

2021 < ŠTEVILKA 2 < APR. MAJ. JUN. < LETNIK XXIX < ISSN 1318-1882

02 UPORABNA INFORMATIKA

U P O R A B N A I N F O R M A T I K A

2021 ŠTEVILKA 2 APR/MAJ/JUN LETNIK XXIX ISSN 1318-1882

▣ Strokovni prispevki

- Tomaž Dular, Julija Lapuh Bele, Rok Pirnat
Odkrivanje prevar z orodji poslovne inteligence 67
- Simona Samida Cerk
Sodobni izzivi izobraževanja na daljavo v okviru medpredmetnega povezovanja in inovativnih učnih okolij 76

▣ Znanstveni prispevki

- Tim Poštuvan, Semir Salkič, Lovro Šubelj
Learning-based link prediction analysis for Facebook100 network 83
- Elena Osrajnik, Sašo Karakatič
Temni vzorci na slovenskih spletnih straneh 95
- Milena Košak Babuder, Blažka Korun, Ema Štarkl, Gaja Nenadović, Dušica Boben, Erika Stanković, Sara Jakop, Luka Vranješ, Karmen Javornik, Jure Žabkar
Računalniško podprto prepoznavanje zgodnjih znakov disleksije 113

▣ Razprave

- Katarina Puc, Niko Schlamberger, Vladimir Batagelj, Tomaž Turk, Ivan Kanič
Pogovor sodelavcev ob 20-letnici Islovarja 129

▣ Informacije

- Iz slovarja** 140

Ustanovitelj in izdajatelj

Slovensko društvo INFORMATIKA
Litostrojska cesta 54, 1000 Ljubljana

Predstavniki

Niko Schlamberger

Odgovorni urednik

Saša Divjak

Uredniški odbor

Andrej Kovači, Evelin Krmac, Ivan Rozman, Jan Mendling, Jan von Knop, John Taylor, Jurij Jaklič, Lili Nemeč Zlatolas, Marko Hölbl, Mirjana Kljajič Borštnar, Mirko Vintar, Pedro Simões Coelho, Saša Divjak, Sjaak Brinkkemper, Slavko Žitnik, Tatjana Welzer Družovec, Vesna Bosilj-Vukšić, Vida Groznik, Vladislav Rajkovič

Recenzentski odbor

Alenka Kavčič, Aljaž Košmerlj, Andrej Brodnik, Andrej Kovačič, Bor Plestenjak, Borut Werber, Borut Žalik, Boštjan Žvanut, Božidar Potočnik, Ciril Bohak, Danijel Skočaj, David Jelenc, Dejan Georgiev, Dejan Lavbič, Denis Trček, Dobravec Tomaž, Domen Mongus, Eva Krhač, Evelin Krmac, Franc Solina, Gregor Weiss, Igor Kononenko, Inna Novalija, Irena Nančovska Šerbec, Ivan Gerlič, Janez Demšar, Jurij Jaklič, Jurij Mihelič, Katarina Puc, Lovro Šubelj, Luka Pavlič, Luka Čehovin, Marina Trkman, Marjan Heričko, Marjan Krisper, Marko Bajec, Marko Hölbl, Martin Vodopivec, Matevž Pesek, Matija Marolt, Mihaela Triglav Čekada, Mirjana Kljajič Borštnar, Mojca Indihar Štemberger, Monika Klun, Peter Trkman, Sandi Gec, Saša Divjak, Slavko Žitnik, Tomaž Erjavec, Uroš Godnov, Uroš Rajkovič, Vida Groznik, Vladislav Rajkovič, Vlado Stankovski, Živa Rant

Tehnični urednik

Slavko Žitnik

Lektoriranje angleških izvlečkov

Marvelingua (angl.)

Oblikovanje

KOFEIN DIZAJN, d. o. o.

Prelom in tisk

Boex DTP, d. o. o., Ljubljana

Naklada

200 izvodov

Naslov uredništva

Slovensko društvo INFORMATIKA
Uredništvo revije Uporabna informatika
Litostrojska cesta 54, 1000 Ljubljana
www.uporabna-informatika.si

Revija izhaja četrtletno. Cena posamezne številke je 20,00 EUR. Letna naročnina za podjetja 85,00 EUR, za vsak nadaljnji izvod 60,00 EUR, za posameznike 35,00 EUR, za študente in seniorje 15,00 EUR. V ceno je vključen DDV.

Revija Uporabna informatika je od številke 4/VII vključena v mednarodno bazo INSPEC.

Revija Uporabna informatika je pod zaporedno številko 666 vpisana v razvid medijev, ki ga vodi Ministrstvo za kulturo RS.

Revija Uporabna informatika je vključena v Digitalno knjižnico Slovenije (dLib.si).

© Slovensko društvo INFORMATIKA

Vabilo avtorjem

V reviji Uporabna informatika objavljamo kakovostne izvirne članke domačih in tujih avtorjev z najširšega področja informatike v poslovanju podjetij, javni upravi in zasebnem življenju na znanstveni, strokovni in informativni ravni; še posebno spodbujamo objavo interdisciplinarnih člankov. Zato vabimo avtorje, da prispevke, ki ustrezajo omenjenim usmeritvam, pošljejo uredništvu revije po elektronski pošti na naslov ui@drustvo-informatika.si.

Avtorje prosimo, da pri pripravi prispevka upoštevajo navodila, objavljena v nadaljevanju ter na naslovu <http://www.uporabna-informatika.si>.

Za kakovost prispevkov skrbi mednarodni uredniški odbor. Članki so anonimno recenzirani, o objavi pa na podlagi recenzij samostojno odloča uredniški odbor. Recenzenti lahko zahtevajo, da avtorji besedilo spremenijo v skladu s priporočili in da popravljeni članek ponovno prejmejo v pregled. Uredništvo pa lahko še pred recenzijo zavrne objavo prispevka, če njegova vsebina ne ustreza vsebinski usmeritvi revije ali če članek ne ustreza kriterijem za objavo v reviji.

Pred objavo članka mora avtor podpisati izjavo o avtorstvu, s katero potrjuje originalnost članka in dovoljuje prenos materialnih avtorskih pravic. Nenaročenih prispevkov ne vračamo in ne honoriramo. Avtorji prejmejo enoletno naročnino na revijo Uporabna informatika, ki vključuje avtorski izvod revije in še nadaljnje tri zaporedne številke.

S svojim prispevkom v reviji Uporabna informatika boste prispevali k širjenju znanja na področju informatike. Želimo si čim več prispevkov z raznoliko in zanimivo tematiko in se jih že vnaprej veselimo.

Uredništvo revije

Navodila avtorjem člankov

Članke objavljamo praviloma v slovenščini, članke tujih avtorjev pa v angleščini. Besedilo naj bo jezikovno skrbno pripravljeno. Priporočamo zmernost pri uporabi tujk in – kjer je mogoče – njihovo zamenjavo s slovenskimi izrazi. V pomoč pri iskanju slovenskih ustreznih priporočamo uporabo spletnega terminološkega slovarja Slovenskega društva Informatika Islovar (www.islovar.org).

Znanstveni članek naj obsega največ 40.000 znakov, strokovni članki do 30.000 znakov, obvestila in poročila pa do 8.000 znakov.

Članek naj bo praviloma predložen v urejevalniku besedil Word (*.doc ali *.docx) v enojnem razmaku, brez posebnih znakov ali poudarjenih črk. Za ločilom na koncu stavka napravite samo en prazen prostor, pri odstavkih ne uporabljajte zamika.

Naslovu članka naj sledi za vsakega avtorja polno ime, ustanova, v kateri je zaposlen, naslov in elektronski naslov. Sledi naj povzetek v slovenščini v obsegu 8 do 10 vrstic in seznam od 5 do 8 ključnih besed, ki najbolje opredeljujejo vsebinski okvir članka. Pred povzetkom v angleščini naj bo še angleški prevod naslova, prav tako pa naj bodo dodane ključne besede v angleščini. Obratno velja v primeru predložitve članka v angleščini. Razdelki naj bodo naslovljeni in oštevilčeni z arabskimi številkami.

Slike in tabele vključite v besedilo. Opremite jih z naslovom in oštevilčite z arabskimi številkami. Vsako sliko in tabelo razložite tudi v besedilu članka. Če v članku uporabljate slike ali tabele drugih avtorjev, navedite vir pod sliko oz. tabelo. Revijo tiskamo v črno-beli tehniki, zato barvne slike ali fotografije kot original niso primerne. Slik zaslonov ne objavljamo, razen če so nujno potrebne za razumevanje besedila. Slike, grafikoni, organizacijske sheme ipd. naj imajo belo podlago. Enačbe oštevilčite v oklepajih desno od enačbe.

V besedilu se sklicujte na navedeno literaturo skladno s pravili sistema APA navajanja bibliografskih referenc, najpogosteje torej v obliki (Novak & Kovač, 2008, str. 235). Na koncu članka navedite samo v članku uporabljeno literaturo in vire v enotnem seznamu po abecednem redu avtorjev, prav tako v skladu s pravili APA. Več o sistemu APA, katerega uporabo omogoča tudi urejevalnik besedil Word 2007, najdete na strani <http://owl.english.purdue.edu/owl/resource/560/01/>.

Članku dodajte kratek življenjepis vsakega avtorja v obsegu do 8 vrstic, v katerem poudarite predvsem strokovne dosežke.

█ Odkrivanje prevar z orodji poslovne inteligence

Tomaž Dular¹, Špela Povrženič¹, Julija Lapuh Bele^{1, 2}, Rok Pirnat¹

¹B2 BI d.o.o.,

²Visoka šola za poslovne vede

tomaz.dular@b2-bi.com, spela.povrzenic@b2-bi.com, julija.bele@b2.eu, rok.pirnat@b2-bi.com

Izvleček

Sistem poslovnega obveščanja mora uporabnikom zagotavljati vrhunsko uporabniško izkušnjo, biti mora prilagodljiv in temeljiti na zanesljivih podatkih. Uporabnikove zbirke podatkov pogosto niso prečiščene, kar je lahko posledica nekonsistentnih poslovnih procesov, napak pri beleženju podatkov ali prevar. Z orodjem Power BI in SQL strežnikom smo pripravili rešitev, ki s pomočjo Benfordovega zakona določi sporne podatke in jih prikaže na grafično všečen način. Interni revizorji tako dobijo odličen pripomoček, s katerim si pomagajo pri analizi poslovanja in odkrivanju napak in prevar.

Ključne besede: odkrivanje prevar, Benfordov zakon, poslovna inteligenca, poslovna analitika

Abstract

Business intelligence systems must be flexible, based on reliable data while providing users with a premium user experience in data analysis. A user's databases are often unrefined, which can be the result of inconsistent business processes, errors in data logging or fraud. Using Power BI and SQL Server, we have developed a solution that, based on Benford's law, determines conflicting information and renders graphically pleasing presentations. Internal auditors are thus provided an excellent tool to help them analyze their operations and detect errors and fraud.

Keywords: Fraud detection, Bedford's law, business intelligence, business analytics

UVOD

Odkrivanje prevar z metodami strojnega učenja je v različnih finančnih institucijah v uporabi že vrsto let (Horan, 2020). S pomočjo naprednih računalniških tehnologij se specializirane skupine strokovnjakov prebijajo skozi milijone transakcij na računih uporabnikov – podjetij in fizičnih oseb. Naložbe v strojno opremo in človeške vire se bankam povrnejo v kratkem času.

Z uvedbo spletnih storitev, kot sta na primer Microsoft Azure AI ali Google Cloud AI Platform, pa za vstop v svet umetne inteligence (AI) niso več potrebne velike investicije. Poleg orodij umetne inteligence obstaja še nekaj drugih možnosti za preverjanje podatkov. Več avtorjev se je lotilo preverjanja računovodskih izkazov (Amiram idr., 2015). Niso pa samo knjigovodski izkazi predmet podrobnih analiz.

Holz (2014) je na primer proučeval kvaliteto podatkov kitajskih državnih statistik o rasti bruto domačega proizvoda in ni potrdil domnev o potvorjenih rezultatih. Zaključuje, da nacionalni zavod za statistiko ne potvarja rezultatov ali pa rezultate prireja skladno s statističnimi zakoni.

V prispevku smo se teme prevar lotili na drugačen način, z uporabo empiričnega Benfordovega zakona in brez zahtevnih modelov strojnega učenja. Cilj je bila priprava informacijske rešitve z orodji poslovne inteligence za odkrivanje prevar, ki bi notranje revizorje v podjetju usmerila na »problematične« podatke. Revizorji namreč ne morejo pregledati celotne množice podatkov, ki jih letno zberejo ali ustvarijo zaposleni v raznih poslovnih procesih. Na osnovi izkušenj podrobno preverijo le podmnožice »problematičnih« podatkov. Naš cilj je bil pripraviti

testne podmnožice s pomočjo rešitve v analitičnem okolju Power BI, ki za teoretično podlago prikazuje pojavnost števk (cifere) v primerjavi z napovedjo Benfordovega zakona.

BENFORDOV ZAKON

Benfordov empirični zakon (Benford, 1938) obravnava verjetnost pojavitve posameznih števk v množici števil. Po občutku bi trdili, da se vse številke pojavljajo enako verjetno. Benford pa trdi, da se nizke številke (1, 2,...) pojavijo pogosteje kot visoke številke (7, 8 in 9). Zakon potrjujejo analize mnogih podatkovnih zbirk, ki ji najdemo v statističnih letopisih. Benford (1938) je svojo domnevo potrdil na dvajsetih zelo raznovrstnih zbirkah podatkov. Zanimivi primeri so analiza časovnih intervalov med zaporednimi potresi, molske mase kemičnih spojin, količina vode v rečnih tokovih, populacija v občini ali državi, rezultati volitev po mestih, analiza vseh dohodkov gospodinjstev znotraj neke države ali mesta, velikost datotek v megabajtih na poljubnem računalniku (Povrženič, 2019). Za nas so posebej zanimive zbirke računovodskih podatkov, ki prav tako sledijo zakonu, kar je dokazal Nigrini (2012).

Na osnovi Benfordovega zakona lahko postavimo trditve: če podatki ne potrjujejo Benfordovega zakona, potem lahko podvomimo v njihovo pravilnost in poštenost. Benfordovega zakona seveda ne moremo uporabljati brez omejitev in na vsaki množici podatkov. Razen tega se moramo zavedati, da je sumljive podatke treba še podrobneje pregledati z drugimi metodami.

Zakon je v bistvu odkril Simon Newcomb (1881). Zato ga imenujemo tudi Newcomb-Benfordov zakon. Ob listanju logaritmskih tabel je opazil, da so vse knjige na začetnih straneh veliko bolj obrabljene in umazane, kakor strani pri koncu (Kossovsky, 2014). V času pred računalniki so bile logaritmske tablice glavni pripomoček za računanje – predvsem za množenje in deljenje večjih števil. Logaritmske tablice so urejene po števkih na prvem mestu - začetne strani vsebujejo logaritme za števila, ki se začnejo z 1, 2. Zaradi umazanih začetnih strani je Newcomb (1881) dobil idejo, da se uporabniki večkrat srečujejo s števili z manjšimi vodilnimi števki, kakor pa z velikimi.

Pojasnimo najprej izraze. Predznaka števila ne upoštevamo. Prva vodilna števka (cifra) v številu je prva neničelna števka na levi strani števila. Za naslednje številke velja podobno: druga vodilna števka je druga levo ležeča števka (njene vrednosti so lahko od 0 do 9), ...

Zakon temelji na empiričnih dokazih. Benford (1938) je ugotovil, da je na proučevanih podatkih iz realnega življenja (npr. hišne številke, število prebivalcev po krajih) kar 30,0 % verjetnost, da je na prvem mestu števila števka 1. Števka 2 se pojavlja na prvem mestu z verjetnostjo 17,6 %, števka 3 pa z verjetnostjo 12,5 %. Verjetnosti z rastjo številke padajo. Če bi bile številke porazdeljene enakomerno, bi imela vsaka od njih verjetnost 11,11 %.

Frekvenco pojavitve številke na prvem mestu oz. verjetnost, da se števka d pojavi na prvem mestu v številu je mogoče izračunati po naslednji formuli

Tabela 1: Benfordov zakon

Zakon	Oznaka številke	Verjetnost pojavitve	Pogoji
Zakon 1. vodilne številke	1. števka = d	$\log\left(1 + \frac{1}{d}\right)$	$d \in \mathbb{Z}$, $d \in \{1, 9\}$
Zakon 2. vodilne številke	2. števka = d	$\sum_{k=1}^{\infty} \log\left(1 + \frac{1}{10k + d}\right)$	$d \in \{0, 9\}$
Zakon 3. vodilne številke	3. števka = d	$\sum_{m=n}^{\infty} \sum_{n=0}^{\infty} \log\left(1 + \frac{1}{100m + 10n + d}\right)$	$d \in \{0, 9\}$
Zakon n-te vodilne številke	n . števka = d	$10^n \sum_{k=1}^{\infty} \log\left(1 + \frac{1}{10k + d}\right)$	$d \in \mathbb{Z}$, $d \in \{0, 9\}$
Kombinacija prvih dveh števk	1. števka = p , 2. števka = q	$\sum_{k=10^{n-2}}^{\infty} \log\left(1 + \frac{1}{10p + q}\right)$	$p \geq 0$ in $q \geq 0$.
Kombinacija prvih treh števk	1. števka = p , 2. števka = q , 3. števka = n	$\log\left(1 + \frac{1}{100p + 10q + r}\right)$	$p \geq 0$ in $q \geq 0$ in $r \geq 0$.
Kombinacija zadnjih dveh števk		$\frac{1}{100}$ oz. 1% - enakomerna porazdelitev	

$$P(d) = \log \left(1 + \frac{1}{d} \right)$$

V tabeli 1 so navedene verjetnosti pojavitve za prve tri vodilne številke in posplošena formula verjetnostne porazdelitve za n-to vodilno številko.

Benfordov zakon lahko posplošimo in prevedemo tudi na drug številski sistem in posledično uporabimo drugo logaritemsko osnovo.

Empirično lahko pokažemo, da ima veliko podatkovnih zbirk iz vsakdanjega življenja logaritemsko porazdeljene vodilne številke pod pogojem, da imajo širok interval podatkov oz. velik red velikosti podatkovne množice. Red velikosti izračunamo iz naslednje formule.

$$\text{Red velikosti} = \log(X_{\text{Max}}) - \log(X_{\text{Min}})$$

Podatkovne zbirke, ki imajo razliko logaritmov skrajnih podatkovnih vrednosti večjo kot 3, imajo ponavadi logaritemsko porazdeljene vodilne številke.

Načelo invariantnosti

Načelo invariantnosti (scale invariance) pravi: ne glede na to, v katerem merskem sistemu enot imamo dane podatke, je naključna spremenljivka enako porazdeljena. Če torej prvotni podatki ustrezajo Benfordovemu zakonu in podatke pretvorimo v katero koli drugo mersko enoto in porazdelitev, bo še vedno ostala Benfordova. Enako velja tudi za podatke, ki niso logaritemski in tudi po pretvarjanju ne bodo.

Testiranje podatkovne zbirke

Za testiranje odmikov v podatkih od Benfordove porazdelitve uporabljamo več testov. S standardni statističnim testom Z testiramo skladnost za posamezno številko, s testom hi-kvadrat pa za celotno podatkovno množico. Zelo dobra mera je tudi odstopanje vsote kvadratov (SSD), s katero za vodilne številke ali kombinacije ugotovimo, ali se zbirka podatkov popolno, sprejemljivo, mejno prilega, ali ne prilega Benfordovi porazdelitvi.

Test ponavljajočih se vrednosti

Test ponavljajočih se vrednosti je dopolnilni test, ki preveri, kolikokrat se v podatkovni množici pojavi določena vrednost. Ta test je bistven pri analizi nepravilnosti v podatkih, saj nam pomaga pri prepoznavanju značilnih količin, ki povzročajo skoke na grafih. V testu se osredotočimo na zelo pogosto

ponavljajoče se podatke v podatkovni zbirki in skušamo odkriti razloge za odstopanja. Pogosto ponavljanje posameznih števil lahko kaže na posebnosti v zbirki.

Odkloni od Benfordove porazdelitve nakazujejo možnosti prevare. Odstopanje od porazdelitve ni dokaz, da je do prevare res prišlo, temveč le usmeritev revizorju, kje naj išče. Ponarejevalec pri navajanju lažnih števil ne sledi Benfordovemu zakonu, temveč nezavedno večkrat uporabi iste kombinacije števk. Takšne kombinacije z uporabo Benfordovega zakona hitro odkrijemo, še posebej, če je izmišljenih števil veliko.

Test prvih vodilnih števk ni zadosten pri odkrivanju prevar. Nujno moramo preveriti tudi teste višjih stopenj in teste kombinacij prvih nekaj in zadnjih nekaj števk. Preveriti moramo tudi test ponavljajočih se vrednosti in odkriti razvojni vzorec števk. Carslaw (1988) je ugotovil, da veliki presežki ali pomanjkanja deleža številke 0 v testih višjih redov pomenijo goljufiva zaokroževanja.

Pri analizi vedno ločeno obravnavamo pozitivne in negativne vrednosti in pri tem pogosto negativne vrednosti izpustimo. To počnemo le zaradi vsebinske interpretacije. Negativni dokumenti so navadno samo stornacije dokumentov in kvarijo porazdelitev. Lahko pa ločeno obravnavamo stornacije in iščemo morebitne prevare. Ločeno obravnavamo tudi podatkovne zbirke glede na vrsto podatkov. Na primer podatke iz glavne knjige analiziramo ločeno po kontih prihodkov, stroškov, prilivov, odливov.

Paziti pa moramo, da v analizi obravnavamo samo surove, neobdelane podatke. Vse vsote, zmnožke in druge agregacije odstranimo iz analize. Prav tako ne smemo analizirati mešanice surovih in agregiranih podatkov.

PRIMERA ANALIZE

Pripravili smo dva primera analize na isti računovodski zbirki podatkov, ki je vsebovala približno štiri milijone zapisov. Zbirka je vsebovala podatke na ravni postavke računa, brez seštevka. Tako smo izpolnili zgoraj navedeni pogoj za analizo, da ne smemo analizirati surovih podatkov skupaj z agregati. Analizirali smo vrednost zaračunane prodaje in zneske avansov. Podatke smo analizirali v Power BI in dobili analitično orodje, s katerim bodo revizorji preverjali skladnost podatkovne zbirke z Benfordovim zakonom. S preprosto dosegljivimi filtri lahko omejujejo

podatke po več dimenzijah: obdobje, države, kupci, stroškovna mesta, referenti, skupine artiklov ...

Z analizo prodajnih vrednosti smo želeli preveriti veljavnost zakona na resnični računovodski zbirki podatkov. Rezultate analize vidimo na grafih od 1 do 4. Stolpce na vseh grafih smo obarvali s štirimi barvami, skladno s štirimi razredi napak, navedeni-

mi v tabeli 2. Z zeleno barvo sta obarvana »popolni« in »sprejemljivi Benford«. Naslednja dva razreda sta obarvana rjavo in rdeče. Rdeča barva označuje popolno neskladje z Benfordovim zakonom. Benfordova porazdelitev je kot referenčna črta narisana z modro barvo.

Tabela 2: **SSD – razredi klasifikacije z velikostjo odstopanj (Kossovsky, 2014)**

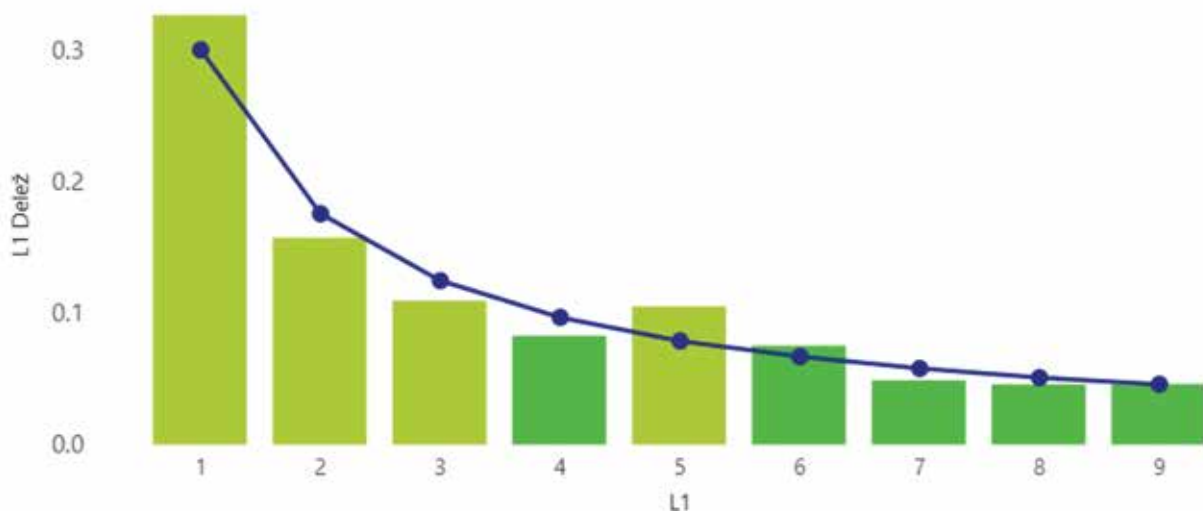
SSD (%)	Popolni Benford	Sprejemljiv Benford	Mejni Benford	Ni Benford
Zakon prve vodilne številke	<2	2-25	25-100	>100
Zakon druge vodilne številke	<2	2-10	10-50	>50
Kombinacija prvih dveh števk	<2	2-10	10-50	>50
Kombinacija zadnjih dveh števk	<4	4-10	40-100	>100
Barve na slikah	■	■	■	■

Analiza neto zaračunanih vrednosti

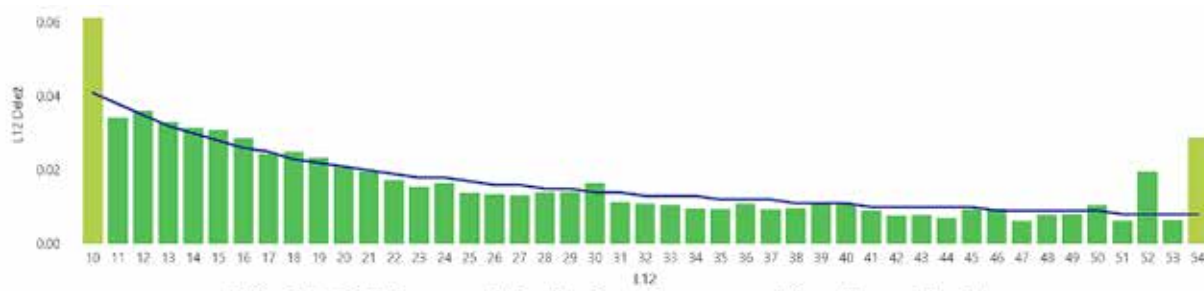
Analiza neto zaračunane vrednosti pokaže skoraj idealno ujemanje z Benfordovim zakonom. Poraz-

delite po zakonu so zapisane v tabeli 1. Ujemanje je skoraj popolno pri prvi številki (slika 1).

Tudi na drugi številki je ujemanje zelo dobro (slika 2).



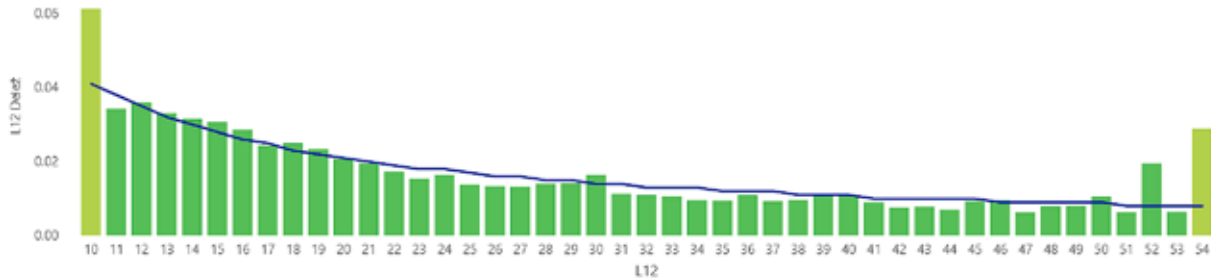
Slika 1: Delež prve vodilne številke pri neto zneskih računov



Slika 2: Delež druge vodilne številke pri neto zneskih računov (izsek)

Tudi kombinacija prve in druge številke kaže skoraj idealno porazdelitev. Na sliki 3, ki prikazuje pogostnost kombinacij prve in druge številke, opazimo

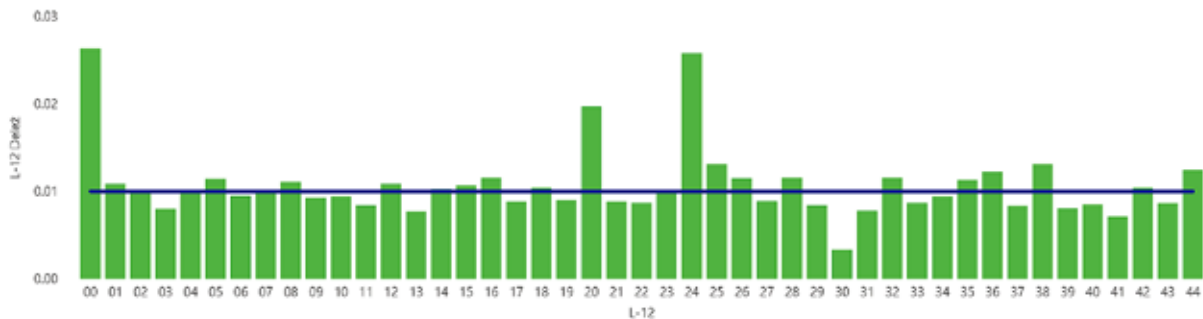
večja odstopanja, vendar ta ne presegajo sprejemljive vrednosti (sprejemljivi Benford).



Slika 3: **Delež kombinacije prve in druge vodilne številke pri neto zneskih računov (izsek)**

Na sliki 4 je porazdelitev zadnjih dveh števk. Za razliko od logaritemske porazdelitve vodilnih števk je tu pričakovana enakomerna porazdelitev. Posa-

mezne vrednosti vidno odstopajo, vendar nobena vrednost SSD ne preseže meja »sprejemljivega Benforda«.



Slika 4: **Delež kombinacije zadnjih dveh števk pri neto zneskih računov (izsek)**

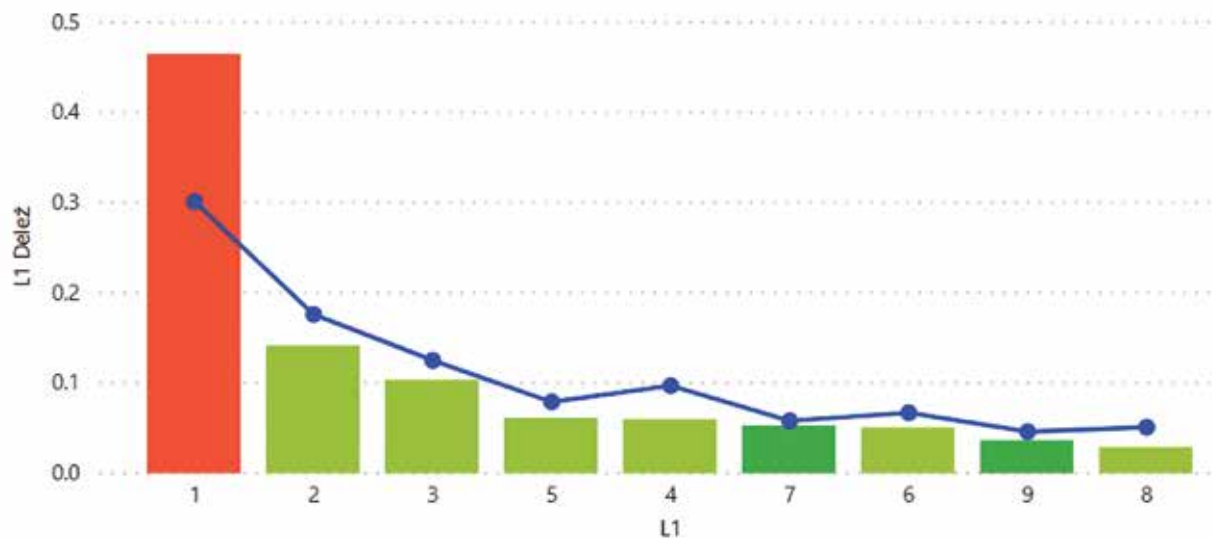
Analiza avansov

Obvladovanje tveganj v poslovanju podjetja zahteva, da kupci iz slabših bonitetnih razredov pred prevzemom blago plačajo vnaprej - avansirajo. V drugi analizi smo zajeli samo skupino kupcev, ki so zavezani k plačilu avansa. Takšni kupci imajo posebno oznako v šifrantu.

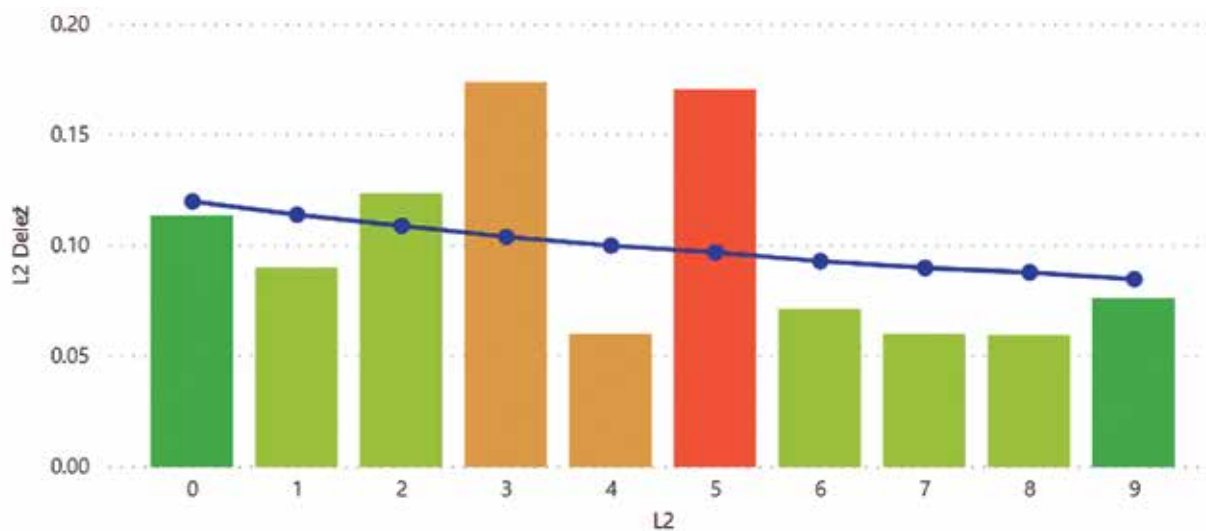
Prva hipoteza: če poslovna politika podjetja zahteva od vseh kupcev enak avans (npr. 50% od zneska računa), bi zaradi pravila invariantnosti dobili Benfordovo porazdelitev kot pri računih.

V drugi hipotezi smo predpostavili, da bi Benfordovo porazdelitev dobili tudi, če bi deleže avansov delili v več razredov (na primer v razrede 50%, 75% in 100% avans).

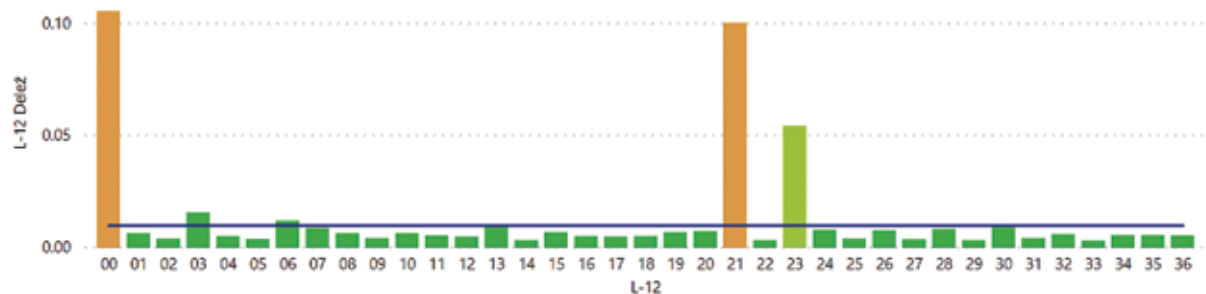
Rezultati, ki so vidni na slikah od 5 in 6, nobene od hipotez ne potrjujejo. Analizo smo predstavili revizorju, ki bo s pregledom dokumentov odgovoril na naslednja vprašanja. Ali je neujemanje z Benfordovo porazdelitvijo posledica kršenja poslovnih pravil? Ali se dogaja to pri vseh kupcih iz slabših bonitetnih razredov, ali je morda kakšen kupec neupravičeno v privilegiranem položaju?



Slika 5: Delež prve vodilne številke pri avansih



Slika 6: Delež druge vodilne številke pri avansih



Slika 6: Delež kombinacije zadnjih dveh števk pri avansih (izsek)

Pri analizi zadnjih dveh števk na sliki 7, pri kombinaciji cifer v dveh primerih opazimo precejšnje odstopanje od Benfordove porazdelitve. Predpostavi-

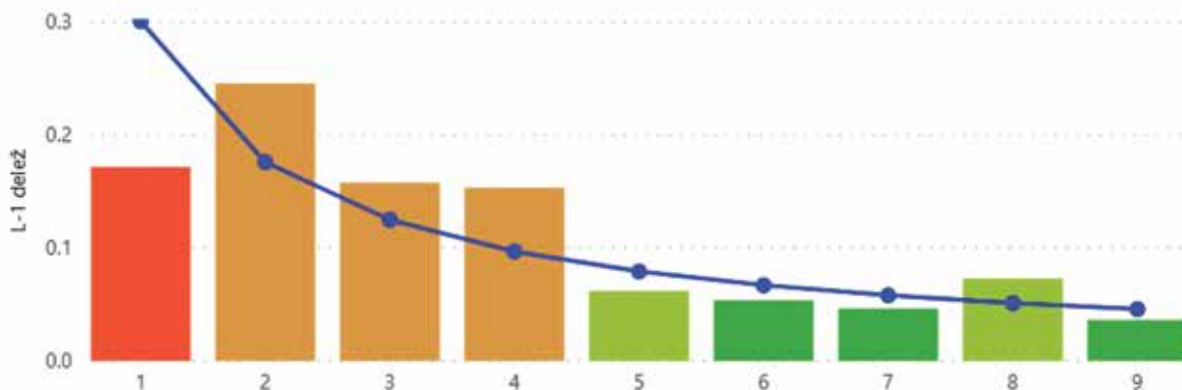
mo lahko, da je razlog odstopanja pri 00 v napačnem zaokroževanju (Carslaw (1988)). Razlog odstopanja pri cifrah 21 ni pojasnjen.

Primer uporabe: analiza testne zbirke podatkov AdventureWorks

Benfordov test lahko uporabimo v različnih zbirkah podatkov. Na spletu najdemo algoritme v več programskih jezikih (npr. https://rosettacode.org/wiki/Benford%27s_law#SQL). Spodnja koda je prirejena za TSQL - jezik za Microsoft SQL strežnik.

```
WITH test AS (
SELECT
  a.digit,
  COUNT(a.digit)*1.0 / b.numbers AS actual,
  CASE WHEN a.digit>0 THEN log(1 + 1 / (a.digit*1.0) , 10 ) END AS expected
FROM
  (SELECT FLOOR(LineTotal/POWER(10,LEN(FLOOR(LineTotal))-1)) AS digit
   FROM [Sales].[SalesOrderDetail]
   WHERE LineTotal>0) a ,
  (SELECT COUNT(*) AS numbers
   FROM [Sales].[SalesOrderDetail]
   WHERE LineTotal>0) b
GROUP BY digit, numbers
)
SELECT *
,SSD=POWER(actual-expected,2)*10000
FROM test
ORDER BY digit
```

Koda 1: Izračun Bedforda po prvi številki



Slika 8: Rezultati Benforda po prvi številki v zbirki AdventureWorks2014

Rezultat poizvedbe, prikazan na sliki 8, potrjuje hipotezo. Rdeče obarvani SSD pokaže, da številka 1 ne sledi Benfordovi porazdelitvi. Številke od 2 do 4 so mejni Benford.

Analizirali smo znano testno podatkovno zbirko AdventureWorks2014, ki je splošno dostopna na spletu. Predpostavili smo, da bomo z Bendfordom dokazali, da ne gre za dejanske podatke. S programom Koda 1 smo analizirali vrednost vrstice LineTotal v tabeli SalesOrderDetail. Program vrne prvo številko (digit), delež ponovitev (actual), pričakovano vrednost (expected) in kvadrat razlike obeh (SSD).

S Kodo 2 smo analizirali še prvi dve številki. Zaradi nekaj vrednosti manjših od deset, smo vse vrednosti pomnožili s sto, kar ne vpliva na rezultat analize [9].

```

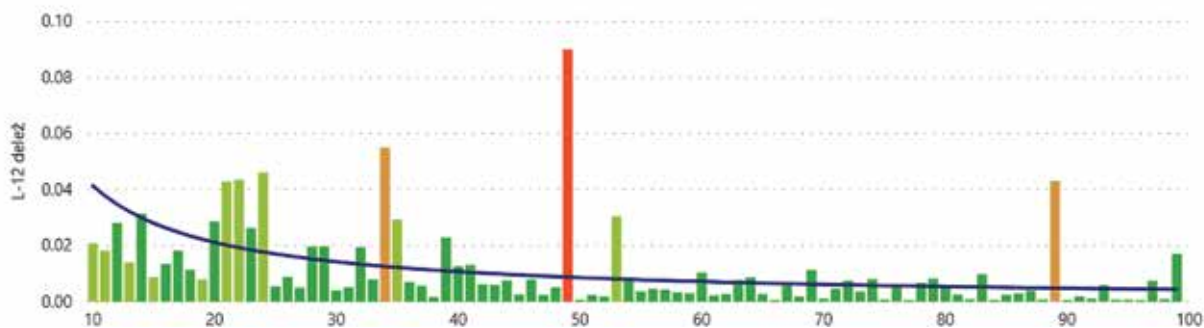
WITH test AS (
SELECT
  a.digit,
  COUNT(a.digit)*1.0 / b.numbers AS actual,
  CASE WHEN a.digit>0 THEN log(1 + 1 / (a.digit*1.0) , 10 ) END AS expected
FROM
  (SELECT FLOOR(LineTotal*100/POWER(10,LEN(FLOOR(LineTotal*100))-2)) AS digit
   FROM [Sales].[SalesOrderDetail]
   WHERE LineTotal>0) a ,
  (SELECT COUNT(*) AS numbers
   FROM [Sales].[SalesOrderDetail]
   WHERE LineTotal>0) b
GROUP BY digit, numbers
)
SELECT *
,SSD=POWER(actual-expected,2)*10000
FROM test

```

Koda 2: Izračun Bedforda po prvi in drugi številki

Rezultat analize na sliki 8, po prvih dveh števkih, nas je presenetil. Na osnovi analize po prvi številki smo pričakovali največje odstopanje (SSD) pri vodilni številki 1. Največja napaka pa je pri števkih 49, čeprav tudi številke med 10 in 19 ne sledijo Benfordu.

Analiza po dveh števkih torej ni le potrditev analize po prvi številki. Primer ponazarja, da je analizo podatkov treba narediti po več števkih. Skupna interpretacija več kazalcev nam šele poda celovito sliko.



Slika 8: Rezultati analize v zbirki AdventureWorks2017 po prvih dveh števkih

Analizo smo nadaljevali za prve tri števke in ugotovili največjo napako pri števkih 499 in 349. Podatke z največjo napako smo izolirali in našli približno deset tisoč zapisov s količino 1 in ceno 4,99 in 49,9. Analizirali smo tudi zadnji dve števki in dobili največja odstopanja pri 99 in 98. Odstopanja bi lahko pojasnili z zakrožitvijo cen na 99 centov in dejstvom, da ima 61% pozicij količino enako ena. Na osnovi opravljene analize trdimo, da ne gre za dejansko zbirko podatkov.

ZAKLJUČEK

Pri iskanju prevar so sicer zelo popularne metode strojnega učenja. S prevarami se lahko uspešno spopademo tudi s preprostejšim matematičnim orodjem. Benfordov empirični zakon porazdelitve števk je uspešen kazalnik nepravilnosti v zbirkah podat-

kov, kar so dokazali številni avtorji in tudi naši primeri. Uporaba zakona ni omejena samo na finančna poročila. Med članki smo zasledili tudi rabo zakona za ločevanje resničnih fotografij od generiranih slik [12]. Avtorji so v zadnjem času z Benfordom preverjali COVID-19 poročila [13] iz Kitajske in vsaj tako potrdili njihovo verodostojnost.

Z analitskim orodjem, kot je Power BI, ali s procedurami SQL je mogoče izdelati računalniškega pomočnika, ki je v rokah revizorjev močno orožje za odkrivanje nepravilnosti v podatkovnih zbirkah. Analiza podatkov z Benfordom ne da dokončne sodbe, ali je podatkovni vir potvorjen. Za dokončno sodbo je treba poznati vsebino podatkov. So pa analize odlični pomočnik, ki usmerja revizorjevo iskanje in tako skrajša čas revizije.

VIRI IN LITERATURA

- [1] Amiram, D., Bozanic, Z., Rouen, E. (2015). Financial statement errors: Evidence from the distributional properties of financial statement numbers- Review of accounting studies, Springer
- [2] Benford, F. (1938). The law of anomalous numbers. Proc. Am. Philos. Soc. 78 (4): 551–572, JSTOR 984802. Pridobljeno 15. 4. 2020, s
- [3] Carslaw, C. (1988). Anomalies in Income Numbers: Evidence of Goal Oriented Behavior, The Accounting Review (1988), 321–327.
- [4] Holz, C. A. (2014). The quality of China's GDP statistics – China Economic Review, Elsevier
- [5] Horan, T. J. (2018). 5 Keys to Using AI and Machine Learning in Fraud Detectionhttps. Pridobljeno 6. 2. 2020, <https://www.fico.com/blogs/5-keys-using-ai-and-machine-learning-fraud-detection>
- [6] Kossovsky, A. E. (2014). Benford's Law: Theory, The General Law Of Relative Quantities, And Forensic Fraud Detection Applications, World Scientific, New Jersey.
- [7] Newcomb, S. (1881). Note on the frequency of use of the different digits in natural numbers. American Journal of Mathematics. 4 (1/4), 39–40. . JSTOR 2369148. Pridobljeno 6.2.2020, s: https://www.jstor.org/stable/2369148?seq=1#metadata_info_tab_contents (potrebna registracija)
- [8] Nigrini, M. (2012). Benford's Law : Applications for Forensic Accounting, Auditing, and Fraud Detection, Wiley Corporate F&A
- [9] Povrženič, Š. (2019). Analiza Benfordovega zakona, magistrsko delo, Univerza v Ljubljani, Fakulteta za matematiko in fiziko.
- [10] Jeffrey Irwin, (2015) The Effective Use of Benford's Law to Assist in Detecting Fraud in U.S. Environmental Protection Agency (EPA) Toxics Release Inventory (TRI) Data.
- [11] Heng Qu, Richard Steinberg, Ronelle Burger, (2002), Abiding by the Law? Using Benford's Law to Examine the Accuracy of Nonprofit Financial Reports, <https://journals.sagepub.com/doi/abs/10.1177/0899764019881510?journalCode=nvsb>
- [21] Nicolò Bonettini, Paolo Bestagini, Simone Milani, Stefano Tubaro, (2020) On the use of Benford's law to detect GAN-generated images , Cornell University, <https://arxiv.org/abs/2004.07682>
- [13]]Koch, Christoffer, Okamura, Ken, Benford's Law and COVID-19 Reporting (April 28, 2020). Available at SSRN: <https://ssrn.com/abstract=3586413> or <http://dx.doi.org/10.2139/ssrn.3586413>
- [41] Aamo Iorliam, Anthony T.S. Ho, Norman Poh, Santosh Tirunagari, Patrick Bours; (2015), 3rd International Workshop on Biometrics and Forensics (IWBF 2015), Data forensic techniques using Benford's law and zipf's law for keystroke dynamics
- [15] Aldwin T. Miranda, World's Distribution of Covid-19 Cases and The Benford's Law, (2020), Cape Comorin, An International Multidisciplinary Double-Blind Peer-reviewed Research JournalSpecial Issue, Volume II Issue V June 2020, ISSN: 2582–1962
- [16] Lucas Silva, Dalson Figueiredo Filho, (2020), Using Benford's law to assess the quality of COVID-19 registerdata in Brazil, Journal of Public Health | pp. 1–4 | doi:10.1093/pubmed/fdaa193

■

Tomaž Dular je po izobrazbi elektro inženir in magister znanosti, s področja računalništva in informatike. Ukvarja se s poslovno analitiko, analizo podatkov, umetno inteligenco, bazami podatkov, poslovnimi informacijskimi sistemi in podatkovnimi integracijami. Deloval je kot inovator v svobodnem poklicu, zasebni raziskovalec. Trenutno deluje kot arhitekt sistema v podjetju MIT informatika in specialist za podatkovno integracijo ter analitik v B2-BI.

■

Julija Lapuh Bele je diplomantka uporabne matematike, magistrica računalništva in informatike ter doktorica znanosti. V zadnjih letih je njeno raziskovalno področje informatika in poslovnih finančah in poslovna analitika, pred tem pa je raziskovala področje uporabnosti programske opreme in sodelovala v razvoju metod in sistema za e-izobraževanje, kjer se je njeno delo navezovalo na implementacijo poslovne analitike v LMS sistem, izboljševanje uporabniške izkušnje in razvoj interaktivnih e-gradiv za področje poslovne matematike ter računalništva in informatike.

■

Rok Pirnat, magister poslovnih ved vodi podjetje B2 BI d.o.o., katerega osnovna dejavnost je izvedba projektov s področja poslovne analitike in raziskovalno razvojna dejavnost. Ukvarja se z razvojem analitičnih modelov v vseh segmentih poslovanja, s čimer pomaga podjetjem, da se razvijajo v pametne organizacije. Rok je predavatelj na številnih poslovnih konferencah in tudi višji predavatelj na visoki šoli.

■

Špela Povrženič je po izobrazbi magistrica finančne matematike. Je poslovna analitičarka. Skozi analitiko podjetjem odkriva vrline, pomanjkljivosti in priložnosti za izboljšave, s tem pa jim pomaga pri rasti in razvoju. Specializirana je na področju analiziranja financ, terjatev in obveznosti. Pohvali se lahko tudi s certifikatom MCSA: BI Reporting, ki potrjuje njeno znanje in izkušnje na področju analiziranja in vizualiziranja podatkov z Excelom in Power BI.

█ Sodobni izzivi izobraževanja na daljavo v okviru medpredmetnega povezovanja in inovativnih učnih okolij

Simona Samida Cerk
OŠ Franceta Bevka Ljubljana, Pohorskega bataljona 1, SI-1113 Ljubljana
simonasamida@gmail.com

Izveleček

Prispevek predstavlja sodobne izzive izobraževanja na daljavo, ki temeljijo na medpredmetnem povezovanju in inovativnih učnih okoljih. V središče postavljamo učenca, ki samostojno razmišlja in raziskuje ter v okviru aktivnih metod znanje povezuje, ga osmišlja in pogloblja.

Medpredmetni pristop je pomemben element sodobnih pogledov na vzgojno-izobraževalni proces, saj predstavlja poglobljeno in sistematično načrtovanje učenja, s katerimi povezujemo vsebine, procese in konceptualno znanje. Pomeni horizontalno in vertikalno povezovanje vsebin, znanj in učnih spretnosti, s katerim vsebino obravnavamo celostno, osvetljeno z različnih vidikov in z različnih predmetnih področij. Še posebej je medpredmetni pristop pomemben pri pouku na daljavo, kjer je potrebno učne vsebine racionalizirati, učence pa motivirati za samostojno in kritično delo.

Inovativna učna okolja podprta z informacijsko-komunikacijsko tehnologijo predstavljajo kontekst, znotraj katerega se odvija učenje v skladu s sodobnimi smernicami didaktike, ki omogočajo celostno učenje, individualizacijo in personalizacijo učnega procesa. Pri izobraževanju na daljavo je inovativno okolje tisto, ki se je sposobno odzvati na razvijanje in spreminjanje izobraževalne prakse z nenehnim razvojem in prilagajanjem.

Ključne besede: informacijsko-komunikacijska tehnologija, inovativno učno okolje, izobraževanje na daljavo, medpredmetni pristop, šola 21. stoletja

Abstract

The paper presents the modern challenges of distance education based on interdisciplinary connections and innovative learning environments. We focus on the student, who thinks and researches independently and, within the framework of active methods, interconnects knowledge, giving it meaning and depth.

The interdisciplinary approach is an important element of modern views on the educational process as it represents the in-depth and systematic planning of learning which is used to connect content, processes, and conceptual knowledge. It enforces the horizontal and vertical integration of content, knowledge and learning skills that are used to handle content holistically, exposed from different aspects and from different subject areas. The interdisciplinary approach is particularly important in distance learning where it is necessary to rationalize the learning content and motivate students for independent and critical work.

Innovative learning environments supported by information and communication technology represent the context within which learning takes place in line with modern guidelines on didactics. It enables integrated learning, individualization and personalization of the learning process. In distance education, an innovative environment is one that can respond to the development and changes in educational practice through continuous development and adaptation.

Keywords: Information and communication technology, innovative learning environment, distance learning, interdisciplinary approach, 21st century school

1 UVOD

Izobraževanje na daljavo je zaradi pandemije postalo povsem nov in globalen pojav svetovne razsežnosti, ki zahteva nujno dostopnost do računalnika in interneta. Vendar pa materialni pogoji, čeprav nujni, niso zadostni. Pri tem je pomembna računalniška in digitalna pismenost tako učiteljev kot učencev. Ko sta zagotovljena ta dva pogoja, je pričakovati, da bo tovrsten pouk imel rezultate (Kodelja, 2020).

Šola 21. stoletja v ospredje postavlja kompetenco kritičnega mišljenja, reševanja problemov, sodelovanja in ustvarjalnosti, ki je podprto z digitalno pismenostjo. Vse bolj postajajo pomembni višji spoznavni procesi, s katerimi naučeno in osmišljeno znanje uporabljamo v različnih situacijah na prožen in ustvarjalen način.

Medpredmetno povezovanje predstavlja pomemben del učenja in poučevanja, saj stremi k povezovanju znanj in doseganju ciljev znotraj enega predmeta in med različnimi predmeti. Učenci pri medpredmetnem pristopu razvijajo interes in motivacijo za učenje ter dosegajo višji učni uspeh. Pri izobraževanju na daljavo se medpredmetni pristop še posebej obrestuje, saj učenje po učnih in tematskih sklopih omogoča racionalizacijo učnega procesa ter sledenje in spremljanje napredka.

Inovativna učna okolja podprta z informacijsko-komunikacijsko tehnologijo so ključna za kakovostno izvedbo pouka na daljavo. Pri tem je na eni strani pomembna računalniška in digitalna pismenost učitelja, ki v digitalno podprtem učnem okolju pri svojem delu preišljeno uporablja sodobne in inovativne didaktične pristope, na drugi strani pa digitalna kompetentnost učenca, da zna in zmore digitalna orodja uporabljati in mu le-ta pomagajo pri doseganju višjih oblik učenja in znanja.

2 ŠOLA 21. STOLETJA

V zadnjem času se postavlja vprašanje, kaj moramo storiti, da bomo ustvarili sodobno šolo, šolo, ki bo učence pripravila na hitro spreminjajočo se družbo in sledila tehnološkim inovacijam ter napredku. V ospredje se postavlja učenca, ki ga z inovativnimi pristopi pripeljemo k samostojnemu razmišljanju in raziskovanju. Pouk v sodobni šoli temelji na celostnem učenju, bogatem učnem okolju, samoregulaciji in samorefleksiji, aktivnih oblikah ter metodah dela, podprt s sodobno informacijsko-komunikacijsko tehnologijo. Nove metode, oblike učenja in poučevanja so za doseganje ciljev ključne in neizbežne.

Za zagotavljanje inovativnosti pri ustvarjanju učnih okolij 21. stoletja je potrebno sodelovati in komunicirati. Vsi deležniki izobraževanja bi se morali združiti, da bi se gonilo inovativnosti odražalo v celotnem izobraževalnem sistemu in ne samo na posameznih področjih. Ustvarjati bi bilo potrebno pogoje, ki spodbujajo inovativnost ter vzpostaviti procese samoevalvacije, ki bi preko smiselnih raziskav, vključevala svet politike in prakso. Z zagotavljanjem skladnosti splošnih izobraževalnih strategij, ki so osredotočene na učenje in povezujejo partnerstva različnih izobraževalnih organov, bi ustvarjali odzivne učne sisteme 21. stoletja (Schleicher, 2019).

Pri poučevanju in učenju v sodobni šoli se poudarja interaktivni vidik. Učitelji naj bi imeli vpogled v to, kako učenci napredujejo, kakšne težave imajo z učenjem, in temu prilagajali učni proces, ki bi zagotavljal zadovoljevanje različnih potreb. To lahko naredi s pomočjo opazovanj, pogovorov in izkazanim delom v najrazličnejših izvedbah (Rutar Ilc, 2014).

2.1 Novi pristopi pri izobraževanju

Šola 21. stoletja upošteva sodobne koncepte znanja, ki poleg vsebinskega znanja v ospredje postavljajo spretnosti in veščine, kot so sposobnost samostojnega iskanja informacij, sposobnost kritičnega presojanja informacij, sposobnost kritičnega mišljenja, sposobnost uporabe informacij v različnih situacijah, sposobnost delovanja v skupini oziroma timu za doseganje skupnih ciljev, sposobnost uporabe informacijsko-komunikacijske tehnologije, ustvarjalnost, inovativnost itd. (Dumont, 2013).

Za uresničevanje sodobnih konceptov znanja po načelih sodobne šole je potrebno zagotoviti vzgojno-izobraževalni načrt, ki ima značilnosti:

- **Usmerjenost na učenca:** učno okolje v ospredje postavlja učenje kot primarno dejavnost in ne predstavlja alternative ključni vlogi učitelja in strokovnjakov za učenje, temveč je odvisno od njih.
- **Strukturiranost in dobro načrtovanje:** načrtovanje je preudarno in zahteva visok nivo profesionalizma, hkrati pa pušča dovolj prostora za raziskovanje in avtonomno učenje.
- **Popolna personalizacija:** učno okolje je občutljivo za individualne in skupinske razlike iz katerih izhajajo učenci glede na njihovo predznanje, motivacijo in sposobnosti. Podana je natančna povratna informacija, ki je personalizirana.

- **Inkluzivnost:** občutljivost za razlike med posamezniki in skupinami, vključuje najšibkejše učence, opredeljuje inkluziven izobraževalni načrt.
- **Socialnost učenja:** učenje poteka v sodelovalnem vzdušju, kar pomeni, da je sodelovanje eksplicitni del učnega okolja in učenje povezano s skupnostjo (Dumont, 2013)

2.2 Sodobne učne metode in oblike dela

Pouk je učinkovit takrat, kadar so učne metode in oblike takšne, da pri učencih spodbujajo miselno aktivnost, samostojno učenje in vključujejo informacijsko-komunikacijsko tehnologijo. Učitelj s pomočjo učnih metod in oblik dela povezuje učne cilje, vsebine, postopke načrtovanja, izvajanja in vrednotenje učnega procesa, ob tem pa upošteva razvojne in individualne značilnosti vsakega učenca. V učnem procesu je zagotovljena konstruktivna povratna informacija, znanje pa je preverjeno in ocenjeno na strokovno ustrezen način (Marentič Požarnik, 2007).

Sodelovalno učenje je v sodobni šoli nadgradnja skupinske učne oblike pri katerem s strukturnim pristopom zagotavljamo uresničevanje osnovnih načel sodelovalnega učenja, kot so: soodvisnost učencev, njihova medsebojna interakcija, heterogenost skupine, vključenost in odgovornost vseh članov skupine ter uporaba ustreznih sodelovalnih veščin (Johnson in Johnson, 1999; Peklaj, 2001)

O sodelovalnem učenju govorimo takrat, ko pouk poteka v manjših skupinah, z namenom doseči skupni cilj. Delo je organizirano tako, da vsak član doseže največji učni učinek in hkrati pomaga drugim, da tudi oni dosežejo kar največ. Osrednje mesto ima interakcija v skupini. Učitelj vodi in usmerja celoten proces, nudi pa tudi pomoč, če jo skupine potrebujejo (Marentič Požarnik, 2010).

Sodelovalno učenje je v primerjavi z individualnim tako iz kognitivnega, socialnega kot tudi iz motivacijskega vidika, empirično dokazano bolj učinkovita oblika učenja kot pa zgolj individualno učenje (Peklaj, 2001).

3 MEDPREDMETNI DIDAKTIČNI PRISTOP

Medpredmetno povezovanje v sodobni in inovativni šoli predstavlja pomemben del učenja in poučevanja, saj je to celosten didaktični pristop, kjer se povezujejo znanja, vsebine in učne spretnosti. Učenci pri medpredmetnem povezovanju razvijajo interes in motivacijo za učenje. Pri tem dosegajo visok učni uspeh

in kažejo večjo radovednost za usvajanje interdisciplinarnih znanj.

Medpredmetno povezovanje je celosten didaktični pristop s katerim horizontalno in vertikalno povežemo znanja, vsebine in učne spretnosti. Spodbuja samostojno in aktivno pridobivanje učnih izkušenj, in poteka v celoviti dejavnosti učenca, vključuje njegove spoznavne, čustvene in telesne funkcije. Do medpredmetnega povezovanja pristopamo na dva načina, in sicer učno-snovno (osredotočanje na učno snov) in procesno-ciljno (osredotočanje na proces učenja). Pri procesnem vidiku medpredmetnega povezovanja se uveljavlja transfer miselnih strategij, kar omogoča aktivno reševanje problemov in uporabo višjih miselnih spretnosti. (Sicherl-Kafol, 2008)

Medpredmetni pristop za učence prinaša mnoge prednosti, kot so celostno dojetje sveta, izkušnjsko učenje, prenašanje usvojenega znanja in procesov na druga področja, individualizacija in diferenciacija pouka. Tovrstno poučevanje in učenje presega okvirje posameznega učnega predmeta, saj omogoča povezovanje in združevanje različnih učnih ciljev v smiselno celoto. Tako celostno učenje zagotavlja povezavo in poenotenje znanja, učence globoko motivira k drugačnemu zaznavanju ustvarjalnemu povezovanju vsebin in oblikovanju novih in drugačnih miselnih modelov, struktur in sistemov (Krnjel, Hodnik Čadež ind., 2008).

Medpredmetno povezovanje upošteva učencevo predznanje, stališča, pričakovanja, čustva in socialno komponento. Uresničuje načelo celovitosti učnega procesa, povezuje učenje z življenjem ob spodbujanju izkušnjskega učenja pri učencih, priznava otrokovo individualnost z upoštevanjem njegovih potreb, interesov, zmogljivosti. Je odprt učni model, ki omogoča učitelju in učencem fleksibilno prilagajanje sprotnim potrebam in situacijam. Od učitelja takšen pristop dela zahteva prožnost, širino v znanju, empatijo in poslušnost za potrebe, interese in sposobnosti učencev (Kuščer, 2000).

▪ Prednosti medpredmetnega povezovanja

Zagovorniki medpredmetnega povezovanja se naslanjajo na dognanja o načinu delovanja človeških možganov in na osnove holizma.

Kadar vsebino obravnavamo bolj celostno in z različnih vidikov, stvarnost doživljamo in zaznavamo kot celoto, ne pa strukturirano po posameznih področjih. Učenci se z medpredmetnim pristopom

naučijo reševanja problemov v realnih situacijah, s kakršnimi se srečujemo skozi vse življenje. Prav s tega vidika se nekatere ključne kompetence, ki jih potrebujemo za življenje in vseživljenjsko učenje, po svoji naravi interdisciplinarne in jih je mogoče zagotavljati prav s kroskurikularnostjo. Medpredmetni pristop tako doda vrednost običajnemu pristopu, saj pouk postane problemski in razvija kritično mišljenje (Pavlič Škerjanc, 2010).

Medpredmetnega pristopa poučevanja se poslužujemo takrat, ko je to smiselno, ko je učna snov primerna in imamo zadovoljive možnosti za izvedbo. Možnosti medpredmetnih povezav zasledimo v učnih načrtih posameznih predmetov. Iz njih lahko razberemo, kateri predmeti se najbolje vežejo med seboj in katere vsebine lahko povežemo, da uresničimo cilje iz učnih načrtov posameznih predmetov (Štemberger, 2008).

▪ **Medpredmetni pristop pri izobraževanju na daljavo**

Medpredmetno povezovanje v okviru izobraževanja na daljavo omogoča racionalizacijo učnih vsebin in osredotočanje na temeljne cilje posameznih učnih predmetov. Učenci pri pouku na daljavo za usvajanje in razumevanje znanja porabijo tudi do trikrat več časa, zato je medpredmetni pristop z ekonomičnega vidika obravnavanja učnih vsebin toliko bolj pomemben. Pri tem je potrebno izpostaviti, da kljub racionalizaciji učnih vsebin le-te obravnavamo celostno in poglobljeno, pri učencih pa dosegamo višjo učno motivacijo in interes za učenje.

Medpredmetni učni pristop je lažje uresničevati od prvega do petega razreda osnovne šole, ko večino učnih predmetov poučuje en učitelj. V šestem razredu in v tretji triadi je za medpredmetni pristop potrebno več skupnega usklajevanja in načrtovanja med posameznimi učitelji.

Pri pouku na daljavo lahko medpredmetno povezovanje izpeljemo pri vseh dnevih dejavnosti, ki so del obveznega programa v osnovni šoli pri katerih je še posebej poudarjeno, da se medpredmetno povezujejo discipline in predmetna področja. Cilj dni dejavnosti je učencem omogočiti utrjevanje in povezovanje znanja pridobljenega pri posameznih predmetih in predmetnih področjih, uporabljanje tega znanja in njegovo nadgrajevanje s praktičnim učenjem v kontekstu medsebojnega sodelovanja in odzivanja na aktualne dogodke v ožjem in širšem družbenem okolju.

Možnosti medpredmetnih povezav je v prvem in drugem triletju zelo veliko, pri tem je pomembno, da povezave učnih vsebin ne nastajajo na silo. Potrebno je upoštevati tudi posebnosti specialno didaktičnega načrtovanja posameznih predmetov, saj lahko v nasprotnem primeru siromašimo proces učenja. Medpredmetno povezovanje bo kakovostno takrat, ko bo smiselno vključevalo vsebine, procese, znanja, spretnosti itd., ki spodbujajo učenje iz različnih zornih kotov posameznih predmetnih področij. Na podlagi tega nastane učni transfer znanj in spretnosti, ki omogočajo razvijanje sistemskega mišljenja in mrežo znanja za ustvarjalno reševanje učnih in življenjskih okoliščin (Sicherl-Kafol, 2008).

▪ **Možnosti medpredmetnega povezovanja pri pouku na daljavo**

Če želimo medpredmetne povezave ustrezno izvajati, je potrebno narediti načrt, v katerem jasno opredelimo cilje povezovanja predmetov ter izdelamo strategije za njihovo doseganje.

Pri medpredmetnem povezovanju govorimo o treh uveljavljenih pristopih. Pri tradicionalnem pristopu sta znanje in razumevanje minimalno povezana tako znotraj predmetov kot med predmeti. Učne poti takega tovrstnega pristopa so linearne in divergentne. Obogateni tradicionalni pristop temelji na vsebinskem oziroma tematskem pristopu povezovanja predmetov, znanje in razumevanje sta povezana znotraj predmetov, učne poti se v okviru učnega procesa združujejo in povezujejo. Integrativni kurikulum pa temelji na konceptualnem pristopu. Učni proces usmerja skupni problem oziroma iskanje odgovora na isto problemsko vprašanje. Prav tako sta znotraj in med predmeti povezana znanje in razumevanje, poučevanje in učenje sta sodelovalna in potekata po krožnih in vijačnih konvergentnih učnih poteh (Pavlič Škerjanc, 2010).

Vzemimo primer medpredmetnega povezovanja družbe, slovenščine, likovne umetnosti, glasbene umetnosti, matematike in gospodinjstva v 5. razredu.

Z dnevi dejavnosti, ki so del obveznega programa, medpredmetno povežemo discipline in medpredmetna področja, ki so vključena v predmetnik osnovne šole. Dneve dejavnosti lahko izvajamo tudi na daljavo, pri čemer izhajamo iz ciljev, ki učencem omogočajo utrjevanje in povezovanje znanja pridobljenega pri posameznih učnih predmetih in predmetnih področjih, uporabljanje tega znanja in nadgra-

Tabela 1: Možnosti medpredmetnega povezovanja

Družba	Slovenščina	Likovna umetnost	Glasbena umetnost	Matematika	Gospodinjstvo
Ljudje v prostoru: spoznavanje naravno-geografskih enot Slovenije, geografskih pojmov, orientacija na zemljevidu	Obravnavava umetnostnih in neumetnostnih besedil (razglednica, recept, književna besedila), spoznavanje narečij	Risanje, slikanje, grafika, kiparstvo, prostorsko oblikovanje (izdelava razglednice, spominka, pustna maska, risanje idrijske čipke)	Spoznavanje s plesov in pesmi slovenskih pokrajin	Računanje zračne razdalje med kraji, računanje relativne višine, zbiranje in obdelava podatkov	Priprava jedi značilnih za posamezno pokrajino, spoznavanje, narodnih noš Slovenije
Ljudje v času: koliščarji, Rimljani, srednjeveška mesta	Obravnavava umetnostnih in neumetnostnih besedil (Deček Brin na domačem kolišču, Bobri)	Izdelava mozaika, ščita legionarja, kolišč	Ustvarjanje ritmov in glasbe s kamni (kamena doba)	Zbiranje in obdelava podatkov, orientacija na časovnem traku zgodovine	Priprava jedi starodavnih ljudstev, spoznavanje načinov tkanja koliščarjev

dnjo s praktičnim učenjem v kontekstu medsebojnega sodelovanja in odzivanja na aktualne dogodke v širšem in ožjem družbenem okolju.

V okviru dni dejavnosti lahko izvedemo medpredmetni dan potep po Sloveniji, kjer učenci preko interaktivnega gradiva oblikovanega v Google slides in Google forms in dela po skupinah preko konferenčnega sistema Zoom sodelujejo v skupinah, odkrivajo naravno-geografske in kulturne značilnosti Slovenije, se učijo slovenskih ljudskih pesmi, ustvarjajo motive slovenskega ljudskega izročila, spoznavajo tradicionalne slovenske jedi, zapisujejo recepte, pretvarjajo količine, računajo razdalje med kraji in se urijo v orientaciji na interaktivnem zemljevidu.

4 INOVATIVNA UČNA OKOLJA

Inovativna učna okolja podprta z informacijsko-komunikacijsko tehnologijo so v središču politik in raziskav na področju sodobnega izobraževanja. Vičič Krabonja in Šverc (v Flogie in Aberšek, 2019) navedata, da se je v šolskem prostoru zaradi hitih sprememb v družbi ter globalizacije na vseh ravneh večina držav članic EU za eno izmed prednostnih nalog na področju izobraževanja postavila doseganje višjih taksonomskih stopenj znanja, doseganje digitalnih kompetenc in kompetenc 21. stoletja. Za ustvarjanje spodbudnih učnih okolij izpostavljata, da potrebujemo multidisciplinarni pristop, ki predstavlja enega izmed pomembnih izzivov in je povezan z vlogo sodobne informacijske tehnologije pri ustvarjanju inovativnih učnih okolij in učnih priložnosti za učence.

Uvajanje prožnih oblik učenja in inovativnih učnih okolij podprtih z informacijsko-komunikacijsko tehnologijo podpirajo razvijanje ključnih kompetenc,

in sicer učenje, poučevanje, vrednotenje, vsebina, kurikulum, organizacija, vodenje in vrednote. Tehnologija pa predstavlja temelj vseh pedagoških sprememb v sklopu inovativnih učnih okolij. Za doseganje vzgojno-izobraževalnih ciljev in kompetenc v sodobni šoli, potrebujemo tudi ustrezen prostor, dostop do izobraževalnih vsebin ter ustrezno didaktično usposobljenost učiteljev. Izpostavljamo celostni pristop v izobraževanju, ki v veliki meri korelira z vzpostavljanjem ustvarjalnih in inovativnih učnih okolij (Flogie in Aberšek, 2019).

Kompetenten učitelj je z uporabo informacijsko-komunikacijske tehnologije uspešnejši, ustvarjalnejši in inovativnejši, saj uporablja veljavne in zanesljive podatke ter se zaveda pravnih in etičnih načel varne in odgovorne uporabe tehnologije (Gerlič, 2000). Različne študije uporabe informacijsko komunikacijske tehnologije v izobraževanju kažejo, da informacijsko-komunikacijska tehnologija olajša učenje otrokom z drugačnimi sposobnostmi in načini učenja, omogoča učinkovitejše učenje in poučevanje z vključevanjem več čutov ter daje možnost boljšega prilagajanja in sledenje pouku (Brečko in Vehovar, 2008).

Orodja za učenje in poučevanje na daljavo

Vzpostavitev ustreznega digitalno podprtega učnega okolja je postala ključna za kakovostno izvedbo pouka na daljavo. Zelo pogosto uporabljeno orodje je spletna učilnica. Moodle je eden izmed programskih paketov za izvajanje e-izobraževanja. Učitelj lahko v spletni učilnici objavlja gradiva, daje domače naloge, pripenja videoposnetke z razlagami, objavlja fotografije izdelkov, s svojimi učenci klepeta v klepetalnici. Učenci v spletni učilnici odpirajo dokumente, ki jih je

pripel učitelj, oddajajo domače naloge, klepetajo v klepetalnici, rešujejo kvize. Učitelj lahko v spletni učilnici spremlja in pregleduje aktivnost učencev, pregleduje domače naloge in daje povratne informacije učencem, spremlja uspešnost reševanja kvizov itd.

Delo v spletni učilnici je lahko zasnovano medpredmetno in v obliki sodelovalnega učenja. Učencem odpremo predmet z naslovom Rimljani. Temo medpredmetno povezujemo, in sicer z matematiko (spoznavanje rimskih števil), slovenščino (neumetnostno besedilo recepti za pripravo rimskih jedi), likovno umetnostjo (izdelava rimskega mozaika), gospodinjstvom (oblačila Rimljanov, priprava jedi). V okviru medpredmetnega povezovanja tako pripravimo mali razredni projekt na temo Rimljanov.

Za spoznavanje te teme uporabimo različna orodja, ki jih nudi spletna učilnica, in sicer spletna povezava do videoposnetkov, kviz, Wordove dokumente, datoteke v PowerPointu. Učence razdelimo v skupine in te opravljajo različne naloge podane v spletni učilnici. Učenci se po skupinah srečujejo v videokonferenčnem sistemu Jitsi meet, ki ga omogoča spletna učilnica. Ko učenci spoznajo in rešijo vse naloge podane v navodilih za delo se preizkusijo v kvizu. Učenci v galerijo pripnejo fotografije likovnih izdelkov, ki so nastale v okviru medpredmetnega povezovanja.

Videokonferenčne storitve so namenjene organizaciji sestankov, predavanj, pa tudi za vzpostavljanje stikov z učenci ali za komunikacijo in sodelovanje med samimi učenci.

Konferenčni sistem Zoom nam omogoča tudi delo po skupinah, in sicer tako, da učence razdelimo po

sobah in tam opravljajo različne naloge. Učencem v orodju Google slides in Google forms pripravimo sobo pobega, ki vsebuje namige iz različnih predmetnih področij. Učenci po skupinah rešujejo izzive in s svojim znanjem ter sodelovanjem poskušajo rešiti uganko oz. geslo, ki jih vodi do končne rešitve. Ko uspešno rešijo uganko se vrnejo v primarno Zoom sobo.

Orodje Kahoot! je namenjeno izdelavi kvizov, razprav, vprašalnikov in je ena izmed atraktivnejših aplikacij za ponavljanje in utrjevanje znanja. Kahoot! omogoča sestavljanje lastnih kvizov ali uporabo že obstoječih, ki jih ustvarjalci delijo z drugimi uporabniki portala. Kviz je mogoče reševati individualno, ali pa v obliki razrednega tekmovanja. Časovna omejitev reševanja posameznega vprašanja, točkovanje in privlačna glasba pripomorejo k napetosti med reševanjem kviza in dvigajo učno motivacijo.

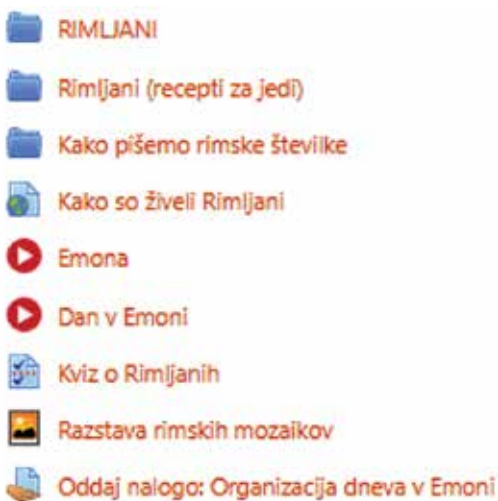
Padlet je storitev, ki omogoča skupno ustvarjanje in rabo. To je tabla, ki jo kreira učitelj, povezavo pa pošlje učencem, ki nato hkrati soustvarjajo – pišejo, delijo fotografije, posnetke in komentarje z drugimi. Vse kar ustvarimo, lahko vidijo vsi sodelujoči, tabla pa se lahko kadarkoli dopolnjuje. Pri pouku na daljavo jo lahko uporabimo za objavljane likovnih izdelkov, videoposnetkov, podajanju navodil za samostojno in skupinsko učenje.

Worldwall je spletno orodje, ki se lahko uporablja za ustvarjanje interaktivnih dejavnosti in dejavnosti za tiskanje. Orodje vsebuje predloge v katerih oblikujemo svoje interaktivne kvize, in križanke, lahko pa uporabimo različico iger kot sta Maze Chase ali Airplane. Spletno orodje pri pouku na daljavo uporabimo za utrjevanje učne snovi na interaktiven in privlačen način.

Učenci lahko pri pouku na daljavo za spoznavanje naravno-geografskih in kulturnih znamenitosti uporabijo mobilno aplikacijo Nexto. Ta nudi obogateno multimedijsko izkušnjo z elementi obogatene resničnosti. Učenci se lahko virtualno podajo na izlet v različne kraje po Sloveniji in preko aplikacije spoznavajo kulturne, naravne in zgodovinske značilnosti določenih krajev. Pri spoznavanju Pirana poslušajo zvoke znamenite sonate skladatelja Tartinija, poslušajo zgodbo o Benečanki ter se preizkušajo v različnih izzivih. Na ta način spoznajo Piran interdisciplinarno.

5 SKLEP

Sodobna šola 21. stoletja upošteva tiste koncepte znanja, ki poleg vsebinskega znanja v ospredje po-



Slika 1: **Odgovori učencev o motivaciji za učenje programiranja**

stavlja tudi spretnosti in veščine, kot so samostojno iskanje informacij, kritično mišljenje, timsko sodelovanje, ustvarjalnost, inovativnost ...

Sodobno in inovativno izobraževanje na daljavo vključuje medpredmetno povezovanje in inovativna učna okolja podprta s sodobno informacijsko-komunikacijsko tehnologijo. Medpredmetni pristop je pri izobraževanju na daljavo pomemben z vidika ekonomičnosti, celostne obravnave učnih vsebin ter povezovanja znanja.

Inovativna in sodobna učna okolja podprta z uporabo različnih informacijsko-komunikacijskih orodij učencem omogočajo poglobljeno usvajanje učne snovi in jim pomagajo pri doseganju višjih oblik učenja ter povezovanja znanja. Učenci so pri učenju motivirani, osredotočeni in željni sprejemati nova znanja. Prožne oblike učenja in inovativna učna okolja podprta z informacijsko-komunikacijsko tehnologijo podpirajo razvijanje ključnih kompetenc 21. stoletja, in sicer učenje, poučevanje, vrednotenje, vsebino, kurikulum, organizacijo, vodenje in vrednote. Za ustvarjanje spodbudnih učnih okolij potrebujemo multidisciplinarni pristop, ki predstavlja enega izmed pomembnih izzivov sodobne in inovativne šole. Za uresničevanje sodobnih konceptov znanja je potrebno zagotoviti vzgojno-izobraževalni načrt, ki je usmerjen na učen-



Slika 5: Reševanje izzivov v mobilni aplikaciji Nexto (<http://proxima.si/sl/project/nexto>, pridobljeno 12. 2. 2021)

ca, vključuje personalizacijo, inkluzivnost in socialnost učenja. Prav to slednje pa je še odprt problem pri izobraževanju na daljavo, kjer učenci še zdaleč nimajo enakih možnosti za usvajanje znanja.

6 LITERATURA

- [1] Brečko, B. N., Vehovar, V. (2008). Informacijsko-komunikacijska tehnologija pri poučevanju in učenju v slovenskih šolah. Ljubljana: pedagoški inštitut.
- [2] Dumont, H. (2013). O naravi učenja: uporaba raziskav za navdih prakse. Ljubljana: Zavod RS za šolstvo.
- [3] Flogie, A., Aberšek, B. (2019) Inovativna učna okolja – vloga IKT. Pridobljeno s <https://en.calameo.com/read/0058307531fae8501fad2>
- [4] Gerlič, I- (2000). Sodobna informacijsko-komunikacijska tehnologija v izobraževanju. Ljubljana: DZS. <http://proxima.si/sl/project/nexto>, pridobljeno 12. 2. 2021
- [5] Izobraževanje na daljavo – Izkušnje za prihodnost. Pridobljeno s https://www.pei.si/wp-content/uploads/2020/06/OM_celota.pdf
- [6] Krnel, D., Hodnik Čadež, T., Potočnik, N., Medved-Udovič, V. (2008). Medpredmetno povezovanje v 1. razredu – večpredmetni delovni učbenik. Naravoslovna solnica letnik 12, št. 3, str. 6-9.
- [7] Kuščer, K. (2000). Integriran pouk: priročnik za obravnavo tematskega sklopa Sadovnjak. Ljubljana: Zavod Republike Slovenije za šolstvo.
- [8] Marentič Požarnik, B. (2007). Čemu potrebujemo širši dogovor o temeljnih učiteljevih zmožnostih/kompetencah. Vzgoja in izobraževanje 38 (5), 44–53.
- [9] Marentič Požarnik, B. (2010). Psihologija učenja in pouka. Ljubljana: DZS.
- [10] Pavlič Škerjanc, K. (2010). smisel in sistem kurikularnih povezav. V: Pavlič Škerjanc in Rutar Ilc (ur.), Medpredmetne in kurikularne povezave, priročnik za učitelje: Ljubljana: Zavod RS za šolstvo.
- [11] Peklaj, C. in sodelavke (2001). Sodelovalno učenje ali Kdaj več glav več ve. Ljubljana: DZS.
- [12] Rutar Ilc, Z. (2014). Formativno spremljanje ali pogled v »črno škatlo«. V Z. Rutar Ilc (Ur.), Vzgoja in izobraževanje (str. 23 – 27). Ljubljana: Zavod RS za šolstvo.
- [13] Schleicher, A. (2019). Šole za učence 21. stoletja: močni vodje, samozavestni učitelji, inovativni pristopi: mednarodno srečanje na vrhu o učiteljskem poklicu. Ljubljana: Šola za ravnatelje.
- [14] Sicherl-Kafol, B. (2008). Medpredmetno povezovanje v osnovni šoli. Didakta: letnik 17, št. 19, str. 7-9.
- [15] Štemberger, V. (2008). Načrtovanje in izvajanje medpredmetnih povezav. V Učitelj v vlogi raziskovalca: akcijsko raziskovanje na področjih medpredmetnega povezovanja in vzgojne zasnovе v javni šoli, ur. Janez Krek, str. 112-130. Ljubljana: Pedagoška fakulteta.

Simona Samida Cerk je magistrica pedagoških znanosti, učiteljica razrednega pouka v OŠ Franceta Bevka v Ljubljani, kjer poučuje peti razred. V svojo pedagoško prakso vnaša inovativne in sodobne pristope poučevanja, vključuje se v projekte na državni in mednarodni ravni. Vključuje se v različne oblike izobraževanja, dejavna je predvsem na področju posodabljanja pouka in sodelovalnega učenja.

LEARNING-BASED LINK PREDICTION ANALYSIS FOR FACEBOOK100 NETWORK

Tim Poštuvan^{1, 2}, Semir Salkić¹, Lovro Šubelj¹

¹ University of Ljubljana, Faculty of Computer and Information Science, Večna pot 113, 1000 Ljubljana, Slovenia

² University of Ljubljana, Faculty of Mathematics and Physics, Jadranska ulica 19, 1000 Ljubljana, Slovenia
tim.postuvan@gmail.com, ss9343@student.uni-lj.si, lovro.subelj@fri.uni-lj.si

Izvleček

Facebook je ena izmed najzanimivejših in najširše uporabljenih socialnih in medijskih platform. Njegovi podatki so konkretno vplivali na raziskovanje socialnih omrežij in na tehnike napovedovanja povezav, ki so nepogrešljive pri analizi omrežij. Članek prvi natančno analizira napovedovanja povezav na Facebook100 omrežju. Na različnih množic značilk preučimo zmogljivost večih algoritmov strojnega učenja. Za pridobitev značilk uporabimo vložitve omrežij in tehnike, ki temeljijo na topologiji omrežja, kot so node2vec in vektor metrik podobnosti. Poleg tega uporabimo tudi značilke vozlišč, ki so na voljo za Facebook 100 omrežje, a le redko najdene pri drugih naborih podatkov. Razložimo tudi uporabljene pristope in nazorno predstavimo rezultate. Na koncu primerjamo naše modele, kjer podamo njihove končne zmogljivosti in klasifikacijske natančnosti.

Ključne besede: napovedovanje povezav, socialna omrežja, klasifikacija, nadzorovano učenje, izbira značilk

Abstract

In social network science, Facebook is one of the most interesting and widely used social networks and media platforms. Its data has significantly contributed to the evolution of social network research and link prediction techniques, which are important tools in link mining and analysis. This paper gives the first comprehensive analysis of link prediction on the Facebook100 network. We study performance and evaluate multiple machine learning algorithms on different feature sets. To derive the features, we use network embeddings and topology-based techniques such as node2vec and vectors of similarity metrics. In addition, we also employ node-based features, which are available for the Facebook100 network, though rarely found in other datasets. The adopted approaches are discussed and results are clearly presented. Lastly, we compare and review the applied models, where overall performance and classification rates are presented.

Keywords: Link prediction, social networks, classification, supervised learning, feature selection

1 INTRODUCTION

Social networks became an important focus in our research. We are witnessing exponential user expansion on social platforms (e.g. Facebook, Twitter and LinkedIn). People are joining these platforms and generating substantial amounts of data, which can reveal interesting clues about user behavior, society and psychology. We can see significant increase in research on the topic of social networks and link prediction. In the last decade we are experiencing a rise in this overlapping topic of network science and

data science, which is primarily used to analyze and understand social networks [Wang et al., 2014]. According to Facebook estimations in March 2020, they have 1.73 billion DAU (Daily Active Users). Much work has been done to understand complexities and challenges of social networks, where considerable knowledge has been obtained ([Pow et al., 2012], [Scott, 1988]). In that manner we are using the Facebook100 dataset (2005), which includes a complete set of people from Facebook networks for 100 different colleges and universities in the USA. This paper

is the first comprehensive study of link prediction techniques applied to the Facebook100 dataset. We are using a learning-based approach with local probabilistic models such as logistic regression. Besides that, we are utilizing standardized link prediction classification models such as: random forests, artificial neural networks and kernel-based models (e.g. SVM). Having in mind that we have two main approaches to link prediction, we chose learning-based approaches instead of similarity-based approaches because we want to compare performance, stability and classification accuracy of multiple classification models instead of showcasing different measures of node proximity. We are using standardized similarity-based methods to derive notable features, which are used as model input. The presented methods are node2vec graph embedding and an ensemble of topology-based metrics such as Jaccard Coefficient, Adamic Adar Coefficient and Preferential Attachment Index. Using these methods, we build classification models which are clearly presented, focusing on performance and generalization, with appropriate discussion and results.

The paper is structured as follows. In section 2 we present related work in the field of link prediction on social networks. Section 3 briefly describes the Facebook100 data. In sections 4 and 5 we present used feature sets and datasets, which are utilized for training and testing. Process and discussion of feature selection for each dataset are stated in the section 6. Sections 7 and 8 give an overview of achieved results, while also providing comprehensive model analysis. Insightful discussion of results which explains interesting nature of the data is explained in section 9. We conclude the paper with an overview in section 10.

2 RELATED WORK

Link prediction has recently become very popular for prediction of future relationships between individuals of social networks. Consequently, a great variety of different approaches were invented. In the past decade, many efforts have been made by psychologists, computer scientists, physicists and economists to solve the link prediction problem in social networks. According to Wang et al. [Wang et al., 2014] there are two ways to predict links: similarity-based approaches and learning-based approaches. Similarity-based approaches calculate a similarity score for every pair of nodes, where higher score means higher pro-

bability that the corresponding nodes will be connected in the future.

Learning-based approaches are treating the link prediction problem as a binary classification task [Hasan et al., 2006]. Therefore, typical machine learning models can be employed for solving the problem. These include classifiers like random forest [Breiman, 2001], multilayer perceptron or support vector machine (SVM) [Cortes and Vapnik, 1995], as well as probabilistic models. The learning-based approaches use non-connected pairs of nodes as instances with features describing nodes and the class label. Pairs of nodes which have potential to become connected are labeled as positive and the others as negative.

Their feature set consists of similarity features from the similarity-based approaches and features derived from domain knowledge (e.g. textual information about members of social networks). Using combination of both can remarkably improve the link prediction performance. Scellato et al. [Scellato et al., 2011] considered social features, place features and global features in location-based social networks for link prediction based on a supervised learning framework.

Both types of approaches rely on various metrics, which use information of nodes, topology of network and social theory to calculate similarity between a pair of nodes. Metrics consist of three categories: node-based, topology-based and social theory-based metrics.

Node-based metrics use the attributes and actions of individuals to assess similarity of node pairs. They are very useful in link prediction; however, it is usually hard to get the data because of privacy issues.

Most metrics are based on the topological information and are called topology-based metrics. They are most commonly used for prediction, because they are generic and domain independent. Topology-based metrics are further divided into the following subcategories: neighbor-based, path-based and random walk-based metrics. Neighbour based metrics assume that people tend to form new relationships with people that are closer to them. The most famous are Common Neighbors [Newman, 2001], Jaccard Coefficient [Salton and McGill, 1986], Adamic Adar Coefficient [Adamic and Adar, 2003] and Preferential Attachment Index [Barabási et al., 2002]. The first three all use the same idea that two nodes are more likely to be connected if they share a lot of common neighbours. On the other hand, Preferential Attachment Index assumes that nodes with higher de-

gree have higher probability of forming new edges.

Neighbor-based metrics capture local neighborhood but do not consider how nodes are reachable from one another. Path-based metrics incorporate this information by considering paths between nodes. They are more suitable to small networks and are not scalable to big networks. Examples of path-based metrics are Local Path [Lü et al., 2009] and Katz metric [Katz, 1953]. Local Path metric makes use of information of local paths with length two and three, while giving more importance to the paths of length two. Katz metric calculates the similarity by summing all the paths connecting the two nodes, giving higher weight to shorter paths.

Social interactions between members of social networks can also be modeled by random walk, which uses transition probabilities from a node to its neighbors to denote the destination of a random walker from the current node. Examples of random walk-based metrics are Hitting Time and SimRank [Jeh and Widom, 2002]. Hitting time metric calculates similarity based on the expected number of steps required for a random walk starting at a node to reach the other node. SimRank metric computes similarity according to the assumption that two nodes are alike if they are connected to structurally similar nodes.

Social theory-based metrics take advantage of classical social theories, such as community, triadic closure, strong and weak ties and homophily, improving performance by capturing additional social interaction information. Liu et al. [Liu et al., 2013] proposed a link prediction model based on weak ties and degree, closeness and betweenness node centralities.

When designing a feature set, the choice of features tremendously influences the performance of link prediction. Sometimes it is hard to find appropriate features, hence it is desirable that an algorithm learns important features on its own. Network embedding methods aim at learning low-dimensional latent representation of nodes in a network. Embeddings should follow the principle that similar nodes in the network have similar embedding representations. The advantage of node embedding as a technique is enormous since it does not require feature engineering by domain experts. Network embeddings methods can be broadly categorized into four classes: methods based on random walks, matrix factorization, neural networks, and probabilistic approaches. For the purpose of this paper methods based on ran-

dom walks are the most relevant.

Methods based on random walks determine similarities using random walks on the original network. The Skip-Gram model, described in Mikolov et al. [Mikolov et al., 2013], is then usually used to generate node embeddings from the random walks. Examples of such methods are DeepWalk [Perozzi et al., 2014] and node2vec [Grover and Leskovec, 2016]. DeepWalk was the first technique for network embeddings, inspired by deep learning. It uses random walks with fixed transition probabilities to measure node similarity, while embeddings are derived using the Skip-Gram model. Node2vec is a generalization of DeepWalk which uses supervised random walks for node neighbourhood exploration. The random walk is controlled by a return parameter p and an in-out parameter q . Then similarly Skip-Gram model is used, but this time approximated via negative sampling, for embedding generation.

Evaluation of the methods plays the crucial role in machine learning task in general. To estimate performance of the link prediction approaches more evaluation criteria exists. While some papers utilize $Precision@N_p$ for a range of N_p values [Zhang et al., 2018], others use AUROC [Grover and Leskovec, 2016].

3 DATA

We study the Facebook social network of friendships at one hundred American colleges and universities at a single moment of time in September 2005 ([Traud et al., 2012], [Traud et al., 2011]). The network consists of one hundred independent networks, where every network corresponds to one university. Friendships are recorded only between people from the same university. Besides the information about friendships, network also contains limited demographic information. The following information is available for each user: student/faculty status flag, gender, major, second major/minor (if applicable), dorm/house, year and high school. Network is unweighted and undirected. The whole network consists of 3.2 million nodes with 23.7 million links between them [Rosi and Ahmed, 2015]. Maximum degree of a single node is approximately 4900 and minimum degree is only 1, with an average of 15. According to statistics network appears to be disassortative but this is only the consequence of its size. It also has high average clustering coefficient 0.097, which is characteristic of social networks.

Tabela 1: **Structural information of train and unseen data**

Dataset	n	m	d	C
Train	4943.5	206247.6	77.26787	0.2808
Unseen	3517.8	140793.2	80.3072	0.2689

Table 1 contains structural measurements for used network sets. All of the presented measurements are averaged over all networks in the corresponding set. The presented values are: average clustering coefficient (C), average degree (d), average number of nodes (n) and average number of edges (m). We can see by the number of nodes and edges that train set is larger than unseen dataset. Clustering coefficient and average degree are high, which is one of expected characteristics of social networks.

Because Facebook100 dataset is enormous, lack of computing power prevented us from considering the

complete dataset for analysis. Therefore, we have decided that we would perform analysis only on a subset of networks. We selected ten networks as seen data and five networks as unseen data. Seen data was used for standard train and test set, where training and testing edges came from the same ten networks. On the other hand, unseen data consisted of five networks, which were not shown to the models during training. We used it to evaluate if our models are transferable to new data.

Degree distribution of the analyzed networks is presented in figures 1 and 2. Degree distribution is plotted on a log-log scale for all networks in both sets respectively. For cleaner overview, we used interpolation (univariate spline) to showcase distribution of all networks. It is visible on both figures that all networks follow power law, which is expected for social networks. Having in mind that we are using real life social networks, it can be concluded that they are scale-free networks by degree distribution results. However, it is interesting to point out existence of big hubs, nodes with very high degree. They are visible on right side of the distribution graph. This is one of the reasons why interpolation at the end of the plot has an unexpected minimum.

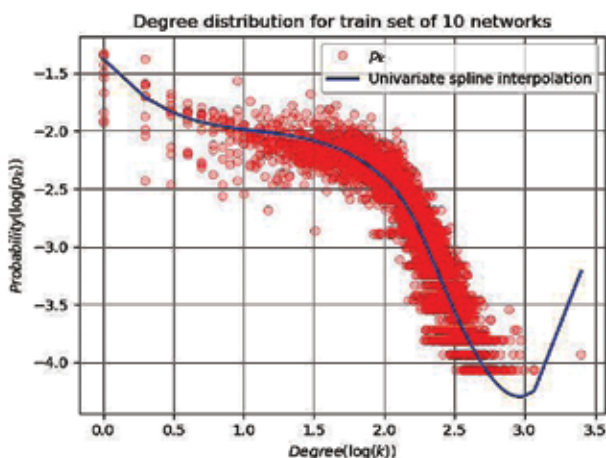


Figure 1: **Degree distribution of train set**

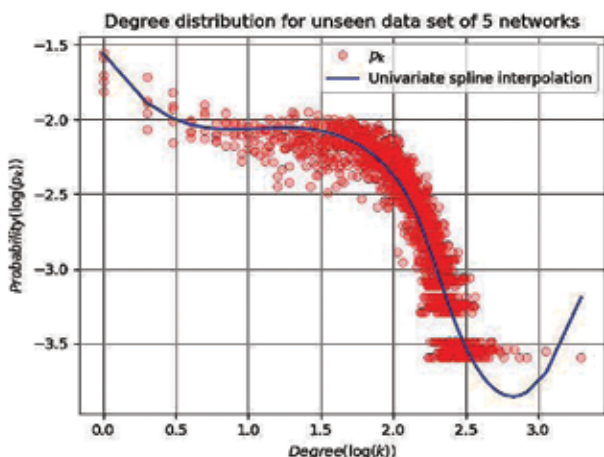


Figure 2: **Degree distribution of unseen data set**

4 FEATURE SET

Feature engineering probably plays the most important role when coping with a machine learning problem. Informative features crucially effect model accuracy; hence the process of feature engineering is usually very time consuming. In learning-based link prediction each pair of nodes is described using a combination of node-based, topology-based and node embedding features, depending on approach. In this paper we are using three different feature sets, from which three datasets were constructed as it will be described in the section 5.

4.1 Node-based features

Node-based features use domain-specific information about individuals. Facebook100 dataset has already a basic set of features, however, not all of them are useful for link prediction task. Almost all features had to be transformed, in order to describe node pairs, instead of individuals. From some features, for example dormitory information, new features had to be created, because otherwise model would not be transferable between networks. Problem arises from

the fact that different universities use different numerations of their dormitories. Considering the above constraints, we derived the following features:

- **same_dorm**: binary value, indicating whether the nodes live in the same dormitory
- **same_year**: binary value, indicating if the nodes started college in the same year
- **year_diff**: numerical value, stating the absolute difference between the years, when the nodes started college
- **high_school_1, high_school_2**: numerical values, stating indices of nodes' high schools
- **major_1, major_2**: numerical values, stating indices of nodes' majors
- **same_faculty**: binary value, indicating whether the nodes have the same faculty status
- **same_gender**: binary value, indicating if the nodes have the same gender

Since networks are undirected, each pair of nodes must be uniquely represented using above features. Representation should not depend on order of the pair, thus major_1 and major_2 are ordered in a way that the value of major_1 is not greater than the value of major_2. The same holds for high_school_1 and high_school_2.

Like the majority of datasets Facebook100 does not contain all information about all individuals. Therefore, missing values had to be handled. We decided that imputing is reasonable only for the feature year_diff, where missing values were substituted with the mean. Values of other features were left intact but as soon as one of the nodes in the pair had a missing value, the corresponding binary values was automatically zero.

4.2 Topology-based features

The most commonly used features for link prediction are topology-based features. They are particularly useful, when you do not have any problem specific information, because they are generic and domain independent. Although Facebook100 dataset has additional domain specific data, topology-based features still have great impact on model accuracy. In this paper we are using the following topology-based features:

- **Jaccard Coefficient** [Salton and McGill, 1986]. Jaccard Coefficient normalizes the size of common neighbors. According to Jaccard Coefficient a pair of nodes is assigned a higher value when the no-

des share a higher proportion of common neighbors relative to total number of their neighbours.

$$JC(x, y) = \frac{|\Gamma(x) \cap \Gamma(y)|}{|\Gamma(x) \cup \Gamma(y)|} \quad (1)$$

where $\Gamma(x)$ is a set of neighbours of node x .

- **Adamic Adar Coefficient** [Adamic and Adar, 2003]. Adamic Adar Coefficient measure is closely related to Jaccard Coefficient. It is calculated as a weighted sum of common neighbours, where common neighbours with fewer neighbours have greater impact. The rationale behind it is that high degree nodes are more likely to occur in common neighbourhood, thus they should contribute less than low degree nodes.

$$AA(x, y) = \frac{1}{\sum_{z \in \Gamma(x) \cap \Gamma(y)} \log|\Gamma(z)|} \quad (2)$$

- **Preferential Attachment Index** [Barabási et al., 2002]. The measure is based on the concept that nodes with higher degree have higher probability of forming new edges.

$$PA(x, y) = |\Gamma(x)|\Gamma(y) \quad (3)$$

- **Resource Allocation Index** [Zhou et al., 2009]. Resource Allocation Index metric is very similar to Adamic Adar Index. The only difference is that Resource Allocation Index punishes high degree nodes more.

$$RAI(x, y) = \frac{1}{\sum_{z \in \Gamma(x) \cap \Gamma(y)} |\Gamma(z)|} \quad (4)$$

4.3 Node embedding features

Network embeddings methods aim to learn low-dimensional latent representation of the nodes in a network. To generate a dataset comprising of every node in a network we are able to use these representations as features. This can be used for a wide variety of tasks such as classification, clustering, link prediction, and visualization. Using node2vec [Grover and Leskovec, 2016] we were able to generate our embeddings dataset.

The key point is that node2vec is based on random node walks performed in a biased manner across the network. With this generic approach we are able to sample any network in a search for vector representation of its structural properties. With

the introduction of search bias α we are able to control our search in breadth-first search or depth-first search manner. If we choose «in-out parameter” (q), walks are more biased to visit nodes further from the start node, thus expressing the nature of exploration. Fixing «return parameter” ($p = 1$) ensures that we are less likely to visit same node twice, which in return adopts the strategy of modern exploration (avoids 2-hop redundancy in sampling).

As stated in the case study by Grover & Leskovec [Grover and Leskovec, 2016] for social structures it is beneficial to tune node2vec hyperparameters to discover communities of nodes which are interacting with each other. Capturing this type of behavior using embedding representation significantly benefits the link prediction task. To find the best set of hyperparameters, we employed grid search over more than 80 different settings that deemed reasonable to us. Each setting was evaluated on network of Caltech using logistic regression and the best combination of hyperparameters was selected. It is important to note here, that we did not consider only AUROC of the logistic regression model, but also the complexity of the embedding. The embedding dimension should be as small as possible while carrying all relevant information. The hyperparameters

that were most consistent with the two criteria: 64 dimensions, 50 walks per node, $q = 0.8$ and 20 nodes in each walk. Since node2vec approach yields embeddings for nodes, we used Hadamard product to express vector representations for edges.

5 DATASETS

Firstly, we had to preprocess all graphs (seen as well as unseen) to obtain train and test node pairs, which will be predicted by models and using which performance will be evaluated. Node pairs can be either positive or negative instances for link prediction task, depending on whether there is an edge between the nodes or not (the nodes are friends or not). For every graph we used the standard approach of generating an incomplete train graph $G_{train} = (V, E_{train})$ from the original graph $G = (V, E)$. The connected node pairs $\{i, j\} \in E \setminus E_{train}$ which are present in the original graph but not included in the train graph, are used as positive instances for link prediction task. Positive instances were sampled randomly from the original graph’s edge set E . We decided to sample 2% of edges in original graph G . Since dataset should contain positive as well as negative instances, we had to obtain also negative instances – pairs of nodes that are not connected by an edge. It is assumed that if two nodes are not connected by an edge,

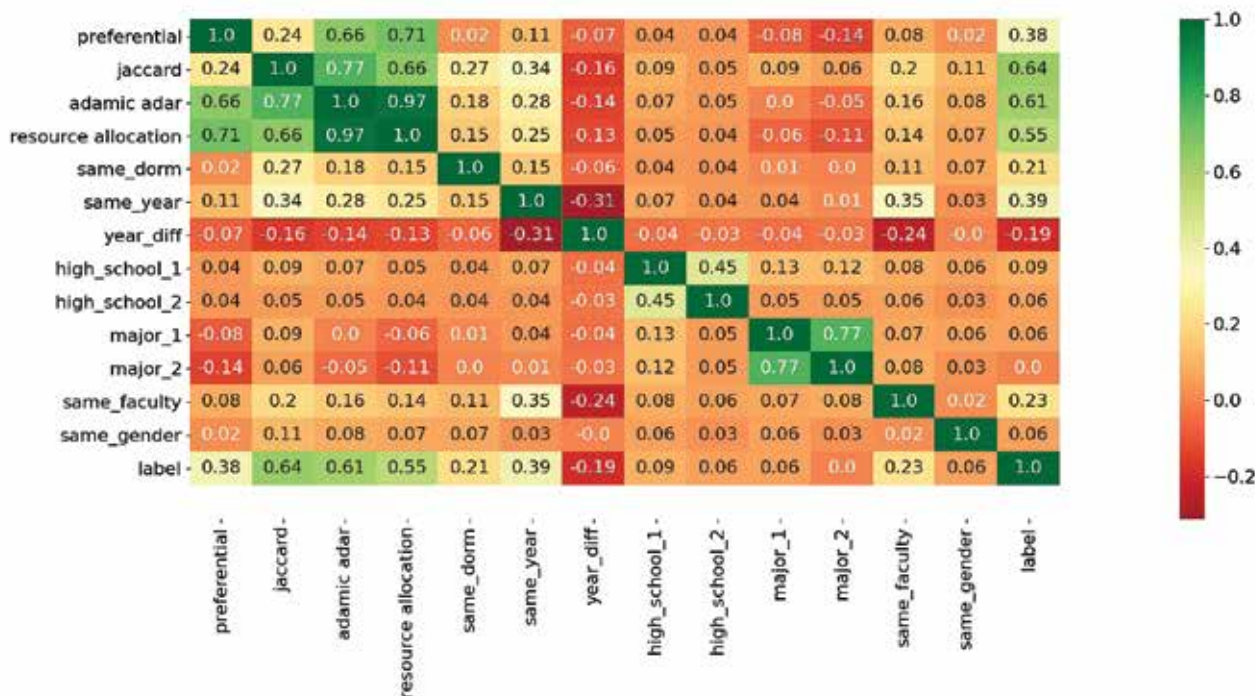


Figure 3: Correlation matrix of node-based and topology-based features

they are not friends. Negative instances were obtained by randomly selecting node pairs $\{i, j\} \notin E$, which are not in the original graph's edgeset. To get a balanced dataset the number of negative instances is the same as the number of positive ones. The rest of the graph (G_{train}) was used to calculate topology-based and node embedding features.

Positive train and test edges were not included in this graph from which features were derived, otherwise we would introduce unjust label information in features.

In our experiments, all unseen data instances were used for testing models' ability to adapt to new graphs. However, seen data was further split into train and test data. We used standard division: 80% of it was used as the train data and the remaining 20% was used as the test data. Within each of the them there is approximately the same number of positive and negative instances.

Using this data three datasets were created: baseline, topological and embedding dataset. Each dataset represents node pairs using a different combination of features. Baseline dataset is the simplest one and contains only topology-based features. A bit more complex is topological dataset, which in addition to the topology-based features makes use of node-based features as well. Node pairs in embedding dataset are described using node-based features and Hadamard product of the corresponding nodes' embeddings.

6 FEATURE SELECTION

Contemporary datasets usually have abundance of data, which is not always relevant to the problem. Hence, datasets should be preprocessed before models are used on them. Preprocessing takes place mainly to reduce the size of the dataset and achieve more efficient analysis, as well as removing redundant features, which have negative impact on the performance of the model. The aim of feature selection is to maximize relevance and minimize redundancy of the features.

Our feature sets are not enormous, thus feature selection was done solely for the sake of performance improvement of the models. We are using recursive feature elimination with cross-validated selection (RFECV) in combination with linear kernel support vector machine (SVM) to get reduced feature sets. This method recursively considers smaller and smaller

sets of features, while after every iteration prunes the least important features according to the chosen model. It belongs to wrapper methods for feature selection, since it appraises subsets of features based on performance of the modelling algorithm. According to Jović et al. [Jovic et al., 2015] wrapper methods have been empirically proven to yield better results than other methods because subsets are evaluated using real models.

6.1 Baseline dataset

The above feature selection method recognized Adamic Adar Coefficient, Jaccard Coefficient and Resource Allocation Index as the most informative features. The most relevant feature is Adamic Adar Coefficient and the least relevant one is Preferential Attachment Index. This is completely coherent with random forest feature importance shown in figure 4. Adamic Adar Coefficient is the most relevant feature, although Jaccard Coefficient has higher correlation with labels, which can be observed in figure 3. All selected features are highly correlated with label, whereas Preferential Attachment Index is not. This is probably the reason why Preferential Attachment Index is the only feature which was not selected. From correlation matrix it is also evident that Adamic Adar Coefficient and Resource Allocation Index are almost perfectly correlated, which is expected because of the similarity in their definitions. Nonetheless, adding it results in a slightly better performance, thus the algorithm decides to keep it.

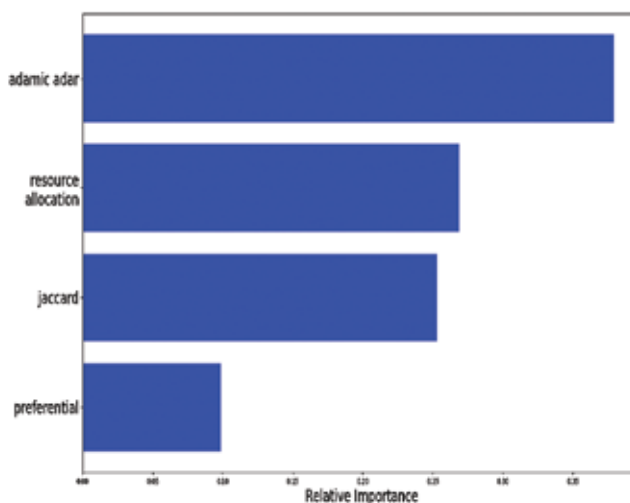


Figure 4: Feature importance of baseline dataset according to random forest classifier

6.2 Topological dataset

The advantages of feature selection are more evident on topological dataset, because it has more features. This time algorithm selected the following features: all four topology-based features, same year, same faculty, same dorm, major_1 and major_2. On figure 5 it is clearly shown that topology-based features are far more important than other node-based features. This is also consistent with correlation matrix, since topology-based features have the highest correlations with labels. They are so informative due to phenomenon called triadic closure. The triadic closure states that in social networks connections tend to form between people who share common friends, which is precisely what these topology-based features are describing. Among the node-based features same year, same faculty and same dorm were selected, all having relatively high correlation with label. Particularly high correlation has same year, which is expected, as college students often form friendships with their classmates. Because of this major_1 and major_2 are also relevant. Feature same faculty exploits the fact that students' and professors' social circles are rarely overlapping.

6.3 Embedding dataset

Feature selection on embedding dataset was especially hard due to artificial features from node2vec. Because hyperparameters of node2vec were carefully tuned, we assumed that node embeddings are optimal. We selected the minimum embedding size which still performed well, while the quality of a set of hyperparameters was evaluated on the link pre-

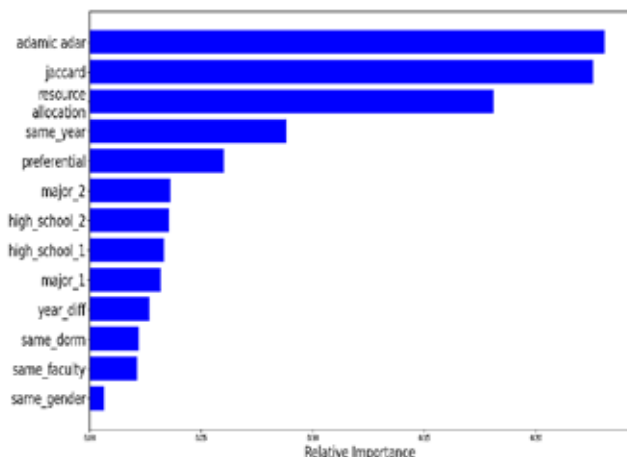


Figure 5: Feature importance of topological dataset according to random forest classifier

diction task. This was the best we could do with available computational resources, since there were too many features to utilize the feature selection method described above. Hence, we filtered only node-based features [Liao et al., 2017] and are using only a few crucial ones: same year and same dorm. We did not select same faculty, although it is more important than same dorm if considered on its own. We decided so, because same faculty has high correlation with same year and correlated features usually have negative impact on performance of the model.

7 RESULTS

Evaluation of our datasets was conducted using an ensemble of classification models. We used simpler models like logistic regression and random forest, as well as more complex ones – support vector machines (SVM) and neural networks (NN), which are capable of modeling more complex non-linear functions. Hyperparameters of all models were optimized as described in section 8. Link prediction task was tested on all three datasets, on test and unseen data, and all aforementioned models. Performance of the models was evaluated using Area Under the Receiver Operating Characteristics (AUROC), which is one of the most common evaluation metrics for link prediction.

Tabela 2: AUROC values for logistic regression, random forest, support vector machine (SVM) and neural network (NN) on test data

Dataset	Logistic regression	Random forest	SVM	NN
Baseline	0.9401	0.9227	0.9628	0.9618
Topological	0.9570	0.9173	0.9639	0.9623
Embedding	0.9365	0.9145	0.9412	0.9389

Table 2 contains AUROC scores for all combinations of the datasets and models on the test data. Similarly, table 3 states the same values, but on unseen data. These tables reveal that support vector machine (SVM) and neural network (NN) are the best models for the link prediction task. Their performance is almost exactly the same, although they are based on completely different concepts. This is indicating that all relevant information from datasets is used for prediction. Only a little worse did logistic regression, which is very surprising, since it is much simpler than SVM and NN. Even more unexpected is that it outperformed random

Tabela 3: AUROC values for logistic regression, random forest, support vector machine (SVM) and neural network (NN) on unseen data

Dataset	Logistic regression	Random forest	SVM	NN
Baseline	0.9263	0.9031	0.9570	0.9560
Topological	0.9478	0.8901	0.9563	0.9538
Embedding	0.9229	0.9047	0.9217	0.9218

forest, which is a non-linear model. We hypothesize that this is a consequence of linear separability of the data, but more about this will be written in section 9. All models appear to be stable, since there is only a slight decrease in performance, when applied to unseen data. Difference is negligible for baseline and topological dataset, whereas noteworthy on embedding dataset.

The models were able to extract more useful information from topological dataset than baseline and embedding ones. Baseline dataset has only a bit worse results, showing that additional node-based features have minimal influence on performance. Difference is visible for logistic regression, whereas SVM and NN have the same score on both datasets. Shockingly, embedding dataset gets the worst results. However, this might be the consequence of the chosen evaluation metric. For example, logistic regression on embedding dataset gets 0.87 F_1 score, while baseline and topological datasets get only 0.76 and 0.79. F_1 score reflects balance between recall and precision, whereas AUROC does not consider the predicted labels, but only measures whether predicted values of negative instances are smaller than predicted values of positive instances. Consequently, metrics are not necessarily coherent, however, we decided to trust AUROC since it is a more standard link prediction metric.

8 MODEL ANALYSIS

Here we used previously obtained data and results to optimize our models. Prior to that, data was standardized to have variance 1.0 and mean 0.0. With this approach, we have done model analysis to interpret best combinations of hyperparameters, which are useful to understand and discover patterns in data.

8.1 Logistic regression

Using grid search cross-validation on logistic regression we saw that different approaches require different configurations. In the case of the baseline approach features equally impact the decision process, which

is reflected in features' coefficients. For this dataset we used ridge regularization (L_2) with default regularization strength of 1.

In the case of topological dataset, we noticed that regularization significantly influences performance. We used lasso regularization (L_1) regularization with immense regularization strength of 1000. In this case regularization is crucial for prevention of overfitting. Having in mind that our features are measurements which are not calculated in a fixed interval, we are benefiting from the L_1 property of data sparsity. Model coefficients are imbalanced, where major study features (i.e. major_1 and major_2) are given low coefficient values, which shows that most of the information is contained in the rest of the features.

In embedding dataset we noticed that addition of node-based data does not have any benefits. This could be the consequence of correlations within features of embedding dataset. Maybe node2vec features contain structural information, using which node-based feature can be well approximated. For embedding dataset L_1 regularization with strength 150 was used. Lasso regularization improves model performance on unseen networks. This is achieved with generalization of the obtained knowledge from social network onto new unseen networks. As expected, coefficients show that all features of embedding vectors are equally important.

8.2 Random forest

Random forest did not perform well on our datasets. We can justify that by the fact that this approach lacks mechanism for regularization. Higher number of dimensions in respect to number of samples (unbalanced training and unseen data) is causing our decision tree models to overfit. Grid search in this case did not yield specific results, as well as tuning of parameters failed to find feature dependent information. This behavior is shown in our model comparison where it is expected to experience better benchmarks on different linear models such as logistic regression and SVM. We noticed that unseen networks' AUROC scores are the lowest over all datasets, therefore we can conclude that random forest model did not respond well to our problem.

8.3 Support vector machine (SVM)

For support vector machine (SVM) only kernels were carefully tuned. Best fit for each dataset was chosen

using grid search. Grid search consisted of linear, polynomial and Gaussian kernel, so the model could work with arbitrary dimensional data. It turned out that for baseline and topological datasets linear kernel was the best option, while embedding dataset required Gaussian kernel.

8.4 Neural network (NN)

Choosing the right hyperparameters for neural networks (NN) was very complicated and tedious task, since neural networks have a lot of different parameters. Nevertheless, correctly setting them can yield much better performance in comparison to other models. For some of the parameters like loss and optimization functions default settings were selected. Because learning-based link prediction is a binary classification task, binary cross-entropy loss function and Adam optimizer were utilized. Hidden and output activation functions were selected using random experimentation. The best results yielded ReLU as hidden activation function and sigmoid as output activation function. Lastly, architecture of neural network had to be defined, which was done using grid search. We tried a great variety of different depths and numbers of nodes per layer, but in the end architectures with only two hidden layer and small number of nodes were the best performing on all datasets.

9 DISCUSSION

Embedding dataset yielded worse results than topological and baseline ones. This could be so because of greater linear separability of topological and baseline

datasets as SVM results from section 7 were implying. It is much easier to train models on linearly separable data than on complex one with a lot of non-linearity, such as node2vec.

Baseline and topological data seem to be well linearly separable, because SVM works best on them when combined with linear kernel. Also, logistic regression performs considerably better than random forest. This is highly unusual for non-linear datasets, but in this case random forest harder adjusts to linear data, while logistic regression is linear by default.

Trying to prove our hypothesis that topological dataset is more linearly separable than embedding one, we visualized data using linear discriminant analysis (LDA). If any linearity exists in the dataset, it should be visible in the low-dimensional space. To visualize data we uniformly sampled 200 samples from test data of both datasets and calculated LDA decision boundary for binary classification.

LDA analysis seems to show that the data is approximately equally linearly separable in topological dataset (figure 6) as well as in embedding one (figure 7). Although, accuracy score of primitive LDA classifier yielded slightly better results on topological test set 87%, in comparison to embedding one which had score of 86.5%. A more concrete evidence of worse linear separability is probably the fact that SVM with Gaussian kernel performed better on embedding dataset than SVM with the linear kernel.

Nonetheless, it appears that embedding dataset is still linearly separable to some degree. This is completely unexpected as node2vec produces inherently non-linear embeddings. These interesting results

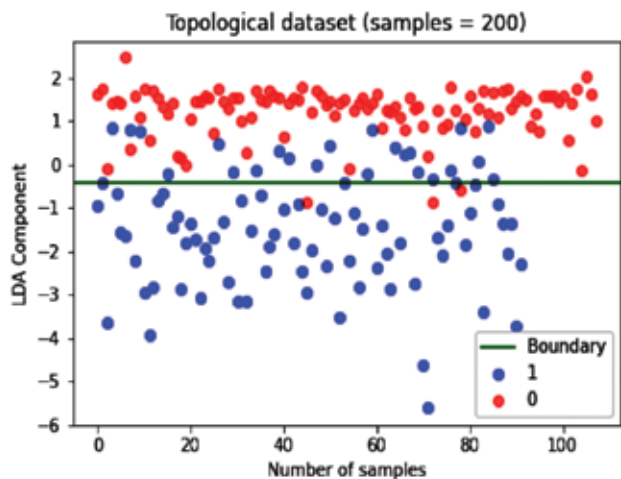


Figure 6: LDA on topological dataset

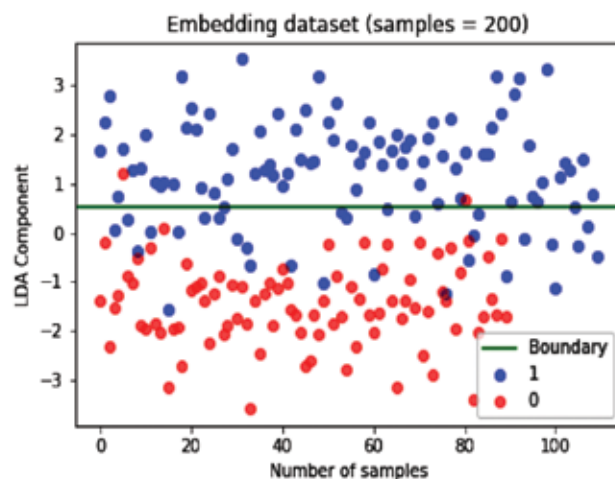


Figure 7: LDA on embedding dataset

pose a new research question that could be examined in future work.

Besides assumed linearity, topological dataset could also give better results due to small number of features (9 features), whereas embedding dataset was significantly more complex (66 features). It is harder to train models on data with a lot of features, because models automatically require more parameters, so worse performance of embedding dataset could also stem from that.

10 CONCLUSION

In the presented paper we conducted the first comprehensive analysis of link prediction on the Facebook100 network. We can conclude that results are unexpectedly good for link prediction task of this nature. After all, we are predicting friendships on completely separated social networks. It is visible that models successfully generalized to unseen data, even though train set is only 50% bigger than unseen data set. Therefore, it is safe to say that our models succeeded in their task.

For optimization of AUROC score baseline and topological approaches are the best. It turns out that simplicity has benefits in terms of higher classification scores. In these two cases node-based features did not really effect performance, except when combined with logistic regression. Although SVM and NN got better results, we recommend logistic regression in combination with topological approach because the model is easier to train and interpret. When very high AUROC scores are important (e.g. link prediction on medical data), we suggest SVM with linear kernel and baseline approach. It gets almost the same results on unseen data, even though it is simpler model than NN.

We have shown that collecting data from multiple social networks yields promising datasets, which can be used for modelling of various predictors in similar social structures. Besides that, in this paper we have shown that use of regularization can be a solution in the case of social networks, when lack of the training data is present. Using this approach, we can obtain data insights globally.

For future work it would be interesting to show how much linearly separable is embedding dataset. Even more fascinating would be to find the true underlying cause of the observed behavior.

11 REFERENCES

- [1] [Adamic and Adar, 2003] Adamic, L. and Adar, E. (2003). Friends and neighbors on the web. *Social Networks*, 25:211–230.
- [2] [Barabási et al., 2002] Barabási, A., Jeong, H., Nédá, Z., Ravasz, E., Schubert, A., and Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and its Applications*, 311(3-4):590–614.
- [3] [Breiman, 2001] Breiman, L. (2001). Machine learning, volume 45, number 1 - springerlink. *Machine Learning*, 45:5–32.
- [4] [Cortes and Vapnik, 1995] Cortes, C. and Vapnik, V. (1995). Support-vector networks. *Mach. Learn.*, 20(3):273–297.
- [5] [Grover and Leskovec, 2016] Grover, A. and Leskovec, J. (2016). Node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 855–864, New York, NY, USA. Association for Computing Machinery.
- [6] [Hasan et al., 2006] Hasan, M. A., Chaoji, V., Salem, S., and Zaki, M. (2006). Link prediction using supervised learning. In *In Proc. of SDM 06 workshop on Link Analysis, Counterterrorism and Security*.
- [7] [Jeh and Widom, 2002] Jeh, G. and Widom, J. (2002). Simrank: A measure of structural-context similarity. In *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '02, page 538–543, New York, NY, USA. Association for Computing Machinery.
- [8] [Jovic et al., 2015] Jovic, A., Brkić, K., and Bogunovic, N. (2015). A review of feature selection methods with applications. pages 1200–1205.
- [9] [Katz, 1953] Katz, L. (1953). A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39– 43.
- [10] [Liao et al., 2017] Liao, L., He, X., Zhang, H., and Chua, T.-S. (2017). Attributed social network embedding. *IEEE Transactions on Knowledge and Data Engineering*, PP.
- [11] [Liu et al., 2013] Liu, H., Hu, Z., Haddadi, H., and Tian, H. (2013). Hidden link prediction based on node centrality and weak ties. *EPL (Europhysics Letters)*, 101:18004.
- [12] [Lü et al., 2009] Lü, L., Jin, C.-H., and Zhou, T. (2009). Similarity index based on local paths for link prediction of complex networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 80:046122.
- [13] [Mikolov et al., 2013] Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013). Efficient estimation of word representations in vector space. *Proceedings of Workshop at ICLR*, 2013.
- [14] [Newman, 2001] Newman, M. E. J. (2001). Clustering and preferential attachment in growing networks. *Physical review. E, Statistical, nonlinear, and soft matter physics*, 64 2 Pt 2:025102.
- [15] [Perozzi et al., 2014] Perozzi, B., Al-Rfou, R., and Skiena, S. (2014). Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '14, page 701–710, New York, NY, USA. Association for Computing Machinery.
- [16] [Pow et al., 2012] Pow, J., Gayen, K., Elliott, L., and Raeside, R. (2012). Understanding complex interactions using social network analysis. *Journal of clinical nursing*, 21:2772–2779.
- [17] [Rossi and Ahmed, 2015] Rossi, R. A. and Ahmed, N. K. (2015). The network data repository with interactive graph analytics and visualization. In *AAAI*.
- [18] [Salton and McGill, 1986] Salton, G. and McGill, M. J. (1986). *Introduction to Modern Information Retrieval*. McGraw-Hill, Inc., USA.

- [19] [Scellato et al., 2011] Scellato, S., Noulas, A., and Mascolo, C. (2011). Exploiting place features in link prediction on location-based social networks. In *KDD*, pages 1046–1054. ACM.
- [20] [Scott, 1988] Scott, J. (1988). Social network analysis. *Sociology*, 22.
- [21] [Traud et al., 2011] Traud, A. L., Kelsic, E. D., Mucha, P. J., and Porter, M. A. (2011). Comparing community structure to characteristics in online collegiate social networks. *SIAM Rev.*, 53(3):526–543.
- [22] [Traud et al., 2012] Traud, A. L., Mucha, P. J., and Porter, M. A. (2012). Social structure of Facebook networks. *Phys. A*, 391(16):4165–4180.
- [23] [Wang et al., 2014] Wang, P., Xu, B., Wu, Y., and Zhou, X. (2014). Link prediction in social networks: the state-of-the-art. *Science China Information Sciences*, 58.
- [24] [Zhang et al., 2018] Zhang, Z., Cui, P., Wang, X., Pei, J., Yao, X., and Zhu, W. (2018). Arbitrary-order proximity preserved network embedding. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '18, page 2778–2786, New York, NY, USA. Association for Computing Machinery.
- [25] [Zhou et al., 2009] Zhou, T., Lü, L., and Zhang, Y.-C. (2009). Predicting missing links via local information. *The European Physical Journal B - Condensed Matter and Complex Systems*, 71:623–630.

■

Tim Poštuvan is in the final year of his bachelor's degree study. He is studying Computer Science and Mathematics at Faculty of Mathematics and Physics and Faculty of Computer and Information Science, University of Ljubljana. His primary research interests lie in machine learning on graphs, however, he is fond of other areas of artificial intelligence as well.

■

Semir Salkić received a bachelor's degree from the Faculty of Electrical Engineering, University of Sarajevo in 2018. He is currently studying Data Science Master Programme at Faculty of Computer and Information Science, University of Ljubljana. He is working as a lead software and embedded engineer at Research Center at Faculty of Applied Sciences Kiel, Germany. His research is focused on practical applications of IoT structures with machine learning.

■

Lovro Šubelj je docent na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Diplomiral je leta 2008 na Fakulteti za matematiko in fiziko in Fakulteti za računalništvo in informatiko ter doktoriral leta 2013 na temo analize velikih omrežij. Je avtor ali soavtor več kot šestdeset znanstvenih prispevkov in patentov ter urednik prestižnih mednarodnih znanstvenih revij. Njegovo preteklo delo je bilo izbrano kot izjemen znanstveni dosežek v Sloveniji ter predstavljeno na uglednih mednarodnih univerzah kot sta Stanford in UCSD. Sodeloval je že pri številnih uspešno zaključenih raziskovalnih in razvojnih projektih v sodelovanju s podjetji Petrol, Celtra, Optilab, Iskratel in drugimi.

Temni vzorci na slovenskih spletnih straneh

Elena Osrajnik¹, Sašo Karakatič¹

¹ Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko, Koroška cesta 46, Maribor, Slovenija
elena.osrajnik@student.um.si, saso.karakatic@um.si

Izvleček

Spletne trgovine poskušajo s številnimi novimi pristopi prepričati potrošnike v nakup izdelka, zato se ti vedno bolj poslužujejo temnih vzorcev, ki na manipulativen način spremenijo želje uporabnika. Temni vzorci so torej načini manipuliranja potrošnikov v nakup izdelkov. V prispevku je pregled sedmih kategorij temnih vzorcev, ki se delijo na nadaljnjih 15 tipov. Empirični del raziskave vsebuje pregled 55 slovenskih spletnih trgovin in analizo uporabe temnih vzorcev v teh trgovinah. Iz zbranih podatkov sta bili narejeni analiza trenutnega stanja slovenskih spletnih trgovin ter primerjava stanja temnih vzorcev na tujih spletnih straneh. Ugotovitve kažejo, da je pojavnost temnih vzorcev na slovenskih straneh primerljiva s pojavnostjo vzorcev na tujih straneh in da se večji del slovenskih spletnih trgovin poslužuje takšnega načina manipulacije potrošnikov.

Ključne besede: manipulacija, oblikovanje spletnih strani, svetovni splet, temni vzorci

Abstract

Online merchants strive to convince consumers to buy their products and resort to the use of strategies for manipulating users' choices. One of the most common strategies is the use of dark patterns, which involve different techniques for manipulating consumers to behave in a specific manner, such as purchasing products. The paper contains an overview of seven categories of 15 different types of dark patterns. We examined 55 Slovenian online shops and analyzed the use of dark patterns on the subject websites. More specifically, our goal was to determine the current state of dark pattern use and compare the results to the use of dark patterns on foreign websites. Our results show that dark patterns occur just as often on Slovenian websites as they do on foreign ones and that most Slovenian online merchants utilize several different dark patterns on their websites to influence consumer behaviour.

Keywords: Manipulation, web page design, World Wide Web, dark patterns

1 UVOD

Izraz «temni vzorci» (angl. dark patterns) je leta 2010 ustvaril ameriški strokovnjak za oblikovanje uporabniške izkušnje, dr. Harry Brignull. Ta je temne vzorce opredelil kot uporabniško izkušnjo (angl. user experience), oblikovano z namenom manipuliranja uporabnika, navadno tako, da uporabnika zavaja v določeno odločitev, za katero se morda brez uporabe temnih vzorcev ne bi odločil (Moran, 2020; Narayanan idr., 2020; Monaco, 2020). Razmah takšnih vzorcev na internetu je bil opazen predvsem po razcvetu

spletne prodaje izdelkov, čeprav so se temni vzorci v podobnih oblikah pojavljali tudi izven spletnega omrežja. Spletne trgovine so začele izrabljati temne vzorce v namene večje prodaje, hitrejšega in enostavnejšega pridobivanja naročnin, doseganja ciljnega števila transakcij ipd., s čimer so zavedno manipulirali s strankami za dobrobit poslovanja podjetja (Moran, 2020; Moser, 2020).

Z naraščanjem povpraševanja po izdelkih narašča tudi njihova ponudba, zato želijo spletni trgovci z iznajdljivimi pristopi povečati prodajo. Takšni pristopi

so med drugimi tudi temni vzorci, ki so ustvarjeni z namenom, da prepričajo potrošnika v nakup izdelkov ali storitev (Gray idr., 2020; Maier in Harr, 2020). Z namenom odkritja, kateri vzorci in v kolikšni meri so prisotni na straneh slovenskih spletnih trgovin, smo raziskali 55 spletnih strani, ki smo jih razdelili v kategorije glede na izdelke, ki jih ponujajo.

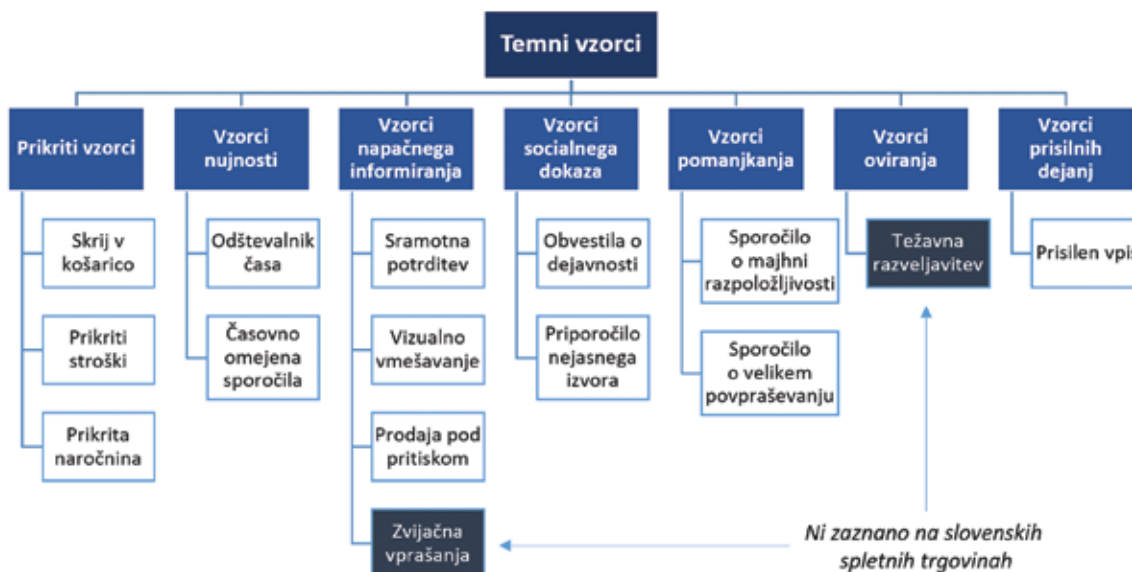
Pred ročnim pregledom spletnih trgovin smo si zastavili par raziskovalnih vprašanj, ki so nam bila v pomoč pri analizi in primerjavi s tujimi spletnimi stranmi. Zanimalo nas je, ali so v slovenskih spletnih trgovinah prisotni vsi tipi temnih vzorcev, ki jih lahko opazimo na tujih spletnih straneh. Prav tako nas je zanimalo tudi, ali je v slovenskih spletnih trgovinah prisotnih več temnih vzorcev, kot jih lahko opazimo na tujih spletnih straneh. Glavni prispevek raziskave je umestitev uporabe temnih vzorcev v slovenskih spletnih trgovinah v primerjavi z mednarodnimi spletnimi stranmi. Prispevek služi raziskovalcem uporabniške izkušnje kot osnovna predstavitev temnih vzorcev in pregled trenutne razširjenosti teh na trgu slovenskih spletnih trgovin. Hkrati pa je namenjen ozaveščanju potrošnikov glede sodobnih pristopov manipulacije njihovega obnašanja na spletu, saj se le z zavedanjem in razumevanjem lahko primerno odzovejo na obstoj teh.

V prispevku smo se glede klasifikacije vzorcev našali na mednarodno raziskavo (Mathur idr., 2019), po kateri smo najdene temne vzorce razdelili v kategorije. Po opravljenem pregledu slovenskih spletnih

strani smo rezultate analizirali in jih primerjali s tujo študijo, na katero smo se opirali pri celotni raziskavi. Ugotovitve smo tudi vizualno predstavili s pomočjo posnetkov zaslona spletnih trgovin in vizualne predstavitev rezultatov v obliki grafov. Ugotovitvam raziskave je sledila diskusija glede deleža najdenih primerov temnih vzorcev in možnih razlogih, zakaj smo zaznali odstopanja v številu pri vzorcih, najdenih na slovenskih straneh, v primerjavi z vzorci v tujih spletnih trgovinah. V zadnjem, sklepnem delu smo povzeli ugotovitve pregleda slovenskih spletnih trgovin in jih primerjali z obstoječo literaturo.

2 TEMNI VZORCI NA SPLETNIH STRANH

Temni vzorci so uporabniški vmesniki, oblikovani tako, da načrtno zavajajo uporabnika pri sprejemanju določenih odločitev. Vzorci izkoriščajo kognitivne pristranskosti, s čimer spodbujajo uporabnike k nakupu nepotrebnih izdelkov ali storitev ter razkrivanju osebnih informacij. S kognitivnimi pristranskostmi oblikovalci vplivajo na nezavedne odločitve potrošnikov in na njihovo nakupno vedenje, pri čemer izkoriščajo hevristični način mišljenja (Di Geronimo idr., 2020; Maier in Harr, 2020; Narayanan idr., 2020; Monaco, 2020). Oblikovanje temnih vzorcev izhaja tudi iz dobrega poznavanja potrošniške psihologije, ki temelji na preučevanju nakupovalnih navad potrošnikov, torej kaj kupujejo, predvsem pa zakaj kupujejo določene izdelke (Moser, 2020). Namen temnih vzorcev je torej manipuliranje uporabnikov



Slika 1: Shema temnih vzorcev

k izvajanju določenih dejanj, ki so nekonsistentna z njihovimi preferencami (Mathur idr., 2019; Luguri in Strahilevitz, 2019; Maier in Harr, 2020).

Za razvrstitev temnih vzorcev v različne kategorije smo uporabili klasifikacijo, ustvarjeno v širši raziskavi raziskovalne skupine Mathur s sodelavci (Mathur idr., 2019). V tej raziskavi so s pomočjo spletnega pajka preiskali več kot enajst tisoč spletnih trgovin, da bi ugotovili, kateri temni vzorci se v njih pojavljajo. Z analizo pridobljenih vzorcev so te razvrstili v 7 širših kategorij ter podrobneje še v 15 vrst. Za vsako vrsto vzorcev so tudi določili, katero kognitivno pristranskost izkorišča pri uporabnikih, in jo opredelili glede na pet dimenzij oblikovanja. Na spodnji Sliki 1 je prikazana shema temnih vzorcev. S sivo barvo so označeni vzorci, ki jih nismo zaznali pri pregledu slovenskih spletnih trgovin.

2.1 Prikriti vzorci

Kategorija prikritih vzorcev (angl. sneaking) se nanaša na temne vzorce, ki poskušajo napačno interpretirati uporabnikova dejanja ter skriti ali preložiti ključne informacije, zaradi katerih bi uporabnik najverjetneje ravnal drugače ali nasprotoval določenim dejanjem. V tej kategoriji so bili temni vzorci razdeljeni v tri vrste: skrij v košarico (angl. sneak into basket), prikriti stroški (angl. hidden costs) in prikrita naročnina (angl. hidden subscription) (Mathur idr., 2019).

SKRIJ V KOŠARICO

Za ta vzorec je značilno, da se v nakupovalno košarico potrošnika priloži dodaten izdelek brez njegovega

dovoljenja, dodaten izdelek pa je promoviran v obliki bonusa ali kot neizogiben strošek. Vzorec izrablja učinek samodejnosti pri potrošnikih, pri čemer prodajalci upajo, da bo uporabnik prezrl dodaten izdelek in končal nakup (Mathur idr., 2019).

Takšni vzorci so glede na razdelitev v dimenzije delno varljivi, saj nepravilno predstavljajo idejo dodajanja izdelka v nakupovalno košarico, ter skrivajoč informacije, saj pred potrošniki načrtno prikrivajo način dodajanja dodatnega izdelka v košarico. Vendar pa ti vzorci niso prikriti, saj lahko potrošnik takoj opazi dodaten izdelek in spozna, da ga je v nakup vključila spletna trgovina (Mathur idr., 2019). Na Sliki 2 je viden temni vzorec skrivanja izdelka v košarico, ki smo ga zasledili v spletni trgovini gizzmo.si. Na sliki je razvidno, da so bili zaradi dovolj visoke vrednosti nakupa v košarico dodani trije izdelki, katerih kupec sam ni izbral in jih prav tako ni mogel odstraniti.

PRIKRITI STROŠKI

Vzorec pred nakupom razkrije dodatno in pogosto nenavadno visoko doplačilo, tik preden potrošnik zaključi z nakupovanjem. Stroški se razkrijejo, tik preden potrošnik plača izdelek, torej zatem, ko je moral vnesti vse informacije o pošiljanju in plačilu ter ko se je strinjal s pogoji uporabe. Ta doplačila pogosto vsebujejo pristojbino za dodatne stroške storitve ali stroške ravnanja. Ta vrsta temnih vzorcev izkorišča kognitivno pristranskost zmote potopljenih stroškov, saj uporabniki navadno sprejmejo in plačajo dodatne stroške, ker so v nakup izdelka vložili veliko časa, energije in volje (Mathur idr., 2019).



Slika 2: Temni vzorec skrij v košarico

Vzorci prikritih stroškov so klasificirani kot vsaj delno varljivi, saj poskušajo pred potrošniki skriti ali minimizirati poslane informacije, prav tako pa jih opredelimo kot skrivajoč informacije, saj prikrivajo ključne informacije pred potrošniki. Tako kot prejšnja vrsta vzorcev tudi ti niso prikriti, saj lahko uporabniki natančno ugotovijo, kdaj so bili v nakupovalno košarico dodani dodatni stroški (Mathur idr., 2019). Na Sliki 3 je prikazan vzorec prikritih stroškov s strani mojacokolada.si, kjer je bil kupec šele pri zadnjem koraku nakupa seznanjen z dodatnimi stroški dostave, ki predhodno niso bili nikjer predstavljeni.

PRIKRITA NAROČNINA

Ta vrsta vzorcev naloži potrošniku ponavljajočo se pristojbino pod lažno krinko enkratne pristojbine ali celo brezplačnega poskusa. Če se uporabniki zavejo takšne manipulacije, se teh dodatnih stroškov zavejo šele čez nekaj dni ali mesecev po nakupu (Mathur idr., 2019).

Glede na klasifikacijo temnih vzorcev se lahko ta vrsta opredeli kot delno varljiva, saj zavaja uporabnike o naravi prvotne ponudbe, ter kot skrivajoč informacije, saj pred uporabniki zadržuje informacije o ponavljajočih se dodatnih pristojbinah (Mathur idr., 2019).

2.2 Vzorci nujnosti

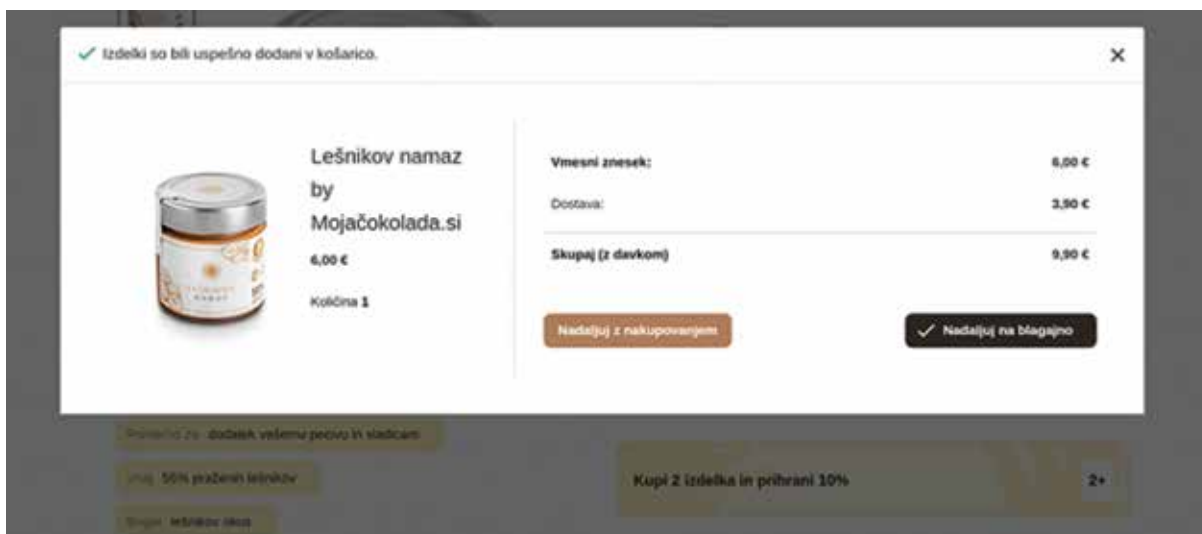
V kategorijo vzorcev nujnosti (angl. urgency) spada tisti temni vzorci, ki potrošniku vsiljujejo skrajni rok prodaje izdelka ali storitve, s čimer pospešujejo

uporabnikov proces odločanja in posledično nakupa. Vzorci so ustvarjeni tako, da izkoriščajo kognitivno pristranskost pomanjkanja, saj uspejo uporabnika prepričati, da je popust ali ponudba bolj privlačna, kot bi bila drugače, prav tako pa mu sporočajo, da bi z nedejavnostjo izgubil potencialen prihranek. Na strani spletne prodaje se navadno pojavljajo v obliki razprodaj, kuponov ali prodaj določenih izdelkov, na straneh košarice pa se pojavljajo v obliki končnega roka o rezervaciji produkta. Temni vzorci so v tej kategoriji razdeljeni na odštevalnike časa (angl. countdown timers) in časovno omejena sporočila (angl. limited-time message) (Mathur idr., 2019). Takšne vzorce nujnosti lahko vidimo na Sliki 4, kjer so vidni popusti za vsak izdelek v rdečem krogu na desni strani izdelka.

ODŠTEVALNIK ČASA

Odštevalnik časa je dinamični indikator skrajnega roka nakupa določenega izdelka, saj odšteva čas, ki ga ima na voljo uporabnik do izteka določene ponudbe. Takšni vzorci postanejo varljivi, če se odštevalnik ponastavi po izteku roka, pri čemer je enaka ponudba še vedno na voljo. Prav tako je varljiv, če se odštevalnik izteče, pa je ponudba, ki naj bi potekla, po roku še vedno veljavna (Mathur idr., 2019).

Po ustvarjeni klasifikaciji bi takšne temne vzorce lahko opredelili kot delno prikrite, saj umetno povečujejo občutek neodložljivosti, ter kot včasih zavajajoče, saj zavajajo uporabnike o ponudbi, ki naj bi že potekla, pa je v resnici še vedno veljavna (Mathur idr., 2019). Na Sliki 5 je predstavljen vzorec odšte-



Slika 3: Temni vzorec prikriti stroški



Slika 4: Temni vzorec prikriti stroški

vanja časa, ki smo ga zasledili na spletni strani postquam.org. Ta vzorec je sporočal o časovno omejeni ponudbi, kjer se je odštevalnik po preteklelem času malo več kot 10 minut ponovno samodejno ponastavil na začetni čas.

ČASOVNO OMEJENA SPOROČILA

Ti vzorci predstavljajo statična nujna sporočila brez dodanega skrajnega roka o izteku ponudbe. Brez podatka, do kdaj velja ponudba, spletna stran zadržuje informacije pred potrošniki in s tem napačno prikazuje naravo ponudbe, saj ni znano, do kdaj bo veljavna (Mathur idr., 2019).

Časovno omejena sporočila lahko opredelimo kot vsaj delno prikrita, pri čemer so zelo podobna odštevalnikom časa, ki v uporabniku spodbujajo

nujnost nakupa, ter kot skrivajoča informacije, saj s tem, ko ne razkrijejo roka izteka ponudbe, ne predstavijo vseh informacij, na podlagi katerih bi si uporabnik lahko ustvaril realno mnenje o nakupu izdelka (Mathur idr., 2019). Na Sliki 6 je vidno časovno omejeno sporočilo, ki smo ga odkrili na strani hajdi.si. Ta najden vzorec prikazuje ponudbo, za katero pa ne vemo kdaj poteče, torej zanjo ni napisan skrajni rok veljavnosti.

2.3 Vzorci napačnega informiranja

Vzorci napačnega informiranja (angl. misdirection) uporabljajo slikovno gradivo, manipulativne zapise s vplivom na emocije za preusmeritev uporabnikov od njihove prvotne odločitve ali pa jim vsilijo drugo odločitev. Delujejo z izkoriščanjem različnih dejavnikov



Slika 5: Temni vzorec odštevalnik časa



Slika 6: Temni vzorec časovno omejeno sporočila

manipulacije in kognitivnih pristranskosti, ne da bi dejansko omejili izbire, ki so na voljo uporabniku. Kategorija je po klasifikaciji razdeljena v štiri vrste vzorcev, sramotno potrditev (angl. confirmshaming), vizualno vmešavanje (angl. visual interference), zvižajna vprašanja (angl. trick questions) in prodajo pod pritiskom (angl. pressured selling) (Mathur idr., 2019).

SRAMOTNA POTRDIČEV

Ta vrsta vzorcev uporablja način izražanja in čustva tako, da vodi uporabnika stran od določene odločitve. Pogosto se pojavlja v prikaznih dialogih, kjer od uporabnika zahteva e-naslov v zameno za popuste, kjer je možnost zavrnitve ponudbe predstavljena kot sramotna odločitev. Takšna besedila so po navadi oblikovana kot: »Ne, hvala, rad bi plačal polno ceno,« ali pa: »Ne, hvala, ne maram varčevati denarja.« S takšno predstavitvijo negativne možnosti ti vzorci izkoriščajo kognitivno pristranskost učinka

uokvirjanja in tudi občutek sramu, ki je učinkovito sredstvo za spreminjanje vedenja (Mathur idr., 2019).

Glede na klasifikacijo se sporočila sramotne potrditve razvrščajo kot nesimetrična, saj je možnost opustitve predstavljena tako, da v potrošnikih vzbuja občutek sramu, ki se mu uporabniki želijo izogniti. Vzorci sramotne opustitve niso prikriti, ker lahko potrošnik jasno vidi in dojame, da je vzorec oblikovan tako, da vpliva na odločitve posameznika (Mathur idr., 2019). Na Sliki 7 je predstavljen vzorec sramotne potrditve, ki smo ga našli na spletni strani fenzy.si. Slika prikazuje način manipuliranja s stranko, saj ima potrošnik možnost predati svoje osebne podatke v zameno za darilno kodo, ali pa zavrniti ponudbo s klikom na gumb »Ne hvala, popusti me ne zanimajo«. Teža večina potrošnikov ne bo storilo rade volje, saj je besedilo napisano tako, da je klik na gumb zastavljen kot sramotna izbira.



Slika 7: Temni vzorec sramotna potrditev

VIZUALNO VMEŠAVANJE

Vzorci vizualnega vmešavanja z jezikovnim slogom in vizualno predstavitvijo vplivajo na odločitev posameznika, in sicer z namenom, da bi se ta odločil za drugo možnost, kot bi se odločil sicer. Vzorec ustvarja videz, da je ena možnost izbire dosti pomembnejša od ostalih in zato bolj privlačna za potrošnika. V nekaterih primerih so ti vzorci oblikovani tako, da s posebnimi vizualnimi učinki pritegnejo pozornost k besedilu, ki predstavlja popuste za določene izdelke. Nekateri vzorci so oblikovani tako, da je možnost zavrnitve ponudbe prikazana z manjšim besedilom ali slabo vidno barvo pisave, s čimer se ustvari iluzija, da ta možnost ni na voljo ali je onemogočena, čeprav uporabnik lahko izbere (Mathur idr., 2019).

To vrsto temnih vzorcev lahko opredelimo kot včasih nesimetrično, saj vzorci v določenih primerih ustvarjajo neenako možnost izbire, pri čemer vodijo uporabnika k izbiri, ki morda ni v skladu z njegovimi željami. Prav tako pa jih lahko opredelimo kot prikrite, saj potrošnik večinoma ne ugotovi, da je vizualna predstavitev vseh možnosti vplivala na njegovo odločitev (Mathur idr., 2019). Na Sliki 8 je viden temni vzorec vizualnega vmešavanja, ki smo ga odkrili na spletni strani malinca.si. Ta je viden na pojavnem okencu »Obvestilce«, kjer je možnost prijave na prejemanje obvestil vizualno bolj izpostavljena, kot zavrnitev prijave, ki izgleda celo, kot da na gumb ni možno klikniti.

ZVIJAČNA VPRAŠANJA

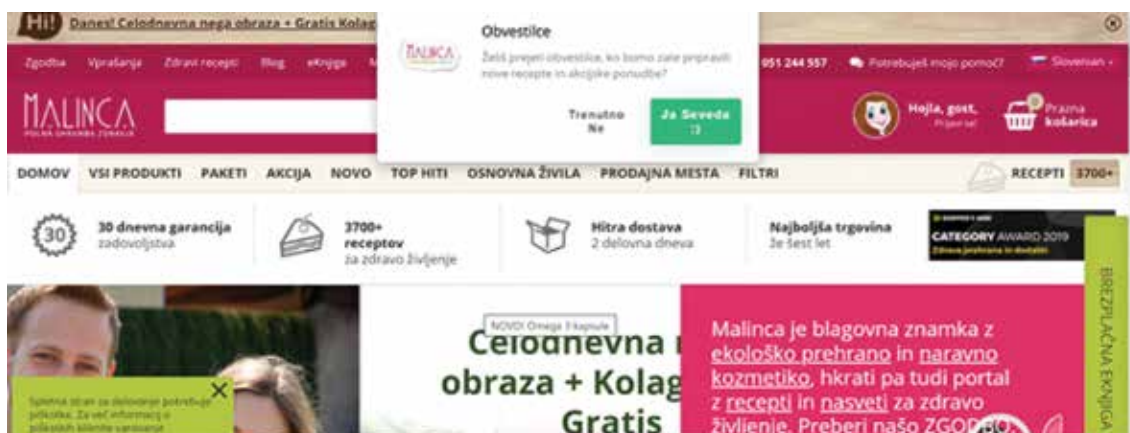
Zvijačna vprašanja uporabljajo nejasen in zapleten jezik, s katerim zmanipulirajo uporabnika v odločitev

za točno določeno izbiro. S takšnim načinom oblikovanja želijo tržniki prepričati uporabnike, da se odločijo za marketinška in promocijska sporočila, tako da s subtilnimi sporočili spremenijo celoten odjavni postopek. Najpogosteje so uporabljena dvojno zanimkana sporočila, kot recimo: »Ne izberite polja, če ne želite prejeti novic,« ali pa je uporabljena negacija besedila, ki spreminja potek določene akcije, kot je potrditev polja za izključitev. Zvijačna vprašanja izkoriščajo kognitivno pristranskost učinka samodejnosti in uokvirjanja, saj uporabniki postanejo bolj dovzetni za vsiljeno izbiro, za katero verjamejo, da je v skladu z njihovimi preferencami (Mathur idr., 2019).

Te temne vzorce lahko opredelimo kot asimetrične, saj je izbira odjave bolj obremenjena kot pa izbira potrditve, prav tako pa so vzorci prikriti, saj uporabnik ne more ugotoviti učinka, ki ga ima njegova odločitev, kar je posledica zavajajoče uporabe jezika (Mathur idr., 2019).

PRODAJA POD PRITISKOM

Vzorec se nanaša na privzete nastavitve situacije ali pogoste taktike prodaje izdelka pod velikim pritiskom, s čimer uporabnika vodijo do nakupa dražjega izdelka, kot ga je nameraval kupiti, ali pa k nakupu dodatnih, podobnih izdelkov. Temni vzorec izkorišča številne kognitivne pristranskosti, kot so učinek samodejnosti, kjer je izdelek izbran samodejno, učinek sidra, kjer je najdražja možnost predstavljena na začetku kot osnova za nadaljnje primerjanje cen, ter pristranskost pomanjkanja, kjer je izdelek predstavljen kot redek ali da je po njem veliko povpraševanje (Mathur idr., 2019).



Slika 8: Temni vzorec vizualno vmešavanje

Glede na klasifikacijo vzorcev lahko vzorce prodaje pod pritiskom označimo kot včasih nesimetrične, saj pritiska na uporabnike, da kupijo dražje možnosti izdelka, ter kot vsaj delno prikrite, saj uporabnik ne ugotovi, da je kupil dražji izdelek, kot ga je nameraval, ker je bila njegova osnova za primerjavo ostalih izdelkov postavljena zelo visoko (Mathur idr., 2019). Na Sliki 9 je predstavljen vzorec prodaje pod pritiskom, ki smo ga zasledili na spletni strani vitapur.si. Na sliki je razvidno, da je kot prvotna vizualno poudarjena in predstavljena dražja različica izdelka (cena 29,95€), cenejša (17,96€) je celo vizualno predstavljena, kot da nanjo ni možno klikniti.

2.4 Vzorci socialnega dokaza

Socialni dokaz (angl. social proof) je kategorija temnih vzorcev, ki zavaja uporabnike k nakupu določenega izdelka zaradi socialnega pritiska drugih potrošnikov, torej ker ostali kupujejo določen izdelek, ga mora tudi potrošnik sam. Posamezniki bodo lastna dejanja presojali glede na odločitve in vedenja drugih, zato ta kategorija vzorcev uporablja sporočila in komentarje drugih potrošnikov kot tehten razlog prepričevanja. Takšni vzorci vplivajo na odločitve potrošnika in jih poskušajo pospešiti, pri čemer izkoriščajo učinek pridružitve, saj se uporabniki pri odločanju zanašajo na mnenja drugih ljudi. Kategorija se deli na dve vrsti, in sicer na obvestila o dejavnosti (angl. activity notifications) in na priporočila nejasnega izvora (angl. testimonials of uncertain origin) (Mathur idr., 2019).

OBVESTILA O DEJAVNOSTI

Obvestila o dejavnosti so pogosto ponavljajoča se sporočila, ki privabljajo pozornost uporabnika in se

pojavi ob določenem izdelku z namenom sporočanja o aktivnostih drugih uporabnikov, ki si ogledujejo ta izdelek ali pa so ga že kupili. Razdeljena so v tri kategorije glede na njihovo vsebino. Lahko so dinamična in ponavljajoča se sporočila, ki sporočajo, da je drug uporabnik ravno kupil izdelek, ki si ga je posameznik ogledoval. Lahko so statična ali dinamična besedilna sporočila, ki sporočajo, koliko ljudi je že dodalo točno ta izdelek v svojo nakupovalno košarico (Mathur idr., 2019).

Obvestila o dejavnosti postanejo zavajajoča, če je predstavljena vsebina, ki vključuje imena, lokacijo in štetje uporabnikov, lažno ustvarjena ali vsebuje neresnična dejstva in informacije. Takšna sporočila so temeljila na naključnih generatorjih, kjer so bili podatki pridobljeni naključno in so se po osvežitvi strani zamenjali, ali pa na vnaprej kodiranih podatkih, ki so po osvežitvi ostali enaki (Mathur idr., 2019).

Po karakteristiki lahko takšne vzorce opredelimo kot delno prikrite, saj uporabnik težko ugotovi njihov učinek na lastna dejanja, če se sporočila pojavljajo na celotni spletni strani. Prav tako pa jih lahko označimo kot včasih zavajajoče, saj je lahko vsebina sporočila varljivo ustvarjena in zavajajoča (Mathur idr., 2019). Na Sliki 10 je vidno obvestilo o dejavnosti, ki smo ga našli na spletni strani popolnapostava.com. V spodnjem levem kotu slike lahko vidimo temni vzorec, ki prikazuje kdo je kupil kateri izdelek in s tem spodbudi potrošnike k nakupu, saj ustvari navidezno veliko povpraševanje po izdelkih.

PRIPOROČILO NEJASNEGA IZVORA

Ta vrsta vzorcev temelji na uporabi priporočil uporabnikov, katerih izvor ali način pridobitve sporočila ni jasen in točno naveden. Priporočila uporabnikov



Slika 9: Temni vzorec prodaje pod pritiskom



Slika 10: Temni vzorec obvestila o dejavnosti

so navadno izmišljena, če na spletni strani ni obrazca za vnos priporočila, ki bi ga potrošnik lahko vpisal, in če se enako sporočilo, vendar z različnimi podatki o imenu in lokaciji, pojavi tudi na drugih spletnih mestih (Mathur idr., 2019). Na Sliki 11 je viden primer priporočila nejasnega izvora na strani konoplja.net. Sicer so predstavljena mnenja strank, ki so navedena z imenom, vendar je način pridobitve teh mnenj nejasen, torej na spletni strani nimajo obrazca, kamor bi stranka sploh lahko vpisala svoje mnenje.

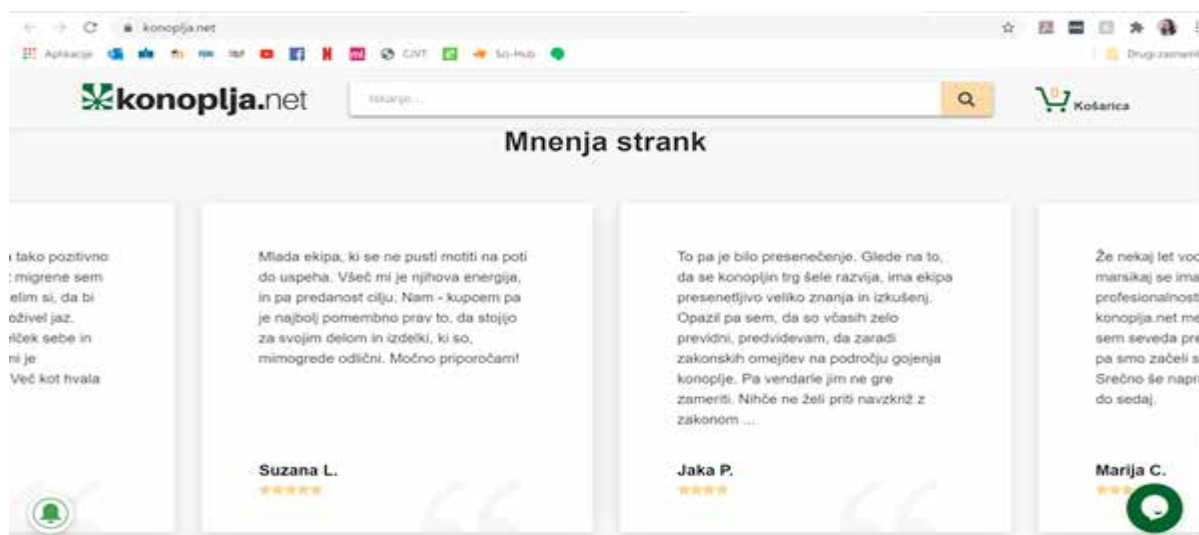
2.5 Vzorci pomanjkanja

V to kategorijo temnih vzorcev (angl. scarcity) spadajo vzorci, ki opozarjajo na omejeno zalogo ali ve-

liko povpraševanje po produktu, s čimer se poveča njegova zaznavna vrednost in zaželenost. Takšni vzorci nakazujejo na omejeno razpoložljivost izdelka ali veliko povpraševanje po njem in da bodo zaloge izdelka hitro pošle. Kategorija se razdeli na dve vrsti vzorcev, na sporočila o majhni razpoložljivosti (angl. low-stock messages) in na sporočila o velikem povpraševanju (angl. high-demand messages) (Mathur idr., 2019).

SPOROČILO O MAJHNI RAZPOLOŽLJIVOSTI

Takšna sporočila sporočajo uporabnikom o omejenih količinah izdelka, prikazujejo njegovo natančno količino zaloge v spletni trgovini ali pa, ali je ta zaloga



Slika 11: Temni vzorec priporočilo nejasnega izvora

majhna oziroma velika. Njihov namen je povečati zaželenost izdelkov in zavesti uporabnike v impulzivno nakupovanje zaradi občutka negotovosti o količini izdelka. Sporočila o majhni razpoložljivosti postanejo zavajajoča, ko prikazujejo naključno število zaloge izdelka ali pa se ta številka po vsaki osvežitvi spletne strani zmanjšuje po določenem vzorcu ali načrtu. Sporočila so prav tako zavajajoča, če vsebujejo besedila o razprodanem izdelku, čeprav izdelka ni bilo več na zalogi v prejšnjih dneh (Mathur idr., 2019).

Po klasifikaciji lahko sporočila o majhni razpoložljivosti opredelimo kot delno prikrite vzorce, saj pri uporabniku ustvarjajo povečan vzgib nakupa, ne da bi se uporabniki tega zavedali. Včasih so takšna sporočila lahko tudi zavajajoča, saj zavajajo uporabnike v mišljenje, da je količina izdelka na zalogi zelo majhna, čeprav v resnici ni, s čimer ustvarjajo neresnično pomanjkanje. Opredelimo pa jih lahko tudi kot skrivajoča informacije, če na spletni strani ni navedena točna količina izdelkov, ki so na zalogi (Mathur idr., 2019). Na Sliki 12 je vidno sporočilo o majhni razpoložljivosti, ki smo ga zasledili na spletni strani mimovrste.com. Pod vsakim izdelkom je prikazano koliko kosov določenega izdelka je še na voljo. Čim nižja je prikazana številka, tem hitreje se bo potrošnik odločil za nakup, saj je prikazano, kot da izdelek ne bo več dolgo časa na voljo.

SPOROČILO O VELIKEM POVPRASEVANJU

Ta vrsta vzorcev nakazuje uporabnikom, da je po iskanem izdelku veliko povpraševanje, s čimer sporočilo namiguje, da bo izdelek kmalu razprodan, zato

mora potrošnik pohiteti z naročilom. Takšni temni vzorci se lahko pojavljajo konsistentno, ne glede na to, kateri izdelek si posameznik ogleduje, ali na produkte, ki so že v nakupovalni košarici. Sporočila o velikem povpraševanju lahko, tako kot sporočila o majhni razpoložljivosti, opredelimo kot delno prikrite vzorce, saj uporabniku ne predstavijo dejanskega stanja, s čimer povečujejo željo po nakupu izdelka (Mathur idr., 2019). Na Sliki 13 je vidno sporočilo o velikem povpraševanju, ki smo ga odkrili na spletni strani aboutyou.si. V zgornjem desnem kotu je prikazano pojavno okno v obliki sporočila, ki kupca opozarja na veliko povpraševanje po izdelkih, ki jih ima sam v košarici. Takšna sporočila kupca spodbudijo k dejanskemu nakupu, saj mu sugerirajo, da je zanimanje za nakup veliko, kar pomeni, da željen izdelek zaradi povečanega povpraševanja kmalu ne bo več na voljo.

2.6 Vzorci oviranja

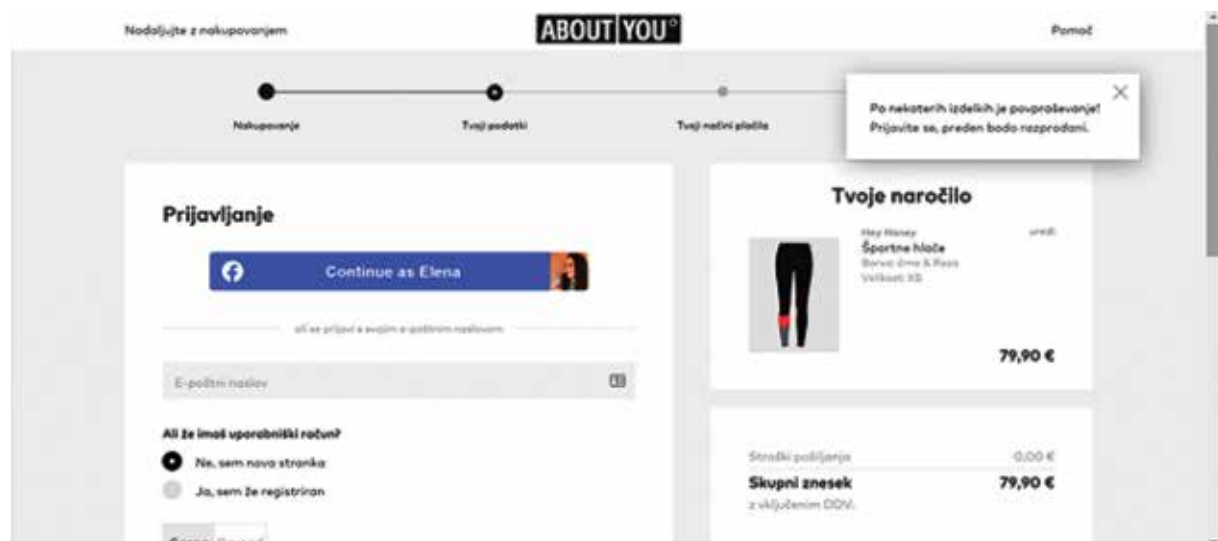
Temni vzorci, ki spadajo v kategorijo vzorcev oviranja (angl. obstruction), naredijo določena dejanja zahtevnejša, kot dejansko so, z namenom odvritve uporabnikov od izvajanja tega dejanja. Vsi takšni vzorci spadajo v eno vrsto, in sicer težavno razveljavitev (angl. hard to cancel) (Mathur idr., 2019).

TEŽAVNA RAZVELJAVITEV

Pri teh vzorcih se uporabniki zelo enostavno naročijo na neko storitev, recimo prejemanje novic nekega spletnega mesta ali plačevanje naročnine, vendar se kasneje zelo težko odjavijo od prejemanja novic in me-



Slika 12: Temni vzorec sporočila o majhni razpoložljivosti



Slika 13: Temni vzorec sporočilo o velikem povpraševanju

sečnega plačila. Velikokrat spletne trgovine ne predstavijo predhodno, da bo odjava napornejša od prijave oziroma da se uporabniki ne bodo mogli odjaviti na enak način, kot so se prijavi (Mathur idr., 2019).

Glede na karakteristike lahko vzorce težavne razveljavitve definiramo kot restriktivne, saj omejujejo možnosti izbire, ki jih lahko uporabimo za preklic določenih storitev. V primerih, kjer način odjave od storitev ni vnaprej predstavljen, pa lahko vzorce označimo kot skrivajoč informacije, saj v sporočilu ni navedeno, kako je odjava težja od prijave (Mathur idr., 2019).

2.7 Vzorci prisilnih dejanj

Vzorci prisilnih dejanj (angl. forced action) zahtevajo od uporabnikov dodatna in izrazita dejanja, da lahko dokončajo neko nalogo. Večinoma se izražajo v obliki dodatnih informacij, ki jih mora uporabnik vnesti na spletno stran v zameno za dokončanje določene dejanja. Ta kategorija ima samo eno vrsto temnih vzorcev, to je prisilen vpis (angl. forced enrollment) (Mathur idr., 2019).

PRISILEN VPIS

Ti vzorci eksplicitno prisilijo uporabnike v strinjanje z marketinškim komuniciranjem spletne trgovine ali pa si morajo uporabniki ustvariti uporabniški račun in na spletno mesto vnesti določene osebne podatke, da lahko nadaljujejo z nekim dejanjem. Ker ti vzorci delujejo na sistemu vse ali nič, torej nadaljevanje z dejanjem je možno le z neko dodatno akcijo, spletne

trgovine od potrošnikov pridobijo dodatne informacije in si ustvarijo bazo prejemnikov novic in oglasov (Mathur idr., 2019).

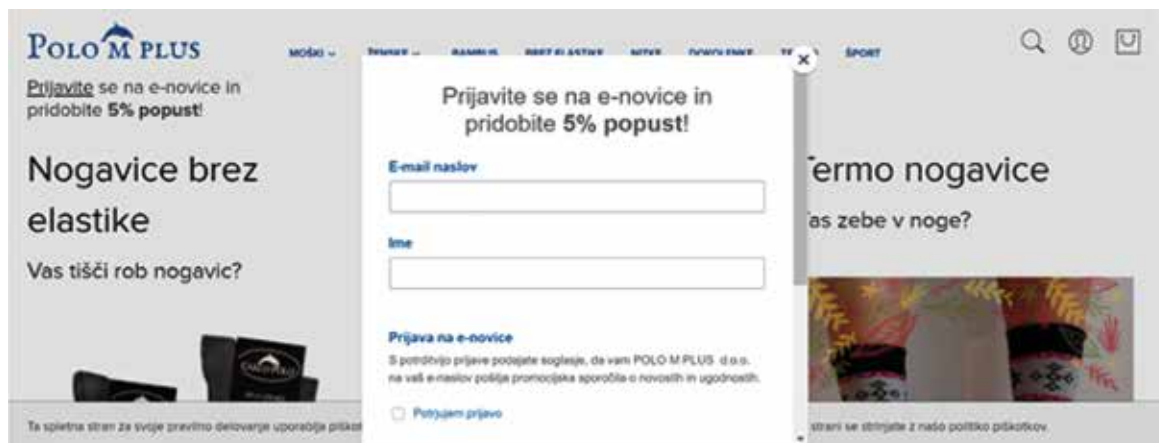
Glede na klasifikacijo lahko vzorce prisilnega vpisa definiramo kot nesimetrične, saj potrebujejo za dokončanje nekega dejanja dodatne naloge, kar ustvarja neenakost izbir. Prav tako so vzorci lahko restriktivni, saj od uporabnika zahtevajo strinjanje z marketinškim komuniciranjem ali ustvarjanje uporabniškega računa potrošnika, čeprav ta kasneje ne bo ničesar kupil (Mathur idr., 2019). Na Sliki 14 je viden vzorec prisilnega vpisa, ki smo ga našli na spletni strani polomplus.si.

3 PREGLED TEMNIH VZORCEV V SLOVENSkih SPLETNIH TRGOVINAH

Sledi predstavitev pregleda uporabe temnih vzorcev v slovenskih spletnih trgovinah. Že v prejšnji sekciji smo vsak vzorec predstavili iz primerov slovenskih spletnih trgovin, kar pomeni, da so vzorci prisotni tudi na slovenskem spletu – raziskava v nadaljevanju pa je služila temu, da smo pregledali, v kolikšni meri so vzorci prisotni na slovenskem tržišču in kako pogosto v primerjavi z vzorci na tujih spletnih straneh.

3.1 Metodologija

Pred začetkom raziskovalnega dela smo izbrali vzorec slovenskih spletnih trgovin, v katerih smo kasneje preverjali prisotnost temnih vzorcev. Trgovine smo izbrali na podlagi njihove kvalitete in poznanosti. Pri tem smo si pomagali z izborom tekmovanj



Slika 14: Temni vzorec prisilen vpis

za Spletnega trgovca iz let 2016, 2017, 2018 in 2019 (Shopper's Mind, 2016; Shopper's Mind, 2017; Gorjan, 2018; Kovačević, 2019), kjer so bile predstavljene najboljše spletne trgovine po izboru strokovne komisije. Prav tako smo pregledali seznam spletnih trgovin v društvu spletnih trgovcev in imenik slovenskih spletnih trgovin, bili pa smo tudi pozorni na mlada podjetja, ki so v zadnjih letih postala popularna in poznana ali pa to trenutno postajajo.

Za lažjo in enostavnejšo analizo temnih vzorcev smo pregledanih 55 spletnih strani razvrstili v osem kategorij glede na vsebino, ki jo trgovine ponujajo. To smo storili tudi zato, da bi iz čim več raznolikih spletnih trgovin pridobili ustrezen in reprezentativen vzorec. Te kategorije so oblačila in obutev, lepota in zdravje, šport, hišni ljubljenci, nakit, elektronika in tehnika, dom in družina ter drugo. Vse pregledane spletne strani in njihova kategorizacija so prikazane v Tabeli 1.

Pregled slovenskih spletnih trgovin je bil izveden v obdobju treh tednov, med 14. 6. 2020 in 5. 7. 2020.

Izveden je bil tako med tednom kot tudi med vikendom. Spletne strani so bile pregledane s pomočjo spletnega brskalnika Google Chrome, verzija 83, pri čemer niso bili uporabljeni nobeni dodatni vtičniki.

3.2 Proces pregleda spletnih strani

Pregled spletnih strani se je začel na začetni strani spletne trgovine, kjer smo bili pozorni predvsem na temne vzorce, kot so odštevalniki časa, časovno omejena sporočila in prisilen vpis. Nadalje smo pregledali spletno stran s ponudbo izdelkov, kjer smo najprej preverili celotno ponudbo, pri čemer smo bili pozorni na temne vzorce, kot so sporočila o majhni razpoložljivosti in velikem povpraševanju ter priporočila nejasnega izvora. Nato smo si ogledali dodatne informacije o posameznem izdelku, torej stran, na kateri je bilo možno naročiti izdelek, določiti njegovo velikost, barvo in količino oz. izvedeti podrobnosti o storitvi. Na tej stopnji pregleda smo preverjali predvsem temne vzorce, kot so obvestila o dejavnosti drugih potrošnikov, vizualno vmešavanje, prodaja

Tabela 1: Kategorizacija pregledanih 55 spletnih strani

Kategorija spletne strani	Spletne strani
Oblačila in obutev	aboutyou.si, alpinashop.si, bibloo.si, fenzy.si, lepsiful.com, papina.si, polomplus.si, ursanina.si
Lepota in zdravje	click2chic.si, ecco-verde.si, licila.si, popolnapostava.com, postquam.org, spleticna.si, vitalabo.si
Šport	bike-center.si, extremevital.com/sl, iglusport.si, intersport.si, optics-trade.eu/si, rossisport.si, tomassport2.si
Hišni ljubljenci	abc-zoo.si, mrpet.si, natis-shop.com/si, platinum.si, zoohit.si
Nakit	fatalka.si, irisimo.si, nakit-kamen.si, slowatch.si, spletna-zlatarna.si, zlatarnacelje.com
Elektronika in tehnika	anni.si, etuizamobi.si, qizzmo.si, outletshop.si, printink.si, sonusart.si, tehnox.si
Dom in družina	emundia.si, hajdi.si, najindom.si, pokolin.si, salonpohistva.si, vitapur.si
Drugo	1nadan.si, cbd-slovenija.si, conrad.si, kmetijskaoprema.si, konoplja.net, malica.si, mimovrste.com, mojacokolada.si, silux.si

pod pritiskom in odštevalniki časa. Ko smo izdelek dodali v nakupovalno košarico, smo opravili pregled košarice, kjer smo bili pozorni na prikrite stroške in vzorce skrivanja v košarico skozi vse faze nakupa, torej od naročila izdelka do vpisa osebnih podatkov in končnega izračuna stroškov. Na tej točki smo tudi zaključili s pregledom spletnih trgovin, saj nismo kupili nobenega izdelka ali se naročili na kakšno storitev.

Slike temnih vzorcev smo zajeli s pomočjo funkcije zajema slike zaslona ter jih kasneje razvrstili po kategorijah in vrstah, ki bile določene na začetku. Ko smo naleteli na določen temni vzorec, smo torej zajeli sliko in jo shranili. Pri vzorcih odštevanja časa smo najprej zajeli sliko zaslona, nato počakali, da se je napisan čas iztekel, in nato ponovno zajeli zaslon. Če je bila ponudba še vedno veljavna po izteku časa, smo bili prepričani, da je odštevalnik časa zavajajoč temni vzorec. Podobno smo naredili tudi pri vzorcih obvestil o dejavnosti drugih potrošnikov, kjer smo prav tako najprej zajeli sliko zaslona, nato počakali par minut in čez nekaj časa ponovno pregledali stanje potrošnikov. Če se stanje ni spremenilo, smo lahko trdili, da so bila ta obvestila o dejavnosti varljiva. Pri sporočilih o majhni razpoložljivosti pa smo pregled opravili nekoliko drugače, in sicer smo pri izdelkih, kjer je bila označena točno določena količina omejene ponudbe, v košarico dodali več izdelkov, kot jih je bilo na voljo. Če smo nakup lahko opravili, kljub temu da smo nakupili več izdelkov, kot je bilo napisano, smo lahko takšne vzorce opredelili kot zavajajoče temne vzorce o majhni razpoložljivosti.

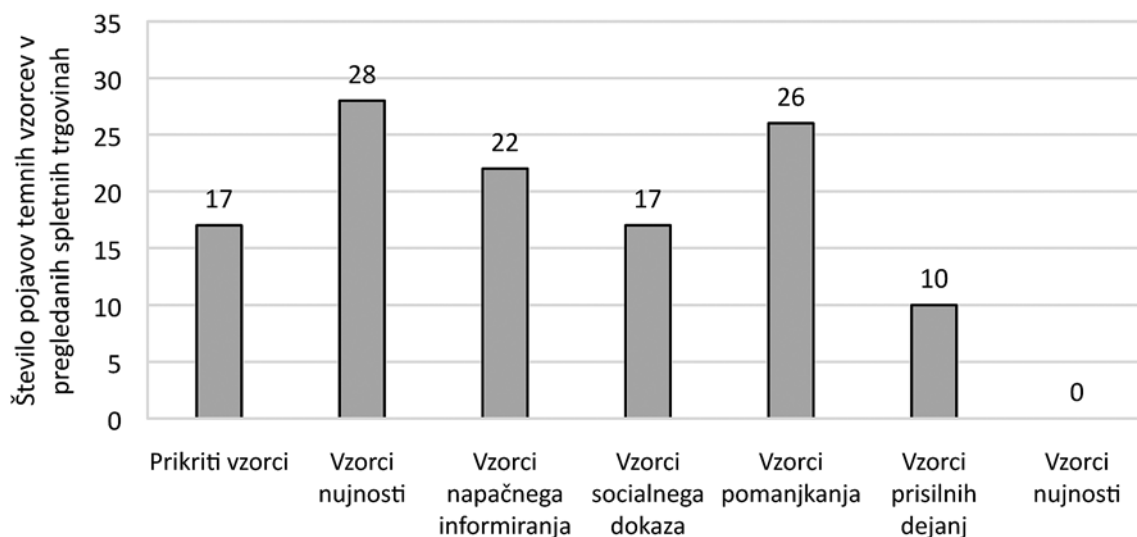
3.3 Rezultati

Po pregledu slovenskih spletnih trgovin smo temne vzorce razvrstili v njihove pripadajoče kategorije in ugotovili, da smo našli 17 prikritih vzorcev, 28 vzorcev nujnosti, 22 vzorcev napačnega informiranja, 17 vzorcev socialnega dokaza, 26 vzorcev pomanjkanja, 10 vzorcev prisilnih dejanj in nobenega vzorca oviranja, kar je prikazano na Grafu 1.

V kategoriji prikritih vzorcev smo zaznali en vzorec skrivanja v košarico, ki smo ga opredelili kot temni vzorec zato, ker se je v košarici pojavil brez vednosti potrošnika. Prav tako smo zaznali 16 primerov prikritih stroškov, navadno dodatnega plačila poštne, ki je bila za prevzem izdelka nujna. Vzorcev prikrite naročnine nismo zaznali na nobeni pregledani spletni strani.

V kategoriji vzorcev nujnosti smo zaznala 11 odštevalnikov časa, ki so odštevali čas do izteka ponudbe ali akcije. Štiri takšne vzorce smo opredelili kot varljive, saj so označevali čas izteka rezervacije izdelka v nakupovalni košarici, nakup katerega je bil mogoč tudi po izteku zapisanega časa, ali pa so prikazovali ponudbo za omejen čas, ki se je naslednji dan ponastavil. V tej kategoriji smo našli tudi 17 časovno omejenih sporočil, ki so prikazovala različna znižanja, vendar brez opredeljenega roka izteka akcije.

V kategoriji vzorcev napačnega informiranja smo razkrili tri vzorce sramotne potrditve, torej sporočila, ki so bila jezikovno oblikovana tako, da potrošnik ni mogel zavrniti sporočila oz. ponudbe, saj je bila ta možnost predstavljena kot sramotna. Prav tako



Graf 1: Razvrstitev temnih vzorcev po kategorijah

smo našli 16 vzorcev vizualnega vmešavanja, kjer je bila predstavljena neenaka izbira možnosti, navadno z različnim oblikovanjem obeh gumbov. Opazili pa smo tudi tri vzorce prodaje pod pritiskom, kjer je bila vnaprej izbrana dražja verzija izdelka, torej večja količina izdelka. Med pregledom nismo opazili nobenega vzorca zvižajčnih vprašanj.

V kategoriji vzorcev socialnega dokaza smo našli šest obvestil o dejavnosti, ki so potrošnika informirale o aktivnostih drugih uporabnikov. Ta sporočila so se navadno prikazala v obliki pojavnih sporočil ali pa kot opozorilo ob izdelku. Opazili pa smo tudi 11 priporočil nejasnega izvora, torej sporočil potrošnikov o nekem izdelku ali spletni strani, vendar pa nismo našli obrazca za vnos mnenja.

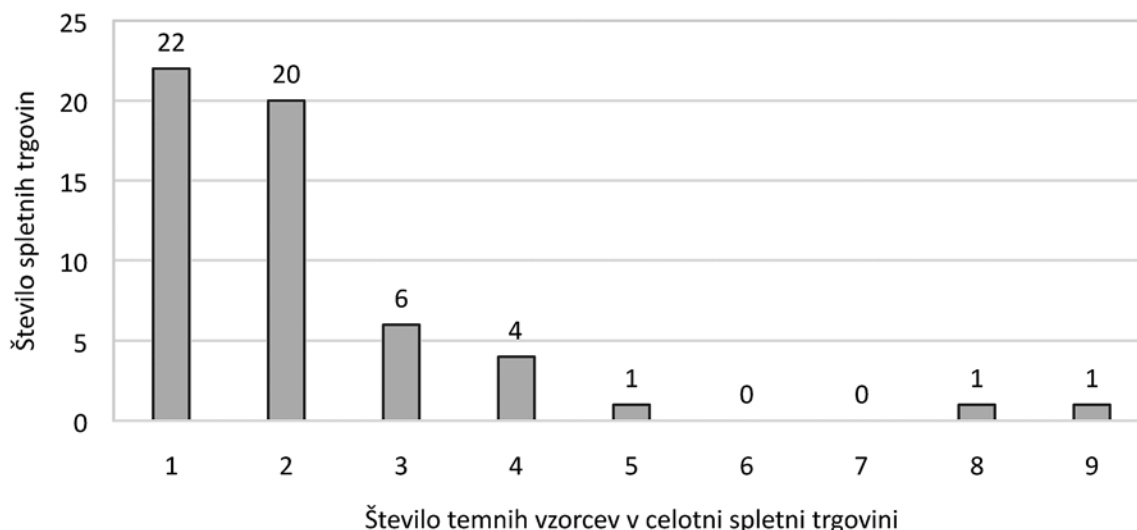
V kategoriji vzorcev pomanjkanja smo odkrili 23 sporočil o majhni razpoložljivosti, ki so nakazovala omejeno količino zaloge ali pa so prikazovala zalogo izdelka. Tukaj smo opazili tudi en vzorec zavajajočega sporočila o majhni razpoložljivosti, ki je prikazoval zalogo izdelka nekonsistentno. Ta se je konstantno zmanjševala, vendar se je ob kliku na osvežitev strani ponovno povečala. V tej kategoriji smo našli tudi tri sporočila o velikem povpraševanju, torej sporočila, ki so nakazovala, da bodo izdelki zaradi velikega števila potrošnikov, ki se zanimajo zanje, hitro pošli.

V kategoriji vzorcev prisilnih dejanj smo zaznali deset vzorcev prisilnega vpisa, ki je od potrošnikov zahteval vpis e-poštnega naslova ali delitev nakupa izdelka na družbenih omrežjih v zameno za popust pri nakupu.

Po zaključenem pregledu smo ugotovili, da ima skoraj polovica preiskanih spletnih trgovin, torej 22 strani, na spletni strani zgolj en temni vzorec. Dvajset spletnih strani je vsebovalo po dva temna vzorca na stran, na šestih straneh smo zasledili po tri vzorce, na štirih straneh pa štiri. Na spletni strani vitapur.si smo opazil pet temnih vzorcev, na strani aboutyou.si smo jih zasledili osem, stran z največ opaženimi temnimi vzorci pa je bila gizzmo.si, na kateri smo našli kar devet temnih vzorcev.

3.4 Primerjava Slovenskih strani s tujimi

Pri pregledu temnih vzorcev smo se opirali na raziskavo avtorjev Mathur in sodelavci (Mathur idr., 2019), ki so s pomočjo spletnega pajka pregledali 11.286 strani spletnih trgovin. Spletni pajek je program, ki samodejno pregleduje spletne strani s slednjem povezavam. Avtorji že sami izpostavijo prednosti in pomanjkljivosti takega pregleda – pregleda se lahko ogromna količina spletnih strani, hkrati pa je tak pregled zelo grob, saj se tudi zaznava temnih vzorcev mora avtomatizirati. Nekateri temni vzorci pa se ne razberejo enostavno s pregledom izvorne kode spletne strani. Z namen po večji točnosti, je naša raziskava izbrala ročni pregled spletnih strani, češar posledica je manjše število pregledanih spletnih strani, hkrati pa večje sigurnosti v najdenih temnih vzorcih. V nadaljevanju je predstavljena analiza primerjave temnih vzorcev, ki smo jih našli v slovenskih spletnih trgovinah, z odkritji temnih vzorcev v tujih spletnih trgovinah.



Graf 2: Pojavnost temnih vzorcev na spletnih straneh.

Ameriški raziskovalci so v kategoriji vzorcev nujnosti odkrili 481 temnih vzorcev, od tega 88 časovno omejenih sporočil in 393 vzorcev odštevalnika časa, ki pa jih niso natančneje opredelili kot zavajajoče ali ne. Sami smo odkrili 11 temnih vzorcev te vrste, od katerih so bili štirje varljivi, saj so omogočili nakup izdelka tudi po izteku ponudbe. Prav tako smo našli 17 časovno omejenih sporočil, ki so bila zelo podobna ameriškim primerom, torej opozorilo na skorajšnji iztek določene ponudbe.

V kategoriji napačnega informiranja je bilo v tuji raziskavi odkritih 270 temnih vzorcev, v naši raziskavi pa smo jih odkrili 22. Za vzorec sramotne potrditve, torej jezikovnega preoblikovanja besedila, kjer je ena možnost predstavljena kot sramotna odločitev, je bilo s spletnim pajkom odkritih 169 primerov. V naši raziskavi pa so bili trije taki primeri, ki so bili vizualno zelo podobni tistim s tujih spletnih strani, saj so z vplivanjem na emocije potrošnika želeli spremeniti njegovo prvotno odločitev. Prav tako je bilo v ameriški raziskavi odkritih 25 vzorcev vizualnega vmešavanja, torej neenake predstavitve dveh ali več možnosti, medtem ko smo jih mi odkrili 16. Vzorcev zvijačnih vprašanj v lastnem pregledu nismo zaznali, zato primerjave te vrste temnih vzorcev ne moremo izvesti, je pa bilo v ameriški raziskavi odkritih devet takšnih vzorcev. Vzorcev prodaje pod pritiskom, torej prodaje dražje verzije določenega izdelka, je bilo 67, v tej raziskavi pa smo odkrili zgolj tri takšne vzorce, ki so bili v oblikovanju zelo podobni vzorcem s tujih spletnih strani, torej je bila dražja opcija že vnaprej izbrana.

V kategoriji socialnega dokaza je bilo na tujih straneh spletnih trgovin odkritih 325 temnih vzorcev, v naši raziskavi pa 17 vzorcev. 313 vzorcev je spadalo v vrsto obvestil o dejavnosti uporabnikov, torej kaj so kupili oz. kaj so si drugi potrošniki ogledovali. V naši raziskavi je bilo 6 vzorcev, ki so bili zelo podobni vzorcem iz ameriške raziskave, torej v obliki pojavnih sporočil ali dodatnih obvestil ob izdelku. Tudi pri tej vrsti vzorcev ameriški raziskovalci niso navedli, če in koliko zavajajočih obvestil o dejavnosti so zaznali, v naši raziskavi pa ni bilo nobenega zavajajočega. Ameriška raziskava je razkrila tudi 12 priporočil nejasnega izvora, torej priporočil drugih uporabnikov o izdelku ali spletni trgovini na splošno, v lastnem pregledu pa smo zasledili 11 takšnih vzorcev, ki so bili prav tako podobne oblike kot vzorci iz izvirne raziskave, torej so bili večinoma v obliki pojavnih sporočil.

Kategorija vzorcev pomanjkanja je v raziskavi tujih spletnih strani zajemala 679 vzorcev, v naši raziskavi pa smo jih našli le 26. Ameriški raziskovalci so odkrili 632 sporočil o majhni razpoložljivosti, torej vzorcev prikaza količine izdelkov na zalogi ali pa omejene zaloge, medtem ko smo jih v slovenskih spletnih trgovinah zaznali 23. Pri tej vrsti vzorcev niso opredelili, koliko, če sploh kaj varljivih vzorcev so odkrili, v naši raziskavi pa smo zaznali en takšen zavajajoč vzorec, kjer se je količina ponudbe spreminjala sorazmerno s pretečenim časom. V izvirni raziskavi so odkrili tudi 47 sporočil o velikem povpraševanju, torej pripis, da je za izdelek povpraševanje zelo veliko, zato naj potrošnik pohiti z naročilom. Naš vzorec spletnih strani je vseboval tri takšne vzorce, kjer je bila vsebina zelo podobna vzorcem na tujih spletnih straneh.

Ker nismo našli nobenega temnega vzorca iz kategorije vzorcev oviranja, tukaj ni mogoče opraviti primerjave, v ameriški raziskavi pa so odkrili 31 vzorcev, kjer se je uporabnik z lahkoto naročil na določeno storitev, a je bila odjava zelo otežena.

V kategoriji vzorcev prisilnih dejanj so ameriški raziskovalci odkrili šest vzorcev prisilnega vpisa, torej oblikovanja profila uporabnika le s predajo osebnih podatkov posameznika. Na slovenskih straneh smo takšnih vzorcev zaznali deset, vendar v malce drugačni obliki, saj smo kot temne vzorce upoštevali tudi primere, kjer je moral posameznik za pridobitev popusta vpisati osebne podatke ali pa nakup izdelka deliti na družbenih omrežjih.

V kategoriji prikritih vzorcev smo v naši raziskavi našli 17 vzorcev, v tuji raziskavi pa je bilo odkritih 26 vzorcev iz te kategorije. V tujih spletnih trgovinah je bilo zaznanih 7 vzorcev skrivanja v košarico, sami pa smo našli zgolj en vzorec, kjer je ponudnik izdelkov v košarico dodal dodaten izdelek, ki ga potrošnik ni mogel odstraniti, je pa bil zastoj. Prav tako smo odkrili 16 vzorcev prikritih stroškov, ki so razkrili dodatne stroške tik pred nakupom, na tujih straneh je bilo teh vzorcev 5. Sami pa nismo odkrili nobenega vzorca prikrite naročnine, medtem ko so na tujih spletnih straneh našli 14 takšnih vzorcev.

4 DISKUSIJA

Graf 3 prikazuje primerjavo deležev strani z določenimi kategorijami temnih vzorcev. Razlike med slovenskimi in tujimi spletnimi trgovinami so hitro razvidne, saj nobenega od temnih vzorcev ni imelo več kot 6 % tujih spletnih trgovin. Po drugi strani pa je

imelo kar 51 % pregledanih slovenskih spletnih strani vzorec nujnosti in 47 % slovenskih spletnih strani vzorec pomanjkanja.

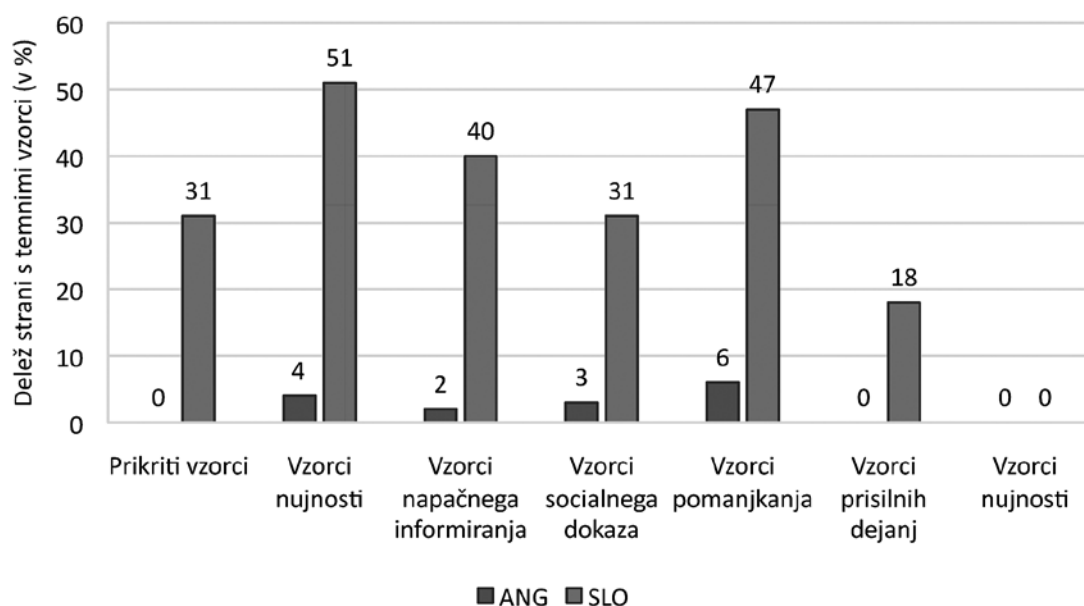
Eden izmed razlogov, zakaj prihaja do takih razlik, je drugačen način iskanja temnih vzorcev. Pri raziskavi tujih spletnih strani je pregled potekal z avtomatiziranim pregledom in tako so se temni vzorci iskali z vnaprej napisanimi pravili iskanja bodisi niza besed bodisi HTML-elementov (Mathur idr., 2019). Pri naši raziskavi pa se je iskanje temnih vzorcev naredilo z ročnim pregledom in interakcijo z vsako spletno stranjo v raziskavi. Avtomatsko iskanje temnih vzorcev je tako omogočilo mnogo večji vzorec pregledanih spletnih strani, je pa potencialno bolj površno pri identifikaciji temnih vzorcev. Že z bolj širokim naborom pravil iskanja vzorcev v izvorni kodi spletnih strani bi se najverjetneje identificiralo več temnih vzorcev. A tudi če bi se nabor pravil temnih vzorcev razširil, je nekatere oblike temnih vzorcev avtomatsko težko odkriti. Določeni elementi spletnih strani se prikažejo v svoji končni obliki šele po grafičnem prikazu spletne strani in v interakciji s spletno stranjo. Take elemente je težko identificirati s preprostim pregledovalnikom izvorne kode in je potrebno vsako spletno stran zagnati v grafičnem brskalniku, ne pa le pregledati kodo. Prav tako je z avtomatskim pregledom nemogoče simulirati interakcijo s spletno trgovino – izbiranje izdelka, dodajanje tega v košarico in sprožitev procesa nakupa.

Po drugi strani sta ročni pregled in interakcija s spletno stranjo izredno zamuden proces, ki oteži pregled reprezentativnega vzorca spletnih trgovin. Kljub temu da je ročni pregled lahko bolj natančen pri identifikaciji temnih vzorcev, še vedno obstaja verjetnost, da so bili določeni temni vzorci spregledani ob pregledu in interakciji s spletno stranjo. Celovit pregled posamezne spletne strani bi zahteval celovito poznavanje in preizkus vseh funkcionalnosti vsake posamezne spletne strani.

Drugi potencialni razlog za razlike v deležu spletnih strani s temnimi vzorci je drugačno časovno obdobje pregleda. V raziskavi tujih spletnih strani (Mathur idr., 2019) se je pregled izvršil avgusta 2018, pregled slovenskih spletnih trgovin pa smo izvedli junija 2020. Pričakovano je, da se prakse, ki lastnikom spletnih strani prinesejo več prometa, več uporabnikov ali večjo stopnjo fluktuacije (angl. turn-over-rate), hitro razširijo. Spletni oblikovalci in razvijalci, ki v svoje produkte vnesejo temne vzorce, ki pozitivno vplivajo na poslovne rezultate lastnika spletne strani, imajo prednost pri prodaji svojih storitev vzpostavitve spletnih strani. Pričakovano je, da se bodo zaradi tega prakse temnih vzorcev pojavile še na več spletnih straneh.

4.1 Omejitev prakse temnih vzorcev na spletnih straneh

Po definiciji temnih vzorcev so ti manipulativni in vplivajo na obnašanje uporabnikov na spletu. Z izpo-



Graf 3: Primerjava deležev strani s temnimi vzorci.

stavitvijo problematike takih manipulativnih praks lahko uporabnike spleta ozavešamo o njihovem obstoju. Šele če se zavedajo obstoja manipulativnih temnih vzorcev, se lahko njihovih posledic izognejo ali jih vsaj zmanjšajo (Kozyreva idr., 2019).

Do določene mere se lahko s temnimi vzorci spopademo s tehnologijo. Kot so protireklamni vtičniki (angl. ad blocker) omejili ali v celoti odstranili prikaz reklamnih obvestil na spletnih straneh, se lahko s podobnimi pristopi izognemo temnim vzorcem, ki ne vplivajo na funkcionalnost (npr. vzorci pomanjkanja ali vzorci nujnosti). Raziskave potrjujejo pričakovanja, da so uporabniki mnogo bolj zadovoljni z uporabo spletnih strani, če te ne vsebujejo reklam (Pujol idr., 2015; Shiller idr., 2018). Vpliv temnih vzorcev na zadovoljstvo uporabnikov je že bil pregledan v številnih raziskavah, ki kažejo povezavo med pojavom temnih vzorcev in uspešnostjo prodaje (Moran, 2020; Narayanan idr., 2020). Vsekakor pa je pozitiven vpliv uporabe temnih vzorcev prisoten pri lastnikih spletnih strani – ti jih namreč ne bi uporabljali, če pozitivnega vpliva ne bi bilo. Kako velik učinek imajo različni temni vzorci na različne metrike uspešnosti spletnih strani in trgovin, pa je potrebno ponovno preučiti z nadaljnjo raziskavo.

S samoregulacijo lastnikov spletnih strani najverjetneje ne bomo dosegli veliko, saj lahko določeni temni vzorci tudi pritegnejo nove uporabnike (npr. z napačnim informiranjem prepričajo v kvaliteto ali uspešnost). Še več, kot že prej omenjeno, je vse temne vzorce zelo težko identificirati s pomočjo tehnologije. Posledično je za spopad s takimi praksami potreben sistematični pristop tudi na drugih področjih (Waldman, 2020; Moran, 2020). Z državnimi ali mednarodnimi regulativami bi se do določene mere lahko omejila uporaba temnih vzorcev. Na nivoju Evropske unije že obstaja direktiva zaščite potrošnikov (Chirita, 2012), ki zahteva izrecno potrditev potrošnika za spreminjanje pogojev poslovanja. V ta sklop spadajo temni vzorci skritih stroškov in skrivanja izdelkov v košarico, saj pri uporabi teh spletnih strani potrošniki niso izrecno izrazili dovoljenja, ampak so bili o spremembi poslovanja (dodan izdelek v košarico ali dodaten skriti strošek) le obveščeni.

5 SKLEP

Po opravljenem pregledu slovenskih spletnih trgovin je temeljna ugotovitev ta, da so temni vzorci že zdaj v veliki meri razširjeni v Sloveniji, saj smo na

vseh pregledanih straneh našli vsaj en temni vzorec. Večinoma so v spletnih trgovinah prisotni po en ali dva vzorca na celotno stran, največ vzorcev, ki smo jih opazili v posamezni spletni trgovini, je bilo devet. Na pregledanih straneh so se največkrat pojavili vzorci iz kategorije vzorcev nujnosti, teh je bilo 28, najpogosteje opažena vrsta vzorcev pa so bila sporočila o majhni razpoložljivosti določenega izdelka, ki so se pojavili 23-krat.

Na pregledanih spletnih straneh pa nismo zasledil popolnoma vseh vrst temnih vzorcev iz predhodno opredeljene kategorizacije, kar je verjetno posledica premajhnega vzorca izbranih spletnih trgovin. Tako nismo opazili vzorcev prikrite naročnine iz kategorije prikritih vzorcev, katerih je bilo v primerjalni ameriški raziskavi opaženih 14. Tudi vzorcev zvičajnih vprašanj iz kategorije vzorcev napačnega informiranja nismo zasledili, čeprav so se na tujih spletnih straneh pojavili devetkrat. Prav tako nismo zasledili nobenega vzorca težavne razveljavitve iz kategorije vzorcev oviranja, ki so se na tujih straneh pojavili 31-krat.

Stran, na kateri smo odkrili največ temnih vzorcev, je bila spletna trgovina gizzmo.si, v kateri smo opazili kar 9 temnih vzorcev. V njej smo zasledili en vzorec skrivanja dodatnega izdelka v košarico, kjer je trgovec dodal darilo k nakupu, in en vzorec prikritih stroškov, kjer je moral potrošnik za prejetje izdelka obvezno plačati še poštnino. Odkrili smo tudi en vzorec zavajajočega odštevanja časa, kjer je bil po poteku določenega časa nakup izdelka še vedno možen, čeprav je odštevalnik nakazoval drugače. Prav tako smo našli en vzorec časovno omejenega sporočila, ki je opozarjal na popuste, vendar brez končnega roka akcije, s čimer je zavajal k čim hitrejšemu nakupu. V tej spletni trgovini smo opazili tudi tri obvestila o dejavnosti drugih potrošnikov. Prvo obvestilo je prikazovalo število uporabnikov, ki si ogledujejo določen izdelek, drugi dve obvestili pa sta obveščali potrošnika, koliko izdelkov je bilo naročenih v zadnjih dveh dneh, eno obvestilo na začetni strani in eno na predstavitveni strani izdelka. Zaznali smo tudi eno sporočilo o majhni razpoložljivosti, ki je prikazovalo količino izdelka, in eno sporočilo o velikem povpraševanju, ki je opozarjalo, da se izdelek hitro prodaja in bo kmalu pošel.

Temni vzorci se na spletu ne pojavljajo zelo dolgo časa, zato se jih večina potrošnikov ne zaveda oz. ne zna prepoznati manipulativnih sporočil in nezavedno podleže njihovem vplivu. Pregled slovenskih

spletnih trgovin je pokazal, da slovenski trgovci vse raje uporabljajo temne vzorce kot oblike manipulacije s potrošniki. Čeprav na slovenskih spletnih straneh nismo uspeli najti vseh vrst temnih vzorcev, ki se pojavljajo v tujih spletnih trgovinah, menimo, da so temni vzorci kar precej razširjeni na slovenskem spletnem trgu. Slovenski spletni trgovci najraje uporabljajo temne vzorce sporočil o omejenih količinah izdelka, s čimer želijo prepričati potrošnike, da je povpraševanje po njem izredno veliko in bo izdelek kmalu razprodan, zato naj potrošniki pohitijo z nakupom. Največ temnih vzorcev se pojavlja na straneh, ki so priljubljene in moderne, kar je vidno tudi na tujih straneh spletnih trgovin.

V prihodnosti se bodo oblike temnih vzorcev nenehno spreminjale glede na potrebe trgovcev, ki bodo želeli ostati konkurenčni na trgu. Prav tako se bodo ustvarjali vedno novi načini manipulacije s potrošniki, katerih se ti ne bodo zavedali. Temni vzorci se trenutno pojavljajo tudi v igričarstvu (Zagal idr., 2013), predvsem pri spletnih nakupih dodatnih virtualnih predmetov v samih igrah. V prihodnje bi se lahko podobne raziskave izvedle tudi na področjih elektronskih sporočil in mobilnih aplikacij ter spletnih strani, ki omogočajo kakršnokoli obliko prodaje, kot so potovalne agencije in spletne prodaje vstopnic oz. prodaja raznih storitev.

6 LITERATURA

- [1] Chirita, A. D. (2012). The impact of Directive 2011/83/EU on consumer rights. Keirse, Samoy and Loos, Cambridge, Intersentia, 65-82.
- [2] Di Geronimo, L., Braz, L., Fregnan, E., Palomba, F., & Bacchelli, A. (2020, April). UI dark patterns and where to find them: a study on mobile applications and user perception. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-14).
- [3] Gray, C. M., Chen, J., Chivukula, S. S., & Qu, L. (2020). End User Accounts of Dark Patterns as Felt Manipulation. arXiv preprint arXiv:2010.11046.
- [4] Kozyreva, A., Lewandowsky, S. and Hertwig, R., 2019. Citizens versus the internet: Confronting digital challenges with cognitive tools.
- [5] Luguri, J., in Strahilevitz, L. (2019). Shining a light on dark patterns. U of Chicago, Public Law Working Paper, (719).
- [6] Maier, M. in Harr, R., 2020. Dark Design Patterns: An End-User Perspective. Human Technology, 16(2).
- [7] Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 1-32.
- [8] Monaco, J. V. (2020, April). Bug or Feature? Covert Impairments to Human Computer Interaction. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (pp. 1-15).
- [9] Moran, N. (2020). Illusion of safety: How consumers underestimate manipulation and deception in online (vs. offline) shopping contexts. Journal of Consumer Affairs, 54(3), 890-911.
- [10] Moser, C. (2020). Impulse Buying: Designing for Self-Control with E-commerce (doktorska disertacija).
- [11] Narayanan, A., Mathur, A., Chetty, M., & Kshirsagar, M. (2020). Dark Patterns: Past, Present, and Future. Queue, 18(2), 67-92.
- [20] Pujol, E., Hohlfeld, O., & Feldmann, A. (2015, October). Annoyed users: Ads and ad-block usage in the wild. In Proceedings of the 2015 Internet Measurement Conference (pp. 93-106).
- [21] Shiller, B., Waldfogel, J., & Ryan, J. (2018). The effect of ad blocking on website traffic and quality. The RAND Journal of Economics, 49(1), 43-63.
- [22] Waldman, A. E. (2020). Cognitive biases, dark patterns, and the 'privacy paradox'. Current opinion in psychology, 31, 105-109.
- [23] Zagal, J. P., Björk, S., & Lewis, C. (2013). Dark patterns in the design of games.
- [24] Shopper's Mind. (30. november 2016). Ecommerce Day 2016 – Zmagovalci tekmovanja Spletni trgovec leta. Pridobljeno 24. februarja 2021 od <https://smind.si/ecommerce-day-2016-zmagovalci-tekmovanja-spletni-trgovec-leta/>
- [25] Shopper's Mind. (28. november 2017). Zmagovalci tekmovanja Spletni trgovec leta 2017. Pridobljeno 24. februarja 2021 od <https://smind.si/zmagovalci-tekmovanja-spletni-trgovec-leta-2017/>
- [26] Gorjan, S. (14. november 2018). Kdo so zmagovalci tekmovanja Spletni trgovec leta 2018? Pridobljeno 24. februarja 2021 od <https://smind.si/zmagovalci-tekmovanja-spletni-trgovec-leta-2018/>
- [27] Kovačević, A. (14. november 2019). Spletni trgovec leta 2019 – to so najboljše spletne trgovine v Sloveniji. Pridobljeno 24. februarja 2021 od <https://smind.si/spletni-trgovec-leta-2019-najboljse-spletne-trgovine-v-sloveniji/>

Elena Osrajnik je magistrska študentka na Fakulteti za elektrotehniko, računalništvo in informatiko Univerze v Mariboru. Diplomirala je na študijskem programu Medijske komunikacije, smer Medijska produkcija.

Dr. Sašo Karakatič je docent na Fakulteti za elektrotehniko, računalništvo in informatiko Univerze v Mariboru. Raziskovalno se ukvarja s področjem umetne inteligence in strojnega učenja ter aplikacijo optimizacijskih pristopov po vzoru narave na področjih transporta in rudarjenja podatkov. skovalno se ukvarja z učno motivacijo, aktivnimi pristopi k poučevanju in psihosocialnimi značilnostmi mladih.

■ Računalniško podprto prepoznavanje zgodnjih znakov disleksije

Milena Košak Babuder¹, Blažka Korun¹, Ema Štarkl², Gaja Nenadović², Dušica Boben³, Erika Stanković⁴, Sara Jakop⁴, Luka Vranješ⁴, Karmen Javornik¹, Jure Žabkar⁴

¹Univerza v Ljubljani, Pedagoška fakulteta, Kardeljeva ploščad 16, 1000 Ljubljana, Slovenija

²Univerza v Ljubljani, Filozofska fakulteta, Aškerčeva cesta 2, 1000 Ljubljana, Slovenija

³Center za psihodiagnostična sredstva d.o.o., Litostrojska 44d, 1000 Ljubljana, Slovenija

⁴Univerza v Ljubljani, Fakulteta za računalništvo in informatiko, Večna pot 113, 1000 Ljubljana, Slovenija

Milena.Kosak-Babuder@pef.uni-lj.si, Jure.Zabkar@fri.uni-lj.si

Izvleček

Disleksija je nevrološko pogojena bralno-napisovalna motnja, ki izvira iz razvojnih posebnosti ali iz posebnosti delovanja osrednjega živčevja. Zanj je značilna šibka sposobnost branja in črkovanja, ki ni v skladu z drugimi sposobnostmi, ter pomanjkljivi avtomatizacija in tekočnost branja in črkovanja. Z namenom čim prejšnjega prepoznavanja disleksije pri otrocih smo razvili spletno aplikacijo, ki vsebuje štiri preizkuse za prepoznavanje zgodnjih znakov disleksije. Aplikacija je dostopna širši javnosti in v primerjavi s klasičnimi pristopi omogoča hitrejšo in bolj množično testiranje; primerna je za testiranje na daljavo, v otroku domačem okolju. V tem prispevku predstavljamo rezultate preliminarnega testiranja aplikacije na manjšem testnem vzorcu.

Ključne besede: disleksija, spletna aplikacija, testiranje

Abstract

Dyslexia is a specific learning disorder, which is neurological in origin. It stems from developmental peculiarities or from specific characteristics of functioning of the central nervous system. It is characterized by poor reading and spelling abilities, which are not in line with other abilities, and by impaired automatization as well as spelling and reading fluency. The purpose of developing this web application, which consists of four tasks, was to provide a freely accessible way of recognising dyslexia in children as early as possible. In comparison to traditional approaches, it makes possible the quicker and more extensive testing. In addition, it is suitable for remote testing in a familiar environment. In this paper, we present the results of preliminary testing of the application.

Keywords: Dyslexia, web application, testing

1 UVOD

Disleksija je kot najpogostejša oblika motenj branja in pisanja ena izmed najbolj raziskanih motenj v skupini specifičnih učnih težav. Za slednje velja, da se razprostirajo na kontinuumu od lažjih do izrazitih ter od kratkotrajnih do tistih, ki trajajo vse življenje. Disleksija je notranje (nevrofiziološko) pogojena bral-

no-napisovalna težava in izvira iz razvojnih posebnosti ali iz posebnosti delovanja osrednjega živčevja. Vključuje skupino raznolikih, a medsebojno povezanih dejavnikov, ki so del posameznika, nanj ter na njegovo delovanje pa vplivajo vse življenje (Raduly Zorgo in sod., 2010). Obstaja kar nekaj vzročnih teorij o disleksiji, a večina raziskovalcev trdi, da je glavni

razlog za težave pri disleksiji pomanjkljivo fonološko zavedanje, ki se kaže kot slabša sposobnost prepoznavanja, razlikovanja in manipuliranja z glasovi ter slabša sposobnost naučiti se ujemanja glasov in črk. To poudarjajo tudi številne opredelitve disleksije, ki vsebujejo podobne komponente. Disleksijo opisujejo kot učno težavo ali pa nevrološko motnjo, ki vpliva na razvoj zmožnosti branja in pisanja (Košak Babuder, 2013). Zaradi značilnih težav natančnega in/ali tekočega prepoznavanja besed, šibkega črkovanja in šibke sposobnosti dekodiranja posamezniki težje usvajajo branje, bralno razumevanje in pisanje.

Težave niso omejene le na branje in črkovanje; prisotne so tudi težave z vzdrževanjem pozornosti, avtomatizacijo novega znanja ter z grobo- in fino-motoričnimi spretnostmi. Ob nevrološko pogojenih razlikah so prisotne tudi kognitivne težave, ki lahko vplivajo na organizacijske spretnosti, zmožnost računanja ter druge kognitivne in čustvene sposobnosti. Posamezniki z disleksijo so lahko izjemno nadarjeni in izvorni pri reševanju različnih vrst problemov, pogosto pa imajo tudi dobre vizualne sposobnosti (Nijakowska, 2016). V populaciji je približno 10 % otrok in mladostnikov z disleksijo. Raziskave navajajo, da je pogostost motnje enaka pri obeh spolih, z nekoliko večjo pogostostjo pri dečkih (Riddick in sod., 2002).

Ponavadi disleksijo zaznamo pri osnovnošolskih otrocih, in sicer zaradi počasnega napredka pri učenju branja in pisanja. Težave se pogosto odražajo v nižjem učnem uspehu in posledično slabši samopodobi, zato je pomembno, da jo čim prej prepoznamo in ustrezno obravnavamo. S tem preprečujemo stigmatizacijo otrok in mladostnikov z disleksijo, spodbujamo pa njihovo lažje vključevanje v družbo in manjšamo težave v odrasli dobi.

V Sloveniji so otroci z blago do zmerno izraženo disleksijo po Zakonu o osnovni šoli («Zakon o osnovni šoli ZOsn-UPB3», 2006) deležni prilagojenih metod in oblik poučevanja ter preverjanja znanja, ti isti z izrazito disleksijo pa so po Zakonu o usmerjanju otrok s posebnimi potrebami («Zakon o usmerjanju otrok s posebnimi potrebami ZUOPP-1», 2011) deležni intenzivnejših prilagoditev in dodatne strokovne pomoči. V procesu prepoznavanja in diagnostičnega ocenjevanja disleksije, za katerega je potrebna multidisciplinarna skupina strokovnjakov (psiholog, specialni in rehabilitacijski pedagog, logoped, učitelj), ločimo več stopenj – od detekcije, klasifikacije, načrtovanja pomoči in spremljanja napredovanja do eval-

vacije (Magajna, 2011). Prvo stopnjo prepoznavanja oseb z disleksijo (detekcija) predstavljajo presejalni preizkusi, katerih namen sta prednostna obravnava identificiranih oseb, ki potrebujejo diagnostično ocenjevanje, ter informiranje posameznikov o verjetnosti pojava disleksije pri njih (Pollak, 2009). Motnja se med posamezniki zelo razlikuje. Presejalni preizkusi omogočajo pri otrocih potrditev disleksije in izvajanje ustrezne obravnave, še preden se pri njih pojavi občutek neuspeha (Snowling, 2013). Testi za odkrivanje disleksije obsegajo testiranje spomina, črkovanja, razumevanja besedila, bralne naloge, hitro poimenovanje, pozornost idr.

1.1 Diagnostično ocenjevanje disleksije

Zgodnja detekcija disleksije je za zagotavljanje ustreznih oblik pomoči in podpore otrokom z disleksijo ključnega pomena. Zaradi večdimenzionalne narave motnje se za učinkovito odkrivanje disleksije uporablja več različnih testov, preizkušenj in baterij. Pomembno je dobro presejalno testiranje, ki loči otroke s tveganjem za nastanek motenj branja in pisanja od otrok brez tveganja.

Za ugotavljanje težav na področju branja in pisanja, s katerimi preverjamo različne elemente branja in pisanja (fonološko zavedanje, hitrost in pravilnost branja, avtomatizacijo branja, bralno razumevanje, pisanje po nareku, pisno izražanje) v Sloveniji uporabljamo:

- Test v motenosti branja in pisanja oz. Šalijev test (Šali, 1971) – test je le delno standardiziran za populacijo otrok drugega razreda;
- SNAP – profil ocene posebnih potreb (SNAP ni test v psihometričnem smislu, temveč pripomoček za zbiranje informacij o otroku, ki so pomembne za ugotavljanje morebitnih težav pri določeni spretnosti) (Weedon in Reid, 2018);
- Enominutni test glasnega branja (Gradišar in Pečjak, 1991);
- Preizkus bralnega razumevanja (Elley in sod., 1995);
- Bralni test (Pečjak, Magajna in Podlessek, 2012) – test je standardiziran merski instrument, ki ugotavlja splošno bralno zmožnost ob koncu prvega triletja;
- Ocenjevalno shemo bralnih zmožnosti otrok od 1. do 3. razreda: OSBZ (Pečjak, Magajna, Podlessek in Potočnik, 2012) – test je standardiziran merski instrument, podatki, zbrani z ocenjevalno shemo,

pa dajo informacijo o tem, katere bralne zmožnosti je otrok že uspel razviti;

- Preizkus tekočnosti branja po modelu, ki temelji na kurikulumu za 2., 3. in 4. razred (Košir, 2011), in
- Test glasovnega zavedanja (Magajna, 1994).

V svetu je računalniško podprto testiranje postalo glavno sredstvo tako za odkrivanje disleksije s presejalnimi testi kot tudi za intervencije, ki so prilagojene učnim težavam in potrebam otrok ter mladostnikov. Uporaba računalniškega sistema predstavlja izjemen dejavnik za izboljšanje običajnih metod odkrivanja disleksije in sproža raziskovanja novih perspektiv v zvezi z identifikacijo posameznikov z disleksijo (Drigas in Politi- Georgousi, 2019).

Rooms (Rooms, 2000) poudarja potencialne koristi uporabe IKT za osnovnošolske otroke z disleksijo in poudarja, da je lahko dostopna in na voljo, ne da bi se otroci z disleksijo počutili drugačne ali izključene; vključeni so veččutni pristopi (slušni, ustni, vizualni, kinestetični) in sistemi za izboljšanje težav otrok z disleksijo (Rooms, 2000). Diagnostično ocenjevanje s pomočjo IKT psihologom in drugim strokovnim delavcem omogoča preprosto in hitro izvedbo pregleda kognitivnih sposobnosti in drugih vitalnih spretnosti (Singleton, 2001). Interaktivna multimedija, navidezna okolja, nevronske mreže, programska oprema, mehka logika, tehnike, ki temeljijo na igrah, in mobilne aplikacije povečajo učinkovitost običajnih postopkov presejanja disleksije, vsak od njih pa ima sofisticirane funkcije, ki olajšajo postopke ocenjevanja (Menghini in sod., 2011).

2 PROBLEM RAZISKAVE Z RAZISKOVALNIMI VPRAŠANJI

V svetu je uporaba računalniško podprtih sistemov za prepoznavanje otrok z disleksijo že dokaj dobro uveljavljena, saj je učiteljem na voljo več različnih programov – od presejalne programske opreme do podrobnejših računalniških ocenjevalnih baterij. Večina računalniških programov za odkrivanje disleksije se opira na ocenjevanje branja in črkovanja ter kognitivnih sposobnosti – kot sta na primer fonološko zavedanje in verbalno pomnjenje – ki podpirajo razvoj pismenosti in ki so na splošno dobri napovedniki disleksije (Singleton in sod., 2009).

Tako klasični testi kot aplikacije imajo svoje prednosti in slabosti. Prednost klasičnih testov je v pri-

sotnosti strokovnjaka, ki izvaja test, hkrati pa otroka opazuje, sproti preverja njegovo razumevanje, prilagaja navodila, da jih otrok razume, opazuje otrokovo pozornost, morebitno utrujenost. Hkrati lahko oseba otroka spodbuja, mu nudi oporo. Slabost klasičnih preizkusov sta predvsem izpostavljenost posameznika in časovna zamudnost, ki ju lahko odpravimo s pomočjo aplikacije. To lahko namreč hkrati rešuje več otrok, zato lahko v kratkem časovnem obdobju ocenimo veliko otrok in presejemo rizične otroke od nerizičnih. Aplikacija ima prednosti tudi z motivacijskega vidika, saj je bolj podobna računalniški igri kot ocenjevanju.

V Sloveniji imamo nekaj preizkusov za ugotavljanje težav na področju branja pri otrocih. Zelo pogosto uporabljamo preizkus SNAP (profil ocene posebnih potreb); gre za uveljavljen pripomoček, ki omogoča sistematičen in izčrpen pregled otrokovih specifičnih učnih težav, med katerimi je tudi disleksija. Vsebuje preizkuse za merjenje kognitivnih in jezikovnih spretnosti, na osnovi rezultatov katerih lahko ugotovljamo rizičnost za disleksijo.

Aplikacija PKP – Disleksija je prvi poskus preverjanja zgodnjih znakov disleksije v elektronski obliki. Otroku dobi takojšnjo povratno informacijo o svojih dosežkih na področju fonološkega zavedanja, branja in bralnega razumevanja, delovnega spomina ter zaporedij. Zamišljena je kot presejalni preizkus, s katerim lahko

prepoznamo posameznike, ki so rizični za pojav disleksije. Če se izkazujejo izrazitejša odstopanja, je predvidena poglobljena diagnostika otrokovih težav na področju branja in pisanja zaradi disleksije.

V okviru tega prispevka smo aplikacijo testirali na neslučajnostnem priložnostnem vzorcu. Naš namen je bil aplikacijo preizkusiti in pridobiti prve preliminarne podatke. Zanimalo nas je:

- kakšni so osnovni parametri opisne statistike in kakšna je razporeditev rezultatov (percentilne vrednosti) pri posameznih preizkusih (fonološkega zavedanja, branja in bralnega razumevanja, delovnega spomina in zaporedij) glede na pravilno rešenih primerov;
- kakšni so osnovni parametri opisne statistike in kakšna je razporeditev rezultatov (percentilne vrednosti) pri posameznih preizkusih (fonološkega zavedanja, branja in bralnega razumevanja, delovnega spomina in zaporedij) glede na povprečen čas reševanja;

- kakšna je težavnost vključenih besedil v preizkusu branja in bralnega razumevanja.

3 METODA

3.1 Spletna aplikacija PKP – Disleksija

Spletna aplikacija PKP – Disleksija deluje v okviru strežniškega sistema na Fakulteti za računalništvo in informatiko; na ta način smo zagotovili ustrezne zmogljivosti aplikacije ob morebitnem hkratnem večjem številu uporabnikov in poskrbeli za varnost podatkov, ki jih aplikacija beleži. Zaledni del aplikacije smo razvili s pomočjo ogrodja Django. Za bazo smo uporabili PostgreSQL, za uporabniški vmesnik pa programski jezik Javascript. Ob vpisu starosti na začetni strani se v bazi ustvari nov zapis za uporabnika. Vsakič, ko uporabnik reši nalogo, se zanj ustvari nov zapis z rešitvijo, časom pričetka reševanja, trajanjem reševanja in morebitnimi drugimi podatki, npr. kolikokrat je popravil rešitev, če je bilo popravljjanje rešitev možno. Razen spola in starosti aplikacija ne beleži osebnih podatkov, kar uporabniku zagotavlja anonimnost.

3.2 Vzorec

Za potrebe raziskave je bil oblikovan neslučajnostni priložnostni vzorec. V vzorec je bilo zajetih 43 otrok tretjega razreda osnovne šole. Starost vključenih otrok je 8 let (39 otrok) oziroma 9 let (4 otroci). Vključena je bila splošna populacija otrok – zajeti so bili tako tisti, ki imajo bralno napisovalne težave, kot tudi tisti, pri katerih tovrstne težave niso bile prepoznane. Otroci so bili iz različnih delov Slovenije (osrednje-slovenska, gorenjska, dolenska in severnoprimska regija). Pri otrocih, vključenih v vzorec, nismo preverjali ali imajo potrjeno disleksijo ali ne, saj je bil namen raziskave ugotoviti, kako naloge rešuje splošna populacija osemletnih otrok. V prihodnje bodo v raziskave vključeni tudi otroci s potrjeno disleksijo.

3.3 Inštrument

V okviru projekta PKP – Disleksija smo razvili aplikacijo za prepoznavanje zgodnjih znakov disleksije, ki omogoča, da otroci sklope nalog oz. preizkuse samostojno rešujejo. Preizkusi vsebujejo pisna in slušna navodila, ki so skrbno pripravljena tako, da zanje predvidevamo, da jih bodo otroci razumeli, kljub temu pa je predvidena tudi možnost, da otrokom pri razumevanju navodil pomagajo starši. Navodilom

sledi kratka demonstracija – predstavitev reševanja, pri kateri otrok dobi nazorno vidno predstavo o preizkusu, ki ga bo reševal. Sledi sklop vaj, med katerimi se odgovori ne točkujejo. Vaje so namenjene zgolj preverjanju otrokovega dojetja navodil – preverjajo, ali je otrok razumel, kako se naloga rešuje. Med vajami lahko kadarkoli ponovno pogledajo tudi navodila.

Aplikacija vsebuje štiri preizkuse, vsak preizkus pa zajema sklop nalog. Pri oblikovanju preizkusov smo sledili protokolu razvoja psiholoških testov po mednarodnih smernicah (npr. različne smernice International Test Commission) in ameriških standardih za pedagoške in psihološke pripomočke (*Standards for educational and psychological testing*, 2014) in vire, ki usmerjajo razvijalce računalniških ali internetnih psiholoških testov. Kot velevajo standardi, so pri razvoju sodelovali strokovnjaki z različnih področij: za disleksijo in razvoj psiholoških pripomočkov ter računalniški programerji. Aplikacija je sestavljena iz štirih preizkusov, ki zahtevajo rabo kognitivnih in jezikovnih spretnosti, ključnih za uspešno branje in pisanje: preizkus koncepta zaporedja, preizkus branja in bralnega razumevanja, fonološkega zavedanja ter delovnega spomina.

Otroci z disleksijo imajo pogosto težave z razumevanjem koncepta zaporedja in posledičnosti, kar se kaže v težavah pri urejanju elementov v zaporedje, priklicu abecede, vrstnega reda dni v tednu in mesecev (Žagar, 2012). Preizkus, ki smo ga oblikovali, preverja otrokovo zmožnost urejanja števil, mesecev in dni v tednu v pravilnem in obratnem vrstnem redu.

Bralno razumevanje zahteva rabo višjih miselnih procesov, kot sta delovni spomin in inhibicija, okrnjeno bralno razumevanje pri otrocih z disleksijo pa dosledno dokazujejo izsledki številnih raziskav (Chen in sod., 2016; Chung in sod., 2020; Lauterbach in sod., 2017). Na podlagi vnaprej izbranih kriterijev, ki pri otrocih napovedujejo kasnejšo pismenost in sposobnost branja, ter upoštevanja tekočnosti in hitrosti branja otrok z disleksijo smo v preizkus vključili štiri besedila. Besedila se razlikujejo po dolžini (dve besedili sta krajši in dve daljši) in po težavnosti (pri vsaki dolžini je eno besedilo preprostejše, eno pa zahtevnejše), predstavljena pa so v obliki zgodbe s temami iz vsakdanjega življenja. Otrok po prebranjem besedilu odgovarja na dihonomne postavke tipa da/ne in postavke izbirnega tipa.

Fonološko zavedanje obsega zavedanje, da je govor sestavljen iz besed, besede pa iz različnih zlogov

in glasov, kar posamezniku posledično omogoča povezovanje črk s pripadajočimi glasovi (Peklaj, 2012). V klasičnih preizkusih tipa papir-svinčnik nalogo navadno izvajamo tako, da testator prebere besedo, iz katere je treba izločiti določen glas ali pa določiti prvi ali zadnji glas v njej, testiranec pa ustno poda odgovor. Za namene spletne aplikacije smo način odgovarjanja prilagodili, saj nismo imeli dostopa do tehnologije, ki bi omogočala zanesljivo avtomatsko prepoznavanje govora, pisno podajanje odgovorov pa bi prineslo še kopico drugih težav, ki bi lahko prispevale k manjši veljavnosti preizkusa. Zato smo se odločili za različico postavk izbirnega tipa.

Številni posamezniki z disleksijo izkazujejo primanjkljaj v delovnem spominu in imajo težave s pomnjenjem informacij, ki niso smiselno povezane (Peklaj, 2012). Izsledki raziskav so najbolj dosledni glede primanjkljajev v verbalni komponenti delovnega in kratkoročnega spomina (Jeffries in Everatt, 2004; Menghini in sod., 2011; Palmer, 2000; Pham in Hason, 2014; Schuchardt in sod., 2013; Siegel in Linder, 1984; Smith-Spark in sod., 2003). Preizkus, ki smo ga oblikovali, predstavlja različico klasičnega preizkusa preverjanja obsega delovnega spomina s slušnimi verbalnimi dražljaji (besedami). Ponovno smo prilagodili način podajanja odgovorov tako, da je omogočeno avtomatsko preverjanje njihove pravilnosti, zato otrok ne podaja odgovorov ustno. Da bi bil preizkus za otroke čim bolj privlačen, smo ga postavili v kontekst hranjenja živali v živalskem vrtu. V nalogi otrok zasliši zaporedje živali, ki jih mora nahraniti, nato pa mora s kliki na sličice živali nahraniti v enakem vrstnem redu, kot so bile predhodno našteje.

V splošnem je struktura preizkusov enotna: vsak preizkus je sestavljen iz navodila, prikaza reševanja, vaj in nalog, ki se točkujejo. Pri oblikovanju navodil smo bili posebej pozorni, da niso predolga, da vključujejo preproste stavčne strukture, a da hkrati zajamejo vse informacije, ki so potrebne za jasnost zahtev naloge. Prikaz reševanja služi seznanitvi z nalogami, vaje pa preverjanju, ali otrok nalogo razume, saj je to za veljavnost meritve ključno. Pri vajah so otroci obveščeni o napačnih odgovorih, pri morebitnem ponovnem reševanju zaradi napake pa se prikaže namig. Pri vsakem sklopu ima otrok pred reševanjem dela, ki se točkuje, možnost večkratnega poslušanja navodil in reševanja vaj. Uporabniški vmesnik smo oblikovali tako, da je za otroke privlačen, da posamičen prikaz ne vsebuje nepotrebnih in motečih dra-

žljajev ter veliko elementov naenkrat, da se informacije prikazujejo zaporedoma in da je barvni kontrast med besedilom in ozadjem ustrezen in nemoteč v skladu s specifičnimi lastnostmi vidnega procesiranja otrok z disleksijo.

3.3.1 Preizkusi

Aplikacija vsebuje naslednje štiri preizkuse:

- *preizkus koncepta zaporedja*: naloge preverjajo otrokovo sposobnost razvrščanja besed in števil v zaporedja.
- *preizkus bralnega razumevanja*: otrok najprej prebere krajše besedilo; sledijo vprašanja, s katerimi naloga preverja otrokovo razumevanje besedila;
- *preizkus fonološkega zavedanja*: preverja ali otrok razume enostavna slušna navodila in pravilno izloči glas iz besede;
- *preizkus delovnega spomina*: preizkus sposobnosti otroka, da si zapomni slišana zaporedja besed.

V tem razdelku bo vsebina preizkusov podrobneje predstavljena.

Preizkus koncepta zaporedja

Pri tem preizkusu smo uporabili dve splošni obliki nalog; prvo za preverjanje koncepta zaporedja besed, drugo pa za preverjanje koncepta številskih zaporedij. Prva splošna oblika naloge je: »Razporedi x od y do α «, pri čemer x določa elemente, ki jih je treba urediti, vrstni red razporejanja pa je določen z y in α , tako da je y začetni element zaporedja, α pa beseda *naprej* ali *nazaj*. Ta oblika naloge velja za elemente, ki so dnevi v tednu in meseci v letu. Primer naloge iz preizkusa:

Razporedi dneve v tednu od torka naprej.
Predvidena rešitev: *torek, sreda, četrtek, petek, sobota*

Druga splošna oblika naloge je: »Razporedi x do α «, pri čemer x določa elemente, ki jih je treba urediti, α pa predstavlja besedno zvezo *od največjega do najmanjšega* ali *od najmanjšega do največjega*. Ta oblika naloge velja za elemente, ki so števila. Primer naloge iz preizkusa:

Razporedi števila od največjega do najmanjšega.
Predvidena rešitev: *12, 10, 8, 6*



Slika 1: Preizkus za preverjanje koncepta zaporedja v nizih

Ob začetku reševanja se na zaslonu prikaže navodilo posamezne naloge in predvaja se njegov zvočni posnetek, nato pa se v pomešanem vrstnem redu prikažejo elementi (Slika 1). Vsak element je izpisan in opremljen z zvočnim posnetkom. Med reševanjem je na ekranu izpisano navodilo naloge, s pritiskom na ikono zvočnika pa lahko otrok zvočni posnetek navodila večkrat posluša.

Preizkus vsebuje osem točkovanih nalog, od katerih štiri vsebujejo zaporedja besed (pri dveh so to dnevi, pri dveh meseci), štiri pa zaporedja števil. Pri polovici nalog, tako pri besedah kot številih, je elemente treba razporediti v pravilni vrstni red, pri polovici pa v obratni vrstni red. Vrstni red prikaza elementov pred razvrščanjem je določen in enak za vse posameznike. Pri nalogah s števili si ta ne sledijo po vrsti kot naravna števila, temveč so izbrana v korakih po dve, v korakih po dve pri vsakem drugem številu ali v različnih korakih.

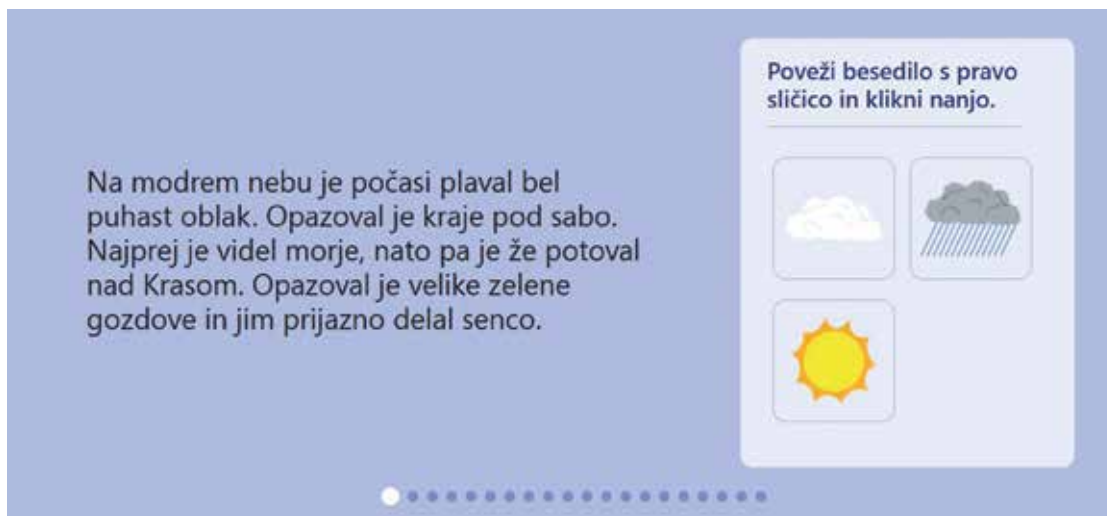
Preizkus branja in bralnega razumevanja

V preizkusu otrok prebere štiri besedila in odgovarja na vprašanja, ki preverjajo bralno razumevanje besedil. Besedilo se prikaže na levi polovici zaslona, otrok pa, ko ga je prebral, s klikom na gumb nadaljuje z odgovarjanjem na postavke (Slika 2). Takrat se zabeleži čas, ki je minil od prikaza besedila do klika na gumb. Nato se na desni polovici zaslona zaporedoma prikazujejo vprašanja, ki so opremljena z zvočnimi posnetki, pri čemer besedilo ves čas ostane vidno. Pod vprašanjem so hkrati prikazani vsi možni odgovori, od katerih je pravilen le eden.

Od štirih vključenih besedil sta dve krajši (35 besed), dve pa daljši (80 besed). Dolžini besedil smo izbrali glede na že obstoječe norme hitrosti branja besedil dolžine 80 besed v slovenščini za osemletne otroke. Pri posamezni dolžini besedila smo oblikovali po eno besedilo, za katero smo predvideli, da bo preprostejše, in eno, za katerega smo predvideli, da bo zahtevnejše.

Bralno razumevanje vsakega od besedil preverjamo s petimi vprašanji, tremi z odgovori izbirnega tipa in dvema z odgovori tipa da/ne, ki so med posameznimi besedili primerljiva in prilagojena tako, da ustrezajo vsebini določenega besedila. Prva naloga pri vsakem besedilu vsebuje odgovore v obliki sličic. Otrok izbere tisto, ki se najbolj ujema z vsebino besedila, pri čemer se izbrani motivi vseh možnih odgovorov skladajo s širšo tematiko besedila. Drugo vprašanje sprašuje, o čem govori besedilo, le da so tokrat podani odgovori besedilni. Tretje vprašanje podrobneje sprašuje, kaj je počel subjekt v besedilu, pri četrtem in petem vprašanju pa je treba oceniti pravilnost oz. napačnost trditve.

Preizkus dopušča možnost vključitve naprave za sledenje pogledu, s katero bi bilo preverjanje hitrosti branja natančnejše, hkrati pa bi lahko bralce razvrščali v skupine po podobnosti vzorcev branja na podlagi strojnega učenja. Epidemiološke razmere v času izvajanja raziskave žal niso dopuščale testiranja v laboratoriju z opremo za sledenje očesnim gibom; te raziskave bomo opravili takoj, ko bo mogoče.



Slika 2: Preizkus za preverjanje hitrosti branja in bralnega razumevanja

Preizkus fonološkega zavedanja

Splošna oblika nalog je: »Beseda je x , x . Kaj dobiš, ko rečeš x brez y ?«, pri čemer je x beseda, iz katere je treba izločiti glas y . Primer naloge iz preizkusa:

Beseda je dres, dres. Kaj dobiš, ko rečeš dres brez d?

Pri vsaki nalogi se zaporedoma zaslišijo trije možni odgovori, od katerih je en pravilen. Otrok mora za vsak odgovor določiti, ali je pravilen ali napačen. Ko določen odgovor označi kot pravnega, nadaljuje z naslednjo nalogo, za vse morebitne preostale, neoznačene odgovore pa se šteje, kot da jih je označil za napačne. Med podajanjem odgovorov pri določeni nalogi lahko s pritiskom na ikono zvočnika na zaslonu ponovno posluša postavko.

Preizkus vsebuje 12 točkovanih nalog. Besede, iz katerih je treba izločiti glas, so štiričrkovne besede z enim soglasniškim sklopom. Soglasniški sklop se v enakomernih deležih besed pojavi na začetku, v sredini in na koncu besede. Glas, ki ga je treba izločiti, je v polovici besed prvi v soglasniškem sklopu, v polovici besed pa drugi. Polovica besed je izmišljenih, polovica pa pravih. Vrstni red nalog je naključen, a pri vsakem reševanju preizkusa isti. Vsaka naloga vsebuje tri možne odgovore, pravilni odgovor pa je v enakomernih deležih števila nalog prvi, drugi ali tretji izmed podanih.

Preizkus delovnega spomina

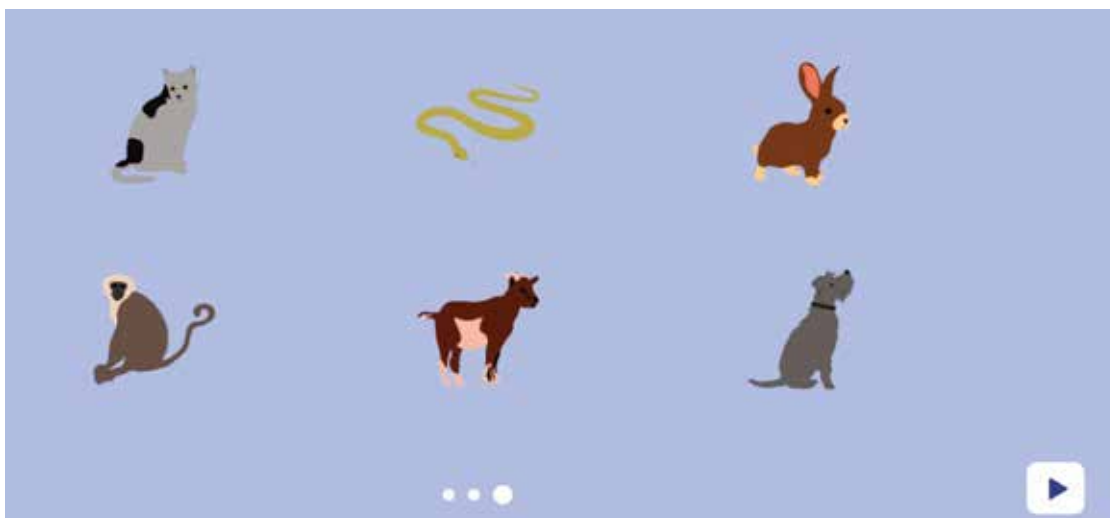
Splošna oblika naloge je: »Nahrani x «, pri čemer je x zaporedje živali. Primer naloge iz preizkusa:

Nahrani konja, kozo, miš.

Zaporedje je podano v obliki zvočnega posnetka ob praznem zaslonu, na katerem se nato pojavi šest sličic različnih živali, ki so enakomerno razporejene v dve vrsti (Slika 3). Živali so prebrane v ritmu ena žival na sekundo. Otrok nato nahrani živali z zaporednimi kliki na sličice živali v istem vrstnem redu, kot jih je slišal.

Ko zaključi s hranjenjem, klikne na puščico, živali izginejo in po treh sekundah se zasliši novo zaporedje živali. Posamezne živali se pojavljajo pri vseh zaporedjih v približno enakomernih deležih, v posameznem zaporedju pa se ista žival ne pojavi več kot enkrat. Prikazane živali so naključno izbrane tako, da nikoli ni prikazanih več vizualno ali semantično podobnih živali (na primer koza in ovca) hkrati. Pred prikazom reševanja naloge je otrok seznanjen z vsemi živalmi tako, da se prikažejo vse sličice iz naloge, opremljene pa so z zvočnim posnetkom in izpisom imena prikazane živali.

V testnem delu se dolžina zaporedja živali stopnjuje. Najkrajša dolžina je dve enoti, najdaljša pa šest enot. Za vsako dolžino preverjamo tri zaporedja, skupaj torej 15 zaporedij. Pred vsako spremembo dolžine zaporedja se predvaja zvočni posnetek, ki otroka na to spremembo opozori. Predvidevali smo, da bo poleg dolžine zaporedja k večji težavnosti nalog prispevala tudi dolžina enot (besed) v posameznem zaporedju. Tako smo pri zaporedjih znotraj ene dolžine in različnih dolžin vključili čedalje več čedalje daljših besed. Najdaljša dolžina imena živali je šest črk, največje število zlogov v imenu pa tri.



Slika 3: Preizkus za preverjanje delovnega spomina

3.4 Potek raziskave

Za pridobivanje vzorca otrok smo se obrnili na ravnatelje, šolske svetovalne delavce in učitelje osnovnih šol po vsej Sloveniji, ki so staršem otrok tretjega razreda posredovali naše vabilo ter pripravljeno soglasje za sodelovanje v raziskavi. Vabilo k sodelovanju so šolski strokovni delavci posredovali tudi med svojimi kolegi in znance na sosednjih šolah. Ob prejetju izpolnjenega soglasja smo staršem posredovali navodila za uporabo in dostop do aplikacije. Otroci so v aplikacijo s preizkusi dostopali z računalnikom. Aplikacija je pripravljena tako, da so lahko otroci pri izpolnjevanju čim bolj samostojni. Starši ali skrbniki so ob posredovanju soglasja dobili dovolj navodil za pomoč otroku, če bi jo ta potreboval. Sodelovanje otrok v raziskavi je bilo prostovoljno. Podatki oz. rezultati so se zbirali anonimno 12 dni (od 17. 11. do 29. 11. 2020). Za predhodno določanje težavnosti besedil v preizkusu branja in bralnega razumevanja smo uporabili spletno aplikacijo berljivosti besedila, ki je nastala v okviru projekta *Za kakovost slovenskih učbenikov (KaUč)*. Za posamezne preizkuse smo izračunali percentilne vrednosti, za opis ključnih značilnosti pa smo uporabili opisno statistiko.

4 REZULTATI Z RAZPRAVO

Otroci so reševali štiri preizkuse, s katerimi smo preverjali uspešnost na področjih, ki so pri disleksiji pogosto ovirana. Preizkusi so preverjali razumevanje koncepta zaporedij, hitrost branja in bralno razumevanje, fonološko zavedanje ter delovni spomin, kar so nekatera od ključnih področij, na katerih ima

jo osebe z disleksijo težave (Magajna in sod., 2015). V raziskavi je sodelovalo 43 otrok tretjega razreda, starih osem in devet let. Zanimalo nas je, kakšni so osnovni parametri opisne statistike in kakšna je razporeditev rezultatov (percentilne vrednosti) pri posameznih preizkusih a) glede na število pravilno rešenih primerov in b) glede na povprečen čas reševanja ter kakšna je težavnost besedil, vključenih v preizkus branja in bralnega razumevanja.

4.1 Število pravilno rešenih primerov

Zanimali so nas osnovni parametri opisne statistike in razporeditev rezultatov (percentilne vrednosti) pri posameznih preizkusih glede na število pravilno rešenih primerov. Otroci, ki imajo svoj rezultat na posameznem preizkusu pod 25. percentilom v populaciji, spadajo v skupino otrok, za katere obstaja tveganje za obstoj izrazitejših učnih težav (Weedon in Reid, 2018). Ker nizki rezultati na teh preizkusih nakazujejo tveganje za pojav disleksije, je pri teh otrocih smiselna poglobljena diagnostična ocena. Pri preizkusu zaporedij je bilo možnih 8 točk (4 točke za del, ki se je nanašal na besedna zaporedja, 4 točke za del, ki se je nanašal na številska zaporedja), pri preizkusu branja in bralnega razumevanja 20 točk (5 točka za vsako od štirih besedil), pri preizkusu fonološkega zavedanja je bilo možnih 12 točk, pri preizkusu delovnega spomina pa 15 točk.

Na področju zaporedij so otroci z najvišjim rezultatom dosegli vseh 8 točk, tisti z najnižjim rezultatom pa polovico možnih točk. Otrokov napredek je glede na vzorec ustrezen, če se njegov rezultat nahaja

Tabela 1: Opisna statistika pri posameznih preizkusih glede na število pravilno rešenih primerov

	Zaporedja					
	Besede	Števila	Skupaj	Bralno razumevanje	Fenološko zavedanje	Delovni spomin
Skupno št. primerov	4	4	8	20	12	15
Minimum	1	2	4	11	4	0
25. percentil		4	7	14	9	5
Mediana (50 percentil)	4	4	8	17	10	7
75. percentil	4	4	8	19	11	9
Maksimum	4	4	8	20	12	13
Aritmetična sredina	3,63	3,79	7,42	16,36	9,73	6,95
Standardni odklon	0,72	0,51	1,02	2,80	1,87	2,74

blizu mediane (8 točk) oz. se nahaja največ 1 točko pod tem rezultatom. Pri preizkusu branja in bralnega razumevanja so otroci z najvišjim rezultatom dosegli vseh 20 točk, otroci z najnižjim rezultatom pa 11 točk. Otrokov napredek je glede na vzorec ustrezen, če se njegov rezultat nahaja okoli mediane (17 točk) oz. se nahaja največ 3 točke pod tem rezultatom. Pri preizkusu fonološkega zavedanja so otroci z najvišjimi rezultati dosegli vseh 12 točk, tisti z najnižjimi pa 4 točke. Otrokov napredek je glede na naš vzorec ustrezen, če se njegov rezultat nahaja blizu 50. percentila (mediane) oz. se nahaja največ 1 točko pod tem rezultatom. Na področju delovnega spomina so otroci dosegli največ 13 od skupno 15-ih točk, eden otrok pa ni dosegel nobene točke. Otrokov napredek je glede na vzorec ustrezen, če se njegov rezultat nahaja okoli mediane (7 točk) oz. največ 2 točki pod tem rezultatom.

Na sliki 4 je prikaz rezultatov točk predstavljen grafično. Glede na to, da je bil vzorec otrok, ki so reševali preizkuse, zelo majhen, v njem pa so bili tako otroci, za katere njihovi starši menijo, da imajo bralno napisovalne težave, ki bi lahko bile povezane z disleksijo, kot tudi otroci brez težav, nismo pričakovali normalne porazdelitve rezultatov, ki je značilna za populacijo. Prav tako tudi težko sklepamo, ali so naloge dovolj težke in dovolj občutljive, da lahko na osnovi rezultatov zaključimo, ali so otrokovi nizki rezultati povezani z disleksijo. Pri preizkusu zaporedja (slika 4, preizkus 1) je več kot polovica otrok dosegla vse možne točke, 10 jih je doseglo le eno točko manj, medtem ko je le pet otrok doseglo 4 do 5 točk. Preizkus je bil glede na rezultate za otroke našega vzorca lahek, kar pa ne pomeni, da ne bi dobro ločil otrok z disleksijo od otrok brez nje. Enako velja tudi za druge preizkuse, za katere je iz grafičnih

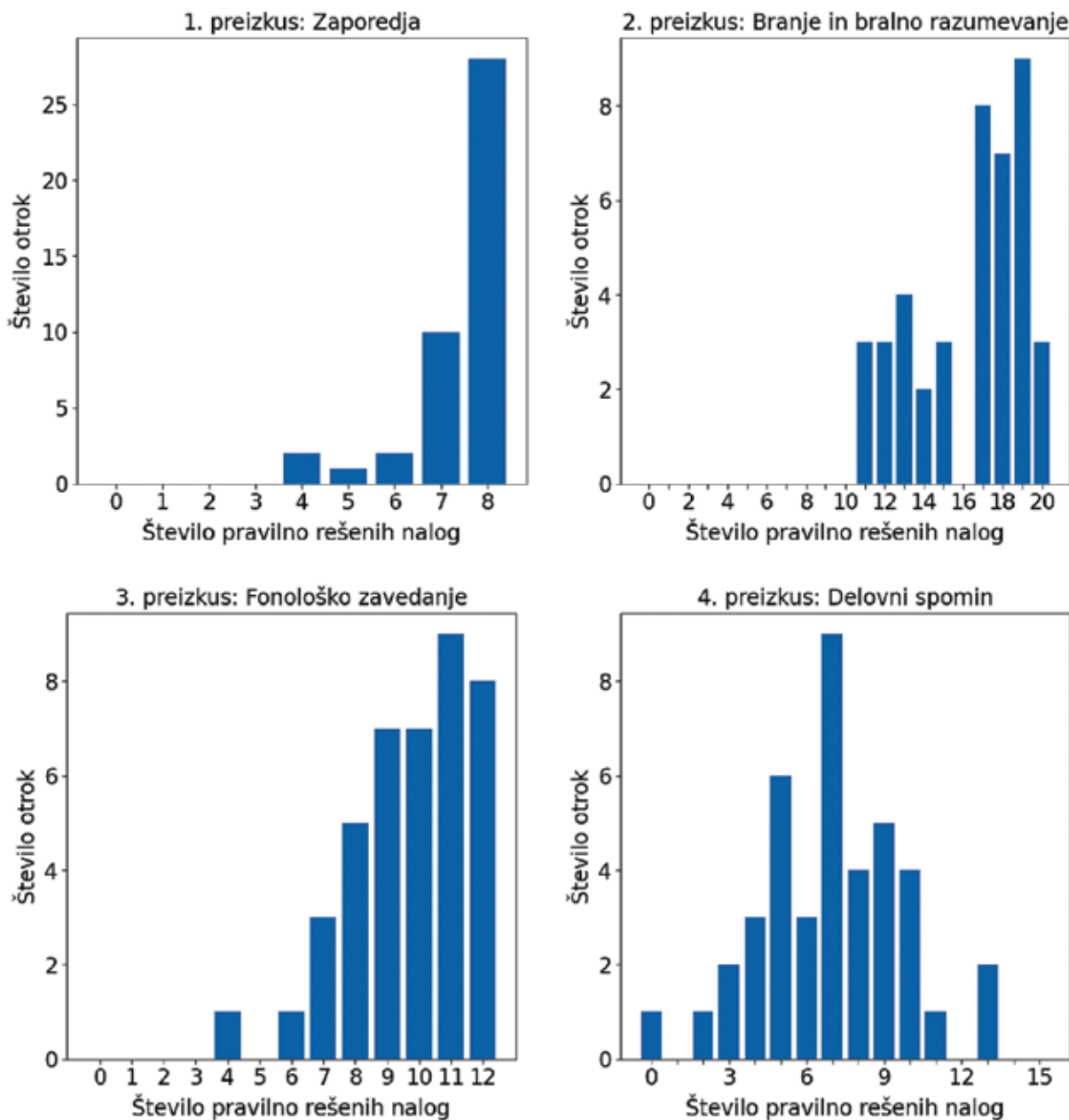
prikazov razvidno (slika 4, preizkusi 2, 3 in 4), da je pri reševanju le majhno število otrok doseglo malo točk. Največjo razpršenost rezultatov lahko vidimo pri preizkusu delovnega spomina (slika 4, preizkus 4), ki ga je reševalo 41 otrok. Največ (9) jih je doseglo 7 točk, 16 otrok je doseglo manj in 16 otrok več kot sedem točk. Le en otrok med njimi ni dosegel nobene točke, dva, ki sta dosegla največ točk, pa sta od skupno 15-ih dosegla 13 točk.

4.2 Povprečen čas reševanja

Zanimali so nas osnovni parametri opisne statistike in razporeditev rezultatov (percentilne vrednosti) pri posameznih preizkusih glede na povprečen čas reševanja.

Rezultati v preglednici 2 prikazujejo povprečen čas, ki so ga otroci porabili za reševanje posamezne naloge vsakega preizkusa. Pri preizkusu koncepta zaporedja je najhitrejši otrok za rešitev posamezne naloge v povprečju potreboval 13,38 sekunde, najpočasnejši pa 59,42 sekunde. Pri preizkusu fonološkega zavedanja je otrok, ki je hitro reševal naloge, porabil za odgovor v povprečju 1,15 sekunde, otrok, ki je bil najpočasnejši pa je za odgovor povprečno potreboval 12,69 sekunde. Pri preizkusu, s katerim smo preverjali delovni spomin, povprečen čas reševanja prikazuje čas, v katerem je otrok rešil nalogo tako, da je kliknil na vse živali v zahtevanem vrstnem redu. Najhitrejši otrok je potreboval za posamezno nalogo v preizkusu v povprečju 3,11 sekunde, najpočasnejši pa 72,38 sekunde.

Ker je preizkus aplikacije potekal samostojno v domačem okolju, otrok med reševanjem nismo mogli nadzirati, kot bi jih lahko na primer pri preizkusu v učilnici. Zato moramo biti pozorni na morebitno



Slika 4: Histogrami prikazujejo število pravilno rešenih primerov za vsak preizkus.

napačno razlago nekaterih časov reševanja. Glede na zelo nizke minimalne čase reševanja pri posameznih preizkusih bi lahko sklepali, da so nekateri otroci rešitve potrdili, ne da bi se trudili za pravilnost rešitev (npr. naključno potrdili ali zavrnili ponujeni odgovor, potrdili odgovor, ne da bi prebrali besedilo). Hkrati ne moremo vedeti razloga za dosežen nadpovprečno dolg čas branja besedil iz drugega preizkusa. Razlog so lahko izrazito počasno branje, počasno procesiranje in obdelava informacij (kar so tudi simptomi disleksije), lahko pa je otrok za reševanje porabil veliko

časa, ker je med reševanjem za nekaj časa prenehal brati besedilo, ni pa prekinil povezave z aplikacijo.

Vpogleda v to, kaj je otrok počel med reševanjem nalog nimamo, lahko pa preverimo pravilnost odgovorov pri otrocih, ki so reševali zelo hitro. Če je otrok reševal nadpovprečno hitro in dosegel dober rezultat, lahko sklepamo, da je naloge reševal korektno, če pa se nadpovprečno hitremu času pridružijo še zelo slabi rezultati, lahko sklepamo, da je otrok naloge reševal nekorektno. Individualnega reševanja posameznikov na podlagi primerjave hitrosti reševanja

Tabela 2: Opisna statistika pri posameznih preizkusih glede na povprečen čas reševanja nalog (v sekundah)

	Zaporedja	Bralno razumevanje			
		Čas reševanje	Čas branja	Fenološko zavedanje	Delovni spomin
Minimum	13.38	2.54	4.05	1.15	3.11
25. percentil	23.17	35.70	14.96	2.12	14.57
Mediana (50 percentil)	29.04	62.36	20.06	3.17	18.19
75. percentil	38.47	90.88	28.57	4.25	25.87
Maksimum	59.42	229.15	63.98	12.69	72.38
Aritmetična sredina	32.07	70.43	22.25	3.56	21.38
Standardni odklon	12.46	49.63	12,06	2.05	12,18

in pravilnosti odgovorov v okviru tega prispevka nismo ocenjevali, saj to ni njegov namen. Pri nadaljnji uporabi aplikacije bo to eden od načinov preverjanja korektnosti reševanja preizkusov.

Dobljeni rezultati kažejo, da bi bilo koristno, da je pri izvajanju reševanja preizkusov za ugotavljanje zgodnjih znakov disleksije z aplikacijo PKP – Disleksija prisotna odrasla oseba, ki nadzoruje potek reševanja (na primer šolska knjižničarka, osebje zdravstvenega doma ali pri domačem reševanju eden izmed staršev, ki je prejel primerna navodila za spremljanje otroka pri reševanju). Ker je predvideni primarni namen končne aplikacije ocena tveganja pri individualnih testirancih in ne statistično zbiranje podatkov, je tudi problem kratkih in dolgih časov reševanja nekoliko lažje obvladljiv; kratek čas skupaj s slabim rezultatom lahko kaže na neustrezno reševanje nalog, dolg čas reševanja pa je lahko tudi od dobrem rezultatu pokazatelj morebitnega tveganja za disleksijo. Če je aplikacija namenjena domači uporabi, moramo tudi na podlagi dobljenih rezultatov pričakovati, da nekateri uporabniki ne bodo sledili navodilom, kar pa bo imelo za posledico tudi napačno oceno tveganja za disleksijo, kar pa ne bo vplivalo na uporabnost aplikacije za tiste, ki jo bodo ustrezno uporabljali v skladu z njenim namenom.

4.3 Težavnost vključenih besedil

Težavnost besedil, vključenih v preizkus branja in bralnega razumevanja, je pomembna, saj morajo biti besedila dovolj enostavna, da so primerna za otroke tretjega razreda. Hkrati morajo besedila po težavnosti ustrezno naraščati, kot smo predvideli, in vsebovati dovolj specifičnih značilnosti, ki bi lahko

pri branju povzročile težave otrokom z disleksijo. Težavnost besedil smo ocenjevali z aplikacijo, ki je nastala v okviru projekta *Za kakovost slovenskih učbenikov (KaUč)*.

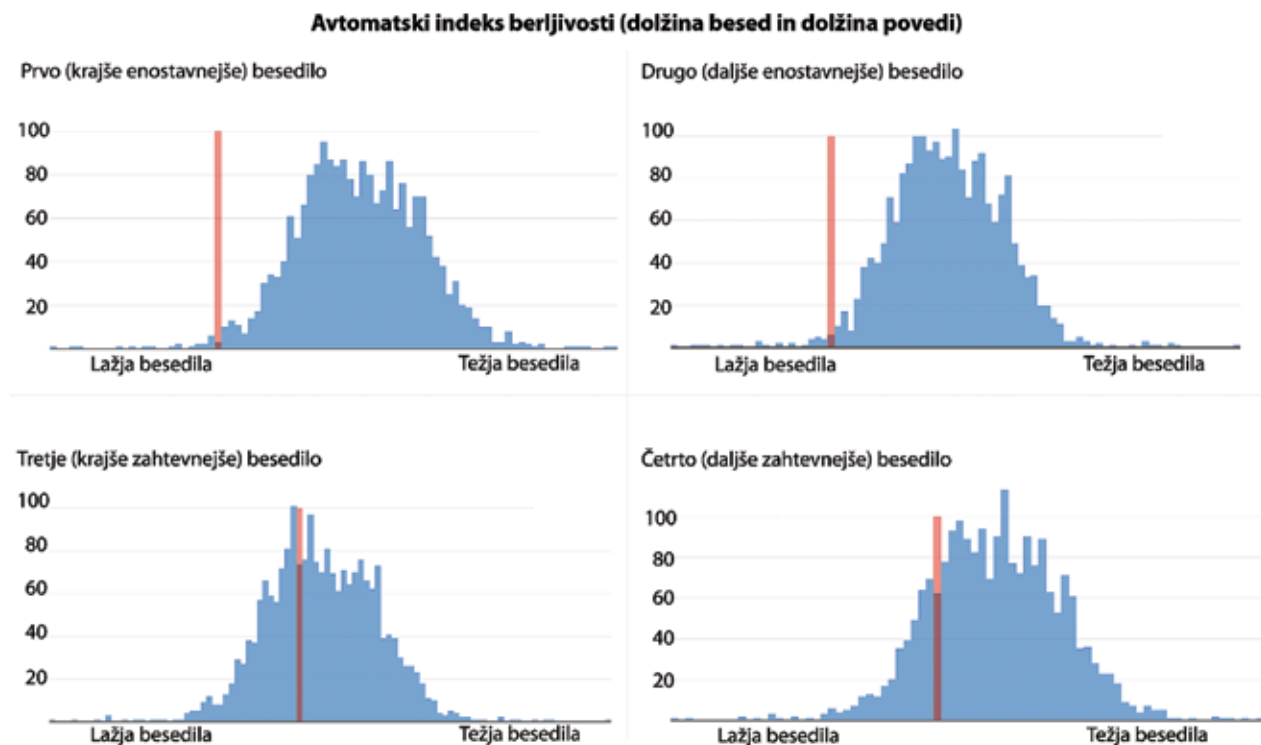
Pri oblikovanju besedil smo upoštevali naslednje kriterije, ki pri branju značilno ovirajo osebe z disleksijo:

- skupno število povedi,
- število eno- in večstavčnih povedi,
- število stavkov v večstavčnih povedih,
- število besed s tremi ali več zlogi,
- število besed s soglasniškimi sklopi,
- število soglasniških sklopov v besedi in
- število soglasnikov v soglasniškem sklopu.

Oblikovali smo štiri besedila, dve krajši (35 besed) in dve daljši (80 besed). Pri posamezni dolžini smo predvideli po eno preprostejšo in eno zahtevnejšo besedilo. V času razvoja aplikacije PKP – Disleksija aplikacija za preverjanje težavnosti besedil v slovenskem jeziku še ni bila na voljo. Zato smo s spletno aplikacijo, nastalo v okviru projekta *Za kakovost slovenskih učbenikov (KaUč)*, preverili že končna oblikovana besedila.

Oblikovana besedila so se izkazala za preprosta in kot takšna primerna za mlajše otroke; tudi njihova težavnost narašča, kot smo predvideli. Glede na avtomatski indeks berljivosti in indeks Coleman-Liau, ki kot kriterij upoštevatata dolžino besed in povedi, se je izkazalo, da sta tako krajše kot daljše zahtevnejše besedilo dosegla višje vrednosti težavnosti kot obe enostavnejši besedili. Slika 5 prikazuje grafične prikaze berljivosti glede na avtomatski indeks berljivosti v aplikaciji *KaUč*.

¹ *KaUč* namreč še nima posebnega ocenjevanja za besedila, napisana za otroke, temveč ocenjuje besedila glede na vsa besedila v slovenskem prostoru, tudi za odrasle.



Slika 5: Grafični prikazi berljivosti glede na avtomatski indeks berljivosti iz aplikacije KaUč za besedila iz preizkusa z branjem in bralnim razumevanjem.

Pri prvem (krajšem in preprostejšem) besedilu je aplikacija izpostavila dve redki besedi, nobena beseda pa ni bila zunaj učbeniškega besedišča. Pri drugem (daljšem in enostavnejšem) besedilu je bilo izpostavljenih 17 redkih besed, šest pa jih je bilo zunaj učbeniškega besedišča. Pri tretjem (krajšem in zahtevnejšem) besedilu je bilo izpostavljenih sedem redkih besed, ena pa je bila zunaj učbeniškega besedišča. Pri četrtem (daljšem in zahtevnejšem) besedilu je aplikacija prav tako izpostavila sedem redkih besed, tri pa so bile zunaj učbeniškega besedišča. Glede na ta dva kriterija je bilo drugo besedilo bolj težavno kot četrto, čeprav je bil naš cilj ravno obraten.

Pri analizi izpostavljenih redkih besed iz drugega besedila smo ugotovili, da je aplikacija kot redke izpostavila nekatere besede, za katere tega nismo pričakovali (npr. muca, tačko, zunaj, mačje, stekla (deležnik glagola steči), košata). Po drugi strani v četrtem besedilu kot redke niso bile označene nekatere besede, ki smo jih v besedilo vključili namerno, da bi ga naredili zahtevnejšega (npr. dvorišča, prehitel, steza, cvetočega, skrbno, priklenila, se podala, pričakali). Čeprav redkost izbranih besed v besedilih nakazuje obratno od naših namer pri oblikovanju besedil, ta

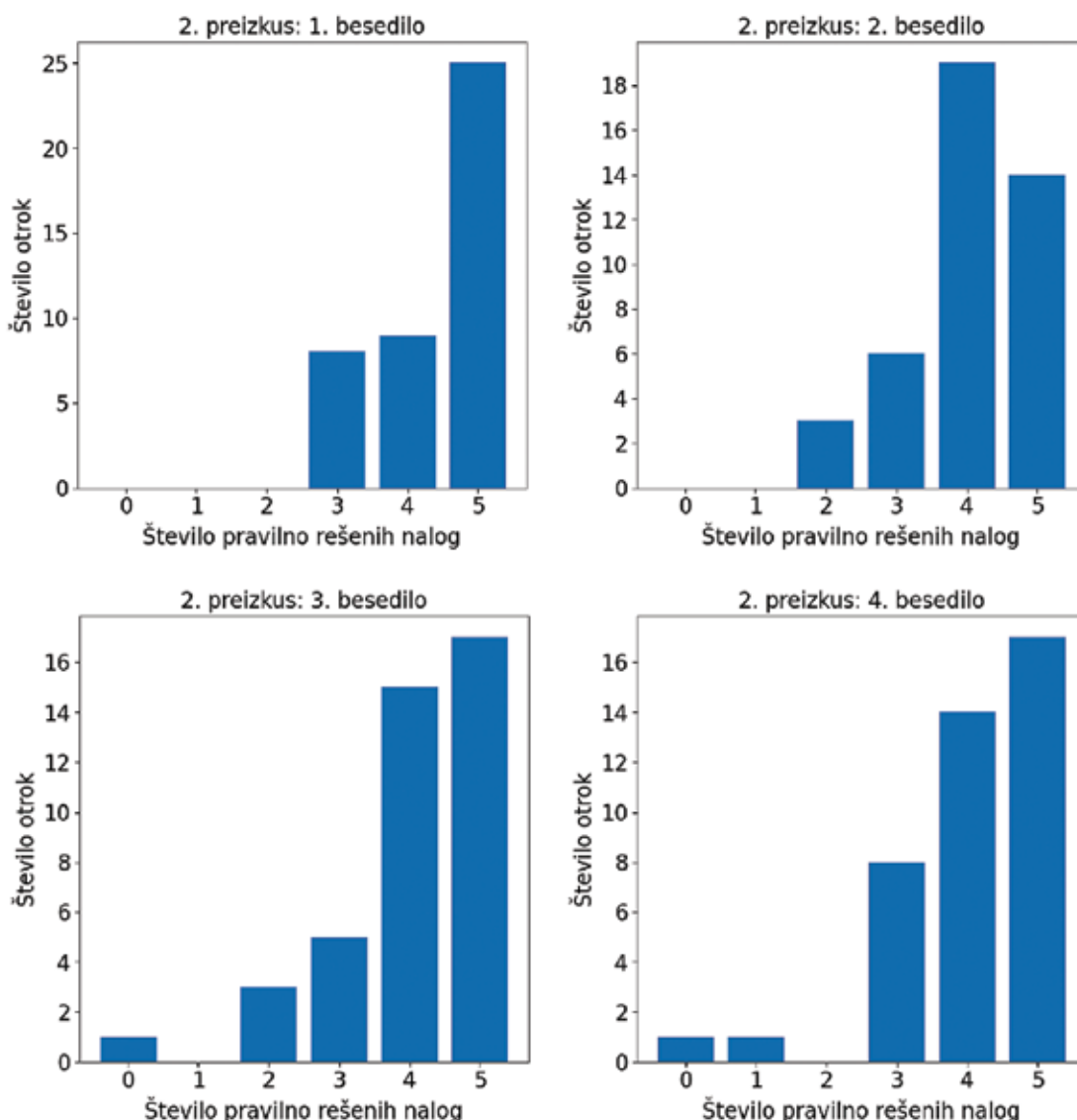
kriterij ni edini, ki otrokom z disleksijo otežuje branje. Bolj pozorni smo bili, da smo v zahtevnejša besedila vključili besede z več soglasniškimi sklopi, pri katerih lahko otroci z disleksijo pogosteje naredijo napako pri branju ne glede na pogostost pojavljanja te besede. Čeprav ostajamo kritični do tega, da bi lahko imela redkost pojavljanja nekaterih označenih besed (npr. muca, tačko, mačje) v splošnih slovenskih besedilih velik vpliv na berljivost in težavnost besedila pri otrocih,¹ bi lahko v prihodnje pri oblikovanju novih besedil bolj upoštevali tudi ta kriterij.

Da je drugo besedilo najzahtevnejše, bi lahko potrdilo tudi dejstvo, da je to edino besedilo, pri katerem je aplikacija izpostavila eno možno težavo pri branju povedi. Izpostavljena je bila namreč poved z več kot 10 besedami. Ob ponovnem pregledu besedil smo ugotovili, da drugo besedilo vsebuje dve povedi, daljši od 10 besed. Vredno se zdi izpostaviti tudi, da pri četrtem besedilu, ki vsebuje štiri povedi, daljše od 10 besed, ni bila izpostavljena nobena. Preverili smo, ali katero od besedil znotraj povedi vsebuje stavke z več kot 10 besedami, pri čemer se je izkazalo, da pri drugem besedilu najdaljši stavek šteje natančno 10 besed, vsi drugi pa so krajši, četrto besedilo pa

vsebuje en stavek znotraj daljše povedi, ki vsebuje 11 besed in en stavek znotraj druge povedi, ki vsebuje 10 besed. Zaradi teh netočnosti v aplikaciji *KaUč* smo se odločili zahtevnosti oblikovanih besedil pri analizi rezultatov posvetiti več pozornosti in preveriti, v kolikšni meri se je težavnost besedil pokazala pri odgovorih otrok v neslučajnostnem vzorcu.

Glede na število pravih odgovorov je iz slike 6 razvidno, da je bilo za otroke v populaciji prvo besedilo res najpreprostejše, saj je na vsa vprašanja pravilno odgovorilo 25 otrok, vsi pa so pravilno odgovorili vsaj na tri vprašanja. Drugo besedilo je od prvega daljše, a smo želeli ohraniti približno enako stopnjo težavnosti kot pri prvem besedilu. Res se število otrok, ki so pravilno odgovorili na vsa vpraša-

nja spusti na 14, največ otrok je doseglo štiri pravilne odgovore, vsi pa so pravilno odgovorili vsaj na dve vprašanja. Otroci s samo dvema pravilnima odgovoroma so bili trije. Tretje in četrto besedilo sta se izkazala za besedila s podobno stopnjo težavnosti, kot smo predvideli pri oblikovanju, kar je razvidno iz na videz zelo podobnih histogramov, v katerih število pravih odgovorov približno eksponentno narašča. Pri obeh besedilih je največ otrok doseglo vse pravilne odgovore, in sicer 17. Podobno število otrok je pravilno odgovorilo na štiri vprašanja, nekoliko bolje jim je šlo pri tretjem, med obema krajšem besedilu. Pri obeh besedilih po en otrok ni odgovoril pravilno na nobeno vprašanje, pri četrtem pa je še en otrok odgovoril pravilno le na eno, kar se pri obeh eno-



Slika 6: Histogrami prikazujejo število pravilno rešenih primerov pri vsakem besedilu v drugem preizkusu.

stavnejših besedilih ni zgodilo. Pri četrtem, daljšem besedilu je tri ali manj pravih odgovorov doseglo 10 otrok, pri tretjem, krajšem pa devet, vendar je bil med njimi delež tistih, ki so pravilno odgovorili na tri vprašanja, večji pri četrtem besedilu.

Drugo besedilo najbolj izstopa tudi med prikazi v histogramih, saj je bilo to edino besedilo, pri katerem je največji delež otrok v populaciji pravilno odgovoril na štiri vprašanja, in ne na vseh pet, kot pri drugih besedilih. Prav tako je število otrok, ki so pravilno odgovorili na vsa vprašanja za tri manjše kot pri obeh predvidoma zahtevnejših besedilih. Odločili smo se podrobneje preveriti, ali je bil slabo reševan celoten preizkus, kar bi potrjevalo tezo o zahtevnosti tega besedila, ali je morda katero izmed vprašanj pretežno in razlog za slabše reševanje verjetno ni zahtevnost besedila. Izkazalo se je, da je na 1., 3., 4. in 5. vprašanje pravilno odgovorilo med 36 in 42 otrok, drugo vprašanje pa je zelo izstopalo, saj je bilo pravih odgovorov približno za polovico manj, in sicer 21. Drugo vprašanje je pri vseh besedilih enako: »O čem govori besedilo? Izberi odgovor, ki se najbolj ujema z vsebino besedila.« Pri vseh besedilih smo predvideli dva odgovora, ki se ujemata z vsebino besedila (pri drugem besedilu je kot pravih predviden odgovor: *O mucinem nočnem sprehodu*), le da se eden od njiju nanaša samo na določen motiv iz besedila in ne na celotno besedilo, zato je ta odgovor predviden kot napačen (*O mačjih prijateljih*). Tretji predvideni odgovor je sicer možno delno povezati z motivi iz besedila, a je glede na celotno besedilo popolnoma napačen (*O spanju*), četrti pa sploh nima povezave s tematiko besedila (*O stavbah*). Glede na vsebino besedila smo bili pri oblikovanju dveh podobnih odgovorov pri drugem vprašanju, ki se nanaša na drugo besedilo, glede na pridobljene podatke prestrogi in bomo morali v prihodnje oblikovati nove odgovore, ki bodo tudi bolje razločili med otroki s težavami in tistimi brez njih.

Iz opisanega lahko sklepamo, da smo ne glede na nepredvideno oceno berljivosti iz aplikacije za oceno berljivosti besedil v slovenskem jeziku, nastale v okviru projekta *KaUč*, besedila oblikovali v skladu s svojim namenom. Pri drugem besedilu je bil glede na rezultate razlog za slabšo reševanost nalog namreč v predvidenih odgovorih na eno izmed vprašanj in ne v neprimerni zahtevnosti samega besedila, ki jih bomo zato v prihodnje po potrebi prilagodili. Kljub temu je aplikacija za ocenjevanje težavnosti besedil odprla nove poglede na oblikovana besedila, ki jih

nismo predvideli. V prihodnje bo lahko v pomoč pri oblikovanju besedil za otroke z disleksijo.

5 ZAKLJUČEK

Aplikacijo za zgodnje odkrivanje znakov disleksije smo razvili za hitrejšo in bolj točno prepoznavanje otrok, pri katerih je prisotno tveganje za pojav disleksije. Z njo želimo prispevati k hitrejšemu in zgodnejšemu prepoznavanju težav na področjih, ki so tipična za disleksijo, ter k pravočasnemu ukrepanju. Z zgodnjim ukrepanjem oz. usmerjeno učno pomočjo lahko preprečimo, da bi se težave poglobile in razširile na vsa učna področja otrokovega delovanja. V Sloveniji podobnih testov za prepoznavanje disleksije v elektronski obliki še ni; obstajajo klasična testiranja otrok ob prisotnosti strokovnjaka, vendar zaradi daljših čakalnih dob in težje dostopnosti (v primerjavi s spletno aplikacijo) obravnava otroka zamuja, s čimer izgubljam dragocen čas za pomoč otrokom z disleksijo. Potrebo po elektronskem testiranju je prepoznala tudi stroka, ki si ob aplikaciji za testiranje želi še razvoja ločene aplikacije z vajami za tiste, pri katerih bi ugotovili znake disleksije.

Ob analizi rezultatov raziskave opazimo, da v nekaterih primerih rezultati predvsem z vidika porabljenega časa izrazito odstopajo v obe skrajnosti. V nekaterih primerih porabljen čas pri posamezni nalogi občutno presega predviden čas, ki bi ga otrok porabil za nalogo tudi v primeru izrazitejših težav, v nekaterih drugih primerih pa je porabljen čas izrazito kratek in menimo, da ni sorazmeren z obsegom oz. z zahtevami naloge. V obeh primerih so posledica rezultati, ki ne dajejo nujno realnih informacij in jih zato ne moremo uporabiti v prvotno predviden namen. Menimo, da bi se opisanemu lahko izognili tako, da bi otrok naloge v aplikaciji reševal ob prisotnosti usposobljene osebe, ki bi mu pomagala pri morebitnih tehničnih težavah, bi imela informacije o morebitnih prekinitvah reševanja in drugih nepredvidenih dejavnikih, o katerih v tej raziskavi nismo imeli informacij, hkrati pa ne bi bila čustveno vpletena v situacijo in otroku ne bi pomagala pri reševanju posameznih nalog.

S preizkusom aplikacije v praksi na neslučajnostnem vzorcu, v katerega nismo namensko vključevali otrok z diagnosticirano disleksijo, smo pridobili prve preliminarne statistične podatke o aplikaciji, s čimer smo dobili vpogled v reševanje oblikovanih preizkusov in nalog. Na podlagi tega bomo ustrezno prilagodili nadaljnje zbiranje podatkov in izpo-

polnjevanje ocene tveganja za disleksijo. Hkrati smo na podlagi pridobljenih podatkov opazili nekatere pomanjkljivosti pri oblikovanju nalog, kar nam bo v prihodnje koristilo pri preoblikovanju nalog tako, da bodo čim boljše diskriminirale med otroki s tveganjem za disleksijo in tistimi brez njega.

Analiza zbranih podatkov je odprla možnosti za nadaljnje raziskovanje, s katerim želimo podrobneje preveriti točnost aplikacije pri odkrivanju znakov disleksije pri splošni populaciji osemletnih otrok. V prihodnja testiranja v nenadzorovanem okolju bomo vključili večje število otrok, načrtujemo pa tudi testiranje v nadzorovanem okolju z uporabo tehnologije za sledenje očesnim gibom, ki ga epidemiološke razmere v času te raziskave niso dopuščale. K sodelovanju bomo povabili tudi otroke z diagnosticirano disleksijo, primerjava njihovih rezultatov z rezultati otrok brez disleksije pa bo ključno pripomogla pri ugotavljanju uporabne vrednosti aplikacije.

6 ZAHVALA

Aplikacijo smo razvili v okviru projekta PKP – Disleksija. Projekt sofinancirata Republika Slovenija in Evropska unija iz Evropskega socialnega sklada.

7 LITERATURA

- [1] Chen, C., Schneps, M. H., Masyn, K. E. & Thomson, J. M. (2016). The effects of visual attention span and phonological decoding in reading comprehension in dyslexia: a path analysis. *Dyslexia*, 22(4), 322–344.
- [2] Chung, K. K. H., Lam, C. B. & Leung, C. O. Y. (2020). Contributions of executive functioning to Chinese and English reading comprehension in Chinese adolescent readers with dyslexia. *Reading and Writing: An Interdisciplinary Journal*.
- [3] Drigas, A. & Politi-Georgousi, S. (2019). ICTs as a Distinct Detection Approach for Dyslexia Screening: A Contemporary View. *International Journal of Online Engineering (iJOE)*, 15(13), 46–60.
- [4] Elley, W. B., Gradišar, A. & Lapajne, Z. (1995). Preizkus bralnega razumevanja. *Educa*. Gradišar, A. & Pečjak, S. (1991). Enominutni test glasnega branja. *Pedagoški inštitut*.
- [5] Jeffries, S. & Everatt, J. (2004). Working memory: Its role in dyslexia and other specific learning difficulties. *Dyslexia*, 10(3), 196–214.
- [6] Košak Babuder, M. (2013). Dijaki z disleksijo v srednji šoli – pomoč in podpora. *Šolsko svetovalno delo : revija za svetovalne delavce v vrtcih, šolah in domovih*, 17 (1/2), 14–21.
- [7] Košir, J. (2011). Formativno ocenjevanje s preizkusom tekočnosti branja, ki temelji na kurikulu. (L. M. in M. Velikonja, Ur.). V L. M. in M. Velikonja (Ur.), *Učenci z učnimi težavami: prepoznavanje in diagnostično ocenjevanje*, Univerza v Ljubljani, Pedagoška fakulteta.
- [8] Lauterbach, A. A., Park, Y. & Lombardino, L. J. (2017). The roles of cognitive and language abilities in predicting decoding and reading comprehension: comparisons of dyslexia and specific language impairment. *Annals of dyslexia*, 67 (3), 201–218.
- [9] Magajna, L. (2011). Prepoznavanje in diagnostično ocenjevanje učnih težav: osnovna izhodišča in pristopi (L. Magajna & M. Velikonja, Ur.). V L. Magajna & M. Velikonja (Ur.), *Učenci z učnimi težavami – prepoznavanje in diagnostično ocenjevanje*, Pedagoška fakulteta, Univerza v Ljubljani.
- [10] Magajna, L., Kavkler, M., Košak Babuder, M., Zupančič Danko, A., Seršen Fras, A. & Rošer Obretan, A. (2015). VII. Otroci s primanjkljaji na posameznih področjih učenja (N. Vovk-Ornik, Ur.). V N. Vovk-Ornik (Ur.), *Kriteriji za opredelitev vrste in stopnje primanjkljajev, ovir oziroma motenj otrok s posebnimi potrebami*, Zavod RS za šolstvo.
- [11] Magajna, L. (1994). *Razvoj bralnih strategij–vloga kognitivnega in fonološkega razvoja ter fonološke strukture jezika* (Doktorska disertacija). Univerza v Ljubljani, Filozofska fakulteta.
- [12] Menghini, D., Finzi, A., Carlesimo, G. A. & Vicari, S. (2011). Working memory impairment in children with developmental dyslexia: Is it just a phonological deficit? *Developmental neuropsychology*, 36(2), 199–213.
- [13] Nijakowska, J. (2016). Grasping dyslexia: Bridging the gap between research and practice. *Selected papers on theoretical and applied linguistics*, 21, 43–58.
- [14] Palmer, S. (2000). Phonological recoding deficit in working memory of dyslexic teenagers. *Journal of Research in reading*, 23(1), 28–40.
- [15] Pečjak, S., Magajna, L. & Podlesek, A. (2012). *Ocenjevalna shema bralnih zmognosti učencev 1.-3. razreda: OSBZ*. Znanstvena založba Filozofske fakultete Univerze v Ljubljani.
- [16] Pečjak, S., Magajna, L., Podlesek, A. & Potočnik, N. (2012). *Bralni test*. Znanstvena založba Filozofske fakultete Univerze v Ljubljani.
- [17] Peklaj, C. (2012). *Učenci z učnimi težavami v šoli in kaj lahko stori učitelj*. Znanstvena založba Filozofske fakultete Univerze v Ljubljani.
- [18] Pham, A. V. & Hasson, R. M. (2014). Verbal and visuospatial working memory as predictors of children's reading ability. *Archives of Clinical Neuropsychology*, 29(5), 467–477.
- [19] Pollak, D. (2009). *Neurodiversity in higher education: Po responses to specific learning differences*. John Wiley & Sons.
- [20] Raduly Zorgo, E., Smythe, I., Gyarmathy, É., Košak Babuder, M., Kavkler, M. & Magajna, L. (2010). *Disleksija- vodnik za tutorje*. Univerza v Ljubljani, Pedagoška fakulteta.
- [21] Riddick, B., Wolfe, J. & Lumsdon, D. (2002). *Dyslexia: A practical guide for teachers and parents*. Routledge. Rooms, M. (2000). Information and Communication Technology and Dyslexia. V J. Towned & M. Turner (Ur.), *Dyslexia in Practice: A Guide for Teachers* (str. 263–272). Springer Science+Business Media.
- [22] Schuchardt, K., Bockmann, A.-K., Bornemann, G. & Maehler, C. (2013). Working memory functioning in children with learning disorders and specific language impairment. *Topics in Language Disorders*, 33(4), 298–312.
- [23] Siegel, L. S. & Linder, B. A. (1984). Short-term memory processes in children with reading and arithmetic learning disabilities. *Developmental psychology*, 20(2), 200–207.
- [24] Singleton, C. (2001). Computer-based assessment in education. *Educational and Child Psychology*, 18(3), 58–74.
- [25] Singleton, C., Horne, J. & Simmons, F. (2009). Computerised screening for dyslexia in adults. *Journal of Research in Reading*, 32(1), 137–152.
- [26] Smith-Spark, J., Fisk, J., Fawcett, A. & Nicolson, R. (2003). Investigating the central executive in adult dyslexics: Evidence from phonological and visuospatial working memory performance. *European Journal of Cognitive Psychology*, 15(4), 567–587.
- [27] Snowling, M. J. (2013). Early identification and interventions for dyslexia: a contemporary view. *Journal of Research in Special Educational Needs*, 13(1), 7–14.

- [28] *Standards for educational and psychological testing.* (2014). American Educational Research Association. Šali, B. (1971). *Test motenosti v branju in pisanju (T-MBP). Zavod SR Slovenije za rehabilitacijo invalidov.* Weedon, C. & Reid, G. (2018). *SNAP-3 : profil ocene posebnih potreb, 3. izdaja : računalniško podprto ocenjevanje in izdelava profila specifičnih učnih težav (5–14 let) : priročnik.* Center za psihodiagnostična sredstva.
- [29] *Zakon o osnovni šoli ZOSn-UPB3.* (2006). *Uradni list RS, št. 81/2006.*
- [30] *Zakon o usmerjanju otrok s posebnimi potrebami ZUOPP-1.* (2011). *Uradni list RD, št. 58/2011.* Žagar, D. (2012). *Drugačni učenci.* Znanstvena založba Filozofske fakultete Univerze v Ljubljani.

■

Dr. Milena Košak Babuder je docentka za področje specialne in rehabilitacijske pedagogike na Pedagoški fakulteti Univerze v Ljubljani. Na Oddelku za specialno in rehabilitacijsko pedagogiko predava predmete, povezane s prepoznavanjem in ocenjevanjem splošnih in specifičnih učnih težav v vseh življenjskih obdobjih, ter predmete, povezane s strategijami in modeli pomoči za njihovo preprečevanje in odpravljanje. Razi- skovalno se ukvarja z inkluzijo oseb s posebnimi potrebami, vplivom splošnih in specifičnih učnih težav na izobraževalno uspešnost učencev, dijakov in študentov ter z razvijanjem strategij in oblikovanjem modelov pomoči in obravnava na teh področjih, še posebej pa raziskuje vpliv disleksije na učenje angleščine kot tujega jezika. Je avtorica in soavtorica več znanstvenih člankov oz. prispevkov v znanstvenih monografijah in študijah.

■

Dušica Boben je profesorica matematike in univ. dipl. psihologinja ter direktorica Centra za psihodiagnostična sredstva – vodilne slovenske založbe za psihološke in sorodne merske pripomočke. Njene izkušnje zajemajo tako razvojno delo in vodenje priredb različnih psiholoških merskih pripomočkov ter usposabljanja zanje kot tudi njihovo uporabo pri otrocih, mladostnikih in odraslih.

■

Ema Štarkl je študentka magistrske stopnje splošnega jezikoslovja na Filozofski fakulteti Univerze v Ljubljani. Diplomirala je na področju teoretičnega jezikoslovja in geolingvistike oz. narečjeslovja na Oddelkih za primerjalno in splošno jezikoslovje ter za slovenistiko. Na področju odkrivanja znakov disleksije s pomočjo IKT je sodelovala na študentskih projektih ŠIPK in PKP.

■

Karmen Javornik je magistrica profesorica specialne in rehabilitacijske pedagogike. Zaposlena je na oddelku za specialno in rehabilitacijsko pedagogiko Pedagoške fakultete Univerze v Ljubljani. Znanja in izkušnje s področja specifičnih učnih težav je poleg študija specialne in rehabilitacijske pedagogike ter različnih dodatnih izobraževanj in izpopolnjevanj pridobivala v službi z izvajanjem dodatne strokovne pomoči, individualne in skupinske pomoči ter med prostovoljnim delom na tem področju.

■

Erika Stanković je magistrska študentka na Fakulteti za računalništvo in informatiko v Ljubljani. Leta 2020 je zaključila študij Računalništva in matematike z diplomom z naslovom Prepoznavanje disleksije pri otrocih s pomočjo analize očesnih gibov. Aktivna je na področju prepoznavanja disleksije, trenutno pa se zanima za metode strojnega učenja in njihovo uporabo.

■

Gaja Nenadović je leta 2020 diplomirala iz psihologije na Filozofski fakulteti Univerze v Ljubljani. Trenutno se zanima predvsem za umestitev psihološkega raziskovanja v širši družbeni kontekst

■

Blažka Korun je študentka 1. stopnje univerzitetnega študija Logopedija in surdopedagogika na Pedagoški fakulteti Univerze v Ljubljani.

■

Sara Jakop je študentka 1. stopnje visoko-strokovnega študija Računalništvo in informatika na Fakulteti za računalništvo in informatiko Univerze v Ljubljani.

■

Luka Vranješ je študent 2. letnika magistrskega študija Podatkovne vede na Fakulteti za računalništvo in informatiko Univerze v Ljubljani.

■

Dr. Jure Žabkar je docent za področje računalništva in informatike na Fakulteti za računalništvo in informatiko Univerze v Ljubljani. Je član Laboratorija za umetno inteligenco in predava predmete povezane s programiranjem in umetno inteligenco ter strojnimi učenjem. Raziskovalno se ukvarja s strojnimi učenjem in podatkovno analitiko, kvalitativnim sklepanjem ter razvojem računalniških sistemov za prepoznavanje disleksije. Je avtor in soavtor več znanstvenih člankov oz. prispevkov v znanstvenih monografijah in študijah.

☒ Pogovor sodelavcev ob 20-letnici Islovarja

Katarina Puc, Niko Schlamberger, Vladimir Batagelj, Tomaž Turk, Ivan Kanič

Slovensko društvo INFORMATIKA

puckatarina@gmail.com, niko.schlamberger@gmail.com, vladimir.batagelj@fmf.uni-lj.si, tomaz.turk@ef.uni-lj.si, ivan.kanic@gmail.com

Izvleček

Prispevek predstavlja dragocen kronološki zapis nastanka in razvoja Islovarja. V pogovoru so podana izhodišča, ki so pred 20 leti privedla do vzpostavitve interaktivnega slovarja za področje računalništva in informatike. Opisani so problemi z izrazoslovjem, na katere so naleteli načrtovalci tega pomembnega spletnega orodja. Razvoj slovarja je v bistvu sledil uvajanju računalništva in informatike v Sloveniji in hkrati doživljal tudi tehnološke spremembe. Iz prvih skromnih poskusov je prerasel v zrelega mladeniča.

Ključne besede: Islovar, spletni slovar, računalništvo in informatika, strokovna terminologija, prevajanje, interaktivnost

Abstract

The article presents an invaluable chronological record of the origin and development of Islovar. The discussion presents the starting points that led to the establishment of an interactive dictionary in the field of computer science and informatics 20 years ago. The problems with terminology encountered by the designers of this important online tool are described. The development of the dictionary essentially followed the implementation of computer science and informatics in Slovenia while at the same time experiencing technological changes. From its first humble attempts, Islovar has grown into a valuable tool.

Keywords: Islovar, online dictionary, computer science and informatics, Slovenian terminology, translation, interactivity

Letos praznuje Islovar 20 letnico svojega nastanka in delovanja. Navadnemu uporabniku se zdi tak slovar nekaj samo po sebi umevnega, dejansko pa lahko pri tem iščemo odgovore na različna vprašanja.

Kako se je porodila ideja o Islovarju?

Kakšni so bili začetki?

Kakšni so bili zgledi?

Kakšna je organiziranost ekipe, ki skrbi za Islovar?

Kako se soočajo sodelavci Islovarja z izzivi, ki jih prinaša buren razvoj IKT? Stalno se pojavljajo nove

stvari, kaj pa morda tudi zastari ali se vsaj spreminja.

Kako vpliva sama sprememba tehnologije na izvedbo Islovarja?

Kako lahko primerjamo Islovar s sorodnimi aplikacijami? Tudi v tujini?

Kakšni so pogledi naprej?

Odgovore na ta vprašanja lahko poiščemo kar v pogovoru s pionirji tega pomembnega orodja. Pogovora so se udeležili Katarina Puc, Niko Schlamberger, Vladimir Batagelj, Tomaž Turk in Ivan Kanič. Zabeležil ga je Saša Divjak



Slika 1: Virtualni sestanek udeležencev pogovora

Katarina:

Iz poročila o delu jezikovne skupine iz leta 2002 lahko preberemo, da smo v marcu 2000 poslali vabilo vsem članom društva. Izvršni odbor društva je nato potrdil ustanovitev jezikovne sekcije in potem smo se sestali v avgustu v ožji skupini in se dogovorili, da bomo skrbeli za strokovni jezik in da bomo izdali terminološki slovar informatike. Spomnim se, da smo se spet sestali decembra. Takrat je bil zraven že Batagelj. Pred tem smo se zbrali na Sredinem seminarju in razpravljali, kakšen naj bi bil slovar. Na decembrskem sestanku smo imeli dva predloga, kako naj bi izgledal. Batagelj je takrat že imel nekaj na internetu in sicer en matematičen in računalniški slovarček. Tam je bilo vse, kar je bilo vidno, že popravljeno. To kar se je delalo in urejalo, pa je bilo skrito za uporabnike. Tomaž pa je prišel z drugo idejo in sicer, da bi bilo vse odprto in da bi bilo vse, kar se dela v Islovarju, javno dostopno in vidno. In smo glasovali. Takrat nas je bilo približno 10. Tomažev predlog je zmagal in sta nato z Jurijem Jakličem delala naprej. Zanimivo je, da je tudi Wikipedija iz tistega časa, iz januarja 2001, in so tudi imeli 20 letnico, ki so jo svečano obeležili. Mislim, da bi bilo prav, da ob naši obletnici tudi mi opremimo Islovar s pasico z napisom »20 let Islovarja«.

Vlado:

Leta 2001 sem za revijo Uporabna informatika napisal članek o slovenskem računalniškem izrazju [Riz], ki je nastal na osnovi mojega predavanja na konferenci Dnevi Slovenske Informatike 2001. Poskusimo na hitro povzeti pomembnejše korake v razvoju slovenskega računalniškega izrazja. Vse skupaj se je začelo z matematiko. Slovensko matematično izrazje je začelo nastajati v 19. stoletju z Močnikom, ki je bil pisec matematičnih učbenikov za celo avstro-ogrsko monarhijo, poskrbel pa je tudi za »domače« knjige. Za to, kar je povezano z računanjem, so za ustrezne izraze poskrbeli matematiki. Matematično terminologijo je v 50 letih zbral in izdal Alojzij Vadnal, profesor na Ekonomski fakulteti [MT]. Po drugi strani je v 50 letih v Moskvi študiral France Križanič in ko se je vrnil domov, je napisal knjigo Elektronski aritmetični računalniki [EAR], ki je najbrž prva slovenska računalniška knjiga. V njej je vpeljal prve slovenske računalniške izraze. S tega vidika je ta knjiga še vedno zanimiva. Sicer opisuje ruske računalnike Strela, a tudi vrsto splošnih računalniških pojmov.

V tistem času je začela izhajati zbirka Sigma, ki jo izdaja DMFA RS. V tej zbirki je bilo objavljenih več knjig, ki prinašajo nova matematična znanja, ki so potrebna v računalništvu. Recimo: Teorija informa-

cij (Rajko Jamnik), Osnove matematične logike (Niko Prijatelj), Numerično reševanje enačb (Zvonimir Bohte) in Verjetnostni račun (Alojzij Vadnal). Tako so v 60 letih matematiki izdali kup knjig, ki so temeljne za računalništvo.

Univerzitetno in raziskovalno računalništvo se je začelo z računalnikom Zuse (1962). Ob tem računalniku so nastale tri skupine. Ena je bila na Štefanu, druga na Matematiki, tretja na Fakulteti za elektrotehniko. Ljudje, ki so delali na Zuseju, so o tem začeli pisati tudi v slovenščini. Tedaj je bila dokaj dejavna tudi terminološka komisija pri Elektrotehniški zvezi. Nekateri slovenski računalnikarji so bili precej dejavni tudi mednarodno. Anton Železnikar in Silvin Leskovar sta uspela »pripeljati« v Ljubljano leta 1971 zelo pomembno konferenco mednarodnega združenja za informatiko IFIP. Konferenca je pomenila precejšnjo vzpodbudo nadaljnjemu razvoju računalništva pri nas. Ob njej se je zgodilo marsikaj. Ena od pomembnih stvari je, da smo v Ljubljani dobili zelo zmogljiv računalnik Cyber. Mislim, da v bližnji okolici ni bilo zmogljivejšega.

Po drugi strani je IFIP vzpodbudil tudi druge dejavnosti. Ob IFIPu je izšla knjiga »Elektronski računalniki«, v kateri so avtorji zbrali takratno računalniško znanje [ELR]. Sestavlja jo izbor poglavij iz računalništva. Za razvoj izrazja pa je zelo pomembno, da so v to knjigo vključili tudi štiri jezični slovarček računalniških izrazov. Zdi se mi, da je tu glavno delo

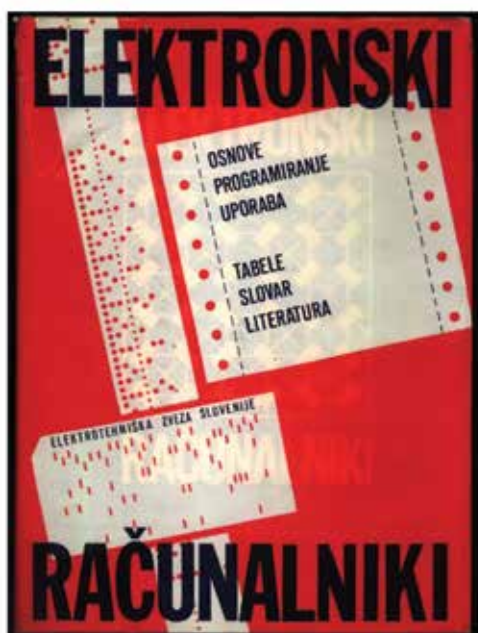
opravila terminološka komisija pri Elektrotehniški zvezi Slovenije.

IFIP je premaknil marsikaj. Na Matematiki smo takrat bili navdušeni nad knjigami, ki jih je napisal D.E. Knuth: The Art of Computer Programming. Egon Zakrajšek je takrat prilagodil predavanja iz računalništva na te knjige, ki so prinašale cel kup novih tematik. Zanje je bilo potrebno najti primerne slovenske izraze.

Leta 1971 se je začel na Matematiki sestajati seminar za numerično in računalniško matematiko, tako imenovani Sredin seminar [Sre]. Na tem seminarju so se zbirali matematiki in računalnikarji od vsepovsod. Na njem smo pogosto odpirali nove teme in ob tem tudi razpravljali o terminologiji: Kako določenim stvarjem lepo rečemo po slovensko?

Druga pomembna stvar je bila, da se je začelo razmišljati o uvajanju računalništva v šole. Po eni strani so na nekaterih fakultetah ob koncu 60-tih in začetku 70-tih let že imeli posamične računalniške predmete, ampak to je bilo pretežno programiranje v fortranu. Po IFIPu pa se je začelo računalništvo razvijati v širšem smislu. Takrat smo bili kar v »špicici« tudi v svetovnem merilu.

Konec 60-tih let se je računalništvo začelo uvajati kot izbirni predmet v nekaterih gimnazijah. Eden od začetnikov je bil Izidor Hafner. Njegov oče je bil znan politik in je spodbudil direktorja Zavoda za šolstvo, da so ustanovili skupino, ki naj bi začela z uva-



Slika 2: Naslovnica knjige Elektronski računalniki in delček slovarčka računalniških izrazov



Slika 3: Priročnik Računalništvo za izbirni predmet v gimnazijah

janjem izbirnega predmeta v srednje šole. Uvajanje je usklajeval Branko Roblek. Najprej so pripravili tečaj za učitelje računalništva. Ob tej priliki je izšla knjiga, nekakšen priročnik, ki naj bi učiteljem pomagal pri tem izbirnem predmetu.

In seveda, če se gre z neko snovjo v šole, mora to biti v slovenščini.

Tudi v tej knjigi je bilo zelo veliko narejenega na področju slovenskega računalniškega izrazja. Pomemben rezultat teh prizadevanj je bil učbenik Uvod v računalništvo, ki sta ga napisala Ivan Bratko in Vladislav Rajkovič.

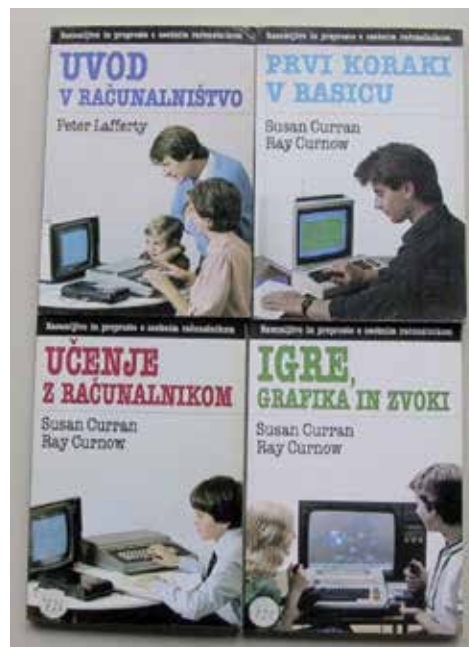
Sredi sedemdesetih let se je kot programski jezik pojavil pascal. Na Matematiki smo ga začeli uporabljati leta 1975. Zanj se je navdušil tudi Boštjan Vilfan. V tistih časih se je tudi začel (leta 1973) študij računalništva na Fakulteti za elektrotehniko [FRI]. Profesor Vilfan se je redno udeleževal Sredinega seminarja in se je odločil, da bo za potrebe študija prevedel dve knjigi/učbenika Nicholasa Wirtha - avtorja pascala. Pri tem je naletel na vrsto terminoloških problemov, o katerih smo večkrat razpravljali. Večina teh stvari je tekla v okviru Sredinega seminarja.

V osemdesetih letih so prišli mikroročunalniki. Z njimi so si lahko računalnike privoščile tudi osnovne šole. Računalništvo se je začelo širiti tudi v osnovne šole. Takrat je pri Zavodu za šolstvo kolega Tomaž Skulj vzpostavil skupino ljudi, ki je začela skrbeti za uvajanje računalnikov v šole. V okviru teh prizadevanj je potekalo več projektov. Mogoče je za osemdeseta leta najbolj znan projekt »Raček«. Ime je okrajšava za »računalniška eksplozija«. V okviru teh projektov so pripravljali razna gradiva, pri čemer je bilo spet potrebno reševati razna terminološka vprašanja. Pred

tem je bilo računalništvo usmerjeno v »ta zaresno« računalništvo na velikih računalnikih. Mikroročunalnike so v začetku gledali bolj postrani. Po kakšnih dveh letih pa smo si nekateri kolegi nabavili Spectrume in smo sprevideli, da imajo te majhne zverinice obetavno prihodnost. Tako smo začeli za šole najprej uporabljati in predavati programski jezik logo. Pri tem se je izkazalo, da ni literature za splošno populacijo.

Mikroročunalniki so računalnike iz računskih centrov razširili na širšo populacijo – postali so hišni računalniki. Imel si ga doma. Ljudje so želeli nekaj prebrati, nekaj izvedeti o teh stvareh. Pojavilo se je nekaj revij, kot so na primer »Bit«, »Moj mikro«, »Monitor« in podobne.

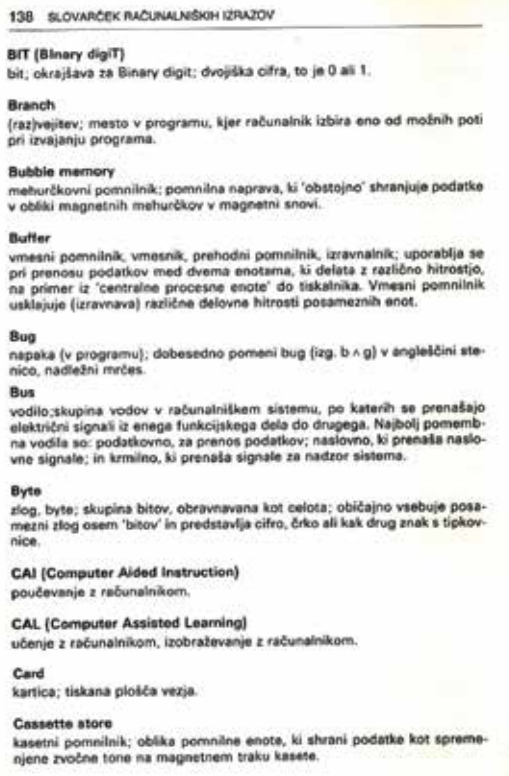
A v začetku gradiva ni bilo. Razne založbe so na hitro nekaj prevajale. ZOTKS (Zveza organizacij za tehniško kulturo Slovenije) se je odločila, da bo prevedla 4 poljudne angleške knjige o računalništvu. Izdala jih je v sodelovanju z založbo Društva matematikov, fizikov in astronomov. Založbo je takrat vodil Ciril Velkovrh, knjižničar na Matematiki. Ko so bili narejeni prevodi teh štirih knjig, je Ciril prosil Bojana Moharja, Jerneja Kozaka, Toma Pisanskega in mene, če vsak recenzira en prevod. Knjige so bile vsebinsko dobre, izkazalo pa se je, da so bili prevodi »neužitni«. To smo tudi povedali. »Za kazen« smo morali knjige sami ponovno prevesti. Takrat smo se poglobili v vprašanja terminologije, vezane na mikroročunalnike. Bolj z uporabami in ne toliko s samim »drobovjem«. Knjige so prikazane



Slika 4: Virtualni sestanek udeležencev pogovora

na sliki 4: Uvod v računalništvo, Prvi koraki v Basicu, Učenje z računalnikom, Igre, grafika in zvoki.

V eni od teh knjig je bil dodan terminološki slovarček z izrazi, ki so značilni za to novo področje računalništva. Košček iz tega slovarčka prikazuje slika 5.



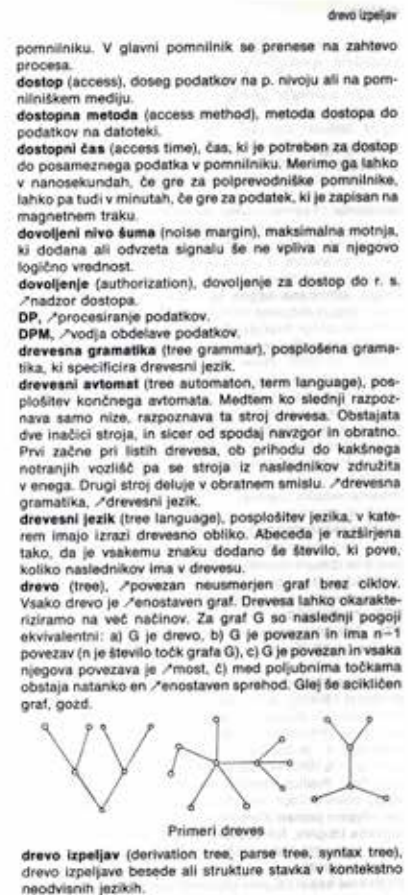
Slika 5: Košček iz slovarčka v eni od knjig ZOTKS



Slika 6: Stran iz knjige o uporabi TeXa in LateXa

Sredi osemdesetih let so izdelali laserski tiskalnik, ki je pomenil revolucijo v založništvu. Priprava knjig se je iz tiskarn selila na računalnike in vsakdo je lahko pripravil svojo knjigo. V matematiki obstaja zelo dober sistem za stavljenje oziroma oblikovanje besedil. Reče se mu TeX. Najpogosteje se uporablja njegova nadgradnja LaTeX. Uporabljati smo ga začeli že sredi osemdesetih. Z Bojanom Golijem sva napisala knjigo o uporabi TeXa in LaTeXa (izšla 1990). Večino terminologije sva povzela po izrazih, ki so jih že desetletja uporabljali stavci v tiskarnah. Dopolniti jo je bilo potrebno z novimi sestavinami, ki jih je prineslo računalniško oblikovanje besedil. Stran iz te knjige prikazuje slika 6.

V tem obdobju je kolega Matjaž Gams z IJS zbral skupino ljudi, ki je pripravljala slovar računalništva. Nekajkrat smo o vprašanih, na katere so naleteli, razpravljali tudi na Sredinem seminarju – na samem seminarju ali zvečer »Pod lipo«. »Gamsov« slovar ali pojmovnik je izšel leta 1985 [RaS]. Košček iz tega slovarja prikazuje slika 7.



Slika 7: Košček iz »Gamsovega« računalniškega slovarja.

Vzporedno so kolegi, ki se ukvarjajo s statistično obdelavo in analizo podatkov, v osemdesetih začeli s pripravo statistične terminologije, ki se delno prekriva z računalniško. Prvo različico [Sta] so izdali leta 1987 in čez čas (2001) še izpopolnjeno različico. Statistična terminologija je zgledno urejena.

V začetku devetdesetih let se je začel projekt Ro (Računalniško opismenjevanje), ki je v bistvu nadaljevanje projektov, kot so Raček in podobni. Skupino je vodil Tomaž Skulj. Imeli smo »Šolski tolar« - nekaj več se je vlagalo v izobraževanje in del tega je šel tudi za računalniško izobraževanje. Pri programu Ro smo med drugim skrbeli tudi za terminologijo. Tako me je Tomaž povezal z ljudmi, ki so se odločili, da bodo prevedli operacijski sistem OS7 za računalnike Mac. Leto in pol smo delali na prevajanju besed iz tega operacijskega sistema. To je bila spet svoja izkušnja, saj cela vrsta pojmov v tem sistemu še ni imela ustaljenih slovenskih izrazov. Je pa zelo pomembno, če jih uporablja splošna populacija, da so v domačem jeziku. To posebej velja za operacijski sistem. Ena od težav je, da obstajajo velike razlike med izrazi, ki jih uporablja recimo Microsoft, in izrazi, ki jih uporabljajo pri Applu. Istim stvarjem pravijo čisto drugače. Ni neke enotne terminologije. Za to so različni razlogi.

To je bila zanimiva izkušnja, ki pa ni najboljše izpadla, ker orodja, ki so jih takrat imeli na razpolago za prilagajanje lokalnim okoljem in prevajanje, niso na primer podpirala sklanjatev in dvojine. Potreben je bil poseben trud, da so bile stvari tako prevedene, da so bile razumljive in pravilne.

Ko je prišel čas za delo jezikovne sekcije na slovarju informatike, smo to, kar se je do takrat zbralo v našem računalniškem slovarčku, ponudili komisiji, da na tem nadaljuje. Sčasoma smo potem naš slovarček ugasnili. Stvari so prešle na novo podporo, ki sta jo ustvarila Tomaž Turk in Jurij Jaklič. To je moj prispevek k pogovoru na vprašanje, kaj se je dogajalo pred začetkom Islovarja.

Katarina:

Tomaž se bo najbrž spomnil, ko smo izbrali 300 izrazov in smo si jih nato on, Jure in jaz razdelili, cel slovarček smo morali pregledati in določiti, kaj bo šlo v naš novi slovar. To je bila koristna vsebina za slovar, ki smo jo dobili kar podarjeno. Seveda pa ni bilo razlag. Bili so le pari slovenskih in angleških izrazov.

Vlado:

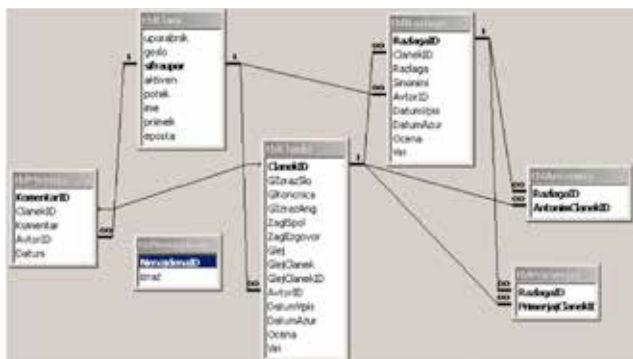
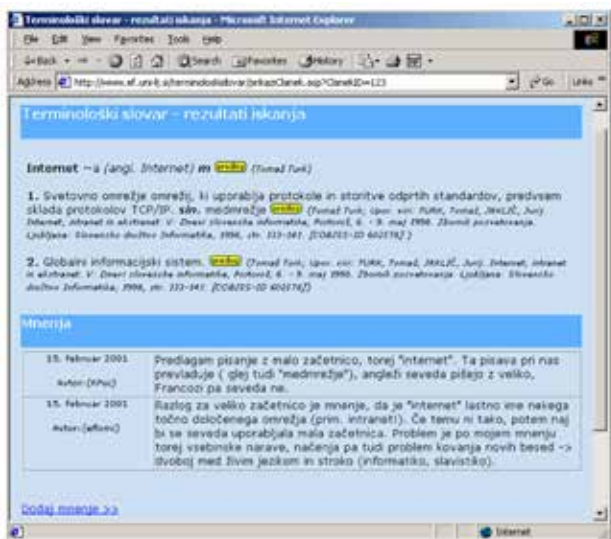
Takrat je bila stvar bolj igračkasto narejena in z Završnikom nisva razmišljala o kakšni resni zadevi. Šlo je le za prevajanje med angleščino in sloveščino (v obe smeri). O razlagah nisva razmišljala. Lahko bi jih dodala, a se tega nisva spomnila.

Tomaž:

Če se malo navežem na to: To so bili še časi, ko smo bili mladostno navdušeni nad takimi stvarmi. Moram pa reči, da so nam ta duh skrbi za slovenski jezik predali ali nekako vcepili naši predhodniki. Ko sem sedaj poslušal Vlada, sem se spomnil na naše kopanje izpred nekaj let po arhivih glede razvoja informatike na Ekonomski fakulteti UL. Že leta 1971 smo imeli predmete s področja informatike, področje sta razvijala prof. Grad in prof. Resinovič, zlasti pa je pomembno to, da že takrat računalnikov nismo razumeli kot da so malo večji kalkulatorji, ampak kot nekaj, kar je zelo pomembno za delovanje organizacij. Da bi razumeli organizacije, sta predlagala sistemski pogled na podjetja, kar so gojile ameriške šole, na primer sistemsko analizo. Pripravljala sta tudi ustrezne učbenike. Seveda tudi mi nismo mogli shajati brez fortrana. Imeli smo tudi srečo na naši fakulteti - ko je bila v 70-tih letih zgrajena nova zgradba za Bežigradom, je v njej namreč dobil svoje mesto tudi Računalniški center univerze. Bili smo dobro povezani. Ko smo se začeli pogovarjati o slovarju, je bil internet že kar na resnem pohodu. Podjetja, ki so dala nekaj nase in na svojo vidnost, so si omislila spletne strani. To so bile v osnovi bolj ali manj predstavitve. Na splošno ni bilo še kakšnih zapletenih spletnih programskih rešitev, kakršne poznamo danes. HTML sva z Jurijem bolje spoznala na enem od seminarjev pri kolegih na matematiki. S seminarja sva jadrno odšla domov na fakulteto in eksperimentirala, dobili smo prvo spletno stran fakultete, in tako dalje.

Kar se tiče rešitve za slovar nismo ravno tekmovali, obe ideji sta bili v redu, osnovna zamisel je bila ta, da tudi uredniki delajo s tem orodjem in da so sestavki takoj objavljeni, brez čakanja na izdajo knjige. Šli smo naprej. Lahko si ogledamo, kako smešno za današnje pojme je izgledal naš prvi Islovar. Za tiste čase je bil kar v redu.

Z Jurijem nama programiranje takega slovarja ni predstavljalo problemov. Podatkovna baza je bila takrat zelo preprosta v primeri z današnjo, ki je veliko bolj kompleksna.



Slika 8: Izgled prvega Islovarja in shema njegove podatkovne strukture

Kot informatika naju je z Jurijem zelo žulilo, kaj pravzaprav uporabniki potrebujejo. Lovila sva se okrog strukture članka slovarskega zapisa. Zgledovali smo se po terminološkem slovarju statistike, ki je takrat nastajal tudi na naši fakulteti. Zadeve okrog tega slovarja je vodila prof. Košmelj, sodelovali pa so tudi drugi profesorji z naše in drugih fakultet: prof. Arh, prof. Ferligoj in drugi. Njihov slovar nam je bil vzor glede strukture članka. Takrat še ne toliko glede načina delovanja urednikov. Katarina bi lahko še kaj povedala, kako smo se lovili glede dela urednikov.

Katarina:

Da, glede dela urednikov smo se tudi lovili. Začeli smo zelo navdušeno, a smo spoznali, da stvar ni tako preprosta. V začetku smo imeli majhno skupino, ki pa se je hitro večala. Ko so ljudje slišali, da ta skupina nastaja, so se pridruževali. Ko smo seštevali, koliko ljudi je sodelovalo pri Islovarju od leta 2001 do danes, je to 46 ljudi, ki so sami strokovnjaki. Nismo

vedeli, kako jih povezati. Delali smo po izkušnjah, nismo pa vedeli, kako se dela slovar. Vzorec nam je res bil Statistični slovar, druga zadeva pa je bila, kako prideš do njega. Omenila bi temeljni kamen iz leta 2003, ko se nam je pridružila Zvonka Leder Mancini, ki je takrat ravno zaključila delo pri Statističnem slovarju in nam je povedala nekaj osnovnih pravil. Potem je naredila prva navodila za delo urednikov. Z njo smo se dogovorili, da smo se sestajali enkrat tedensko. To je bila skupina, ki smo jo poimenovali Slovaropisna skupina. To je bila manjša skupina, ne vsi uredniki skupaj. Vsi, ki smo bili pripravljeni delati, smo se dobivali vsakih 14 dni v računalniški učilnici na Statističnem uradu. Bilo nas je kakšnih 15-20, tako skupino pa je nemogoče poenotiti. Nekajkrat se nam je zgodilo, da smo se razšli brez nekega rezultata, pri uredniškem delu pa je potrebno soglasje. Prej, v sami razpravi to ni nujno, na koncu pa se je treba odločiti, kako je. Uredniško delo se je s časom spreminjalo. V skupinah smo prišli do neke hierarhije. Manjše skupine smo poimenovali strokovne skupine, ena pa je bila slovaropisna skupina. To je ostalo še do danes. Strokovna skupina se dobiva po svojem urniku. Pregleda oziroma sestavi neko zbirko. Slovaropisna skupina potem to zbirko dokončno uredi. Zaradi pandemije v zadnjem času delujemo na virtualnih sestankih, s čimer prihranimo precej časa..

V Islovar je doprinesel vsakdo, kar je pač mogel. Ni nujno, da je vsak prinesel 15 novih izrazov, lahko tudi le nekaj,3 do 4. In sestankov se je udeležil, ko je pač mogel. Na tak način je stvar toliko časa delovala. Ker, če ti rečejo, da moraš poleg službe še narediti to in to, ne bi šlo. Tega ne bi bilo, ker je vse delo bilo v glavnem ljubiteljsko. Toliko o delu urednikov. Nekaj bi še rekla o Zvonki. Zvonka je bila pomembna pomoč. Ne le, da je bila pripravljena delati z nami redno vsak teden. Ona je vse, kar smo mi naredili, prej pregledala, bila je zelo kritična in nas včasih okarala, kaj smo pripravili. Odkar je ni več – žal nas je zapustila in umrla – jo zelo pogrešamo. Je pa delala vse na papirju. Ni znala delati z računalnikom. Včasih smo jo morali prepričevati, da je na papirju drugače, kot je na spletu.

Mislím, da naj kaj pove še Niko, ki je bil med ustanovitelji in nas je ves čas podpiral. Društvo nas je ves čas podpiralo. Brez društva ne bi bilo Islovarja. Tudi v okviru društva smo imeli promocijske dogodke. Prvi poskusni snopič slovarja smo predstavili na konferenci v Portorožu.

Niko:

Ker sem bil vprašan, se spodobi, da odgovorim. Moja oris ne bo tako zgodovinski in enciklopedičen kot ta prof. Batagelja, ampak bolj usmerjen na začetke Islovarja v društvu. Prvič sem bil izvoljen za predsednika društva INFORMATIKA leta 1997, občutek za jezik so mi privzgojili profesorji v srednji šoli, za kar sem jim hvaležen, zavest, da je treba na področju strokovnega jezika informatike kaj narediti, pa sem imel že precej pred nastankom Islovarja. Predsedniški mandat je omogočil, da ne ostane le pri zavedanju, temveč da bo mogoče tudi kaj narediti. Mimgrede - omenjen je bil svetovni kongres informatike leta 1971 v Ljubljani. Letos bo 50 let tega kongresa. Upam, da se bo tega spomnil še kdo razen nas v društvu. Z mojo intimno zaskrbljenostjo za strokovni jezik informatike se začne, ko smo (kot se je temu takrat reklo) združevali delo v podjetju Intertrade, ki je imelo zastopstvo za IBM. Njegovega prispevka k strokovnemu izrazju informatike v slovenščini in takratni srbohrvaščini predvsem v poslovnem delu nikakor ne bi smeli spregledati. Glede tega je imel veliko vlogo in pomen Izobraževalni center podjetja Intertrade v Radovljici, kjer so se usposabljali bodoči programerji in drugi bodoči računalniški strokovnjaki iz vse Jugoslavije.

Zgodovino matematično-znanstvenega izrazja informatike je skoraj enciklopedično orisal prof. Batagelj in je zelo dobro urejeno. Za to gre zasluga verjetno predvsem fakultetama za matematiko in elektrotehniko. Poslovno-strokovno izrazje pa je bilo razmeroma slabše obdelano. To sem opazil, ko sem pisal skripta za tedanjo Višjo upravno šolo, ki je imela v študijskem programu tudi predmet Računalništvo. Skripta so imela naslov »Uvod v programiranje«. V bistvu to niso bile osnove fortrana, kobola, PL/1 ali kakega drugega programskega jezika, saj je bilo jasno, da upravni delavci ne bodo programirali, razumeti pa so morali pomen in namen računalniške obdelave podatkov in okolja, da se bo lahko sporazumeval z razvijalci programov. Vsebina so bile torej računalniške naprave, operacijski sistemi, podatkovne strukture in komunikacije, skratka okolje, v katerem se računalniški programi razvijajo in delujejo. Mislim, da mi je takrat to razmeroma dobro uspelo s tem, da sem se dokaj namučil, ko sem iskal primerne slovenske izraze za vse te, takrat še razmeroma nove pojave.

Možnost za sistematični pristop za razvijanje stro-

kovnega jezika je v društvu obstajala. Statut društva daje možnost ustanavljati različne sekcije. Ena, ki je že delovala, je bila Sekcija za operacijske raziskave. Prišlo mi je na misel, da bi bilo dobro narediti tudi nekaj na področju strokovnega jezika, ker se takrat s tem sistematično ni še nihče ukvarjal, niti SAZU niti fakultete. Skratka, to je bila za društvo očitno niša v smislu izpolnjevanja poslanstva društva. Takrat sva se začela pogovarjati s Katarino. Ona je imela strokovno znanje, da je razumela cilj, a tudi težave in probleme. Jaz sem imel podobno razumevanje zaradi okoliščin, ki sem jih pojasnil. Takrat sva se dogovorila, da bo Slovensko društvo INFORMATIKA začelo razvijati strokovni jezik; na kakšen način, je bilo še razmeroma nejasno in odprto, vendar začetek je bil. Jaz sem celo predlagal geslo, pod katerim naj bi se to dogajalo, in mi je zelo žal, da ni objavljeno v Islovarju, ker se mi še danes zdi primerno in ilustrativno za namen, za katerega smo Islovar sploh ustvarili. Takrat sem predlagal geslo »Odličen strokovni jezik za strokovno odličnost«. Eno brez drugega ni mogoče. Mislim, da to izhodišče velja še danes.

Kakorkoli, s Katarino sva bila dovolj propulzivna, da sva animirala nekaj kolegic in kolegov. O tem je bilo povedanega že kar precej. Bi pa vendar spomnil še na nekaj stvari. V tistih prvih začetkih nismo bili prav prepričani, kaj naj bi bil rezultat našega dela. Vedeli smo, da moramo strokovni slovenski jezik informatike razvijati in čistiti, ampak v kakšni obliki naj bi se to zgodilo, na začetku ni bilo še jasno. Delo sekcije za jezik se je začelo tako, da smo se sestajali in obravnavali izbrane izraze in se s konsenzom odločili za pomen in definicijo vsakega izraza posebej. Verjetno smo se takrat odločili tudi za osnovno zgradbo slovarja: za vsako iztočnico mora biti navedena izgovorjava, angleška ustreznica in definicija. Ta pristop se je izkazal kot dober, delo urednikov poteka v bistvu na ta način še danes in tudi zgradba Islovarja je še vedno taka. Pri našem delu nam je bil Statistični urad Republike Slovenije zelo naklonjen in nam je omogočil uporabo računalniške učilnice za delo in sestanke. Seveda pa zgolj obravnavanje izrazov ni mogel biti rezultat našega dela in vprašanje je bilo, v kakšni obliki naj bi delo predstavili. Strinjali smo se, da naj nastane terminološki slovar informatike, ki bi izšel v predvidoma v knjižni obliki in našli smo tudi podjetje, ki je izkazalo interes za sodelovanje. Izkazalo pa se je, da je imelo drugačne ambicije in odločili smo se, da gremo sporazumno narazen.

Odločitve nismo obžalovali, saj je bila edina realna alternativa internetni slovar. Verjetno smo se takrat tudi dogovorili za njegovo ime – I(za internetni)slovar, ki ga nosi še danes. Če bi bila knjiga takrat izšla, bi pri tem ostalo, kot se s knjigami dogaja - za razliko od spletnega slovarja, ki je lahko tekoče posodabljan. Seveda pa je to pomenilo, da je moralo društvo najti finančna sredstva za razvoj programja za internetni slovar. No, ta denar smo našli in ne le enkrat, danes je razpoložljiva že tretja verzija Islovarja.

Ko se spominjam preteklih dogodkov in prispevkov, ki so jih dali moji prijatelji in kolegi, ne bi smeli spregledati nekdanjih in sedanjih predsednikov sekcije. Katarina Puc je bila prva predsednica, dr. Batagelj za njo, za njim pa dr. Turk. Vsem trem gre vsa pohvala in zahvala, da so prispevali dobršen del svojega časa in nemalo tudi svojega življenja za razvoj slovarja. Prof. Turk je že povedal, kako se je ideja programske podpore Islovarja razvijala in kako je napredoval od prve do sedanje tretje verzije. Nisem edini, ki razmišlja tudi o četrti verziji, ki jo bo treba prej ali slej spraviti v življenje. Upam, da bomo pri tem deležni pomoči, nasvetov in sodelovanja tudi v prihodnje.

Ko že omenjam zgodovino, naj se rahlo dotaknem tudi prihodnosti. Na zadnjem sestanku sekcije je dr. Turk povedal, da se z mesta predsednika sekcije umika in je predlagal, da ga na tem mestu zamenja mag. Jože Kranjc, ki je na to pristal, za kar se mu prijazno zahvaljujem.

Katarina:

Kako sem jaz ugotovila potrebo po slovenskih izrazih? Bila sem tehnična urednica in lektorica revije *Uporabna informatika* od ustanovitve revije leta 1993. To je trajalo 10 let. Takrat sem dognala, da avtorji priznavajo, da bi veliko lažje pisali v angleščini kot v slovenščini. Velikokrat so bili v besedilu kar angleški izrazi. Tako sem iz prakse dognala, da kljub temu, kar je bilo do takrat narejeno, mnogih pravih slovenskih izrazov še ni, ker se tehnologija tako hitro spreminja.

Ivan:

Multidisciplinarnost je že bila omenjena. Že pogled v Islovar pokaže jasno sliko povezave na različna področja. V začetku je že bila tendenca oziroma naša želja, da bi ob posameznih terminih imeli oznako, s katerega strokovnega področja izraz prihaja. Ugotovili smo, da je to po eni strani težko, ker je lahko neka zadeva zanimiva na dveh ali treh področjih. Še

bolj pa zato, ker nikakor nismo mogli priti do neke sheme nekakšnega mini tezaura ali kakorkoli bi že rekli tej shemi. Drugi dokaz te multidisciplinarnosti je, kar ste že omenjali, prihod kolegice Zvonke. Takrat se je k izredno bogatemu znanju strokovnjakov s področja informatike dodalo še znanje terminologa in slovaropisca. Na ta način se je bistveno povečala kakovost oblikovanja izrazov tam, kjer še niso bili popolnoma dodelani, kjer smo jih nekoliko pilili in včasih prilagajali ali si jih na novo »izmislili«. Bistveno se je izboljšala kakovost strukture slovarskega gesla in slovarja v celoti. Pri tem ne smemo pozabiti, da je slovar postal zelo pogosto referenca pri strokovnem delu. Zelo pogosto srečamo Islovar kot vir informacije ali kot priporočilo kot »pa poglej v Islovar«. Do tega, da tudi v družbenih omrežjih prevajalcev pogosto rečejo: »no, saj to je pa tako ali tako lepo razloženo v Islovarju«. Ta vrzel, ki je nastala z Zvonkino smrtjo, je nenadomestljiva. Nekoliko pomoči smo vendarle dobili. Obrnili smo se na terminološko sekcijo pri ZRC SAZU, kjer so bili prijazno pripravljene pomagati na več načinov. Eden je bil ta, da so nam pripravili seminarček o tem, kako naj se oblikuje terminološki slovar, kako izgleda terminologija in na katere pasti je treba paziti oziroma, kam se je treba bolj zazreti. Do tega, da lahko tudi pri njih pričakujemo pomoč v smislu dodatne redakcije gradiva, ki ga mi pripravimo, oni pa ga pregledajo in odpravijo najhujše napake. Pri ostalem pa svetujejo. Se mi zdi, da je tudi to pot v prihodnost, da bomo imeli tudi s terminološkega vidika vsaj manjšo pomoč.

Niko:

Dotaknil bi se še dveh stvari: prispevek dr. Turka je nesmrten v tem, da je izdelal metodo, kako priti do novega izraza. To je obsežen postopkovni blok diagram. Metoda je izvirna in je nisem zasledil nikjer drugje. Druga stvar je to, kar je sedaj omenil Ivan. V več razpisih državnih organov s področja informatike in informacijske tehnologije je bilo navedeno, da je treba strokovne izraze razumeti tako, kakor so definirani v Islovarju. To je veliko priznanje, vendar pa, ko to vemo, ni mogoče razumeti, zakaj država ne najde denarja, da bi razvoj Islovarja podprla tudi finančno. Večkrat smo se namreč prijavi na javne razpise ministrstva za kulturo za razvoj Islovarja s skromnimi zahtevki reda nekje okoli dva do tri tisoč evrov, kar je daleč od tistih milijonov, ki jih sedaj država namenja za jezikovne tehnologije. Bili smo uslišani

mogoče dvakrat. Enkrat smo celo prejeli odgovor, da je naš zahtevek za financiranje presešel cenzus. To se pravi: dosegli smo dovolj točk za financiranje, v nadaljevanju sklepa pa je bilo povedano, da ministrstvo kljub temu na razpis prijavljenega projekta razvoja Islovarja ne bo financiralo. Ne razumem, od kod jim ta »pogum« in razlogi za takšno odločitev. Mislim, da ima tu država do nas kar velik dolg.

Tomaž:

No, glede metode, kako priti do slovarskega sestavka, sem bolj ali manj le narisal tisto, kar smo se skupaj dogovorili in naučili, posebej smo zapisali navodila urednikom. Nekako vzporedno s tem smo razvili novo programsko rešitev, to je drugo verzijo Islovarja, ki je še danes živ na naslovu <http://stari-islovar.org>, zdaj pa imamo Islovar 3.0, ki smo ga razvili v sodelovanju s Filozofsko fakulteto UL, bolj konkretno s prof. Špelo Vintar. Na področju slovaristike se je namreč v zadnjih desetletjih veliko spremenilo in

to je bilo treba upoštevati. Treba je tudi povedati, da Islovar ne nastaja v enem kabinetu, na eni katedri, na eni fakulteti, v enem podjetju, ampak nastaja s sodelovanjem ljudi z različnih institucij v našem prostoru.

VIRI:

- [Riz] Vladimir Batagelj: Razvoj slovenskega računalniškega izražaja - osebni pogled. *Uporabna informatika* 9(2001/2)2, 95-99.
- [Sre] Vladimir Batagelj: Sredin seminar: Seznam predavanj 1971-2021. April 2021. Seminarji: 375, 376, 406, 568, 575, 582, 603, 610, 620, 623, 624, 672, 694, 790, 820, 824, 828, 891, 918, 922, 1021, 1025, 1069, 1115, 1117, 1126. <http://vlado.fmf.uni-lj.si/sreda/sreda1300.pdf>
- [RaS] Matjaž Gams: Računalniški slovarček. CZ, Ljubljana 1985.
- [Sta] Blaženka Košmelj s sodelavci: Statistična terminologija. RSS, Ljubljana 1987.
- [EAR] France Križanič: Elektronski aritmetični računalniki. Knjižnica Sigma, MK, Ljubljana 1960.
- [MT] Alojzij Vadnal: Matematična terminologija. DZS, Ljubljana 1953/1974.
- [EIR] Elektronski računalniki. Uredil Spiller-Muys F. Elektrotehniška zveza Slovenije, Ljubljana 1971.
- [FRI] FRI 20 - 20 let Fakultete za računalništvo in informatiko Univerze v Ljubljani. 29. junij 2016 http://eprints.fri.uni-lj.si/3655/1/Zbornik_FRI20_web_100.pdf

■

Katarina Puc diploma na Filozofski fakulteti s predmetov a)francoski jezik s književnostjo, b)angleški jezik s književnostjo. Magisterij poslovne politike in organizacije na Ekonomski fakulteti Univerze v Ljubljani, z delom: Ekonomski in organizacijski kriteriji za odločanje o uporabi tehnologije v pisarniških sistemih. Pomembnejše delovne izkušnje: izobraževanje, uredništvo in tehnično uredništvo knjig, revij, zbornikov, prevajanje književnih del in strokovnih besedil iz angleščine, francoščine in nemščine v slovenščino, lektoriranje strokovnih besedil. Pobudnica ustanovitve jezikovne sekcije pri Slovenskem društvu Informatika. Urednica spletnega slovarja informatike Islovar.

■

Niko Schlamberger je diplomiral na Fakulteti za strojništvo v Ljubljani. Večino poklicne poti deluje na področju informatike in računalništva, delovne izkušnje obsegajo programiranje, sistemsko analizo in razvijanje računalniških rešitev, predavanja (strokovna in na višji šoli), izvajanje usposabljanja na področju informatike, svetovanje, vodenje projektov, organiziranje in izvedbo nacionalnih in mednarodnih konferenc, vodenje in upravljanje ter mednarodno sodelovanje. Objavil je preko 60 strokovnih člankov s področja informatike, organizacije in upravljanja v državi in na mednarodnih konferencah v tujini, uredil in souredil je več strokovnih publikacij s področja registrov in klasifikacij. Bil je podpredsednik svetovnega združenja informatikov International Federation for Information Processing (IFIP), sekretar in predsednik evropskega združenja informatikov Council of European Professional Informatics Societies (CEPIS), koordinator regionalnega združenja IT STAR. Je aktualni predsednik Slovenskega društva INFORMATIKA in eden od urednikov Islovarja.

■

Vladimir Batagelj je zaslužni profesor Univerze v Ljubljani. Bil je profesor za diskretno in računalniško matematiko na FMF. Raziskovalno se ukvarja predvsem s teorijo grafov, algoritmi na grafih in omrežjih, kombinatorično optimizacijo, analizo podatkov (razvrščanje v skupine, analiza in prikaz velikih omrežij) in uporabo IT v izobraževanju. Z Andrejem Mrvarjem je soavtor svetovno uveljavljenega programa Pajek za analizo velikih omrežij. V mednarodnih revijah je objavil večje število člankov in je soavtor nekaj knjig: *Generalized blockmodeling* (CUP 2005), *Exploratory Social Network Analysis with Pajek* (CUP 2005, 2011, 2018), *Understanding Large Temporal Networks and Spatial Networks* (Wiley 2014), *Advances in Network Clustering and Blockmodeling* (Wiley 2020). Za svoje dosežke je prejel več priznanj. Je častni član DMFA RS in izvoljeni član ISI (International Statistical Institute).

■

Tomaž Turk je redni profesor na Ekonomski fakulteti Univerze v Ljubljani in predstojnik Katedre za poslovno informatiko in logistiko. Je nosilec predmetov na področju razvoja informacijskih sistemov in programskih rešitev, komunikacijskih omrežij, managementa informatike ter poslovnih simulacij. Raziskovalno se ukvarja s privzemanjem informacijske tehnologije, z ekonomiko informatike, managementom telekomunikacijskih

omrežij, informacijsko družbo ter optimizacijskimi pristopi. Sodeloval je na več nacionalnih in mednarodnih projektih. Objavlja predvsem v revijah *Technological Forecasting & Social Change*, *Telecommunications Policy*, *Computer Standards and Interfaces*, *Mathematics and Computers in Simulation*, *International Journal of Industrial Ergonomics* ter *Computer Communications*. Bil je podpredsednik raziskovalnega projekta COST Action 298 'Participation in the Broadband Society', ki ga je financirala Evropska znanstvena fundacija. V letih 2013-2017 je bil na Ekonomski fakulteti prodekan za študijske zadeve.

■

Ivan Kanič je po osnovni izobrazbi jezikoslovec, vendar je celotno poklicno pot posvetil bibliotekarstvu in terminologiji, zaposlen je bil v Centralni ekonomski knjižnici EF UL in Narodni in univerzitetni knjižnici. Na Oddelku za bibliotekarstvo, informacijsko znanost in knjigarstvo FF UL je deloval kot zunanji predavatelj za področje avtomatizacije knjižnic ter izgradnje in uporabe elektronskih virov. V različnih projektih je sodeloval pri snovanju, razvoju in vrednotenju informacijskih sistemov in uporabniških vmesnikov za bibliografske, kataložne in terminološke podatkovne zbirke ter izobraževanju njihovih uporabnikov. Je urednik in soavtor bibliotekarskih terminoloških slovarjev ter avtor in skrbnik terminološke podatkovne zbirke za področje bibliotekarstva. v Ljubljani.

Iz Islovarja

Islovar je spletni terminološki slovar informatike, ki ga objavlja jezikovna sekcija Slovenskega društva INFORMATIKA na naslovu <http://www.islovar.org>. Slovar je javno dostopen za vpoglede in vnašanje novih izrazov. Tokrat objavljamo izbor novejših izrazov, ki smo jih uredili v tem letu.

avtonómni robót m (*angl. autonomous robot*) robot, ki opravlja vedenje ali naloge z visoko stopnjo samostojnosti; prim. avtomatizacija vožnje, avtonomno vozilo

digitálni dvójnik -ega -a m (*angl. digital twin*) navidezna klonirana predstavitev fizičnega predmeta ali procesa v realnem času

elektrónski račún -ega -a m (*angl. electronic invoice*) račun, ki je bil izdan, poslan in prejet v elektronski obliki; prim. E-račun

É-račún -a m (*angl. eInvoice*) elektronski račun v Sloveniji, ki je v standardizirani obliki e-Slog; prim. elektronski račun

e-Slóg -a m (*angl. e-Slog*) standard, razvit v okviru Gospodarske zbornice Slovenije in usklajen z evropskim standardom za izdajanje elektronskih naročilnic, potrditev naročil, izdajanje dobavnic in računov

goljufíva téhnična podpóra -e -e -e ž (*angl. technical support fraud*) izvajanje lažne varnostne ali tehnične podpore z namenom pridobitve podatkov od žrtve

hektivízem -zma m (*angl. hacktivism*) hekersko gibanje, ki za promocijo svojih političnih stališč uporablja informacijsko-komunikacijsko tehnologijo hekerska dejanja za promocijo stališč ali doseganje političnih ciljev

informácijska várnost -e -i ž (*angl. information security*) varnost podatkov in informacijskih sistemov pred zlorabo, uničenjem; sin. kibernetška varnost;

prim. varnost informacijskega sistema, INFOSEC skupno delovanje ljudi, ukrepov, procesov in tehnologij za zaščito podatkov in informacijske komunikacijske opreme; sin. kibernetška varnost; prim. varnost informacijskega sistema, INFOSEC

kibernétska hígiéna -e -e ž (*angl. cyber hygiene*) postopki, ki bi jih morali redno izvajati za zaščito računalniških sistemov in podatkov, npr. arhiviranje podatkov, preprečevanje nepooblaščenih dostopov, osveščanje uporabnikov

nabírálni robót -ega -a (*angl. harvesting bot, harvesting robot*) računalniški program za nabiranje naslovov; sin. nabiralni program

ôzko grlo -ega -a s (*angl. bottleneck*) komponenta z najmanjšo zmogljivostjo v celotnem sistemu

prestréznik gésel -a -- m (*angl. password dumper*) škodljiva programska koda, ki prestreza avtentikacijske podatke, npr. gesla; sin. kopirnik gesel, snemalnik gesel

protokól oddáljenega namízja -a -- -- m (*angl. remote desktop protocol, RDP*) protokol, ki uporabniku omogoča povezavo z drugim računalnikom prek omrežne povezave; sin. RDP; prim. navidezno zasebno omrežje

upárjenje -a s (*angl. pairing*) vzpostavljanje brezžične povezave med dvema združljivima enotama, npr. povezava Bluetooth, infrardeča povezava postopek, ki omogoča, da med seboj komunicirata dve napravi ali več naprav zaradi neposredne izmenjave podatkov, npr. pametna ura in telefon

Izpitni centri ECDL

ECDL (European Computer Driving License), ki ga v Sloveniji imenujemo evropsko računalniško spričevalo, je standardni program usposabljanja uporabnikov, ki da zaposlenim potrebno znanje za delo s standardnimi računalniškimi programi na informatiziranem delovnem mestu, delodajalcem pa pomeni dokazilo o usposobljenosti. V Evropi je za uvajanje, usposabljanje in nadzor izvajanja ECDL pooblaščen ustanova ECDL Fundation, v Sloveniji pa je kot član CEPIS (Council of European Professional Informatics) to pravico pridobilo Slovensko društvo INFORMATIKA. V državah Evropske unije so pri uvajanju ECDL močno angažirane srednje in visoke šole, aktivni pa so tudi različni vladni resorji. Posebno pomembno je, da velja spričevalo v 148 državah, ki so vključene v program ECDL. Doslej je bilo v svetu izdanih že več kot 11,6 milijona indeksov, v Sloveniji več kot 17.000, in podeljenih več kot 11.000 spričeval. Za izpitne centre v Sloveniji je usposobljenih osem organizacij, katerih logotipe objavljamo.



Strokovni prispevki

Tomaž Dular, Julija Lapuh Bele, Rok Pirnat
ODKRIVANJE PREVAR Z ORODJI POSLOVNE INTELIGENCE

Simona Samida Cerk
SODOBNI IZZIVI IZOBRAŽEVANJA NA DALJAVO V OKVIRU
MEDPREDMETNEGA POVEZOVANJA IN INOVATIVNIH UČNIH OKOLIJ

Znanstveni prispevki

Tim Poštuvan, Semir Salkič, Lovro Šubelj
LEARNING-BASED LINK PREDICTION ANALYSIS FOR FACEBOOK100 NETWORK

Elena Osrajnik, Sašo Karakatič
TEMNI VZORCI NA SLOVENSKIH SPLETNIH STRANEH

Milena Košak Babuder, Blažka Korun, Ema Štarkl, Gaja Nenadović, Dušica Boben,
Erika Stanković, Sara Jakop, Luka Vranješ, Karmen Javornik, Jure Žabkar
RAČUNALNIŠKO PODPRTO PREPOZNAVANJE ZGODNIH ZNAKOV DISLEKSIIJE

Razprave

Katarina Puc, Niko Schlamberger, Vladimir Batagelj, Tomaž Turk, Ivan Kanič
POGOVOR SODELAVCEV OB 20-LETNICI ISLOVARJA

Informacije

IZ ISLOVARJA

ISSN 1318-1882



9 771318 188001