2.1. Ugrađena mobilna rješenja

Za izvođenje na samom mobilnom uređaju danas postoje dva dominantna besplatna rješenja. Google ML Kit Text Recognition v2 radi potpuno na uređaju (Android i iOS) i nudi pet modela pisama: latinicu, kineski, devanagari, japanski i korejski; hrvatski je službeno na popisu podržanih jezika, ali ćirilica nije podržana, što je konkretan tehnički jaz za višejezične deklaracije jugoistočne Europe (srpska ćirilica, makedonski) [112]. Model za latinicu u nepovezanoj (unbundled) inačici na Androidu dodaje aplikaciji svega oko 260 kB (model se isporučuje kroz Google Play Services), dok je na iOS-u riječ o oko 38 MB po pismu [112].

Appleov Vision okvir (VNRecognizeTextRequest) ugrađen je u operacijski sustav pa aplikaciji ne dodaje ništa na veličinu; nudi brzi način rada za obradu u stvarnom vremenu i precizni neuronski način, uz automatsku detekciju jezika. U iOS-u 18 podržava 18 jezika, uključujući ruski i ukrajinski (dakle ćirilicu), ali hrvatski nije među njima. Latinični hrvatski tekst prepoznaje se latiničnim modelom, no bez jezične korekcije [113]. Od iOS-a 26 dostupan je i zahtjev RecognizeDocumentsRequest koji vraća strukturirani prikaz dokumenta s tablicama, listama i detektorima podataka – izravno primjenjivo na nutritivne tablice [114]. Ova asimetrija platformi (ML Kit podržava hrvatski bez ćirilice; Vision podržava ćirilicu bez hrvatskoga) snažan

je argument za hibridnu strategiju OCR-a po platformama, ili za vlastiti model preko prenosivih izvedbenih okruženja.

Za izvođenje vlastitih modela na uređaju dostupna su izvedbena okruženja LiteRT (nekadašnji TensorFlow Lite) s ubrzanjem na neuronskim procesorima [115], Core ML (ugrađen u Appleove sustave), ONNX Runtime Mobile s osnovnim paketom od svega nekoliko megabajta [116] te ExecuTorch iz PyTorch ekosustava [117]. Kvantizacija na 8-bitne cijele brojeve tipično smanjuje model četverostruko uz dvostruko do trostruko ubrzanje [115]. U tom je kontekstu posebno zanimljiv PP-OCRv5 mobile: izvorni PP-OCR sustav nakon kvantizacije zauzima svega 3,5 MB i na mobilnom procesoru postiže latencije reda stotinjak milisekundi za detekciju i ~11 ms po retku za prepoznavanje [118].

4.2.2. Usporedba obrade na uređaju i u oblaku

Alternativa klijentskoj obradi jesu komercijalni oblak-servisi: Google Cloud Vision (1,50 USD na 1.000 slika nakon besplatne kvote), AWS Textract (1,50 USD na 1.000 stranica) i Azure AI Vision Read (oko 1,00 USD na 1.000 transakcija) [119, 120]. Pri milijunima skenova mjesečno to su troškovi reda desetak tisuća USD mjesečno, dok je marginalni trošak obrade na uređaju nula. Važniji je, međutim, argument privatnosti: fotografije deklaracija rutinski hvataju usputne osobne podatke (ruke, lica, račune, interijere, EXIF geolokaciju), koji su osobni podaci u smislu Opće uredbe o zaštiti podataka; obrada na uređaju udžbenička je primjena načela smanjenja količine podataka (čl. 5. st. 1. (c) GDPR-a) i tehničke i integrirane zaštite podataka (čl. 25.), a potpuno uklanjanje i pitanje prijenosa podataka u treće zemlje [121]. Obrada na uređaju dodatno omogućuje rad u stvarnom vremenu na slici kamere i rad bez mrežne veze, što je relevantno u trgovinama sa slabim signalom. S druge strane, oblak-servisi u pravilu bolje obrađuju složene rasporede teksta, no recenzirana izravna mjerenja pokazuju da je na teškoj domeni ambalaže razlika među sustavima mala [104]. Sažetak usporedbe dan je u tablici 4.2.

Tablica 4.2. Usporedba OCR obrade na klijentskoj strani i u oblaku

Kriterij	Na uređaju (on-device)	U oblaku (cloud)
Trošak po skeniranju	~0	~0,001–0,0015 USD
Latencija	milisekunde, stvarno vrijeme	mrežni prijenos + obrada
Rad bez mreže	da	ne
Privatnost (GDPR)	podaci ne napuštaju uređaj	prijenos i obrada kod trećih strana
Točnost na složenim rasporedima	niža	u pravilu viša
Veličina aplikacije	+0,26–38 MB (ovisno o platformi)	bez dodatka

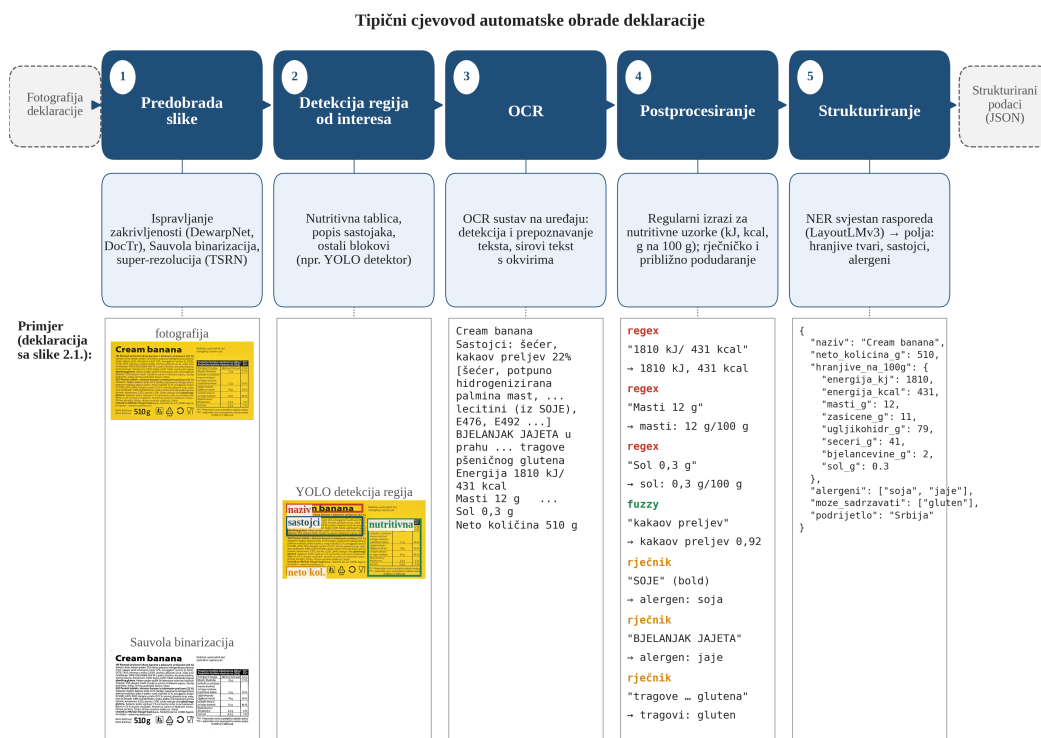
4.2.3. Multimodalni jezični modeli kao alternativa

Multimodalni veliki jezični modeli (engl. Vision-Language Models, VLM) mogu izravno iz fotografije ekstrahirati strukturirane podatke deklaracije. Na standardiziranom benchmarku OCRBench v2 većina velikih modela ipak ostaje ispod 50 od 100 bodova [122], a najjeftinije komercijalne inačice cjenovno su usporedive s namjenskim oblak OCR servisima. Ključni rizik su halucinacije: VLM-ovi na degradiranim slikama samouvjereno izmišljaju uvjerljiv tekst jer se pretjerano oslanjaju na jezične priore [123]; za nutritivne vrijednosti halucinirana znamenka opasnija je od klasične OCR pogreške jer je tečna i ničim označena. Istodobno, specijalizirani mali OCR modeli s nekoliko milijuna parametara nadmašuju milijardne VLM-ove na čistoj transkripciji [108]. Za sustave na uređaju u nastajanju su i mali VLM-ovi (Gemini Nano kroz ML Kit GenAI, Appleov Foundation Models okvir, otvoreni SmolVLM) [124, 125, 126], no puni OCR dokumenata za sada nije njihova deklarirana sposobnost. Racionalna arhitektura stoga je “mali namjenski OCR na uređaju + jezično postprocesiranje”, uz VLM najviše kao oblak-rezervu za teške slučajeve.

4.2.4. Cjevovod obrade deklaracije

Iz pregledane literature proizlazi tipičan cjevovod (engl. pipeline) automatske obrade deklaracije, prikazan konceptualno na slici 4.1: (1) predobrada slike: ispravljanje perspektive i zakrivljenosti (sustav DewarpNet smanjuje OCR pogrešku u prosjeku za 42% [127]; DocTr dodatno ispravlja i osvjetljenje [128]), lokalna adaptivna binarizacija (Sauvola nadmašuje globalnu Otsu metodu na neravnomjerno osvjetljenoj ambalaži [129]) te super-rezolucija sitnog teksta (TSRN na skupu TextZoom podiže točnost prepoznavanja za više od 13% [130]); (2) detekcija regija od

interesa: razlikovanje nutritivne tablice, popisa sastojaka i ostalih blokova, npr. YOLO detektorom kakav u produkciji koristi Open Food Facts [131]; (3) OCR odabranim sustavom na uređaju; (4) postprocesiranje: regularni izrazi za nutritivne uzorke (“kJ”, “kcal”, “g na 100 g”), rječničko i približno (fuzzy) podudaranje prema taksonomiji sastojaka, koje empirijski donosi i do 36% veću F1 mjeru [104, 132]; (5) strukturiranje: prepoznavanje imenovanih entiteta svjesno rasporeda (layout-aware), gdje model LayoutLMv3 fino ugođen na 3.000 anotiranih tablica u produkciji Open Food Factsa postiže F1 mjeru 0,96 [133]. Referentna vrijednost za cijeli zadatak je skup POIE s 3.000 fotografija nutritivnih tablica: detekcija teksta postiže F1 0,979, ali end-to-end ekstrakcija entiteta samo 0,795. Strukturiranje, a ne detekcija, usko je grlo [134].



Slika 4.1. Tipični cjevovod automatske obrade deklaracije: predobrada, detekcija regija, OCR, postprocesiranje i strukturiranje podataka

4.3. Automatsko prepoznavanje jezika i obrada višejezičnog teksta

4.3.1. Identifikacija jezika i pisma

Identifikacija jezika (engl. Language Identification, LID) preduvjet je ispravne obrade višejezične deklaracije. Standardni alati su fastText LID model sa 176 jezika (u komprimiranoj inačici manji od 1 MB, što ga čini prikladnim za mobilne uređaje) [135], Googleovi CLD2/CLD3 te novi Glo-tLID s pokrivenošću preko 1.600 jezika, koji posebno dobro pokriva niskoresursne jezike [136]. Ključni problem za deklaracije jest duljina teksta: popis sastojaka je kratak, telegrafski i pun

stručnih naziva, a usporedne evaluacije pokazuju da točnost većine detektora naglo pada na tekstovima kraćima od stotinjak znakova, pri čemu je fastText najbolji kompromis točnosti, brzine i pokrivenosti [137]. Dodatni je izazov razlikovanje blisko srodnih jezika, hrvatskoga, srpskoga i bosanskoga, na kojima ni najbolji modeli nisu pouzdani na razini kratkog popisa sastojaka [138]. Praktična posljedica: LID na deklaracijama treba kombinirati s determinističkim signalima, prije svega eksplicitnim jezičnim oznakama (“HR:”, “SRB:”, “MK:”) koje proizvođači sami tiskaju, te s identifikacijom pisma, koja je za razlikovanje latinice, ćirilice i grčkog pisma deterministička preko Unicode raspona. Za identifikaciju jezika na uređaju dostupan je i ML Kit Language Identification sa 100+ jezika [139].

4.3.2. Segmentacija višejezičnih blokova

Tipična višejezična deklaracija (“HR: Sastojci... / SRB: Sastojci... / MK: Састојки...”) nije pravi primjer miješanja jezika unutar rečenice (engl. code-switching), nego niz sekvencijalnih jednojezičnih blokova, pa je problem bliži segmentaciji dokumenta. Ipak, literatura o code-switchingu nudi upotrebljive tehnike identifikacije jezika na razini riječi i detekcije granica jezika [140, 141]. Za deklaracije se prirodno nameće kombinacija triju signala: detekcija jezičnih prefiksa regularnim izrazima (jezično neovisna), klizni LID prozor preko OCR redaka te prostorno grupiranje po blokovima na temelju okvira (bounding boxova) koje OCR vraća. Vrijedna su “sidra” i E-brojevi aditiva prema Uredbi (EZ) br. 1333/2008: oni su jezično neovisni identifikatori identični u svim jezičnim blokovima, pa služe za poravnavanje blokova i provjeru konzistentnosti prijevoda [142].

4.3.3. Povezivanje i normalizacija višejezičnih podataka

Nakon segmentacije, istovjetne informacije na različitim jezicima treba povezati i normalizirati. Open Food Facts održava višejezičnu taksonomiju sastojaka s oko 80.000 unosa, s hijerarhijom, prijevodima i sinonimima po jezicima, koja istodobno služi kao leksikon za podudaranje i kao implicitni prijevodni most među jezicima [143]; njihov produkcijski sustav Robotoff popis sastojaka izdvaja fino ugođenim modelom XLM-RoBERTa, jezik detektira fastTextom, a rezultat prihvaća tek ako je najmanje 60% sastojaka prepoznato u taksonomiji [144]. Alternativa eksplicitnom prijevodu (ML Kit On-Device Translation s 59 jezika [145], Appleov Translation okvir) jesu višejezični rečenični umeci (embeddings) poput modela LaBSE, koji pokriva 109 jezika i omogućuje međujezično podudaranje istovjetnih popisa sastojaka bez prijevoda [146], što je jeftinije i robusnije za provjeru konzistentnosti višejezičnih blokova. Za normalizaciju prema vanjskim bazama služe i ontologije hrane poput FoodOn, nastale iz tezaurusa LanguaL [147]. Kod nutritivnih tablica normalizacija uključuje i preračunavanje mjernih sustava: natrij u sol (sol = natrij $\times$ 2,5), kcal u kJ te vrijednosti po porciji u vrijednosti na 100 g, što je nužno za usporedbu

proizvoda iz različitih regulatornih režima (v. poglavlje 3).

4.3.4. Prepoznavanje imenovanih entiteta u domeni hrane

Za pretvorbu prepoznatog teksta u strukturirane podatke koriste se metode prepoznavanja imenovanih entiteta (engl. Named Entity Recognition, NER) prilagođene domeni hrane. Referentni skup TASTESet s više od 13.000 entiteta pokazuje da suvremeni modeli postižu prosječnu F1 mjeru oko 0,95 na receptima [148], a pregledni radovi sustavno obrađuju NER u računalnoj obradi hrane [149]. U domeni deklaracija, studija ekstrakcije halal/haram sastojaka kombinacijom OCR-a i NER-a izvještava NER F1 mjeru 0,967 [150], a Open Food Facts za ispravljanje OCR-om izvučenih popisa sastojaka koristi fino ugođeni otvoreni veliki jezični model koji je nadmašio komercijalne modele i smanjio broj neprepoznatih sastojaka u bazi za 11% [151]. Sveobuhvatan pregled cijelog područja detekcije i prepoznavanja etiketa namirnica daju Guimarães i suradnici [152].

4.4. Postojeća rješenja i srodne aplikacije

4.4.1. Open Food Facts i Yuka

Open Food Facts (OFF) najveća je otvorena kolaborativna baza prehrambenih proizvoda: više od 4 milijuna proizvoda iz 150 zemalja, podaci pod licencom Open Database License [4]. Korisnici skeniraju crtični kod i fotografiraju etikete, a strojno učenje (sustav Robotoff) iz fotografija izvlači sastojke, nutritivne vrijednosti i kategorije. OCR se, važno je istaknuti, izvodi na poslužitelju putem Google Cloud Visiona, dakle nije klijentski niti radi u stvarnom vremenu [144]. Za hrvatsko tržište baza je, međutim, vrlo oskudna: svega oko 7 tisuća proizvoda, manje od 0,2% globalne baze [153], što kvantificira problem pokrivenosti regionalnih tržišta.

Yuka je najpoznatija komercijalna aplikacija za ocjenjivanje prehrambenih proizvoda, s više od 80 milijuna korisnika u 14 zemalja (podaci s početka 2026.) [154]. Ocjena od 0 do 100 kombinira nutritivnu kvalitetu (60%, metodom temeljenom na Nutri-Scoreu), aditive (30%, prema vlastitoj klasifikaciji rizika) i ekološki uzgoj (10% bonusa) [155]. Učinak na tržište je dokumentiran: prema anketama, 78% francuskih proizvođača uzima Yuka ocjenu u obzir pri reformulaciji proizvoda [154]. Metodologija je, s druge strane, kritizirana kao arbitrarna (ponderi 60/30/10, bonus za *bio* bez čvrstih dokaza, jedan “rizični” aditiv ruši ocjenu neovisno o dozi) [156]. Za temu ovog rada ključno je da Yuka ne pokriva niti jednu državu zapadnog Balkana, upravo regiju s najizraženijim višejezičnim deklaracijama.

4.4.2. Ostale aplikacije i infrastruktura

Od ostalih rješenja valja spomenuti MyFitnessPal (praćenje prehrane, skener crtičnog koda od 2022. dostupan samo u plaćenju inačici [157]), njemačke Fddb i CodeCheck (potonji s više od 100 milijuna skenova [158]), američke aplikacije za alergene Fig i Spoonful koje provjeravaju etikete prema profilu korisnika [159, 160], te Too Good To Go s oznakom “često dobro i nakon” koja adresira pogrešno tumačenje datuma minimalne trajnosti [161]. Znanstvene studije učinka pokazuju da aplikacije za skeniranje hrane mogu nadmašiti oznake na prednjoj strani ambalaže u poticanju zdravijih izbora, osobito kod potrošača s niskim povjerenjem u institucije [162], premda eksperimentalni dokazi o promjeni stvarnog ponašanja ostaju nekonzistentni [163]; sustavni pregledi randomiziranih pokusa ipak nalaze konzistentna poboljšanja prehrambenih navika [164].

Infrastrukturno, područje se ubrzano mijenja migracijom s linearnih crtičnih kodova na dvodimenzionalne: inicijativa GS1 Sunrise 2027 predviđa da do kraja 2027. maloprodajne blagajne u 48 zemalja skeniraju 2D kodove, a standard GS1 Digital Link ugrađuje globalni identifikator proizvoda (GTIN) u web-adresu koja može voditi do deklaracije, e-etikete i podataka o sljedivosti [165, 166]. EU je već propisala prvu masovnu primjenu QR kodova kao nositelja obveznih informacija: od prosinca 2023. vina mogu nutritivnu deklaraciju i popis sastojaka dati elektronički putem e-etikete [24]. Za aplikaciju poput ScanEata to znači dvostruku strategiju: parsiranje strukturiranih 2D kodova za novu generaciju ambalaže, uz OCR koji ostaje nužan za milijune postojećih proizvoda i za neovisnu provjeru tvrdnji s fizičke etikete.

4.4.3. Analiza praznina (GAP analiza)

Iz pregleda proizlaze četiri praznine postojećih rješenja. Prvo, sva se popularna rješenja primarno oslanjaju na lanac *crtični kod – baza podataka*; kada proizvoda nema u bazi, slijedi ručni unos ili odgođena poslužiteljska obrada fotografija. Drugo, pokrivenost regionalnih tržišta je slaba (oko 7 tisuća hrvatskih proizvoda u OFF-u; Yuka bez ijedne balkanske države) [153, 154]. Treće, višejezičnost: nijedna analizirana aplikacija ne segmentira višejezične blokove niti provjerava njihovu međusobnu konzistentnost, a vodeći on-device OCR (ML Kit) ne podržava ćirilicu [112]. Četvrto, nijedna potrošačka aplikacija ne provjerava sukladnost deklaracije s propisima države (prisutnost obveznog jezika, obveznih podataka, isticanja alergena). Upravo te četiri praznine definiraju prostor doprinosa nadogradnje aplikacije ScanEat opisane u 5. poglavlju.

4.5. Metode za kontinuirano poboljšanje i analizu prikupljenih podataka

Dosadašnji pregled u ovom poglavlju obuhvatio je tehnologije kojima se deklaracija pretvara u strukturirane podatke. Smjerovi doktorskog istraživanja izloženi u potpoglavlju 5.5 počinjavaju i na

metodama kojima se prikupljeni podaci kontinuirano provjeravaju i iz kojih sustav uči, kao što su detekcija anomalija i provjera konzistentnosti strukturiranih podataka, aktivno učenje iz ljudskih korekcija, učenje na uređaju i federativno učenje te dovršavanje grafova znanja.

Polazište svih navedenih metoda je korpus podataka prikupljen dosadašnjim radom aplikacije. Tablica 4.3 uspoređuje javne skupove podataka za prepoznavanje i ekstrakciju podataka s deklaracijama, obrađene u potpoglavljima 4.1.2 i 4.2.4, s korpusom kojim autor raspolaže (v. potpoglavljja 5.1 i 5.4).

Tablica 4.3. Usporedba postojećih skupova podataka za analizu deklaracija s korpusom aplikacije ScanEat

Skup podataka	Veličina	Jezici/pisma	Tip anotacije	Javno dostupan
POIE [134]	3.000 fotografija nutritivnih tablica	nije navedeno u radu	entiteti nutritivnih vrijednosti (end-to-end ekstrakcija)	da
HalalBench [104]	1.043 slike deklaracija	14 jezika	ekstrakcija sastojaka	da
Južnoafrička studija [103]	1.628 fotografija ambalaže	nije navedeno u radu	transkripcija teksta (CER)	nije navedeno u radu
OFF anotirane nutritivne tablice [133]	3.000 anotiranih tablica	nije navedeno u radu	strukturirane nutritivne vrijednosti	nije navedeno u radu
Korpus ScanEat [5]	fotografije za preko 50.000 proizvoda	višejezične deklaracije tržišta Hrvatske i regije	ručno provjereni strukturirani podaci na razini polja (sastojci, alergeni, nutritivne vrijednosti)	ne (planirana objava anonimiziranog poduzorka, v. potpoglavljje 5.4)

Iz usporedbe je vidljivo da postojeći skupovi sadrže najviše tri tisuće primjera i da nijedan ne pokriva višejezične deklaracije jugoistočne Europe. Korpus aplikacije ScanEat po opsegu ih nadmašuje za više od reda veličine i jedini za svaki proizvod sadrži ručno provjerene strukturirane podatke na razini polja, pa objava njegova anonimiziranog poduzorka (v. potpoglavljje 5.4) izravno adresira prazninu identificiranu u potpoglavljju 4.1.2.

4.5.1. Detekcija anomalija i provjera konzistentnosti strukturiranih podataka

Zapis o proizvodu nastao ekstrakcijom iz deklaracije strukturirani je (tablični) podatak: skup polja s brojčanim vrijednostima (nutritivne vrijednosti) te kategoričkim i tekstualnim vrijednostima (sastojci, alergeni, kategorija). Detekcija anomalija (engl. anomaly detection) u takvim podacima obuhvaća dvije komplementarne skupine metoda. Prvu čine statistički i reprezentacijski pristupi koji anomalne zapise otkrivaju kao odstupanja od naučene strukture podataka: metoda dubokih izolacijskih šuma proširuje klasične izolacijske šume nasumičnim neuronskim reprezentacijama i time skalabilno bodovanje izolacijom prenosi na nelinearne i visokodimenzionalne podatke [167], dok samonadzirani pristupi temeljeni na kontroliranoj korupciji značajki uz samu detekciju daju i objašnjenje na razini pojedine varijable, dakle informaciju o tome koje je polje zapisa sumnjivo [168].

Drugu skupinu čine eksplicitna integritetska pravila. Formalizam zabranjenih ograničenja (engl. denial constraints) izražava pravila oblika “ne postoji zapis ili par zapisa koji istodobno zadovoljava zadane predikate”, čime se prirodno kodiraju i pravila među poljima istog zapisa i pravila među zapisima. Za automatsko otkrivanje takvih pravila iz podataka postoje učinkoviti algoritmi [169], provjera povreda nad velikim skupovima izvediva je u gotovo linearnom vremenu [170], a inkrementalni validatori povrede otkrivaju već pri unosu ili izmjeni zapisa, što odgovara trenutačnoj kontroli kvalitete pri svakom novom skeniranju [171].

Za ovu su domenu obje skupine izravno primjenjive. Provjere vjerodostojnosti iz potpoglavlja 5.2, energetska vrijednost izračunata iz makronutrijenata prema pretvorbenim faktorima iz Priloga XIV. Uredbe (EU) br. 1169/2011 [1] ili očekivani rasponi vrijednosti unutar kategorije proizvoda, upravo su zabranjena ograničenja, odnosno statistička odstupanja unutar kategorije, a prvi smjer doktorskog istraživanja (v. potpoglavlje 5.5) te provjere poopćuje: nekonzistentnosti među jezičnim blokovima iste ambalaže, odstupanja proizvoda od usporedivih proizvoda iste kategorije te promjene deklaracije u vremenu formuliraju se kao problemi detekcije anomalija i provjere konzistentnosti nad korpusom strukturiranih zapisa.

4.5.2. Aktivno učenje i učenje iz ljudskih korekcija

Aktivno učenje (engl. active learning) obuhvaća metode u kojima model sam odabire primjere čije bi označavanje najviše smanjilo njegovu nesigurnost, čime se trošak ručne anotacije usmjerava onamo gdje donosi najveću korist; suvremeni pregledi sistematiziraju strategije upita, među kojima je najraširenije uzorkovanje po nesigurnosti (engl. uncertainty sampling) [172]. Šire područje strojnog učenja s čovjekom u petlji (engl. human-in-the-loop machine learning) razlikuje aktivno učenje, interaktivno strojno učenje i podučavanje modela te opisuje upravo radne tijekove u kojima ljudske korekcije kontinuirano poboljšavaju model [173]. Obradeni su i praktični izazovi

produksijske primjene: kalibracija pouzdanosti pri odabiru nestabilnih primjera [174], dinamičko podešavanje pragova nesigurnosti pri pomaku koncepta u kontinuiranom toku podataka [175] te robusnost na pogrešne ljudske oznake [176].

Upravo taj metodski okvir formalizira treći, nosivi smjer doktorskog istraživanja (v. potpoglavlje 5.5): stupnjevani model pouzdanosti iz potpoglavlja 5.2 svaku korisničku i operatorsku korekciju pretvara u novi označeni primjer, a povijesni parovi sirovog OCR izlaza i operatorski ispravljenih podataka za preko 50.000 proizvoda čine polazni skup na kojem se strategije odabira primjera mogu vrednovati i prije produksijske primjene. Pomak koncepta pritom nije hipotetski: asortiman proizvoda, dizajn ambalaže i regulatorni zahtjevi mijenjaju se u vremenu, pa je dinamičko upravljanje petljom učenja bitan dio problema.

4.5.3. Učenje na uređaju i federativno učenje

U potpoglavlju 4.2 obrađeno je izvođenje (inferencija) modela na mobilnom uređaju. Novija literatura pokazuje da je na uređaju izvedivo i treniranje. Kvantizacijski svjesno skaliranje, rijetka ažuriranja parametara i namjenski laki motori za treniranje omogućuju dotreniranje modela unutar svega 256 kB memorije [177], a filtriranje gradijenata smanjuje memorijske i računske zahtjeve treniranja konvolucijskih mreža na uređaju [178]. Time postaje ostvariva prilagodba modela lokalnim podacima korisnika bez njihova prijenosa na poslužitelj.

Federativno učenje (engl. federated learning) taj pristup poopćuje: model se trenira nad podacima raspršenima po uređajima korisnika, a poslužitelju se šalju samo ažuriranja parametara, ne i sami podaci; suvremeni pregledi sistematiziraju otvorena pitanja generalizacije, robusnosti i pravednosti [179], a za vizualne je zadatke pokazano da na uspjeh pri heterogenim podacima bitno utječe i sam izbor arhitekture modela [180]. U kontekstu ovoga rada učenje na uređaju i federativno učenje metodološki su nastavak privatnosne argumentacije iz potpoglavlja 4.2.2 [121]: drugi smjer doktorskog istraživanja (v. potpoglavlje 5.5) predviđa ih kao mogući mehanizam personaliziranog učenja individualnih osjetljivosti, kod kojega zdravstveno obojeni podaci o smetnjama korisnika ne bi napuštali uređaj. S obzirom na otvorena pitanja heterogenosti podataka i troška komunikacije, taj se pravac tretira kao perspektiva daljnjeg rada, a ne kao težište istraživanja.

4.5.4. Dovršavanje grafa znanja i predviđanje veza

Graf model proizvod–sastojak–alergen–kategorija (v. potpoglavlje 5.1) primjer je grafa znanja (engl. knowledge graph). Metode ugrađivanja grafova znanja (engl. knowledge graph embedding) uče vektorske reprezentacije entiteta i relacija iz kojih se predviđaju nedostajuće veze (engl. link prediction) [181], a strojno učenje za pročišćavanje grafova znanja obuhvaća dopunu nedostajućih

informacija, uklanjanje redundancije i razrješavanje nekonzistentnosti [182]. Za ovu su domenu posebno zanimljivi pristupi koji uz strukturu grafa koriste i tekstualne opise entiteta predtreniranim jezičnim modelima [183], jer su čvorovi grafa sastojaka upravo višejezični nazivi, te korekcije pristranosti prema slabo povezanim entitetima [184], jer su regionalni sastojci i njihovi sinonimi u grafu tipično rijetki. Time se metodski utemeljuje četvrti smjer doktorskog istraživanja (v. potpoglavlje 5.5): automatsko predlaganje veza sastojak–alergen i hijerarhijskih veza među sastojcima, uz validaciju čovjekom u petlji. Domenski resursi na koje se takav graf oslanja postoje, a to su ontologija hrane FoodOn [147] i višejezična taksonomija sastojaka Open Food Factsa [143, 144].

5. APLIKACIJA SCANEAT I PRIJEDLOG NADOGRADNJE

5.1. Postojeća arhitektura i funkcionalnosti

Aplikacija ScanEat razvijena je kao mobilna aplikacija za analizu pretpakiranih prehrambenih proizvoda, dostupna na tržištu Hrvatske i regije. Sastoji se od izvornih (nativnih) mobilnih aplikacija za Android (Kotlin) i iOS (Swift), poslužiteljskog REST API-ja izvedenog u PHP-u (radni okvir Symfony) na infrastrukturi AWS te web-aplikacije za administraciju (backoffice) koja koristi isti API. Podaci se pohranjuju u graf bazi podataka Neo4j: graf model prirodno opisuje višestruke veze proizvod–sastojak–alergen–kategorija, pa se detekcija alergena svodi na obilazak veza od sastojaka prema alergenima, a pronalaženje alternativnih proizvoda na pretragu unutar zajedničkih kategorija. U trenutku pisanja rada aplikacija broji više od 150.000 korisnika, oko 50.000 proizvoda u bazi i više od 2 milijuna skeniranja [5]. Autor je ovoga rada suizrađivač aplikacije ScanEat, a opisani korpus podataka i fotografija nastao je u sklopu njezina komercijalnog rada.

Osnovni tijek korištenja polazi od skeniranja crtičnog koda proizvoda. Ako je proizvod u bazi, aplikacija prikazuje potpuni sastav i nutritivne vrijednosti, a *osobni detektor* uspoređuje sastojke s profilom korisnika: podržano je više od 200 alergena i intolerancija te profili vezani uz kronične bolesti (dijabetes, hipertenzija, giht), trudnoću i dojenje te veganski odnosno vegetarijanski način prehrane. Aplikacija dodatno prati službene objave o povlačenju proizvoda s tržišta i korisnicima šalje push obavijesti, čime se izravno adresira problem sporog dosega opoziva do potrošača (v. potpoglavlje 3.3). Aplikacija se može koristiti anonimno, a napredne funkcionalnosti (tražilica baze, obiteljski profili, prijedlozi alternativnih proizvoda) dio su premium pretplate.

Ključan je za ovaj rad postupak dodavanja proizvoda kojih nema u bazi. Korisniku se u tom slučaju nudi da ambalažu čitko fotografira sa svih strana; fotografije se prenose na poslužitelj i ulaze u red obrade u backofficeu. Ondje se nad njima izvodi OCR na poslužiteljskoj strani, čiji rezultat operater ručno provjerava i dopunjuje (pristup čovjek-u-petlji): ispravno prepoznata polja se potvrđuju, a pogrešno prepoznata ili neprepoznata unose ručno. Po završetku obrade korisnik dobiva obavijest da je proizvod dostupan u bazi. Nusproizvod je tog radnog tijeka, važan za daljnje istraživanje, korpus korisničkih fotografija ambalaže s ručno provjerenim strukturiranim podacima za sve proizvode u bazi (v. potpoglavlje 5.4). Backoffice sučelje za validaciju OCR rezultata prikazano je na slici 5.1.

Slika 5.1. Backoffice sučelje za validaciju OCR rezultata aplikacije ScanEat

Taj poluautomatski postupak, iako osigurava visoku kvalitetu podataka, ima tri temeljna nedostatka koja motiviraju predloženu nadogradnju. Prvo, spor je: od skeniranja do dostupnosti proizvoda prolazi vrijeme naknadne obrade (tipično nekoliko dana), pa korisnik u trgovini ostaje bez trenutačnog odgovora upravo u trenutku odluke o kupnji. Drugo, ne skalira se: propusnost sustava ograničena je ljudskim radom u backofficeu, a trošak obrade raste linearno s brojem novih proizvoda, što je posebno izraženo na regionalnom tržištu s velikim brojem proizvoda izvan globalnih baza (v. potpoglavlje 4.4). Treće, fotografije ambalaže, koje rutinski hvataju i usputne osobne podatke, prenose se i obrađuju na poslužitelju, što je s gledišta zaštite podataka nepovoljnije od obrade na uređaju (v. potpoglavlje 4.2). Nadogradnja opisana u nastavku upravo te tri točke adresira premještanjem OCR-a na klijentsku stranu, uz zadržavanje provjerenog načela čovjeka-u-petlji, ali izvedenog tako da korisnika ne vraća čitanju fizičke etikete.

5.2. Automatsko OCR skeniranje na klijentskoj strani

Predložena nadogradnja zamjenjuje poslužiteljski OCR s odgođenom ručnom validacijom automatskim cjevovodom obrade na klijentskoj strani, utemeljenim na nalazima iz 4. poglavlja, uz temeljnu arhitektonsku odluku da se OCR izvodi izravno na mobilnom uređaju korisnika. Razlozi su, kako je obrazloženo u potpoglavlju 4.2, četverostruki: privatnost (fotografije ne napuštaju uređaj, čime se sustavno primjenjuju načela smanjenja količine podataka i integrirane zaštite podataka iz GDPR-a [121]), trošak (marginalni trošak skeniranja je nula, bez poslužiteljskih troškova OCR servisa), latencija (obrada u stvarnom vremenu na slici kamere, s trenutačnom povratnom informacijom korisniku) i dostupnost (rad bez mrežne veze u trgovini).

Predloženi cjevovod obrade prikazan je na slici 5.2 i sastoji se od pet koraka:

(1) *Kontinuirano inkrementalno skeniranje*: umjesto snimanja pojedinačnih fotografija,

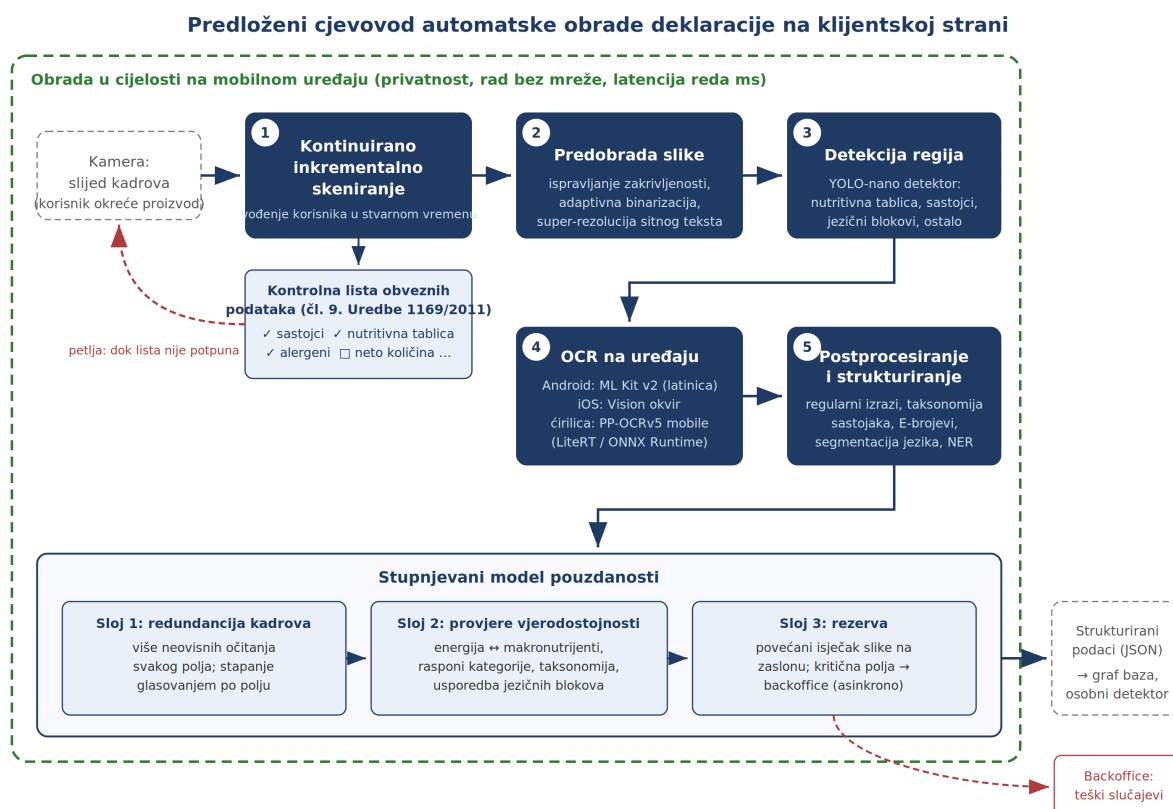
korisnik uključi kameru i okreće proizvod u ruci, a sustav u stvarnom vremenu obrađuje slijed kadrova i inkrementalno prikuplja podatke s deklaracije. Za svaku regiju od interesa (nutritivna tablica, popis sastojaka, neto količina, rok trajanja) zadržava se rezultat iz kadra s najvećom pouzdanošću, a rezultati iz više kadrova stapaju se glasovanjem na razini polja, čime se ublažava temeljni problem sitnog teksta: umjesto jedne fotografije na kojoj sve mora biti čitko, sustav bira najbolji kadar za svaku regiju zasebno. Sučelje pritom prikazuje kontrolnu listu obveznih podataka izvedenu iz čl. 9. Uredbe (EU) br. 1169/2011 [1] i vodi korisnika (“okrenite proizvod”, “približite kameru popisu sastojaka”) sve dok sva polja nisu prikupljena, nakon čega se skeniranje automatski dovršava. Ovakav pristup pretvara snimanje iz diskretnog čina u interaktivnu *petlju kamera–OCR–korisnik* i moguć je isključivo uz OCR na klijentskoj strani, jer zahtijeva latencije reda milisekundi po kadru (v. potpoglavlje 4.2).

(2) Predobrada: ispravljanje perspektive i zakrivljenosti površine, lokalna adaptivna binarizacija te po potrebi super-rezolucija sitnog teksta [127, 129, 130].

(3) Detekcija regija od interesa: lagani detektor objekata (npr. YOLO klase nano, po uzoru na produkcijski model Open Food Factsa [131]) razlikuje nutritivnu tablicu, popis sastojaka i ostale blokove.

(4) OCR na uređaju: na Androidu ML Kit Text Recognition v2 (podržava hrvatski, model dodaje ~260 kB [112]), na iOS-u Vision okvir (ugrađen u sustav [113]); za ćirilične blokove, koje ML Kit ne pokriva, kao dopunski sustav predlaže se mobilna inačica PP-OCRv5 (oko 14 MB za ćiriličnu skupinu jezika) izvedena kroz LiteRT ili ONNX Runtime Mobile [107, 116].

(5) Postprocesiranje i strukturiranje: regularni izrazi za nutritivne uzorke, približno podudaranje prema višejezičnoj taksonomiji sastojaka Open Food Factsa [143] i E-brojevima aditiva [142], te lagani NER model za granice entiteta; polja s niskom pouzdanošću označavaju se korisniku na potvrdu, čime se zadržava čovjek-u-petlji za kritične podatke poput alergena.



Slika 5.2. Predloženi cjevovod automatske obrade deklaracije u aplikaciji ScanEat

Ključna projektna odluka jest stupnjevani model pouzdanosti umjesto ručne provjere. Budući da je smisao aplikacije upravo osloboditi korisnika čitanja sitno tiskane deklaracije, provjera točnosti ne smije se svesti na to da korisnik uspoređuje prepoznati tekst s ambalažom. Umjesto toga, točnost se osigurava u tri sloja koja ne zahtijevaju čitanje etikete. Prvi je sloj redundancija samog skeniranja: kontinuirana obrada slijeda kadrova daje više neovisnih očitavanja svakog polja, koja se stapaju glasovanjem, pa se slučajne OCR pogreške međusobno poništavaju. Drugi je sloj automatska provjera vjerodostojnosti: energetska vrijednost provjerava se prema deklariranim makronutrijentima (pretvorbeni faktori iz Priloga XIV. Uredbe (EU) br. 1169/2011 [1]), nutritivne vrijednosti prema očekivanim rasponima kategorije proizvoda, sastojci podudaranjem s višejezičnom taksonomijom [143] i jezično neovisnim E-brojevima [142], a kod višejezičnih ambalaža i međusobnom usporedbom jezičnih blokova, koji su prirodna redundancija istih informacija. Treći je sloj rezerva za polja koja ni nakon toga nemaju dovoljnu pouzdanost: korisniku se takvo polje prikaže kao povećani isječak snimljene fotografije uz prepoznatu vrijednost, pa usporedbu obavlja na zaslonu, na povećanoj slici, a ne na fizičkoj etiketi, ili se, za kritična polja poput alergena, proizvod korisniku odmah prikaže s jasno označenim statusom privremene obrade, a sporna polja asinkrono proslijeđuju postojećem backoffice kanalu na operatorsku provjeru. Backoffice time iz obveznog koraka za svaki proizvod postaje rezerva za mali udio teških slučajeva, a svaka korekcija, korisnička na zaslonu ili operatorska u backofficeu, i dalje predstavlja novi označeni primjer za poboljšanje modela (v. potpoglavlje 5.5), što je bitno i s obzirom na dokumentirane

stope pogreške OCR sustava na ambalaži [103, 104].

5.3. Prepoznavanje višejezičnih ambalaža

Druga komponenta nadogradnje adresira središnji problem regionalnog tržišta: deklaracije s blokovima na više jezika. Predloženi modul kombinira signale opisane u potpoglavlju 4.3:

(1) identifikacija pisma po Unicode rasponima prepoznatog teksta (latinica/ćirilica/grčko pismo) kao deterministički prvi filtar.

(2) detekcija eksplicitnih jezičnih oznaka (“HR:”, “SLO:”, “SRB:”, “BIH:”, “MK:”, “AL:”) regularnim izrazima.

(3) klasifikacija jezika kratkih segmenata komprimiranim fastText modelom (manji od 1 MB, prikladan za uređaj [135]), uz svjesno ograničenje nepouzdanog razlikovanja hrvatskoga, srpskoga i bosanskoga [138]. U tim se slučajevima blok označava kao skupina srodnih jezika, a razlučuje se tek eksplicitnom oznakom.

(4) prostorno grupiranje redaka u blokove prema OCR okvirima.

Nakon segmentacije slijedi povezivanje blokova: istovjetni sastojci u različitim jezičnim blokovima povezuju se preko taksonomije sastojaka (koja sadrži prijevode i sinonime po jezicima [143]), preko jezično neovisnih sidara (E-brojevi, postotci, brojčane vrijednosti) te, eksperimentalno, višejezičnim rečeničnim umecima poput LaBSE za slučajeve izvan taksonomije [146]. Povezivanje omogućuje dvije funkcionalnosti kojih u postojećim aplikacijama nema: prikaz deklaracije korisniku na njegovu jeziku neovisno o tome koji je blok fotografiran, te provjeru konzistentnosti: upozorenje kada se popisi sastojaka ili alergeni među jezičnim blokovima razlikuju, što je dokumentirani stvarni problem višejezičnih etiketa (v. potpoglavlje 3.3).

Treća funkcionalnost izravno povezuje tehnički i regulatorni dio rada: provjera sukladnosti deklaracije s propisima države. Na temelju pravila kodificiranih iz 3. poglavlja (obvezni jezik po državi, popis obveznih podataka iz čl. 9. Uredbe (EU) br. 1169/2011 [1], zahtjev isticanja alergena, format nutritivne tablice) modul za odabranu državu provjerava: postoji li blok na propisanom jeziku; jesu li prisutni svi obvezni podaci; jesu li alergeni istaknuti; te jesu li nutritivne vrijednosti u očekivanom formatu. Takva provjera, kao pomoćni alat, može koristiti i potrošačima i malim proizvođačima i uvoznicima pri samokontroli, uz jasnu ogradu da ne predstavlja pravno mjerodavan nadzor.

5.4. Metodologija vrednovanja i očekivani rezultati

Vrednovanje predložene nadogradnje planira se na tri razine. Prva razina je točnost OCR-a i ekstrakcije. Za razliku od tipične situacije u literaturi, gdje javnih skupova fotografija ambalaže

s višejezičnim deklaracijama praktički nema (v. potpoglavlje 4.1), ovo istraživanje polazi od jedinstvenog postojećeg resursa: kroz dosadašnji rad aplikacije prikupljen je korpus korisničkih fotografija ambalaže za preko 50.000 proizvoda s tržišta Hrvatske i regije, snimljenih u realnim uvjetima (trgovina, kućanstvo; različiti uređaji, osvjetljenja i kutovi), pri čemu za svaki proizvod postoje i ručno provjereni strukturirani podaci (sastojci, alergeni, nutritivne vrijednosti) nastali operatorskom validacijom poslužiteljskog OCR-a u backofficeu (v. potpoglavlje 5.1). Taj korpus predstavlja skup podataka označen na razini polja kakav u literaturi za ovu domenu ne postoji ni približnog opsega. Za potrebe vrednovanja korpus je potrebno dopuniti dvjema razinama anotacije koje postojeći radni tijek nije bilježio: prostornim oznakama regija na fotografijama (okviri nutritivne tablice, popisa sastojaka i jezičnih blokova, potrebni za vrednovanje detekcije regija i segmentacije po jezicima) te oznakama jezika pojedinih blokova; planira se anotacija reprezentativnog poduzorka od najmanje nekoliko tisuća proizvoda, stratificiranog po kategorijama, materijalima ambalaže i broju jezika. Budući da su postojeće fotografije statične, za vrednovanje kontinuiranog skeniranja (v. potpoglavlje 5.2) dodatno će se prikupiti manji skup videosekvenci skeniranja. Objava anonimiziranog poduzorka korpusa kao javnog referentnog skupa, uz prethodno uklanjanje usputnih osobnih podataka s fotografija i uz pravnu analizu uvjeta objave, planirani je samostalan znanstveni doprinos. Mjere: stopa pogreške znakova i riječi (CER, WER) za OCR; preciznost, odziv i F1 mjera po polju za ekstrakciju (usporedivo s referentnim vrijednostima POIE [134] i produkcijskim sustavom Open Food Factsa [133]); točnost identifikacije jezika i segmentacije blokova po jezicima.

Druga razina je usporedba s ručnim unosom: mjerenje vremena unosa proizvoda i stope pogrešaka unosa za skupinu ispitanika, ručno naspram automatski s potvrdom, čime se kvantificira temeljna teza rada da automatsko skeniranje zamjenjuje ručni rad uz manju stopu pogreške. Treća razina je tehnička izvedivost na uređaju: latencija po kadru i po skeniranju, potrošnja memorije i energije na reprezentativnom rasponu uređaja (od ulaznih do vodećih modela), s ciljem obrade u stvarnom vremenu.

Očekivani rezultati uključuju: funkcionalan prototip klijentskog OCR cjevovoda integriran u ScanEat; prvi javno dostupan anotirani skup višejezičnih deklaracija jugoistočne Europe (kao samostalan znanstveni doprinos); kvantificiranu usporedbu on-device OCR sustava na toj domeni (ML Kit, Vision, PP-OCRv5), koja u literaturi nedostaje [104]; te prototip provjere sukladnosti deklaracija s propisima kao poveznicu računalne i regulatorne dimenzije problema. Identificirani rizici su niska točnost OCR-a na sitnom tekstu (ublažava se vodenim snimanjem i super-rezolucijom), nepouzdana razlikovanje srodnih jezika (ublažava se eksplicitnim oznakama i grupiranjem) i trošak dodatne prostorne anotacije poduzorka (ublažava se poluautomatskim predanotiranjem postojećim detektorima uz ručnu korekciju te činjenicom da terenske oznake polja za sve proizvode već postoje iz backoffice validacije).

5.5. Smjerovi doktorskog istraživanja

Opisana nadogradnja predstavlja inženjersku osnovu na kojoj se planiraju četiri izvorna znanstvena doprinosa doktorskog istraživanja.

Četiri se smjera pritom ne planiraju razrađivati ravnomjerno. Kao nosivi smjer doktorskog istraživanja izdvaja se treći, samopoboljšavajući sustav ekstrakcije utemeljen na aktivnom učenju nad postojećim korpusom korekcija, jer se najizravnije oslanja na postojeće resurse (v. potpoglavlje 5.4) i najbrže dovodi do mjerljivih rezultata. Prvi i četvrti smjer razrađuju se kao sekundarni, a drugi kao perspektiva daljnjeg rada; hijerarhija je naznačena uz svaki smjer u nastavku.

Prvi je smjer automatsko prepoznavanje pogrešaka i nesukladnosti u deklaracijama. Modul sukladnosti iz potpoglavlja 5.3 provjerava formalne zahtjeve (prisutnost obveznog jezika i obveznih podataka), no masovno skeniranje otvara mogućnost dubljih, sadržajnih provjera: otkrivanje nutritivnih vrijednosti koje su međusobno nekonzistentne (npr. energetska vrijednost koja ne odgovara deklariranim makronutrijentima prema pretvorbenim faktorima iz Priloga XIV. Uredbe (EU) br. 1169/2011 [1]) ili statistički odudaraju od usporedivih proizvoda iste kategorije; otkrivanje razlika u popisima sastojaka i alergena među jezičnim blokovima iste ambalaže, koje su dokumentirano najslabija karika višejezičnih etiketa (v. potpoglavlje 3.3); te otkrivanje promjena deklaracije u vremenu – kada proizvođač izmijeni sastav proizvoda bez promjene crtičnog koda i vizualnog identiteta ambalaže, što je u praksi najčešći uzrok zastarjelih podataka u kolaborativnim bazama. Poseban je slučaj potonjeg automatska detekcija smanjivanja neto količine uz nepromijenjeni crtični kod i cijenu proizvoda (“shrinkflacija”, v. potpoglavlje 3.1.1 [20]), koju vremenski niz skeniranih neto količina čini izravno mjerljivom. Ponovljena skeniranja istog proizvoda različitih korisnika pritom čine vremenski niz deklaracija, pa se baza pretvara u sustav kontinuiranog nadzora tržišta odozdo (engl. crowdsourced market surveillance), komplementaran inspekcijском nadzoru čiji su kapaciteti ograničeni (v. potpoglavlje 3.3.2). U hijerarhiji istraživanja ovaj se smjer tretira kao sekundaran: oslanja se na infrastrukturu i korpus nosivoga trećeg smjera, a metodski se temelji na postupcima detekcije anomalija i provjere konzistentnosti iz potpoglavlja 4.5.1.

Drugi je smjer personalizirano učenje individualnih osjetljivosti. Osobni detektor u postojećoj inačici radi deterministički: korisnik označi poznate alergene, a sustav ih pronalazi u sastavu. Doktorsko istraživanje proširit će taj model učenjem iz podataka o ponašanju: korisnik kroz povijest skeniranja implicitno bilježi što konzumira, a aplikacija mu omogućuje da uz proizvod zabilježi smetnju ili reakciju. Nad tako nastalim podacima, uz graf model proizvod-sastojak-alergen koji je za to prirodno prikladan, može se učiti model koji identificira sastojke statistički povezane sa smetnjama pojedinog korisnika i generira hipoteze o netoleranciji na sastojke koje korisnik sam ne bi povezao (npr. pojedini aditivi ili skriveni izvori alergena). Metodološki je riječ o problemu učenja iz slabo označenih i nepotpunih podataka s izraženim učincima zbunjujućih

varijabli (engl. confounding), a etički o pomoćnom alatu koji generira hipoteze za razgovor s liječnikom, ne o dijagnostici, što će biti eksplicitno ugrađeno u dizajn. S obzirom na osjetljivost zdravstvenih podataka, istraživanje će razmotriti i federativno učenje odnosno učenje na uređaju, čime se personalizacija metodološki povezuje s privatnosnom arhitekturom iz potpoglavlja 5.2. Zbog pravnih i etičkih preduvjeta obrazloženih u nastavku, ovaj se smjer tretira kao perspektiva daljnjeg rada, a ne kao težište doktorskog istraživanja.

Personalizirano učenje individualnih osjetljivosti dodatno pretpostavlja rješavanje triju pitanja koja prethode svakom eksperimentu. Prvo, podaci o smetnjama i reakcijama korisnika podaci su o zdravlju, dakle posebna kategorija osobnih podataka, pa njihova obrada zahtijeva izričitu pravnu osnovu (u pravilu izričitu privolu korisnika za tu svrhu) i procjenu učinka na zaštitu podataka. Drugo, istraživanje nad takvim podacima pretpostavlja prethodno odobrenje nadležnog etičkog povjerenstva. Treće, mora biti uređeno pitanje vlasništva i kontrole nad podacima: korisnik mora zadržati mogućnost uvida, izvoza i brisanja svojih zapisa te povlačenja privole bez gubitka osnovnih funkcionalnosti aplikacije. Postojeća ograda prema kojoj sustav generira hipoteze za razgovor s liječnikom, a ne dijagnozu, ostaje temeljno projektno načelo i dopunjuje se navedenim preduvjetima.

Treći je smjer samopoboljšavajući sustav ekstrakcije: svaka korisnička korekcija automatski predloženih podataka (stupnjevani model pouzdanosti iz potpoglavlja 5.2) i svaka operatorska korekcija u backofficeu predstavljaju novi označeni primjer. Polazište je pritom iznimno povoljno: iz dosadašnjeg rada backofficea već postoje povijesni parovi sirovog OCR izlaza i operatorski ispravljenih podataka za preko 50.000 proizvoda, nad kojima se model tipičnih OCR pogrešaka u ovoj domeni može učiti i prije prvog klijentskog skeniranja. Formalizacija te petlje kao aktivnog učenja, odabir primjera za ručnu provjeru prema nesigurnosti modela, periodičko dotreniranje la-ganih modela za detekciju regija i NER te mjerenje stope poboljšanja u produkciji, istraživačko je pitanje samo po sebi, a istodobno rješava i problem troška anotacije skupa podataka identificiran u potpoglavlju 5.4. Polazni skup povijesnih parova sirovog OCR izlaza i operatorskih korekcija već postoji, metodski okvir dan je u potpoglavlju 4.5.2, a rezultati su mjerljivi standardnim mjerama ekstrakcije iz potpoglavlja 5.4.

Četvrti je smjer dovršavanje grafa znanja o hrani. Graf model proizvod–sastojak–alergen–kategorija u bazi Neo4j (v. potpoglavlje 5.1) nije samo pohrana, nego i predmet učenja: metodama predviđanja veza (engl. link prediction) nad grafom mogu se automatski predlagati nedostajuće veze sastojak–alergen, primjerice za sinonime, regionalne nazive i skrivene izvore alergena koje ručno održavana taksonomija još ne pokriva, kao i nedostajuće hijerarhijske veze među sastojcima. Svako novo skenirano polje sastojaka pritom obogaćuje graf, a predložene veze validiraju se istim mehanizmom čovjeka-u-petlji, čime se ovaj smjer prirodno nadovezuje na treći. Time bi se višejezična taksonomija iz potpoglavlja 5.3 iz statičkog leksikona razvila u samodopunjavajući graf znanja specifičan za regionalno tržište. I ovaj se smjer tretira kao sekundaran: prirodno se nadovezuje na nosivi treći smjer, a metodski okvir predviđanja veza dan je u potpoglavlju 4.5.4.

5.6. Istraživački potencijal i očekivani doprinosi daljnjeg doktorskog istraživanja

Na temelju izloženih smjerova i metodologije vrednovanja, očekivani znanstveni doprinosi doktorskog istraživanja, svi u polju računarstva, jesu:

1. anotirani skup podataka višejezičnih deklaracija jugoistočne Europe: dopuna produkcijskog korpusa prostornim i jezičnim anotacijama te objava anonimiziranog poduzorka kao prvog javno dostupnog skupa te vrste (v. potpoglavlja 5.4 i 4.5)
2. kvantitativna usporedba OCR sustava na klijentskoj strani (ML Kit, Vision, PP-OCRv5) na domeni višejezičnih deklaracija, kakva u literaturi nedostaje (v. potpoglavlje 5.4)
3. cjevovod ekstrakcije koji uči iz korisničkih i operatorskih korekcija: formalizacija petlje čovjeka-u-petlji kao aktivnog učenja, s vrednovanjem strategija odabira primjera na povijesnim korekcijama i mjerenjem stope poboljšanja u produkciji (v. potpoglavlja 5.5 i 4.5.2)
4. metoda segmentacije, povezivanja i provjere konzistentnosti višejezičnih blokova deklaracije, uključujući razlikovanje blisko srodnih jezika i dvaju pisama te poravnavanje preko jezično neovisnih sidara (v. potpoglavlja 5.3 i 4.3)

Modul provjere sukladnosti deklaracija s propisima (v. potpoglavlje 5.3) u ovom se popisu ne navodi kao znanstveni doprinos: riječ je o kodifikaciji regulatornih pravila u strojno čitljiv oblik, dakle o primjeni sustava koja motivira rad i povezuje njegovu računalnu i regulatornu dimenziju, ali ne predstavlja izvorni znanstveni doprinos u polju računarstva.

6. ZAKLJUČAK

U radu je dan sustavan pregled područja ambalaže pretpakiranih prehrambenih proizvoda s dvaju komplementarnih gledišta: regulatornoga i računalnoga, te je na tom temelju predložena nadogradnja aplikacije ScanEat automatskim OCR skeniranjem deklaracija na klijentskoj strani i automatskim prepoznavanjem višejezičnih ambalaža.

Pregled regulatornog okvira pokazao je da je označavanje pretpakirane hrane u EU detaljno harmonizirano Uredbom (EU) br. 1169/2011, ali da tri područja ostaju izvor stvarnih problema: jezična pravila, koja članak 15. prepušta državama članicama, pa proizvođači pribjegavaju gusto tiskanim višejezičnim etiketama; označavanje ambalaže za razvrstavanje otpada, koje je do donošenja Uredbe (EU) 2025/40 fragmentirano po državama (Triman u Francuskoj, obvezni kodovi materijala u Italiji, različiti sustavi povratne naknade); te pridržavanje propisa, gdje sustavi brzog uzbunjivanja i inspekcijski nalazi pokazuju da su neoznačeni alergeni i nesukladne deklaracije trajno među najčešćim uzrocima opoziva proizvoda. Usporedba s državama izvan EU-a (Codex Alimentarius, SAD, Ujedinjeno Kraljevstvo, Kanada, Švicarska, Japan, Kina, zaljevske države, Euroazijska unija, Čile) pokazala je duboke strukturne razlike: u formatu nutritivne deklaracije, popisima alergena, jezičnim zahtjevima i shemama označavanja na prednjoj strani ambalaže, koje svaki sustav automatske analize deklaracija mora eksplicitno modelirati po jurisdikcijama.

Pregled tehnologija pokazao je da je deklaracija na ambalaži težak problem prepoznavanja teksta u sceni: na javnim benchmarcima fotografija ambalaže i najbolji otvoreni OCR sustavi bez predobrade i postprocesiranja postižu neupotrebljive rezultate, dok produkcijski sustavi (Open Food Facts) visoke točnosti postižu tek slojevitim cjevovodom s detekcijom regija, rječničkim postprocesiranjem i strukturiranjem svjesnim rasporeda. Na klijentskoj strani zrela besplatna rješenja postoje (ML Kit, Appleov Vision okvir), ali s asimetričnom podrškom jezicima i pismima koja upravo za višejezične deklaracije jugoistočne Europe (latinica + ćirilica) zahtijeva hibridnu strategiju. Identifikacija jezika na kratkim popisima sastojaka i razlikovanje blisko srodnih jezika ostaju otvoreni problemi koje treba rješavati kombinacijom statističkih i determinističkih signala.

Analiza postojećih aplikacija otkrila je četiri praznine: ovisnost o crtičnom kodu i pokrivenosti baze, slabu pokrivenost regionalnih tržišta, potpuno zanemarivanje višejezičnosti deklaracija te nepostojanje provjere sukladnosti deklaracije s propisima. Predložena nadogradnja aplikacije ScanEat adresira sve četiri: klijentski OCR cjevovod uklanja ovisnost o bazi i ručnom unosu uz očuvanje privatnosti, modul za višejezične ambalaže segmentira i povezuje jezične blokove te provjerava njihovu konzistentnost, a modul sukladnosti kodificira regulatorna pravila po državama u automatsku provjeru.

Daljnje istraživanje, kao priprema doktorskog rada, usmjerit će se na tri pravca: dopunu,

anonimizaciju i objavu anotiranog skupa višejezičnih deklaracija jugoistočne Europe izvedenog iz produkcijskog korpusa aplikacije, kakvoga u literaturi nema; kvantitativnu usporedbu klijentskih OCR sustava i vlastitih laganih modela na toj domeni; te formalizaciju provjere sukladnosti deklaracija kao strojno čitljivih pravila po jurisdikcijama, uključujući praćenje nadolazećih promjena (primjena Uredbe (EU) 2025/40 od kolovoza 2026., harmonizirane oznake ambalaže od 2028., migracija na 2D kodove do 2027.). Time bi se aplikacija ScanEat od alata za informiranje potrošača razvila u platformu koja povezuje potrošače, proizvođače i regulatorni okvir na višejezičnom regionalnom tržištu. Povrh toga, kao izvorni znanstveni doprinosi doktorskog rada planiraju se automatsko otkrivanje sadržajnih pogrešaka, nesukladnosti i vremenskih promjena deklaracija masovnim skeniranjem (uključujući detekciju shrinkflacije), personalizirano učenje individualnih osjetljivosti iz povijesti skeniranja, formalizacija korisničkih korekcija kao petlje aktivnog učenja te automatsko dovršavanje grafa znanja o sastojcima i alergenima. Pritom se, u skladu s hijerarhijom uspostavljenom u potpoglavlju 5.5, kao nosivi smjer izdvaja formalizacija korisničkih i operatorskih korekcija kao petlje aktivnog učenja nad postojećim korpusom, prvi i četvrti smjer razrađuju se kao sekundarni, a personalizirano učenje individualnih osjetljivosti kao perspektiva daljnjeg rada uvjetovana pravnim i etičkim preduvjetima. Očekivani znanstveni doprinosi konsolidirani su u potpoglavlju 5.6.

LITERATURA

- [1] EUR-Lex: *Uredba (EU) br. 1169/2011 Europskog parlamenta i Vijeća o informiranju potrošača o hrani*, s Interneta, <https://eur-lex.europa.eu/eli/reg/2011/1169/oj>, 11. srpnja 2026.
- [2] EUR-Lex: *Uredba (EU) 2025/40 Europskog parlamenta i Vijeća o ambalaži i ambalažnom otpadu (PPWR)*, s Interneta, <https://eur-lex.europa.eu/eli/reg/2025/40/oj>, 11. srpnja 2026.
- [3] Europska komisija: *RASFF — Rapid Alert System for Food and Feed*, s Interneta, https://food.ec.europa.eu/food-safety/rasff_en, 11. srpnja 2026.
- [4] Open Food Facts: *Discover Open Food Facts*, s Interneta, <https://world.openfoodfacts.org/discover>, 11. srpnja 2026.
- [5] Scan Eat d.o.o.: *ScanEat – aplikacija za analizu pretpakiranih prehrambenih proizvoda*, s Interneta, <https://scaneat.app/>, 20. kolovoza 2026.
- [6] EUR-Lex: *Direktiva 94/62/EZ o ambalaži i ambalažnom otpadu (pročišćeni tekst)*, s Interneta, <https://eur-lex.europa.eu/legal-content/HR/TXT/?uri=CELEX:01994L0062-20180704>, 11. srpnja 2026.
- [7] EUR-Lex: *Uredba (EZ) br. 1935/2004 o materijalima i predmetima koji dolaze u dodir s hranom*, s Interneta, <https://eur-lex.europa.eu/eli/reg/2004/1935/oj>, 11. srpnja 2026.
- [8] EUR-Lex: *Odluka Komisije 97/129/EZ o utvrđivanju identifikacijskog sustava za ambalažne materijale*, s Interneta, <https://eur-lex.europa.eu/legal-content/HR/TXT/?uri=CELEX:31997D0129>, 11. srpnja 2026.
- [9] ISO: *ISO 7000-1135 (Möbiusova petlja) i norma EN ISO 14021:2016 o samodeklariranim okolišnim tvrdnjama*, s Interneta, <https://www.iso.org/obp/ui#iso:grs:7000:1135>, 11. srpnja 2026.
- [10] PRO Europe: *The Green Dot Trademark*, s Interneta, <https://www.pro-e.org/the-green-dot-trademark>, 11. srpnja 2026.
- [11] EUR-Lex: *Direktiva Vijeća 76/211/EEZ o usklađivanju zakonodavstava država članica u odnosu na pakiranje određenih pretpakovina označenih masom ili obujmom*, s Interneta, <https://eur-lex.europa.eu/eli/dir/1976/211/oj>, 11. srpnja 2026.

- [12] Plastics Technology: *In Sustainable Packaging, the Word is 'Monomaterial'*, s Interneta, <https://www.ptonline.com/blog/post/in-sustainable-packaging-the-word-is-monomaterial>, 11. srpnja 2026.
- [13] European Bioplastics: *Guidelines for the use of the Seedling logo (EN 13432)*, s Interneta, <https://www.european-bioplastics.org/guidelines-for-the-use-of-the-seedling-logo/>, 11. srpnja 2026.
- [14] EUR-Lex: *Uredba Komisije (EZ) br. 450/2009 o aktivnim i inteligentnim materijalima i predmetima koji dolaze u dodir s hranom*, s Interneta, <https://eur-lex.europa.eu/legal-content/HR/TXT/?uri=CELEX:32009R0450>, 11. srpnja 2026.
- [15] Europska komisija: *Food information to consumers — legislation*, s Interneta, https://food.ec.europa.eu/food-safety/labelling-and-nutrition/food-information-consumers-legislation_en, 11. srpnja 2026.
- [16] EUR-Lex: *Provedbena uredba Komisije (EU) 2018/775 o pravilima za primjenu članka 26. stavka 3. Uredbe (EU) br. 1169/2011*, s Interneta, https://eur-lex.europa.eu/eli/reg_impl/2018/775/oj, 11. srpnja 2026.
- [17] EUR-Lex: *Direktiva 2011/91/EU o oznakama ili znakovima za identifikaciju serije kojoj hrana pripada*, s Interneta, <https://eur-lex.europa.eu/eli/dir/2011/91/oj>, 11. srpnja 2026.
- [18] EUR-Lex: *Uredba (EZ) br. 1924/2006 o prehrambenim i zdravstvenim tvrdnjama koje se navode na hrani*, s Interneta, <https://eur-lex.europa.eu/eli/reg/2006/1924/oj>, 11. srpnja 2026.
- [19] Narodne novine: *Zakon o informiranju potrošača o hrani*, NN 56/13, 14/14, 56/16, 32/19, s Interneta, <https://www.zakon.hr/z/593/zakon-o-informiranju-potrosaca-o-hrani>, 11. srpnja 2026.
- [20] EUR-Lex: *Direktiva 2007/45/EZ o utvrđivanju pravila o nazivnim količinama za pretpakirane proizvode*, s Interneta, <https://eur-lex.europa.eu/eli/dir/2007/45/oj>, 11. srpnja 2026.
- [21] Europski parlament: *Legislative Train Schedule — Proposal for a harmonised mandatory front-of-pack nutrition labelling*, s Interneta, <https://www.europarl.europa.eu/legislative-train/theme-a-european-green-deal/file-mandatory-front-of-pack-nutrition-labelling>, 11. srpnja 2026.

- [22] FoodNavigator: *Has the European Commission abandoned Nutri-Score?*, s Interneta, <https://www.foodnavigator.com/Article/2025/03/04/has-the-european-commission-abandoned-nutriscore/>, 11. srpnja 2026.
- [23] Europska komisija: *Date marking and food waste prevention*, s Interneta, https://food.ec.europa.eu/food-safety/food-waste/eu-actions-against-food-waste/date-marking-and-food-waste-prevention_en, 11. srpnja 2026.
- [24] Europska komisija: *Alcohol labelling*, s Interneta, https://food.ec.europa.eu/food-safety/labelling-and-nutrition/food-information-consumers-legislation/alcohol-labelling_en, 11. srpnja 2026.
- [25] EUR-Lex: *Uredba (EU) 2026/471 o izmjeni pravila o označavanju proizvoda od vinove loze*, s Interneta, <https://eur-lex.europa.eu/eli/reg/2026/471/oj>, 11. srpnja 2026.
- [26] EUR-Lex: *Direktiva (EU) 2024/1438 o izmjeni direktiva o medu, voćnim sokovima, voćnim džemovima i dehidriranom mlijeku*, s Interneta, <https://eur-lex.europa.eu/eli/dir/2024/1438/oj>, 11. srpnja 2026.
- [27] Europska komisija: *Obavijest Komisije C(2026) 3702 — Guidance document for Regulation (EU) 2025/40*, s Interneta, [https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=intcom:C\(2026\)3702](https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=intcom:C(2026)3702), 11. srpnja 2026.
- [28] EUROOPEN: *12 August 2026: Ready or not, here comes the PPWR!*, s Interneta, <https://www.euopen-packaging.eu/news/12-august-2026-ready-or-not-here-comes-the-ppwr/>, 11. srpnja 2026.
- [29] Europska komisija: *Language and presentation of food information*, s Interneta, https://food.ec.europa.eu/food-safety/labelling-and-nutrition/food-information-consumers-legislation/language-and-presentation-food-information_en, 11. srpnja 2026.
- [30] Légifrance: *Code de la consommation, article R412-7*, s Interneta, https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000032807834, 11. srpnja 2026.
- [31] Bundesministerium der Justiz: *Lebensmittelinformations-Durchführungsverordnung (LMIDV)*, § 2, s Interneta, https://www.gesetze-im-internet.de/lmidv/__2.html, 11. srpnja 2026.
- [32] FASFC/AFSCA: *Food labelling legislation (Belgium)*, s Interneta, <https://www.fasfc.be/food/legislation/food-labelling-legislation>, 11. srpnja 2026.

- [33] Ruokavirasto: *Pakolliset elintarviketiedot (MMMä 834/2014)*, s Interneta, <https://www.ruokavirasto.fi/elintarvikkeet/elintarvikeala/pakkausmerkinnat-ja-markkinointi/pakolliset-elintarviketiedot/>, 11. srpnja 2026.
- [34] Food Safety Authority of Ireland: *Food Information (FIC) — General Provisions (S.I. No. 556/2014)*, s Interneta, [https://www.fsai.ie/enforcement-and-legislation/legislation/food-legislation/food-information-fic-\(labelling\)/general-provisions](https://www.fsai.ie/enforcement-and-legislation/legislation/food-legislation/food-information-fic-(labelling)/general-provisions), 11. srpnja 2026.
- [35] AESAN: *Norma general de etiquetado (Real Decreto 1334/1999)*, s Interneta, https://www.aesan.gob.es/AECOSAN/web/seguridad_alimentaria/subdetalle/norma_general_etiquetado.htm, 11. srpnja 2026.
- [36] Gazzetta Ufficiale: *Decreto Legislativo 15 dicembre 2017, n. 231 (sankcije za Uredbu 1169/2011)*, s Interneta, <https://www.gazzettaufficiale.it/eli/id/2018/2/8/18G00023/sg>, 11. srpnja 2026.
- [37] Hooley Brown: *Scandimix labelling: a quick guide for consumer brands*, s Interneta, <https://www.hooleybrown.com/blog-post/scandimix-labelling-a-quick-guide-for-consumer-brands>, 11. srpnja 2026.
- [38] EUR-Lex: *Obavijest Komisije o pitanjima i odgovorima o primjeni Uredbe (EU) br. 1169/2011 (2018/C 196/01)*, s Interneta, [https://eur-lex.europa.eu/legal-content/HR/TXT/?uri=CELEX:52018XC0608\(01\)](https://eur-lex.europa.eu/legal-content/HR/TXT/?uri=CELEX:52018XC0608(01)), 11. srpnja 2026.
- [39] FoodDrinkEurope: *Guidelines on Legibility of Labelling*, s Interneta, https://www.fooddrinkeurope.eu/wp-content/uploads/2022/10/06672-Guidelines-on-Legibility-of-Labelling_45.pdf, 11. srpnja 2026.
- [40] Consumer Council (Hong Kong): *Study on Legibility of Food Labels*, s Interneta, <https://www.consumer.org.hk/en/press-release/p-551-study-on-legibility-of-food-labels>, 11. srpnja 2026.
- [41] Narodne novine: *Zakon o gospodarenju otpadom*, NN 84/21, 142/23, s Interneta, https://narodne-novine.nn.hr/clanci/sluzbeni/2021_07_84_1554.html, 11. srpnja 2026.
- [42] Narodne novine: *Pravilnik o ambalaži i otpadnoj ambalaži, plastičnim proizvodima za jednokratnu uporabu i ribolovnom alatu koji sadržava plastiku*, NN 137/23, s Interneta, https://narodne-novine.nn.hr/clanci/sluzbeni/2023_11_137_1864.html, 11. srpnja 2026.
- [43] Narodne novine: *Uredba o naknadi gospodarenja otpadom i povratnoj naknadi*, NN 137/24, s Interneta, https://narodne-novine.nn.hr/clanci/sluzbeni/2024_11_137_2258.html, 11. srpnja 2026.

- [44] Fond za zaštitu okoliša i energetske učinkovitost: *Otpadna ambalaža unutar sustava povratne naknade*, s Interneta, <https://www.fzoeu.hr/hr/otpadna-ambalaza-unutar-sustava-povratne-naknade/9313>, 11. srpnja 2026.
- [45] European Environment Agency: *Waste management country profile — Croatia*, s Interneta, <https://www.eea.europa.eu/en/topics/in-depth/waste-and-recycling/municipal-and-packaging-waste-management-country-profiles-2025>, 11. srpnja 2026.
- [46] Europska komisija: *Izješće ranog upozorenja COM(2023) 304 o ostvarenju ciljeva recikliranja (IP/23/3105)*, s Interneta, https://ec.europa.eu/commission/presscorner/detail/en/ip_23_3105, 11. srpnja 2026.
- [47] Légifrance: *Décret n° 2021-835 du 29 juin 2021 relatif à l'information des consommateurs sur la règle de tri (Triman/Info-tri)*, s Interneta, <https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000043714227>, 11. srpnja 2026.
- [48] EUWID Recycling: *Triman ECJ case: Commission notice of action published*, s Interneta, <https://www.euwid-recycling.com/news/policy/triman-ecj-case-commission-notice-of-action-published-260526/>, 11. srpnja 2026.
- [49] UL Solutions: *Italy: Environmental Labeling Requirements for Packaging*, s Interneta, <https://www.ul.com/news/italy-environmental-labeling-requirements-packaging>, 11. srpnja 2026.
- [50] Stiftung Zentrale Stelle Verpackungsregister: *Packaging Act (VerpackG) / LUCID register*, s Interneta, <https://www.verpackungsregister.org/en/foundation-authority/packaging-act>, 11. srpnja 2026.
- [51] Boletín Oficial del Estado: *Real Decreto 1055/2022, de envases y residuos de envases*, s Interneta, <https://www.boe.es/buscar/act.php?id=BOE-A-2022-22690>, 11. srpnja 2026.
- [52] Paragraf.rs: *Pravilnik o deklarisanju, označavanju i reklamiranju hrane, Sl. glasnik RS 19/2017 s izmjenama*, s Interneta, <https://www.paragraf.rs/propisi/pravilnik-o-deklarisanju-oznacavanju-i-reklamiranju-hrane.html>, 11. srpnja 2026.
- [53] Agencija za sigurnost hrane BiH: *Pravilnik o pružanju informacija potrošačima o hrani, Sl. glasnik BiH 68/13*, s Interneta, <https://www.fsa.gov.ba/>, 11. srpnja 2026.
- [54] Agencija za hrana i veterinarstvo (S. Makedonija): *Pravilnik za informacii povrzani so hranata, Sl. vesnik RM 150/15*, s Interneta, <http://fva.gov.mk/mk/zakon-bezbednost-hranata/pravilnik/2015-150-pravilnik-informacii-povrzani-hranata>, 11. srpnja 2026.

- [55] eCFR: *21 CFR 101.15 — Food; prominence of required statements*, s Interneta, <https://www.ecfr.gov/current/title-21/chapter-I/subchapter-B/part-101/subpart-A/section-101.15>, 11. srpnja 2026.
- [56] Fedlex: *Verordnung des EDI betreffend die Information über Lebensmittel (LIV)*, SR 817.022.16, s Interneta, <https://www.fedlex.admin.ch/eli/cc/2017/158/de>, 11. srpnja 2026.
- [57] USDA Foreign Agricultural Service: *China: Prepackaged Food Labeling Standards Finalized (GB 7718-2025)*, s Interneta, <https://www.fas.usda.gov/data/china-prepackaged-food-labeling-standards-finalized>, 11. srpnja 2026.
- [58] GCC Standardization Organization: *GSO 9:2022 — Labeling of prepackaged food stuffs*, s Interneta, <https://www.gso.org.sa/store/standards/GSO:810825/GSO%209:2022?lang=en>, 11. srpnja 2026.
- [59] Schmidt & Schmidt: *TR CU 022/2011 — Food products in part of their marking (EAEU)*, s Interneta, <https://schmidt-export.com/eac-certification/tr-cu-0222011-food-marking>, 11. srpnja 2026.
- [60] USDA Foreign Agricultural Service: *Serbia: FAIRS Country Report Annual (RB2024-0006)*, s Interneta, https://apps.fas.usda.gov/newgainapi/api/Report/DownloadReportByFileName?fileName=FAIRS+Country+Report+Annual_Belgrade_Serbia_RB2024-0006.pdf, 11. srpnja 2026.
- [61] Vlada Crne Gore: *Uredba o informisanju potrošača o hrani, Sl. list CG 22/18*, s Interneta, <https://wapi.gov.me/download/679d5474-071e-487b-89e2-e9545dda017e?version=1.0>, 11. srpnja 2026.
- [62] Schoenherr: *North Macedonia: New Consumer Protection Act*, s Interneta, <https://www.schoenherr.eu/content/north-macedonia-new-consumer-protection-act>, 11. srpnja 2026.
- [63] AKU (Albanija): *VKM nr. 434, 11.7.2018, Për etiketimin e ushqimeve dhe informimin e konsumatorëve*, s Interneta, <https://aku.gov.al/wp-content/uploads/2020/06/VKM-nr.-434-date-11.7.2018.pdf>, 11. srpnja 2026.
- [64] FAO/WHO Codex Alimentarius: *General Standard for the Labelling of Prepackaged Foods CXS 1-1985 (Rev. 2018, 2024)*, s Interneta,

<https://www.fao.org/fao-who-codexalimentarius/codex-texts/list-standards/en/>, 11. srpnja 2026.

- [65] World Health Organization: *47th session of the FAO/WHO Codex Alimentarius Commission adopts new standards*, s Interneta, <https://www.who.int/news-room/events/detail/2024/11/25/default-calendar/47th-session-of-the-FAO-WHO-Codex-Alimentarius-Commission-adopts-new-standards>, 11. srpnja 2026.
- [66] World Trade Organization: *The WTO and the FAO/WHO Codex Alimentarius*, s Interneta, https://www.wto.org/english/thewto_e/coher_e/wto_codex_e.htm, 11. srpnja 2026.
- [67] Federal Register: *Food Labeling: Revision of the Nutrition and Supplement Facts Labels*, 81 FR 33742, s Interneta, <https://www.federalregister.gov/documents/2016/05/27/2016-11867/food-labeling-revision-of-the-nutrition-and-supplement-facts-labels>, 11. srpnja 2026.
- [68] U.S. Food and Drug Administration: *The FASTER Act: Sesame Is the Ninth Major Food Allergen*, s Interneta, <https://www.fda.gov/food/food-allergies/faster-act-sesame-ninth-major-food-allergen>, 11. srpnja 2026.
- [69] USDA Agricultural Marketing Service: *National Bioengineered Food Disclosure Standard*, s Interneta, <https://www.ams.usda.gov/rules-regulations/national-bioengineered-food-disclosure-standard>, 11. srpnja 2026.
- [70] U.S. Food and Drug Administration: *Front-of-Package Nutrition Labeling (proposed rule)*, s Interneta, <https://www.fda.gov/food/nutrition-food-labeling-and-critical-foods/front-package-nutrition-labeling>, 11. srpnja 2026.
- [71] legislation.gov.uk: *Regulation (EU) No 1169/2011 (assimilated law version)*, s Interneta, <https://www.legislation.gov.uk/eur/2011/1169/body>, 11. srpnja 2026.
- [72] Food Standards Agency: *Labelling guidance for prepacked for direct sale (PPDS) food products (Natasha's Law)*, s Interneta, <https://www.food.gov.uk/business-guidance/labelling-guidance-for-prepacked-for-direct-sale-ppds-food-products>, 11. srpnja 2026.
- [73] GOV.UK: *'Not for EU' labelling for retail products across Great Britain: Policy update*, s Interneta, <https://www.gov.uk/government/publications/not-for-eu-labelling-for-retail-products-across-great-britain-policy-update>, 11. srpnja 2026.
- [74] Health Canada: *Nutrition labelling: Front-of-package nutrition symbol*, s Interneta, <https://www.canada.ca/en/health-canada/services/food-nutrition/nutrition-labelling/front-package.html>, 11. srpnja 2026.

- [75] Canadian Food Inspection Agency: *Implementation of the 2016 nutrition labelling amendments*, s Interneta, <https://inspection.canada.ca/en/food-labels/labelling/2023-12-08>, 11. srpnja 2026.
- [76] Food Standards Australia New Zealand: *P1067 — Health Star Rating System*, s Interneta, <https://www.foodstandards.gov.au/food-standards-code/proposals/p1067-health-star-rating-system>, 11. srpnja 2026.
- [77] Consumer Affairs Agency (Japan): *Japan's Food Labelling System*, s Interneta, https://www.caa.go.jp/en/policy/food_labeling/assets/food_labeling_cms204_240425_01.pdf, 11. srpnja 2026.
- [78] Cambridge University Press: *Chile: Front-of-Package Warning Labels and Food Marketing*, Journal of Law, Medicine & Ethics, s Interneta, <https://www.cambridge.org/core/journals/journal-of-law-medicine-and-ethics/article/chile-frontofpackage-warning-labels-and-food-marketing/2B4ED75FFBD5A965E60D471D8763D884>, 11. srpnja 2026.
- [79] Europska komisija: *2023 Annual Report — Alert and Cooperation Network (RASFF)*, s Interneta, https://food.ec.europa.eu/document/download/911d49f2-b3ef-4752-8ea3-5f20dbbe9945_en?filename=acn_annual-report_2023.pdf, 11. srpnja 2026.
- [80] Europska komisija: *2024 Annual Report — Alert and Cooperation Network (RASFF)*, s Interneta, https://food.ec.europa.eu/document/download/a47b9d6a-9b47-4b57-a1ca-35e5bbfa837f_en?filename=acn_annual-report_2024.pdf, 11. srpnja 2026.
- [81] Soon, J. M. i dr.: *Food allergen recalls in the United Kingdom: A critical analysis of reported recalls from 2016 to 2021*, Food Control, Vol. 144, 2022.
- [82] Food Safety Authority of Ireland: *FSAI Annual Report 2024*, s Interneta, <https://www.fsai.ie/publications/fsai-annual-report-2024>, 11. srpnja 2026.
- [83] Campden BRI: *Evaluation of food allergen information, labelling and unintended food allergen presence in imported prepacked foods purchased online in the UK*, Food Control, Vol. 163, 2024.
- [84] GOV.UK / Food Standards Agency: *Major annual retail survey leads to action on Dubai-style chocolate, goat meat and slush ice drinks*, s Interneta, <https://www.gov.uk/government/news/major-annual-retail-survey-leads-to-action-on-dubai-style-chocolate-goat-meat-and-slush-ice-drinks>, 11. srpnja 2026.
- [85] Državni inspektorat Republike Hrvatske: *Godišnje izvješće o radu Državnog inspektorata za 2023. godinu*, s Interneta, <https://dirh.gov.hr/>, 11. srpnja 2026.

- [86] Europska komisija (DG SANCO): *Guidance document with regard to the setting of tolerances for nutrient values declared on a label*, s Interneta, https://food.ec.europa.eu/system/files/2016-10/labelling_nutrition-vitamins_minerals-guidance_tolerances_1212_en.pdf, 11. srpnja 2026.
- [87] Santos, M. i dr.: *Compliance of declared vs. analysed values with EU tolerance limits for mandatory nutrients in prepacked foods*, Food Chemistry, Vol. 302, 2020.
- [88] Europska komisija: *Distance selling (čl. 14. Uredbe (EU) br. 1169/2011)*, s Interneta, https://food.ec.europa.eu/food-safety/labelling-and-nutrition/food-information-consumers-legislation/distance-selling_en, 11. srpnja 2026.
- [89] Europska komisija: *EU coordinated action: Online offered food (2017)*, s Interneta, https://food.ec.europa.eu/food-safety/eu-agri-food-fraud-network/eu-coordinated-actions/online-offered-food-2017_en, 11. srpnja 2026.
- [90] Europska komisija: *Horse meat 2013-14: Test results — EU coordinated action*, s Interneta, https://food.ec.europa.eu/food-safety/eu-agri-food-fraud-network/eu-coordinated-actions/horse-meat-2013-14/test-results_en, 11. srpnja 2026.
- [91] Europska komisija: *EU coordinated action 'From the Hives' (Honey 2021-2022)*, s Interneta, https://food.ec.europa.eu/food-safety/eu-agri-food-fraud-network/eu-coordinated-actions/honey-2021-2022_en, 11. srpnja 2026.
- [92] Food Safety News: *EU sees rise in food safety alerts and fraud investigations (Agri-Food Fraud Network 2023)*, s Interneta, <https://www.foodsafetynews.com/2024/09/eu-sees-rise-in-food-safety-alerts-and-fraud-investigations/>, 11. srpnja 2026.
- [93] Tesseract OCR projekt: *Tesseract User Manual*, s Interneta, <https://tesseract-ocr.github.io/tessdoc/>, 11. srpnja 2026.
- [94] Zacharias, E.; Teuchler, M.; Bernier, B.: *Image Processing Based Scene-Text Detection and Recognition with Tesseract*, arXiv:2004.08079, 2020.
- [95] Zhou, X. i dr.: *EAST: An Efficient and Accurate Scene Text Detector*, Proc. IEEE CVPR, str. 5551-5560, 2017.
- [96] Baek, Y. i dr.: *Character Region Awareness for Text Detection*, Proc. IEEE CVPR, str. 9365-9374, 2019.
- [97] Liao, M. i dr.: *Real-time Scene Text Detection with Differentiable Binarization*, Proc. AAAI, str. 11474-11481, 2020.
- [98] Shi, B.; Bai, X.; Yao, C.: *An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition*, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 39, No. 11, str. 2298-2304, 2017.

- [99] Baek, J. i dr.: *What Is Wrong With Scene Text Recognition Model Comparisons? Dataset and Model Analysis*, Proc. IEEE ICCV, str. 4715-4723, 2019.
- [100] Du, Y. i dr.: *SVTR: Scene Text Recognition with a Single Visual Model*, Proc. IJCAI, str. 884-890, 2022.
- [101] Li, M. i dr.: *TrOCR: Transformer-based Optical Character Recognition with Pre-trained Models*, Proc. AAAI, str. 13094-13102, 2023.
- [102] Kim, G. i dr.: *OCR-free Document Understanding Transformer*, Proc. ECCV, str. 498-517, 2022.
- [103] Nagayi, M. i dr.: *Evaluating OCR performance on food packaging labels in South Africa*, arXiv:2510.03570, 2025.
- [104] Arief, H.: *HalalBench: A Multilingual OCR Benchmark for Food Packaging Ingredient Extraction*, arXiv:2604.22754, 2026.
- [105] JaideAI: *EasyOCR*, s Interneta, <https://github.com/JaideAI/EasyOCR>, 11. srpnja 2026.
- [106] PaddleOCR: *Introduction to PP-OCRv5*, s Interneta, <https://www.paddleocr.ai/main/en/version3.x/algorithm/PP-OCRv5/PP-OCRv5.html>, 11. srpnja 2026.
- [107] PaddleOCR: *PP-OCRv5 Multi-Languages*, s Interneta, https://www.paddleocr.ai/main/en/version3.x/algorithm/PP-OCRv5/PP-OCRv5_multi_languages.html, 11. srpnja 2026.
- [108] Cui, C. i dr.: *PaddleOCR 3.0 Technical Report*, arXiv:2507.05595, 2025.
- [109] Mindee: *docTR — Document Text Recognition*, s Interneta, <https://github.com/mindee/doctr>, 11. srpnja 2026.
- [110] Datalab: *Surya OCR*, s Interneta, <https://github.com/datalab-to/surya>, 11. srpnja 2026.
- [111] OpenMMLab: *MMOCR*, s Interneta, <https://github.com/open-mmlab/mmodocr>, 11. srpnja 2026.
- [112] Google: *ML Kit Text Recognition v2*, s Interneta, <https://developers.google.com/ml-kit/vision/text-recognition/v2>, 11. srpnja 2026.
- [113] Apple: *VNRecognizeTextRequest — Vision framework*, s Interneta, <https://developer.apple.com/documentation/vision/vnrecognizetextrequest>, 11. srpnja 2026.
- [114] Apple: *Read documents using the Vision framework (RecognizeDocumentsRequest)*, WWDC25, s Interneta, <https://developer.apple.com/videos/play/wwdc2025/272/>, 11. srpnja 2026.

- [115] Google: *TensorFlow Lite is now LiteRT*, s Interneta, <https://developers.googleblog.com/tensorflow-lite-is-now-litert/>, 11. srpnja 2026.
- [116] Microsoft: *ONNX Runtime Mobile*, s Interneta, <https://onnxruntime.ai/docs/tutorials/mobile/>, 11. srpnja 2026.
- [117] PyTorch: *Introducing ExecuTorch 1.0*, s Interneta, <https://pytorch.org/blog/introducing-executorch-1-0/>, 11. srpnja 2026.
- [118] Du, Y. i dr.: *PP-OCR: A Practical Ultra Lightweight OCR System*, arXiv:2009.09941, 2020.
- [119] Google Cloud: *Vision API Pricing*, s Interneta, <https://cloud.google.com/vision/pricing>, 11. srpnja 2026.
- [120] Amazon Web Services: *Amazon Textract Pricing*, s Interneta, <https://aws.amazon.com/textract/pricing/>, 11. srpnja 2026.
- [121] European Data Protection Board: *Guidelines 4/2019 on Article 25 Data Protection by Design and by Default*, v2.0, s Interneta, https://www.edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_201904_data_protection_by_design_and_by_default_v2.0_en.pdf, 11. srpnja 2026.
- [122] Fu, L. i dr.: *OCRBench v2: An Improved Benchmark to Evaluate Large Multimodal Models on Visual Text Localization and Reasoning*, Proc. NeurIPS Datasets & Benchmarks, 2025.
- [123] Liu, Z. i dr.: *Seeing is Believing? Mitigating OCR Hallucinations in Multimodal Large Language Models*, arXiv:2506.20168, 2025.
- [124] Google: *On-device GenAI APIs in ML Kit (Gemini Nano)*, s Interneta, <https://android-developers.googleblog.com/2025/05/on-device-gen-ai-apis-ml-kit-gemini-nano.html>, 11. srpnja 2026.
- [125] Apple: *Foundation Models framework*, s Interneta, <https://developer.apple.com/documentation/foundationmodels>, 11. srpnja 2026.
- [126] Marafioti, A. i dr.: *SmolVLM: Redefining small and efficient multimodal models*, arXiv:2504.05299, 2025.
- [127] Das, S. i dr.: *DewarpNet: Single-Image Document Unwarping With Stacked 3D and 2D Regression Networks*, Proc. IEEE ICCV, str. 131-140, 2019.
- [128] Feng, H. i dr.: *DocTr: Document Image Transformer for Geometric Unwarping and Illumination Correction*, Proc. ACM Multimedia, str. 273-281, 2021.

- [129] Sauvola, J.; Pietikäinen, M.: *Adaptive document image binarization*, Pattern Recognition, Vol. 33, No. 2, str. 225-236, 2000.
- [130] Wang, W. i dr.: *Scene Text Image Super-Resolution in the Wild*, Proc. ECCV, str. 650-666, 2020.
- [131] Open Food Facts: *nutrition-table-yolo (YOLOv8n model card)*, s Interneta, <https://huggingface.co/openfoodfacts/nutrition-table-yolo>, 11. srpnja 2026.
- [132] Nguyen, T. T. H.; Jatowt, A.; Coustaty, M.; Doucet, A.: *Survey of Post-OCR Processing Approaches*, ACM Computing Surveys, Vol. 54, No. 6, str. 1-37, 2021.
- [133] Open Food Facts: *nutrition-extractor (LayoutLMv3 model card)*, s Interneta, <https://huggingface.co/openfoodfacts/nutrition-extractor>, 11. srpnja 2026.
- [134] Kuang, J. i dr.: *Visual Information Extraction in the Wild: Practical Dataset and End-to-end Solution*, Proc. ICDAR, str. 36-53, 2023.
- [135] Joulin, A.; Grave, E.; Bojanowski, P.; Mikolov, T.: *Bag of Tricks for Efficient Text Classification (fastText language identification)*, s Interneta, <https://fasttext.cc/docs/en/language-identification.html>, 11. srpnja 2026.
- [136] Kargaran, A. H.; Imani, A.; Yvon, F.; Schütze, H.: *GlottLID: Language Identification for Low-Resource Languages*, Findings of EMNLP, str. 6155-6218, 2023.
- [137] Moillic, J.: *Language Identification for very short texts: a review*, s Interneta, <https://medium.com/besedo-engineering/language-identification-for-very-short-texts-a-review-c9f2756773ad>, 11. srpnja 2026.
- [138] Burchell, L. i dr.: *OpenLID-v3: Improving the Precision of Closely Related Language Identification*, arXiv:2602.13139, 2026.
- [139] Google: *ML Kit Language Identification*, s Interneta, <https://developers.google.com/ml-kit/language/identification>, 11. srpnja 2026.
- [140] Winata, G. i dr.: *The Decades Progress on Code-Switching Research in NLP: A Systematic Survey on Trends and Challenges*, arXiv:2212.09660, 2022.
- [141] Kumar, S. i dr.: *Beyond Monolingual Assumptions: A Survey of Code-Switched NLP in the Era of Large Language Models*, arXiv:2510.07037, 2025.
- [142] EUR-Lex: *Uredba (EZ) br. 1333/2008 o prehrambenim aditivima*, s Interneta, <https://eur-lex.europa.eu/eli/reg/2008/1333/oj>, 11. srpnja 2026.
- [143] Open Food Facts: *Ingredients taxonomy*, s Interneta, <https://wiki.openfoodfacts.org/Ingredients>, 11. srpnja 2026.

- [144] Open Food Facts: *Robotoff — Architecture and ingredient detection*, s Interneta, <https://openfoodfacts.github.io/robotoff/introduction/architecture/>, 11. srpnja 2026.
- [145] Google: *ML Kit On-Device Translation*, s Interneta, <https://developers.google.com/ml-kit/language/translation>, 11. srpnja 2026.
- [146] Feng, F.; Yang, Y.; Cer, D.; Arivazhagan, N.; Wang, W.: *Language-agnostic BERT Sentence Embedding*, Proc. ACL, str. 878-891, 2022.
- [147] Dooley, D. M. i dr.: *FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration*, npj Science of Food, Vol. 2, No. 23, 2018.
- [148] Wróblewska, A. i dr.: *TASTEset — Recipe Dataset and Food Entities Recognition Benchmark*, arXiv:2204.07775, 2022.
- [149] Zhou, X. i dr.: *Revisiting named entity recognition in food computing: enhancing performance and robustness*, Artificial Intelligence Review, Vol. 57, 2024.
- [150] Pratama, R. i dr.: *Named-Entity Recognition and Optical Character Recognition for Detecting Halal Food Ingredients: Indonesian Case Study*, Proc. 10th CITSM, IEEE, 2022.
- [151] Arancio, J.: *How Did Open Food Facts Fix OCR-Extracted Ingredients Using Open-Source LLMs?*, s Interneta, <https://towardsdatascience.com/how-did-open-food-facts-use-open-source-llms-to-enhance-ingredients-extraction-d74dfe02e0e4/>, 11. srpnja 2026.
- [152] Guimarães, V.; Nascimento, J.; Viana, P.; Carvalho, P.: *A Review of Recent Advances and Challenges in Grocery Label Detection and Recognition*, Applied Sciences, Vol. 13, No. 5, 2871, 2023.
- [153] Open Food Facts: *Open Food Facts — Hrvatska*, s Interneta, <https://hr.openfoodfacts.org/>, 11. srpnja 2026.
- [154] FoodNavigator: *Yuka drives reformulation but shuts out food makers*, s Interneta, <https://www.foodnavigator.com/Article/2026/02/09/yuka-drives-reformulation-but-shuts-out-food-makers/>, 11. srpnja 2026.
- [155] Yuka: *Press kit*, s Interneta, <https://yuka.io/wp-content/uploads/presskit/us/yuka-presskit.pdf>, 11. srpnja 2026.
- [156] Langer, A.: *Yuka App Review: Scan or Scam?*, s Interneta, <https://abbylangernutrition.com/yuka-app-review-scan-or-scam/>, 11. srpnja 2026.

- [157] Droid-Life: *MyFitnessPal Puts Barcode Scanner Behind Premium Paywall*, s Interneta, <https://www.droid-life.com/2022/08/24/myfitnesspal-puts-barcode-scanner-behind-premium-paywall/>, 11. srpnja 2026.
- [158] Scandit: *CodeCheck Success Story*, s Interneta, <https://www.scandit.com/resources/case-studies/code-check/>, 11. srpnja 2026.
- [159] Fig — Food Is Good: *Fig: Food scanner for dietary restrictions*, s Interneta, <https://foodisgood.com/>, 11. srpnja 2026.
- [160] Spoonful: *Spoonful — Food scanner app*, s Interneta, <https://spoonfulapp.com/>, 11. srpnja 2026.
- [161] Too Good To Go: *Look, Smell, Taste, Don't Waste — date labelling initiative*, s Interneta, <https://www.toogoodtogo.com/en-gb/look-smell-taste>, 11. srpnja 2026.
- [162] Cornudet, C. i dr.: *When Food Scanner Apps Outperform Front-of-Pack Nutrition Labels: A Conditional Process Model to Foster Healthier Food Choices in Times of Growing Distrust*, Psychology & Marketing, 2025.
- [163] Sarnataro, K. i dr.: *How a food scanner app influences healthy food choice*, Appetite, Vol. 203, 2024.
- [164] Wang, Y. i dr.: *The use of internet-based smartphone apps consistently improved consumers' healthy eating behaviors: a systematic review of randomized controlled trials*, Frontiers in Digital Health, Vol. 6, 2024.
- [165] GS1 US: *What is GS1 Sunrise 2027?*, s Interneta, <https://www.gs1us.org/industries-and-insights/by-topic/sunrise-2027>, 11. srpnja 2026.
- [166] GS1: *GS1 Digital Link Standard: URI Syntax*, s Interneta, <https://ref.gs1.org/standards/digital-link/uri-syntax/>, 11. srpnja 2026.
- [167] Xu, H.; Pang, G.; Wang, Y.; Wang, Y.: *Deep Isolation Forest for Anomaly Detection*, IEEE Transactions on Knowledge and Data Engineering, 2023. DOI: 10.1109/TKDE.2023.3270293
- [168] Mok, C.; Kim, S. B.: *Corruption-based anomaly detection and interpretation in tabular data*, Pattern Recognition, 2025. DOI: 10.1016/j.patcog.2024.111149
- [169] Pena, E. H. M.; Porto, F.; Naumann, F.: *Fast Algorithms for Denial Constraint Discovery*, Proceedings of the VLDB Endowment, 2022. DOI: 10.14778/3574245.3574254
- [170] Liu, Z.; Deep, S.; Fariha, A.; Psallidas, F.; Tiwari, A.; Floratou, A.: *Rapidash: Efficient Detection of Constraint Violations*, Proceedings of the VLDB Endowment, 2024. DOI: 10.14778/3659437.3659454

- [171] Kaminsky, Y.; Pena, E. H. M.; Naumann, F.: *Incremental Detection of Denial Constraint Violations*, Proceedings of the VLDB Endowment, 2024. DOI: 10.14778/3717755.3717761
- [172] Li, D.; Wang, Z.; Chen, Y.; Jiang, R.; Ding, W.; Okumura, M.: *A Survey on Deep Active Learning: Recent Advances and New Frontiers*, IEEE Transactions on Neural Networks and Learning Systems, 2025. DOI: 10.1109/TNNLS.2024.3396463
- [173] Mosqueira-Rey, E.; Hernández-Pereira, E.; Alonso-Ríos, D.; Bobes-Bascarán, J.; Fernández-Leal, Á.: *Human-in-the-loop machine learning: a state of the art*, Artificial Intelligence Review, 2023. DOI: 10.1007/s10462-022-10246-w
- [174] Han, Y.; Liu, D.; Shang, J.; Zheng, L.; Zhong, J.; Cao, W.; Sun, H.; Xie, W.: *BALQUE: Batch active learning by querying unstable examples with calibrated confidence*, Pattern Recognition, 2024. DOI: 10.1016/j.patcog.2024.110385
- [175] Martins, V. E.; Cano, A.; Barbon Junior, S.: *Meta-learning for dynamic tuning of active learning on stream classification*, Pattern Recognition, 2023. DOI: 10.1016/j.patcog.2023.109359
- [176] Kwak, B.; Kim, Y.; Kim, Y. J.; Hwang, S.; Yeo, J.: *TrustAL: Trustworthy Active Learning Using Knowledge Distillation*, Proc. AAAI, 2022. DOI: 10.1609/aaai.v36i7.20688
- [177] Lin, J.; Zhu, L.; Chen, W.-M.; Wang, W.-C.; Gan, C.; Han, S.: *On-Device Training Under 256KB Memory*, Proc. NeurIPS, 2022. DOI: 10.52202/068431-1667
- [178] Yang, Y.; Li, G.; Mărculescu, R.: *Efficient On-Device Training via Gradient Filtering*, Proc. IEEE CVPR, 2023. DOI: 10.1109/CVPR52729.2023.00371
- [179] Huang, W.; Ye, M.; Shi, Z.; Wan, G.; Li, H.; Du, B.; Yang, Q.: *Federated Learning for Generalization, Robustness, Fairness: A Survey and Benchmark*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024. DOI: 10.1109/TPAMI.2024.3418862
- [180] Qu, L.; Zhou, Y.; Liang, P. P.; Xia, Y.; Wang, F.; Adeli, E.; Fei-Fei, L.; Rubin, D.: *Rethinking Architecture Design for Tackling Data Heterogeneity in Federated Learning*, Proc. IEEE CVPR, 2022. DOI: 10.1109/CVPR52688.2022.00982
- [181] Cao, J.; Fang, J.; Meng, Z.; Liang, S.: *Knowledge Graph Embedding: A Survey from the Perspective of Representation Spaces*, ACM Computing Surveys, 2024. DOI: 10.1145/3643806
- [182] Subagdja, B.; Shanthoshigaa, D.; Wang, Z.; Tan, A.-H.: *Machine Learning for Refining Knowledge Graphs: A Survey*, ACM Computing Surveys, 2024. DOI: 10.1145/3640313

- [183] Wang, L.; Zhao, W.; Wei, Z.; Liu, J.: *SimKGC: Simple Contrastive Knowledge Graph Completion with Pre-trained Language Models*, Proc. ACL, 2022. DOI: 10.18653/v1/2022.acl-long.295
- [184] Shomer, H.; Jin, W.; Wang, W.; Tang, J.: *Toward Degree Bias in Embedding-Based Knowledge Graph Completion*, Proc. WWW, 2023. DOI: 10.1145/3543507.3583544

POPIS OZNAKA I KRATICA

API	aplikacijsko programsko sučelje (engl. Application Programming Interface)
CER	stopa pogreške znakova (engl. Character Error Rate)
CFIA	Kanadska agencija za inspekciju hrane (engl. Canadian Food Inspection Agency)
EAEU	Euroazijska ekonomska unija
engl.	engleski
EU	Europska unija
FDA	američka Agencija za hranu i lijekove (engl. Food and Drug Administration)
FIC	informiranje potrošača o hrani (engl. Food Information to Consumers)
FOP	označavanje na prednjoj strani ambalaže (engl. Front-of-Pack)
GCC	Vijeće za suradnju zaljevskih država (engl. Gulf Cooperation Council)
GDPR	Opća uredba o zaštiti podataka (engl. General Data Protection Regulation)
GTIN	globalni broj trgovačke jedinice (engl. Global Trade Item Number)
LID	identifikacija jezika (engl. Language Identification)
LSTM	duga kratkoročna memorija (engl. Long Short-Term Memory)
NER	prepoznavanje imenovanih entiteta (engl. Named Entity Recognition)
NN	Narodne novine
NRV	referentna hranjiva vrijednost (engl. Nutrient Reference Value)
OCR	optičko prepoznavanje znakova (engl. Optical Character Recognition)
OFF	Open Food Facts
ONNX	otvoreni format za razmjenu neuronskih mreža (engl. Open Neural Network Exchange)
PFAS	per- i polifluoroalkilne tvari (engl. Per- and Polyfluoroalkyl Substances)
PPWR	Uredba o ambalaži i ambalažnom otpadu (engl. Packaging and Packaging Waste Regulation)
QR	kod brzog odziva (engl. Quick Response)
QUID	količinsko označavanje sastojaka (engl. Quantitative Ingredient Declaration)
RASFF	Sustav brzog uzbunjivanja za hranu i hranu za životinje (engl. Rapid Alert System for Food and Feed)
VLM	vizualno-jezični model (engl. Vision-Language Model)
WER	stopa pogreške riječi (engl. Word Error Rate)
WTO	Svjetska trgovinska organizacija (engl. World Trade Organization)
YOLO	detektor objekata u jednom prolazu (engl. You Only Look Once)

SAŽETAK

Rad daje pregled područja ambalaže pretpakiranih prehrambenih proizvoda s regulatornoga i računalnoga gledišta te predlaže nadogradnju mobilne aplikacije ScanEat. U regulatornom dijelu obrađeni su zahtjevi označavanja hrane u Europskoj uniji (Uredba (EU) br. 1169/2011), novi okvir za ambalažu i ambalažni otpad (Uredba (EU) 2025/40), jezični zahtjevi pojedinih država i praksa višejezičnih deklaracija, nacionalni propisi odabranih država članica i država izvan EU-a (Codex Alimentarius, SAD, Ujedinjeno Kraljevstvo, Kanada, Švicarska, Japan, Kina i dr.) te dokumentirani problemi pridržavanja propisa, među kojima prednjače neoznačeni alergeni kao vodeći uzrok opoziva proizvoda. U računalnom dijelu pregledno su obrađene tehnologije optičkog prepoznavanja znakova (OCR) s naglaskom na izvođenje na klijentskoj strani (mobilnom uređaju), identifikacija jezika i pisma na kratkim tekstovima, segmentacija višejezičnih deklaracija te ekstrakcija strukturiranih podataka (sastojci, alergeni, nutritivne vrijednosti), uz analizu postojećih aplikacija i baza podataka. Na temelju identificiranih praznina, ovisnosti postojećih rješenja o crtičnom kodu, slabe pokrivenosti regionalnih tržišta, zanemarivanja višejezičnosti i nepostojanja provjere sukladnosti, predložena je arhitektura nadogradnje aplikacije ScanEat: klijentski OCR cjevovod (vođeno snimanje, predobrada, detekcija regija, OCR, postprocesiranje i strukturiranje), modul za prepoznavanje i povezivanje višejezičnih blokova deklaracije te modul provjere sukladnosti deklaracije s propisima pojedine države. Definirana je metodologija vrednovanja koja uključuje izgradnju anotiranog skupa višejezičnih deklaracija jugoistočne Europe, mjere točnosti ekstrakcije te usporedbu s ručnim unosom podataka. Ocrtani su i smjerovi doktorskog istraživanja: automatsko otkrivanje nesukladnosti i promjena deklaracija, personalizirana detekcija osjetljivosti, aktivno učenje iz korisničkih korekcija i dovršavanje grafa znanja o hrani.

Ključne riječi: ambalaža, pretpakirani prehrambeni proizvodi, deklaracija, označavanje hrane, OCR, obrada na klijentskoj strani, višejezičnost, sukladnost s propisima, mobilna aplikacija