

KREIRANJE EFEKTIVNE KOMUNIKACIJE SA KORISNICIMA KROZ TEHNIKE VEB RUDARENJA I NJEGOVA PRAVNA REGULATIVA

Mirjana Tankosić⁵³, Ana Grbić⁵⁴

Originalni naučni rad

UDK: 004.738.5:339.138
004.6:658.89

Apstrakt

Savremene kompanije svoje poslovne aktivnosti započinju otkrivanjem, analiziranjem i razumevanjem potreba i želja potrošača, a završavaju ih isporučivanjem očekivane vrednosti iz koje proističe satisfakcija potrošača.

Cilj istraživanja ovog rada je analiza korišćenja tehnika prikupljanja podataka o potrošačima putem veba i izdvajanja potrebnih obrazaca informacija u veliku bazu podataka. Obzirom da se ovim tehnikama može doći i do ličnih podataka potrošača analiziraće se da li je to u skladu sa pravnim i etičkim normama.

Praktični pristup rada se sastoji od izdvajanja informacija o ponašanju korisnika u poseti veb lokaciji i obrade podataka pomocu inteligencije rojeva, kako bi se pronašlo optimalno rešenje i rezultati za efektivnu komunikaciju sa korisnicima. Ceo ovaj proces se koristi za poboljšanje stranica i prilagođavanje novim korisnicima i njihovim preferencijama.

U doba razvijenog konkurenetskog okruženja, informacije o potrošačima su ključna karika uspešnog marketinškog poslovanja jedne kompanije. Jedan od izvora koji može pružiti dodatne informacije o segmentaciji potrošača je veb rudarenje.

Ključne reči: komunikacija, digitalni marketing, pravna regulativa, veb rudarenje

JEL: M29, M30, K38, M24

Uvod

Savremenim kompanijama postalo je veoma važno da upoznaju potrebe i želje potrošača. Na ovaj način dolazi se do zaključka da su za savremena kompanije najvažniji faktor potrošači. U početku su kompanije lansirale proizvode sa najboljim performansama kako bi zauzele što veci deo tržišta. Kada su dostigli maksimalne performanse bilo je potrebno privući kupce na nov način. Komunikacija kompanija

⁵³ Docent, dr Mirjana Tankosić, Higher Colleges of Technology, Faculty of Applied Media, Fujairah Women's Campus, Al Qasar/Kuwait Road, Al Faseel, Fujairah, UAE, Telefon: +9712 206 9343, E-mail: mtankosic@hct.ac.ae

⁵⁴ Docent, dr Ana Grbić, Visoka škola socijalnog rada, Terazije 34, Beograd, Telefon: +381 11 404 08 39, E-mail: ana.grbic@asp.edu.rs

sa potrošačima je prestala da bude ograničena bez obzira na udaljenost i nacionalne granice (Tankosić, 2013). Sa tradicionalnim načinom fokusiranja na proizvod kompanije, počeli su da pružaju potrošačima pažnju i osluškuju potrebe tržišta. Na početku „osluškivanja tržišta“ kompanije su koristile zastarele metode ispitivanja uzorka. Tokom godina postalo je jasno da ispitanci prilikom popunjavanja upitnika nisu davali najiskrenije odgovore. Informaciona tehnologija je izazvala duboke promene u ekonomiji, kulturi i politici (Tankosić, Grbić, Krivokapić, 2020).

Inovacije u bilo kom sistemu su organizovane i svrshodne aktivnosti usmerene na stvaranje promena (Tankosić, Trifunović, 2013). Sa razvojem tehnologije, marketinški stručnjaci su prepoznali mogućnost primene nauke u svojim istraživanjima. Kako bi saznali što više informacija o svojim potrošačima, potrebno je samo da imaju pristup vebu jer se tu već dobровoljno nalaze sve informacije o potrošačima. Nove tehnologije i digitalizacija pomažu razvoju veb rudarenja. Prikupljanjem informacija o potrošačima putem interneta može se doprineti razvoju tradicionalnog i digitalnog marketinga bilo koje kompanije.

Internet tehnologija se brzo razvila u protekloj deceniji, evoluirajući u komunikaciju, umrežavanju i informacionim sadržajima. Jedan od glavnih aspekata Internet tehnologije je World Wide Web (WWW) koji pruža značajnu količinu informacija dostupnih svima. Dinamički deo WWW-a je interakcija korisnika sa veb stranicama, prikupljanje informacija i sadržaja. Postoji mnogo veb stranica koje su popularne i relevantnije od drugih, ali, na osnovu pretraživanja korisnika, važnost informacija je u potpunosti zasnovana na preferencijama korisnika. Ključne reči pokreću upit u određenom pretraživaču, i kao rezultat, obezbediće listu najbolje ocenjenih veb stranica prema kriterijumima (Shang, Li, 2013). Otkrivanje znanja sadržanog u podacima (Knowledge Discovery in Data), je oblast koja se oslanja na dospinu mašinskog učenja kao dela veštačke inteligencije, sa ciljem ekstrakcije potencijalno korisnih i interesantnih saznanja iz podataka velikog obima. Tradicionalno, tehnike rudarenja podataka (Data Mining), bile su primenjene na podatke koji su se kroz vreme prikupili u relacionim ili transakcionim bazama ili skladištima podataka preduzeća, ali se one, u principu, mogu primeniti na bilo koju vrstu repozitorijuma podataka, pa i na veb podatke. (Bošnjak, Grlević, Bošnjak, 2019).

Metode

U cilju razumevanja ponašanja korisnika/potrošača, potrebno je analizirati modele zasnovane na znanju koji prikazuju sve tražene informacije i podatke. Kako bi se postigla velika količina sadržaja u brzom odgovoru na određeni upit, tehnologija Veb Mining-a se pojavljuje u tehničkoj pozadini i društvenom okruženju (Morris, Teevan, Bush, 2008). Glavna svrha ove tehnike je da predstavi generisane informacije o veb lokaciji kroz strukturu linkova i korisničkih preferencija.

Tehnike veb rudarenja mogu značajno poboljšati onlajn marketinške aktivnosti. Veb rudarenje se sastoji od rudarenja veb sadržaja i korišćenja veba (Dai, Wang, Shu, 2017). Kopanje veb sadržaja je povezano sa sadržajem veb stranice, bazom podataka i

pozadinskim procesom. Veb Usage Mining je aktivnost koja otkriva nove informacije i znanje analizom veb strukture, sadržaja i evidencije korišćenja. To je deo metode Data Mining koja koristi podatke, ali u ovom slučaju, Veb Usage Mining koristi podatke iz veb resursa, uzima sva zanimljiva ponašanja, obrasce i informacije. Pošto je znanje glavni deo ove tehnike, ono se može koristiti kao ključna veza za optimizaciju i poboljšanje brzine i preciznosti odgovora veb sadržaja korisnicima.

U ovom radu data je analiza i prezentacija teoretskih osnova Swarm Intelligence (SI) kao optimizacije Web Usage Mining-a (WUM) za potrebe onlajn marketinških aktivnosti, implementacija svitac (firefly) algoritma i rezultati SI pruža efikasan način za pronalaženje optimalnog rešenja. Budući da se bavi prirodnim i veštačkim sistemima, da bi se definisao model koji može da optimizuje i kreira rešenje, potrebno je izvršiti poređenje WUM i SI kako bi se prikupile sve informacije koje su neophodne za eksperimentalni deo poboljšanja. Analitičkom metodom se obuhvatila analiza sadržaja relevantne naučne i stručne literature u ovoj oblasti (Grbić, Tankosić, 2022).

Pretraživači pomažu korisnicima da traže informacije na vebu, ali su efikasni samo kada korisnici imaju jasno znanje o tome šta traže. Za preporuku veb stranica koristi se pristup zasnovan na grafikonu koji identificuje one korisnike koji otkrivaju zanimljive stranice, na osnovu njihovog znanja i proračuna da su stranice adekvatne. Analizom aktivnosti pregledanja ranijih korisnika mogu se identifikovati nove zanimljive stranice i zatim preporučiti novim korisnicima koji pokazuju isto interesovanje.

Za rangiranje važnosti određenih stranica na osnovu pretrage, ključne su veze i analiza sadržaja. Ovaj deo je neophodan za optimizaciju i može se koristiti. Brojčana težina se dodeljuje na osnovu različitih parametara. Cilj ovog rada je da predstavi važne aspekte pretrage veb sadržaja, analizira model rudarenja upotrebe veba i predstavi optimizaciju koja se može koristiti kao analitika rezultata za buduće veb stranice kako bi imali bolji uvid u preferencije korisnika i obezbedili kvalitet, ali takođe obezbedili bolje rezultate pretraživanja veb korisnicima i omogućila efikasnija komunikacija sa korisnicima sadržaja veba (potrošačima).

Definisanje koncepta rudarenja upotrebe veba

Kopanje upotrebe veba (Web Usage Minning) je deo tehnike rudarenja podataka koja otkriva obrasce ponašanja na osnovu podataka sa veba. Proces rudarenja upotrebe veba je podeljen na (Jafari, Sabzchi, 2018): 1. priprema podataka; 2. otkrivanje obrazaca; 3. analiza šablona.

Priprema podataka kreira datoteku sesije servera u kojoj je svaka sesija definisana kao niz različitih tipova zahteva jednog korisnika koji poseti stranicu (Ishikawa, 2018). Otkrivanje šablona prati obrasce, klasteri korišćenja i klasifikaciju korisnika za svaku sesiju koju su korisnici omogućili (Nguyen, Phung, Adams, 2018). Svi podaci i informacije izvučeni iz procesa Veb Usage Mininga mogu se koristiti za personalizaciju veba, modifikacije sajtova i otkrivanje poslovne inteligencije (Shoval,

Maidel, Shapira, Taieb-Maimon, 2018). Takođe može biti korisno za keširanje veb servera jer cie smanjiti broj zahteva i uravnotežiti opterećenje (Bonchi, 2018).

Prethodna obrada je suštinski deo istraživanja korišćenja veba jer svi podaci koji su dostupni na vebu nisu strukturirani i treba ih transformisati tako da analitički deo može da proizvede rezultate korisničkih obrazaca. Neobrađeni podaci se prikupljaju i pretvaraju u skup korisničkih profila kako bi se izvršila faza otkrivanja šablonu (Dong, 2013). U prethodnoj obradi podataka, podaci se čiste, korisnički dnevnik se identificuje u svakoj sesiji i vrši se prepoznavanje putanje.

Otkrivanje šablonu izdvaja ponašanje korisnika na osnovu podataka koji su formatirani u fazi pre obrade. Zbog ovog važnog koraka svi podaci moraju biti prethodno pripremljeni i konvertovani kako bi se mogli koristiti kao ulaz. Znanje dobijeno ovim procesom se koristi u statističke svrhe sajtova i komunikaciju s korisnicima. Dve važne tehnike u ovoj fazi koje se razmatraju u eksperimentalnom delu rada su: statistička analiza i grupisanje.

Statistička analiza dobija fajl sesije iz veb log-a i svi podaci se koriste za kreiranje izveštaja o stranicama korisnika, prosečnom vremenu posete stranice, uobičajenim unosima. Ovo znanje može biti korisno za praciene dana sa velikim prometom na sajtu, za promenu funkcije sajta za korisnike ili za poboljšanje bezbednosnih aspekata sajta. Druga tehnika, grupisanje, koristi se za grupisanje korisnika ili podataka koji su u određenoj meri slični (Jafari, Sabzchi, 2018). Pošto je klasifikacija zasnovana na aspektu sličnosti (korisnik/podaci), ovo se može koristiti za formiranje aktivnosti za nove korisnike koji imaju identičnu percepciju.

Analiza šablonu izvlači sva pravila, obrasci i statistiku iz rezultata otkrivanja šablonu, filtrira nerelativne podatke i kreira statističke rezultate definisane sesije. Ovo je kombinacija mehanizma upita za znanje kao što je SQL, grafički obrasci i ukupni rezultati podataka.

Evaluacija veb stranice inteligencije rojeva

Inteligencija roja (Swarm Intelligence – SI) je termin povezan sa sistemom koji prikuplja ponašanje decentralizovanih veštačkih sistema po analogiji sa stvornim svetom gde kolektivno ponašanje roja može dovesti do intelligentnog ponašanja pojedinca prikazanog od strane grupe (Dzitac, 2017). Ova kolektivna inteligencija predstavlja sposobnost grupe da reši više problema nego njeni pojedinačni članovi (Zhu, 2010). Ovi sistemi su jednostavnii agenti koji komuniciraju jedni sa drugima, u isto vreme u interakciji sa okruženjem. Oni prate pravila bez centralizovane kontrole.

Algoritmi zasnovani na metaheuristici simuliraju prirodni fenomen i mogu se optimizovati i implementirati u različitim oblastima kako bi se kreirala bolja rešenja i rezultati. Procesi zasnovani na prirodi se oponašaju za rešavanje problema optimizacije. Inteligencija roja zasnovana je na pojedincima koji prikazuju kolektivnu inteligenciju. Ovo ponašanje se može videti u kolonijama mrava (Jovanović, Tuba, 2010) kolonijama medonosnih pčela (Tuba, Bačanin, 2014) itd.

Svitac (firefly) algoritam (FA) predstavlja jednu od najnovijih metaheuristika inteligencije rojeva (Bačanin, Tuba, 2014). Prvobitno ga je predložio Jang 2008. godine (Jang, 2008) i primenjen na probleme optimizacije. Originalni svitac algoritam je kreiran da vodi ponašanje svitaca, ali i da oponaša ponašanje bljeskalice. Svaki svitac se krecie ka svetlosti. Ovo je glavni deo algoritma koji je definisan cilnjim funkcijama. Ovo treperenje se koristi za privlačenje svitaca kao signalizirani sistem za parenje i za plen. Pošto svici nemaju globalni pogled na okolinu, njihova veza se zasniva na intenzitetu svetlosti, što je slično korisniku veba koji se krecie po mreži bez informacija o ruti koju prate drugi korisnici sa istim preferencijama (Dzitac, 2017).

Da bi se primenio i implementirao svitac algoritam, treba uzeti u obzir tri idealizovana pravila (Bačanin, Tuba, 2014):

1. Svi svici su bespolni; privlačnost nije zasnovana na polu;
2. Privlačnost je definisana i direktno proporcionalna njihovom sjaju i manje sjajniji svitac cie se kretati ka svetlijem. Osvetljenost se povećava kako se razdaljina između svitaca smanjuje;
3. Sjaj svitaca je određen vrednošću funkcije cilja.

Kada je u pitanju funkcija cilja $f(x)$, ona se koristi za definisanje i kodiranje osvetljenosti svitaca. Napravljen je da prikaže intenzitet svetlosti na lokaciji x . Na taj način, $I(x) = f(x)$. Važan deo je pokret. Kretanje se zasniva na privlačnosti (Bačanin, Tuba, 2014), a kada je svitac j privlačniji od svica I, svitac I se krecie prema j (Jang, 2017):

$$x_i(t) = x_i(t) + \beta_0 r^{-\gamma r_{i,j}^2} (x_j - x_i) + \alpha(\text{rand} - 0.5) \quad (1)$$

Gde β_0 predstavlja privlačnost $r=0$, α je parametar randomizacije, rand je slučajni broj ravnomerno raspoređen između 0 i 1; a rastojanje između svitaca i i j je $r_{i,j}$ (Bačanin, Tuba, 2014).

Pojednostavljenje algoritma pseudo koda svitac je predstavljeno u nastavku (Bačanin, Tuba, 2014):

```

Generate initial population of fireflies xi,
(i = 1, 2, 3, ..., FN)
Light intensity Ii at point xi is defined by f(x)
Define light absorption coefficient γ
Define number of iterations IN
while (t < IN) do
for (i = 1 to FN) do
for (j = 1 to i) do

```

```

if( $I_j < I_i$ ) then
    Move firefly  $j$  towards firefly  $i$  in  $d$  dimension
    Evaluate new solution, replace worst with better
    solution, and update light intensity
end if
end for
end for
Rank all fireflies, find the current best, and move them
randomly
end while

```

(Jafari, Sabzchi, 2018)

Tokom implementacije, algoritam treba da pronađe najkraci put do sadržaja stranice i da prati ponašanje veb korisnika i njihove preferencije (Dorigo, Stützle, 2014). Praćenjem prosečnih preferencija i povratnih informacija korisnika, ovo može zajedno da meri stopu veb stranice za sledecie korisnike sa sličnim preferencijama i proizvede rezultat boljih informacija i podataka o pretrazi. Definisanje stope se tretira kao „težina“ stranice - jedan od parametara za optimizaciju.

Algoritam se zasniva na otkrivanju veza na veb stranicama da bi se izračunala matrica jačine veze i definisanje informacija sa putanje korisnika. Kreiraju se pravila učenja kako bi se promenio tok pretrage i obezbedilo bolje rešenje.

Kako bi se otkrio broj pregleda stranice, trebalo bi da postoji procena i analiza za generisanje parametara za preferencije i aktivnosti korisnika veba. Ovaj proces je od suštinskog značaja za sve modele zasnovane na znanju i u ovom slučaju, modele za korišćenje veba. Pristup se sastoji od (Sajwan, Acharya, Bhargava, 2018): upravljanja korisničkim profilom; identifikovanja optimalnog profila najbližeg suseda; podudaranja profila.

Upravljanje korisničkim profilom prikuplja profile korišćenja korisnika sa date veb stranice. Svaki korisnik ima odgovarajući čvor i na taj način se svaki korisnik razlikuje od ostalih korisnika. Mora se uzeti u obzir da dva različita korisnika posećuju istu veb stranicu. Da bismo generisali parametre, prvo treba da postoji identifikacija korisnika, URL stranice koju je posetio korisnik, vremenska oznaka korisnika koji posecije stranicu, i što je najvažnije, da se vide i zahtevaju teme određene stranice koje se sviđaju korisnicima.

Sa identifikacijom ovih činjenica, sledeći korak je analiza i izračunavanje optimalnog profila najbližeg suseda. Ovaj proces omogućava pronalaženje optimalnog rešenja za korisnike koji su najbliži trenutno aktivnim korisnicima (Sajwan, Acharya, Bhargava, 2018). Sličnosti vode izbor najbližeg suseda. Ovo se može poboljšati i identifikovati tehnikom SI. Sa SI se mogu prikupiti važne informacije: relativna pozicija, funkcija vremenske distance, težina ivice aktivnog korisnika i veb stranice.

Sa svim prikupljenim informacijama, možemo započeti proces podudaranja profila. Težina stranice je definisana, zajedno sa ranijim profilima korisnika. Podudaranje se može opisati kao izračunavanje udaljenosti između dva različita profila korišćenjem iste funkcije vremenske distance, i na taj način se prikupljaju i čuvaju nove informacije i znanja za buducie korisnike, sugerując analitički deo stranice i obezbeđujući adekvatne informacije o veb sajtu za korisnika sa sličnim interesovanjem. SI može da obradi bilo koju vrstu izbora suseda.

Poslednji korak je preporučiti korisnicima koja stranica odgovara njihovim potrebama. Ceo ovaj proces se koristi za poboljšanje stranica i prilagođavanje korisnicima i njihovim željama. Ovo je suštinski deo u svrhu optimizacije, za unapređenje postojećeg modela i pružanje najboljih rešenja i rezultata. Sledeći podnaslov rada objašnjava koncept problema optimizacije i formulu za optimizaciju tokom implementacije.

Evaluacija predloženog modela rudarenja upotrebom veba

Optimalni model povratne sprege koristi se za procenu težine stranice. Povratne informacije se sastoje od posete korisnika određenim stranicama na veb lokaciji nakon čega se analizira i čuva njihova aktivnost pregledanja. Povratne informacije su povezane sa preferencijama, tako da je potrebno pratiti odgovor. Da bi se prikupili svi podaci neophodni za rešavanje problema, prvo treba da postoji objašnjenje definicije problema (Dai, Wang, Shu, 2017).

Definicija modela se sastoji od modela povratnih informacija o aktivnostima i preferencijama korisnika. Ako pretpostavka modela povratne informacije korisnika FU ima elemente $FU=\{V, T, L, M, R\}$ gde je $V=\{v_1, v_2 \dots v_n\}$ aktivnosti korisnika pretraživanja, $T=\{t_1, t_2 \dots t_n\}$ predstavlja vreme u kojem korisnik pretražuje veb stranice, $L=\{l_1, l_2 \dots l_n\}$ označava klikove na veb stranice, $M=\{m_1, m_2 \dots m_n\}$ predstavlja ponašanje koje korisnik procenjuje na veb stranicama, a $R=\{r_1, r_2 \dots r_n\}$ označava ponašanje dok odgovara na veb stranice (Dai, Wang, Shu, 2017).

Ako korisnik poseti k stranicu, tada je vrednost aktivnosti pretraživanja v_i podešena na k . Ako korisnik ne pregleda ovu veb stranicu u datom trenutku, tada se aktivnosti pretraživanja postavljaju $v_i=0$ i vremenska vrednost $t_i=0$. Najbolja ocena se sprovodi uz procenu broja klikova (Saremi, Abedin, Kermani, 2008). Učestalost klikova se procenjuje na osnovu vremenske vrednosti, na početku l_i je 0, a kada korisnik klikne prvi put vrednost je podešena na $l_i=l_i+1$.

Kada je u pitanju ponašanje korisničke evaluacije M , veb stranica je važan faktor u kreiranju modela zbog specifikacije upita koji se analizira. Vrednost m_i nije ista, i varira od $1 \dots n$.

Početna vrednost ponašanja korisnika kao odgovor na veb stranicu je 0, ako korisnik odgovara ponašanju, vrednost r_i se postavlja na 1, a nastavlja se sa brojem aktivnosti korisnika. Kada se sakupe sve vrednosti od korisnika i veb stranica, može se definisati težina \hat{u}_i itih veb stranice [6].

Formula veb stranice i model se mogu predstaviti kao (Dai, Wang, Shu, 2017):

$$\omega_i = \lambda_1 \times \frac{1}{v_i + 1} + \lambda_2 \times \frac{t_i}{t_{\max}} + \lambda_3 \times l_i + \lambda_4 \times m_i + \lambda_5 \times r_i \quad (2)$$

Gde t_{\max} predstavlja maksimalno vreme koje korisnik očekuje da proveđe na stranici, sa svim prethodnim faktorima u proceni. Korisnik je fleksibilan da menja i postavlja različite vrednosti svih pet faktora u modelu povratne sprege. Važan faktor je vreme koje korisnik proveđe u pregledanju dokumenata. Na taj način korisnik definiše značaj dokumenata za njega.

Rezultati istraživanja

U nastavku rada bice prikazani praktični, empirijski rezultati rudarenja korišćenja veba koji mogu poboljšati onlajn marketinške aktivnosti i rezultate.

Praktični pristup se sastoji od izdvajanja informacija o ponašanju korisnika koji posećuju veb lokaciju i obrade podataka pomoći SI (Swarm Intelligence) da bi se pronašlo optimalno rešenje i rezultati. Gugl analitika je implementirana na veb sajtu Fakulteta za kompjuterske nauke u Beogradu. Sajt predstavlja glavni izvor informacija o ustanovi i svim aktivnostima koje su u vezi sa njom. Ovaj sajt sadrži vesti, akademske događaje, diplomske i postdiplomske informacije o studiranju, kao i informacije o nastavničkom i administrativnom osoblju. Takođe, uključuje platforme za nekoliko pod-lokacija koje pripadaju Fakultetu kompjuterskih nauka i istraživačkim centrima.

Prikupljanje podataka o pretraživanju korisnika i period validacije postavljeni su od oktobra do novembra 2021. godine što je prikazano na slici 1. Gugl analitika je značajan i koristan alat za dijagnosticiranje stranica i za uvid koji sadržaj stranice korisnici najviše posećuju kako bi se poboljšao sajt za nove korisnike. Pregled sadržaja sajta i stranica je pracién i prikupljen u periodu od 23. oktobra do 23. novembra, što se može videti na dijagramu na slici 1. Broj pregleda stranica je povećan u novembru na 829 pregleda dnevno. Ukupan broj pregleda stranice je 8.308, jedinstvenih prikaza stranice 6.159, zajedno sa prosečnim vremenom na stranici do 1 minuta i 24 sekunde. Maksimalan broj stranica na veb lokaciji je 143, ali u svrhu optimizacije, prvih 10 se prikupljaju i analiziraju pošto su one najposećenije što je ilustrovano na slici 3.

Slika 1. Pregled sadržaja sajta i pretrage



Izvor: Istraživanje autora

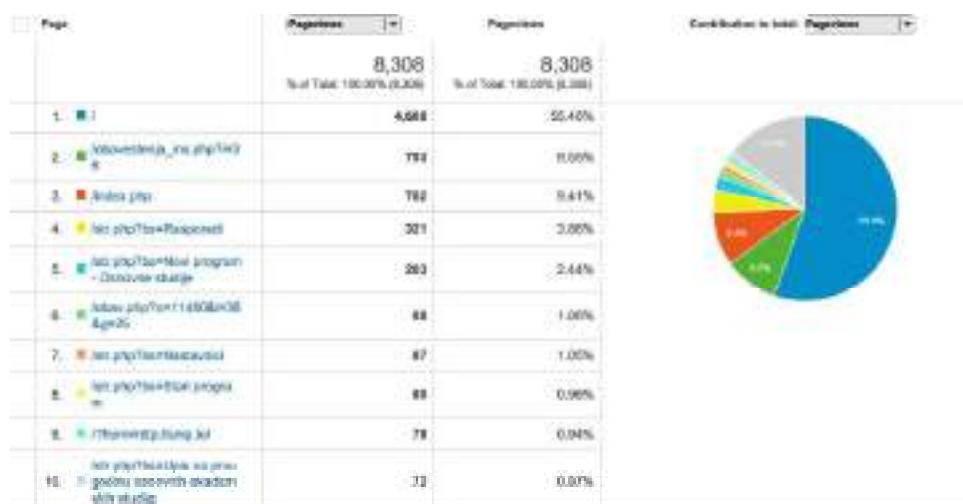
Najvažniji deo prikupljanja korisničkih podataka je ponašanje korisnika prilikom njihovog pretraživanja sajta. Ovaj deo je važan jer može da prikaže rezultate i preporuči korisnicima koja stranica odgovara njihovim potrebama. Ceo ovaj proces se koristi za poboljšanje stranica i prilagođavanje novim korisnicima i njihovim željama. Ovo je suštinski deo u cilju optimizacije, kako bi se unapredio postojeći model i obezbedila najbolja rešenja i rezultati kao što se može videti na slici 2. Početna stranica je indeksna stranica predstavljena zelenom bojom, a tok prikazuje iteraciju prikaza po korisniku. Ovakvim ponašanjem, podaci se prikupljaju i implementiraju u algoritam kako bi se rezultati prikazali za veb lokaciju i buducie korisnike.

Slika 2. Tok ponašanja korisnika



Izvor: Istraživanje autora

Slika 3. Najposećenije stranice



Izvor: Istraživanje autora

Implementacija i eksperimentalni rezultati svitac algoritma za Web Usage Mining

Softverski okvir je razvijen za Web Usage Mining pomocu svitac algoritma. Frejmwork obuhvata metode za parametre koji se koriste kao ulaz, firefli optimizovan kod za optimizaciju i metode za sortiranje veb stranica koje su definisane težinom (najposećenije stranice). Metoda je testirana na 10 veb stranica sa sajta Fakulteta za kompjuterske nauke, prikupljenih od strane Google analitike. GUI (Graphical User Interface) koji je kreiran za ovaj okvir sastoji se od parametara: broj svitaca;

maksimalni ciklus, vreme izvođenja, seme, alfa, gama, betaMin, InputFiles kao izvor podataka.

Slika 4: GUI svitac algoritam za Web Usage Mining



Izvor: Istraživanje autora

Početna populacija svitaca se prvo deli na SP jednake podpopulacije koristeći sledeći izraz:

$$SP = \frac{SN}{PN} \quad (3)$$

gde je SP broj podpopulacije, SN je broj rešenja (broj agenata svitaca u populaciji) i PN je broj veb stranica u uzorku podataka. Dakle, u ovom slučaju, vrednosti za SP, SN i PN su 6, 60 i 10, respektivno.

Slika 5: Rezultati svitac optimizacije



Izvor: Istraživanje autora

Kreiranje efikasne komunikacije sa korisnicima uz pomoć veb podataka

Svako preduzecje, kao privredni subjekt, bilo veliko ili malo, sastoje se od mnogo elemenata. Jedina konstanta u svim preduzeciima su konzumenti - potrošači, a bez njih, preduzecja ne postoje. Zato održavanje stalne komunikacije sa potrošačima nije samo element koji mora postojati u svakom poslu, već i temelj efikasne komunikacije sa javnošću.

Poboljšanje korisničkog iskustva postaje imperativ za preduzecia da ostanu ispred konkurenčije. Da bi poboljšali korisnička iskustva, preduzecia treba da steknu kritične uvide u ponašanje kupaca/korisnika koji posećuju veb-sajt, uključujući šta im je potrebno i gde se nalaze na putu kupovine. Kada su dostupni podaci u realnom vremenu, ovi uvidi pružaju sveobuhvatan pregled korisnika koji prevazilazi njihovo ime, uzrast, pol ili druge jednostavne „signale“ namere kupovine ili drugog ciljanog ponašanja.

Implementacija svitac algoritma za Veb Usage Mining, iz priloženog istraživanja u radu, pokazuje da može omogućiti kompanijama efikasnije upravljanje svojim interakcijama sa trenutnim i potencijalnim korisnicima. Kompanije na osnovu podataka koje dobiju izdvajanjem informacija o ponašanju korisnika koji posećuju veb lokaciju i obrade podataka pomoći inteligencije rojeva, mogu da pripreme relevantnije, pravovremeno i personalizovane kampanje koje će privući buduće kupce i zadržati postojeće. Neke druge prednosti uključuju sledeće:

- Povećava efikasnost i efektivnost prodaje, istovremeno poboljšavajući transparentnost marketing povrata ulaganja (ROI – Return on investment).

- Preduzećia mogu da ojačaju lojalnost kupaca tako što bolje razumeju kupce i na taj način poboljšaju kvalitet svojih usluga.
- Pruža praktične uvide zasnovane na podacima koji podržavaju i pomažu u optimizaciji ključnih odluka o proizvodu.
- Kompanije mogu da predvide efikasnost kampanje i da prilagode svoje marketinške strategije na različitim kanalima kako bi vodile strategije unakrsne i dodatne prodaje.

Ovi uvidi pomažu u poboljšanju radnih tokova usluga za korisnike i usmeravaju buduće marketinške kampanje kako bi se poboljšalo zadovoljstvo kupaca, kao i korisničko iskustvo.

Pravna regulativa – zaštita ličnih podataka i privatnosti na Internetu

Zaštita podataka je skup strategija i procesa koje korisnici interneta mogu koristiti da obezbede privatnost, dostupnost i integritet ličnih podataka. Strategija zaštite podataka je od velikog značaja za svaku organizaciju koja prikuplja, rukuje ili čuva osetljive podatke. Sa naglim razvojem digitalne tehnologije i interneta, ovo pravo ozbiljno je dovedeno u pitanje. U Zakonu o zaštiti ličnih podataka Srbije, utvrđuje se da je poverenik dužan da sačini i javno objavi na svojoj internet stranici listu vrsta radnji obrade prikupljenih podataka. Povelja Evropske unije o osnovnim pravima garantuje da svako ima pravo na zaštitu ličnih podataka u svim aspektima života: kod kuće, na poslu, prilikom kupovine, kada su na lečenju, u policijskoj stanici ili na internetu (European Commision, 2016). U digitalnom dobu, prikupljanje i čuvanje ličnih podataka je od suštinskog značaja. Podatke koriste sva preduzeća koja postoje na internetu i imaju svoje zvanične veb-sajtove. Ponašanje na internetu je predmet regulacije nacionalnog prava. Međutim, zbog karakteristika interneta i njegove globalne rasprostranjenosti redefinišu se i domeni zaštite, jer je elektronska komunikacija internacionalnog karaktera, pa samim tim i zaštita mora da bude međudržavno orijentisana (Dimitrijević, 2010). Jedan od načina da se korisnici zaštite je da se upoznaju sa uslovima korišćenja i politikom privatnosti pre nego što počnu da pretražuju veb.

Zaključak

U ovom radu je predložena metaheuristička optimizacija svitac algoritma da bi se odredilo optimalno rešenje za rudarenje korišćenja veba i uz pomoć tih podataka kreirala efikasna komunikacija sa budućim i postojećim korisnicima. Rudarstvo korišćenja veba je izazovna tehnika, ali je uspešno rešena metaheuristikom rojeve inteligencije čime je olakšano grupisanje ili segmentacija publike na osnovu sličnih obrazaca u prikupljenim podacima. Upotreba veba je nedavno porasla u velikoj meri. Veb nije samo izvor informacija, već je i odlična platforma za uspešno vođenje poslovanja. Neke od glavnih upotreba veba u poslovanju kompanije su oblasti elektronske trgovine i digitalnog marketinga. Jedna od tehnologija koja služi kao alat u

ovim oblastima je veb rudarenje pomoću koga se može izvršiti izdvajanje potrebnih obrazaca informacija u veliku bazu podataka (Yadao et al., 2022). Na osnovu ovih podataka može se uspešnije segmentirati tržište, što će u velikoj meri doprineti kreiranju efikasne komunikacije sa korisnicima. S obzirom na to da je internet nezaobilazno sredstvo poslovanje svake kompanije, napori su usmereni ka postizanju visokog stepena zaštite privatnosti korisnika interneta, što bi trebalo da omoguci da njihovo pretraživanje interneta i poslovanje bude bezbedno. U okviru buducieg istraživanja, drugi algoritmi rojeva inteligencije bi mogli biti prilagođeni za rešavanje problema u originalnoj i modifikovanoj implementaciji kreiranja marketing strategije.

Literatura

1. A. Brun, S. Castagnos, and A. Boyer, "From Community Detection to Mentor Selection in Rating-Free Collaborative Filtering", *Advances in Multimedia*, 2017, pp.1 -19.
2. Achar, C., So, J., Agrawal, N., Duhachek, A., What we feel and why we buy: the influence of emotions on consumer decision-making, *Current Opinion in Psychology*, Vol.10, 2016, pp.166-170.
3. A. Grbić, M. Tankosić, "Socijalno-pravni aspekti i etika roditeljstva u republici Srbiji", *Megatrend Revija: međunarodni časopis za primenjenu ekonomiju*. - ISSN 1820-3159 (God. 19, br. 1, 2022, str. 67-84)
4. D. Dong, "Exploration on Web Usage Mining and its Application", *International Workshop on Intelligent Systems and Applications*, 2013, pp.1
5. Dimitrijević, P. Pravo informacione tehnologije, 2010, Sven, Niš
6. European Commision (2016): Reform of EU data protection rules Direktiva Data Protection 95/46/EZ. Dostupno na: http://ec.europa.eu/justice/data-protection/reform/index_en.htm (17.8.2022)
7. F. Bonchi, "Web log data warehousing and mining for intelligent Web caching", *Data and knowledge engineering*, 2018, Vol. 39, Issue 2, pp. 165–189.
8. H. Qahri Saremi, B. Abedin, and A. Meimand Kermani, "Website structure improvement: quadratic assignment problem approach and ant colony meta-heuristic technique," *Applied Mathematics and Computation*, vol. 195, no. 1, pp. 285–298, 2008.
9. H. Ishikawa, "An intelligent Web recommendation system: A Web usage mining approach", *Proceeding ISMIS '02 Proceedings of the 13th International Symposium on Foundations of Intelligent Systems*, 2016, Vol. 13, pp. 342-350.
10. I. Dzitac, Advanced AI Techniques for Web Mining, Mathematical methods, COMPUTATIONAL TECHNIQUES, NON-LINEAR SYSTEMS, INTELLIGENT SYSTEMS, 2017.
11. I. Brajević and M. Tuba, "An upgraded artificial bee colony algorithm (ABC) for constrained optimization problems", *Journal of Intelligent Manufacturing*, vol. 24, no. 4, pp. 729-740, Aug 2013.
12. L. Dai, W. Wang, W. Shu, An Efficient Web Usage Mining Approach Using Chaos Optimization and Particle Swarm Optimization Algorithm Based on Optimal

- Feedback Model, Hindawi Publishing Corporation Mathematica Problems in Engineering Volume 2 013, Article ID 3 4 0 4 8 0, August 2017.
13. M. Jafari, Farzad Soleymani Sabzchi, Applying Web Usage Mining Techniques to Design Effective Web Recommendation Systems: A Case Study, ACSIJ Advances in Computer Science: an International Journal, Vol. 3, Issue 2, No.8, March 2018.
 14. M. Dorigo and T. Stützle. *Ant Colony Optimization*. MIT Press, Cambridge, MA, 2004.
 15. M. Sajwan, K. Acharya, S. Bhargava, Swarm Intelligence Based Optimization for Web Usage Mining in Recommender System, International Journal of Computer Applications Technology and Research Volume 3– Issue 2, 11 9 - 124, 2018, ISSN: 2319–865.
 16. M. R. Morris, J. Teevan, and S. Bush, “Enhancing collaborative web search with personalization: groupization, smart splitting, and group hit-highlighting,” in Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW ’08), pp. 481–484, San Diego, Calif, USA, November 2008
 17. M. Tuba, and N. Bacanin, "Artificial bee colony algorithm hybridized with firefly metaheuristic for cardinality constrained mean-variance" The Scientific World Journal, Volume 2014.
 18. M. Tuba and R. Jovanovic, “Improved ant colony optimization algorithm with pheromone correction strategy for the traveling salesman problem”, International Journal of Computers, Communications & Control, vol. 8, issue 3, pp. 477–485, Jun 2013.
 19. M. Subotic, and M. Tuba, "Parallelized Multiple Swarm Artificial Bee Colony Algorithm (MS-ABC) for Global Optimization," Studies in Informatics and Control, vol. 23, no. 1, pp. 117-126, 2014.
 20. N. Bacanin, M. Tuba, I. Strumberger, “RFID Network Planning by ABC Algorithm Hybridized with Heuristic for Initial Number and Locations of Readers”, 2015 17th UKSIM-AMSS International Conference on Modelling and Simulation, IEEE DOI 10.1109/UKSim, 2015.
 21. N. Bacanin, M. Tuba, “Firefly Algorithm for Cardinality Constrained Mean-Variance Portfolio Optimization Problem with Entropy Diversity Constraint”, The Scientific World Journal, Volume 2014, Article ID 721521, Hindawi Publishing Corporation, May 2014.
 22. P. Shoval, V. Maidel, B. Shapira, and M. Taieb-Maimon,"Ontological content-based filtering for personalized newspapers: A method and its evaluation", Online Information Review, 2018, Vol. 34, No. 5, pp. 729-756.
 23. Plassmann H., Ramsøy T.Z., Milosavljevic M., Branding the brain: A critical review and outlook, Journal of Consumer Psychology, Vol. 22, 2012, pp. 18-36.
 24. R. Jovanovic and M. Tuba, “An ant colony optimization algorithm with improved pheromone correction strategy for the minimum weight vertex cover problem”, Applied Soft Computing, vol. 11, issue 8, pp. 5360–5366, Dec 2011.
 25. R. Jovanovic and M. Tuba, “Ant colony optimization algorithm with pheromone correction strategy for the minimum connected dominating set problem”, Computer Science and Information Systems (ComSIS), vol. 10, no. 1, pp. 133–149, Jan 2013..

26. Sharma, N., Patterson, P.G., The impact of communication effectiveness and service quality on relationship commitment in consumer, professional services, *Journal of Services Marketing*, Vol. 13, 1999, pp.151-170.
27. Shivani Yadao, A. Vinaya Babu, Midhunchakkavarthy Janarthanan, and Amiya Bhaumik, "A Combined Approach of Web Content Mining and Neural Networks for Predicting the societal impact of covid-19 through twitter", *Journal of Pharmaceutical Negative Results*, pp. 58–74, Oct. 2022
28. T. Nguyen, D. Phung, B. Adams, T. Tran, and S. Venkatesh, "Classification and Pattern Discovery of Mood in Weblogs", *Advances in Knowledge Discovery and Data Mining*, 2018, Vol. 6119, pp. 283-290.
29. Tankosic, M., Influence of Media to Social Behavior and the Changes that have been Caused by New Media, *Proceedings of the 5th International Conference on Applied Economics, Business and Development (AEBD '13)*, Chania, Crete Island, Greece, 2013, pp. 143-148.
30. Tankosic, M., Grbic A., Krivokapic Z., "The Marginalization and Exploitation of Women in Media Industry", Globalization and Its Impact on Violence Against Vulnerable Groups, IGI Global publisher of timely knowledge, 2020, pp.75-94
31. Tankosic, M., Trifunovic, D., Development of Cultural Institutions through Implementation of Service Innovation, *Proceedings of the 5th International Conference on Applied Economics, Business and Development (AEBD '13)*, Chania, Crete Island, Greece, 2013, pp. 161-166.
32. Tankosic, M., Ivetic, P., Mikelic, K., Managing internal and external Communication in a Competitive Climate via EDI concept, *International Journal of Communications, COPE Ethics Guidelines*, Vol. 2, 2017, pp. 1 - 6, 2367-8887.
33. X.-S. Yang, *Nature-Inspired Metaheuristic Algorithms*, Luniver Press, 2008.
34. X. S. Yang, "Multiobjective firefly algorithm for continuous optimization", *Engineering with Computers*, Vol. 29, No. 2, pp. 175-184, Mar 2017.
35. Y. Shang and L. Li, "Precision evaluation of search engines," *World Wide Web: Internet and Web Information Systems*, vol. 5, pp. 159–173, 2013
36. Yan-fei Zhu Overview of swarm intelligence Computer Application and System Modeling (ICCASM), 2010 International Conference on Taiyuan

Datum prijema (Date received): 23.11.2022.

Datum prihvatanja (Date accepted): 29.12.2022.

CREATION EFFECTIVE COMMUNICATIONS WITH USERS THROUGH THE TECHNIQUES WEB MINING AND ITS LEGAL REGULATION

Mirjana Tankosić,⁵⁵ Ana Grbić⁵⁶

Abstract

Modern companies start their business activities by discovering, analyzing and understanding the needs and desires of consumers, and end them by delivering the expected value that results in consumer satisfaction.

The aim of the research of this work is to analyze the use of techniques for collecting data about consumers via the web and extracting the necessary forms of information into a large database. Given that these techniques can also be used to access the personal data of consumers, it will be analyzed whether this is in accordance with legal and ethical norms.

The practical approach of the work consists of extracting information about the behavior of users visiting the website and processing the data with the help of swarm intelligence, in order to find the optimal solution and results for effective communication with users. This entire process is used to improve the pages and adapt to new users and their preferences.

In the age of a developed competitive environment, information about consumers is a key link in a company's successful marketing business. One source that can provide additional information on consumer segmentation is web mining.

Keywords: communication, digital marketing, legal regulation, web mining

JEL: M29, M30, K38, M24

Introduction

It has become very important for modern companies to know the needs and wants of consumers. In this way, the conclusion is reached that the most important factor for modern companies are consumers. In the beginning, companies launched products with the best performance in order to capture as much of the market as possible. When they reached maximum performance, it was necessary to attract customers in a new way. Communication between companies and consumers has ceased to be limited regardless of distance and national borders (Tankosić, 2013). With the traditional way of focusing on the company's product, they began to provide consumers with attention and listen to the needs of the market. At the beginning of "market listening" companies used outdated sample testing methods. Over the years,

⁵⁵ Assistant Professor, Mirjana Tankosić, PhD, Higher Colleges of Technology, Faculty of Applied Media, Fujairah Women's Campus, Al Qasar/Kuwait Road, Al Faseel, Fujairah, UAE, Phone: +9712 206 9343, E-mail: mtankosic@hct.ac. yes

⁵⁶ Assistant Professor, PhD Ana Grbić, College of Social Work, Terazije 34, Belgrade, Phone: +381 11 404 08 39, E-mail: ana.grbic@asp.edu.rs

it became clear that respondents did not give the most honest answers when filling out the questionnaire. Informational technology is caused deep changes in economy, culture and politics (Tankosić, Grbić, Krivokapić, 2020).

Innovations in any system are organized and purposeful activities aimed at creating changes (Tankosić, Trifunović, 2013). With the development of technology, marketing experts recognized the possibility of applying science in their research. In order to find out as much information as possible about their consumers, they only need to have access to the web because all information about consumers is already there voluntarily. New technologies and digitization help the development of web mining. By collecting information about the consumer through the Internet, you can contribute to the development of traditional and digital marketing of any company.

Internet technology has developed rapidly in the past decade, evolving in communication, networking and information content. One of the main aspects of Internet technology is the World Wide Web (WWW) which provides a significant amount of information available to everyone. The dynamic part of the WWW is user interaction with web pages, gathering information and content. There are many websites that are popular and more relevant than others, but, based on user searches, the importance of information is entirely based on user preferences. Keywords will trigger a query in a particular search engine, and as a result, will provide a list of the best rated web pages according to the criteria (Shang, Li, 2013). Discovery knowledge contained in data (Knowledge Discovery in Date), is area which relies on the will reach mechanical engineering learning like works skills intelligence, with extractions potentially useful and interesting one's knowledge from data big volume. Traditionally, techniques mining data (Data Mining), were applied on the data which one are through the time collected ¼ but in relational or transactional bases or team composition data company, but those, in principle, can apply on the three were which one type of repository data, so and on the web data. (Bošnjak, Grlević, Bošnjak, 2019).

Methods

In order to understand user/consumer behavior, it is necessary to analyze knowledge-based models that display all required information and data. In order to achieve a large amount of content in a quick response to a certain query, Web Mining technology appears in the technical background and social environment (Morris, Teevan, Bush, 2008). The main purpose of this technique is to present generated information about a website through a structure of links and user preferences.

Web mining techniques can significantly improve online marketing activities. Web mining consists of mining web content and using the web (Dai, Wang, Shu, 2017). Web content mining is related to web page content, database and background process. Web Usage Mining is an activity that uncovers new information and knowledge by analyzing web structure, content, and usage records. It is part of the Data Mining method that uses data, but in this case, Web Usage Mining uses data from web

resources, takes all the interesting behaviors, patterns and information. Since knowledge is the main part of this technique, it can be used as a key link to optimize and improve the speed and accuracy of web content response to users.

This paper provides an analysis and presentation of the theoretical foundations of Swarm Intelligence (SI) as an optimization of Web Usage Mining (WUM) for the needs of online marketing activities, the implementation of the firefly algorithm and the results of SI provides an efficient way to find the optimal solution. Since it deals with natural and artificial systems, in order to define a model that can optimize and create a solution, it is necessary to compare WUM and SI in order to gather all the information that is necessary for the experimental part of the improvement. Analytical com method included analysis contents relevant science and professional literature in this one area (Grbić, Tankosić, 2022).

Search engines help users search for information on the web, but they are only effective when users have a clear knowledge of what they are looking for. A graph-based approach is used to recommend web pages that identifies those users who discover interesting pages, based on their knowledge and calculation that the pages are adequate. By analyzing the browsing activity of previous users, new interesting pages can be identified and then recommended to new users who show the same interest.

To rank the importance of certain pages based on search, links and content analysis are key. This part is required for optimization and can be used. Numerical weight is assigned based on various parameters. The aim of this paper is to present the important aspects of web content search, analyze the web usage mining model and present optimization that can be used as result analytics for future web pages to have better insight into user preferences and ensure quality, but also provide better web search results. users and enabled more effective communication with web content users (consumers).

Defining the concept of web usage mining

Web Usage Mining is part of a data mining technique that discovers patterns of behavior based on data from the web. The web usage mining process is divided into (Jafari, Sabzchi, 2018): 1. data preparation; 2. pattern detection; 3. pattern analysis.

Data preparation creates a server session file in which each session is defined as a series of different types of requests from a single user visiting a page (Ishikawa, 2018). Pattern detection tracks patterns, usage clusters, and user classification for each session enabled by users (Nguyen, Phung, Adams, 2018). All data and information extracted from the Web Usage Mining process can be used for web personalization, site modifications, and business intelligence discovery (Shoval, Maidel, Shapira, Taieb-Maimon, 2018). It can also be useful for web server caching as it will reduce the number of requests and balance the load (Bonchi, 2018).

Pre-processing is an essential part of web usage research because all data available on the web is unstructured and needs to be transformed so that the analytical part can produce results of user patterns. The raw data is collected and converted into a set of user profiles to perform the pattern discovery stage (Dong, 2013). In data preprocessing, the data is cleaned, the user log is identified in each session, and path recognition is performed.

Pattern detection extracts user behavior based on data that has been formatted in the pre-processing stage. Because of this important step, all data must be pre-prepared and converted in order to be used as input. The knowledge obtained through this process is used for statistical purposes of the sites and communication with users. Two important techniques in this phase that are considered in the experimental part of the work are: statistical analysis and clustering.

Statistical analysis obtains the session file from the web log and all data is used to create reports on user pages, average page visit time, common entries. This knowledge can be useful for monitoring high traffic days on the site, for changing the functionality of the site for users, or for improving the security aspects of the site. Another technique, clustering, is used to group users or data that are somewhat similar (Jafari, Sabzchi, 2018). Since the classification is based on the aspect of similarity (user/data), this can be used to form activities for new users who have an identical perception.

Pattern analysis extracts all rules, patterns and statistics from pattern detection results, filters non-relative data and creates statistical results of the defined session. This is a combination of knowledge query engines such as SQL, graphical patterns, and aggregate data results.

An evaluation of the Swarm Intelligence website

Swarm Intelligence (SI) is a term associated with a system that collects the behavior of decentralized artificial systems by analogy with the real world where the collective behavior of a swarm can lead to the intelligent behavior of an individual displayed by the group (Dzitac, 2017). This collective intelligence represents the group's ability to solve more problems than its individual members (Zhu, 2010). These systems are simple agents that communicate with each other, at the same time interacting with the environment. They follow the rules without centralized control.

Algorithms based on metaheuristics simulate natural phenomena and can be optimized and implemented in different fields to create better solutions and results. Nature-based processes are emulated to solve optimization problems. Swarm intelligence is based on individuals displaying collective intelligence. This behavior can be seen in ant colonies (Jovanović, Tuba, 2010), honey bee colonies (Tuba, Bačanin, 2014), etc.

The firefly algorithm (FA) represents one of the latest metaheuristics of swarm intelligence (Bačanin, Tuba, 2014). It was originally proposed by Yang in 2008

(Yang, 2008) and applied to optimization problems. The original firefly algorithm was created to guide the behavior of fireflies, but also to mimic the behavior of a flash. Each firefly moves towards the light. This is the main part of the algorithm which is defined by the objective functions. This flickering is used to attract fireflies as a mating and prey signaling system. Since the scrolls do not have a global view of the environment, their connection is based on light intensity, which is similar to a web user navigating the network without information about the route followed by other users with the same preferences (Dzitac, 2017).

In order to apply and implement the firefly algorithm, three idealized rules should be taken into account (Bačanin, Tuba, 2014):

1. All scrolls are sexless; attraction is not based on gender;
2. Attractiveness is defined and directly proportional to their brightness, and the less brilliant firefly will move towards the brighter. The brightness increases as the distance between the candles decreases;
3. The brightness of the fireflies is determined by the value of the objective function.

When it comes to the objective function $f(x)$, it is used to define and encode the brightness of the sconces. It is made to show the light intensity at location x . Thus, $I(x) = f(x)$. The important part is the movement. Movement is based on attraction (Bačanin, Tuba, 2014), and when firefly j is more attractive than firefly i , firefly moves towards j (Yang, 2017):

$$x_i(t) = x_i(t) + \beta_0 r^{-\gamma r_{ij}^2} (x_j - x_i) + \alpha(\text{rand} - 0.5) \quad (1)$$

Where β_0 represents the attraction $r = 0$, α is the randomization parameter, rand is a random number evenly distributed between 0 and 1; and the distance between the scrolls i, j is $r_{i,j}$ (Bačanin, Tuba, 2014).

Simplification algorithm pseudo code Firefly is presented in below (Bačanin, Tuba, 2014):

```

Generate initial population of fireflies  $xi$ ,
( $i = 1, 2, 3, \dots, FN$ )
Light intensity  $Ii$  at point  $xi$  is defined by  $f(x)$ 
Define light absorption coefficient  $y$  define number of iterations  $IN$ 
while ( $t < IN$ ) do
for ( $i = 1$  to  $FN$ ) do
for ( $j = 1$  to  $i$ ) do
if ( $Ij < Ii$ ) then
Move firefly  $j$  towards firefly in  $d$  dimension
Evaluate new solution, replace worst with better solution, and update

```

```

light intensity
end if
end for
end for
Rank all fireflies, find the current best, and move them randomly
end while

```

(Jafari, Sabzchi, 2018)

During implementation, the algorithm should find the shortest path to the content of the page and monitor the behavior of web users and their preferences (Dorigo, Stützle, 2014). By tracking average user preferences and feedback, this can collectively measure the website's rate for subsequent users with similar preferences and produce better search information and data output. Defining the rate is treated as the "weight" of the page - one of the parameters for optimization.

The algorithm is based on discovering links on web pages to calculate the link strength matrix and define information from the user journey. Learning rules are created to change the search flow and provide a better solution.

In order to find out the number of page views, there should be an evaluation and analysis to generate parameters for web user preferences and activities. This process is essential for all knowledge-based models and in this case, models for web usage. The approach consists of (Sajwan, Acharya, Bhargava, 2018): user profile management; identifying the optimal profile of the nearest neighbor; profile matching.

User profile management collects user usage profiles from a given website. Each user has a corresponding node and thus each user differs from other users. It must be taken into account that two different users are visiting the same web page. In order to generate the parameters, first there should be the identification of the user, the URL of the page visited by the user, the timestamp of the user visiting the page, and most importantly, to see and request the topics of a particular page that users like.

With the identification of these facts, the next step is the analysis and calculation of the optimal nearest neighbor profile. This process enables finding the optimal solution for users who are closest to currently active users (Sajwan, Acharya, Bhargava, 2018). Similarities guide the selection of the nearest neighbor. This can be improved and identified by the SI technique. With SI, important information can be collected: relative position, time-distance function, edge weight of active user and web page.

With all the information gathered, we can start the profile matching process. Page weight is defined, along with previous user profiles. Matching can be described as calculating the distance between two different profiles using the same temporal distance function, and thus collecting and storing new information and knowledge for future users, suggesting an analytical part of the page and providing adequate

information about the website for users with similar interests. SI can handle any type of neighbor selection.

The final step is to recommend to users which site suits their needs. This whole process is used to improve the pages and adapt them to users and their wishes. This is an essential part for the purpose of optimization, to improve the existing model and provide the best solutions and results. The following subtitle of the paper explains the concept of the optimization problem and the formula for optimization during implementation.

Evaluation of the proposed mining model using the web

An optimal feedback model is used to estimate the weight of the page. Feedback consists of a user's visit to certain pages on a website after which their browsing activity is analyzed and stored. Feedback is linked to preferences, so the response needs to be monitored. In order to collect all the data necessary to solve the problem, there should first be an explanation of the problem definition (Dai, Wang, Shu, 2017).

The model definition consists of a feedback model of user activities and preferences. If the user feedback model assumption FU has elements $FU = \{V, T, L, M, R\}$ where $V = \{v_1, v_2, \dots, v_n\}$ of user search activity, $T = \{t_1, m_2, \dots, m_n\}$ represents the behavior that the user evaluates on web pages, and $R = \{r_1, r_2, \dots, r_n\}$ denotes the behavior while responding to web pages (Dai, Wang, Shu, 2017).

If the user visits the k th page, then the search activity value v and is set to k . If the user does not view this web page at a given time, then the search activity is set to $v=0$ and time value $t=0$. The best evaluation is carried out by evaluating the number of clicks (Saremi, Abedin, Kermani, 2008). The frequency of clicks is estimated based on the time value, at the beginning l is 0, and when the user clicks the first time the value is set to $l=1+1$.

When it comes to user evaluation behavior M , the web page is an important factor in the creation of the model due to the specification of the query being analyzed. The value of m and is not the same, and varies from $1\dots n$.

Initial value behavior user like the answer on the web pages is 0, if the user fit behavior, value r sets on 1, a continues with number activities user. When collect all values from the user and web page, can define weight and weight etc web pages [6].

The website formula and model can be represented as (Dai, Wang, Shu, 2017):

$$\omega_i = \lambda_1 \times \frac{1}{v_i + 1} + \lambda_2 \times \frac{t_i}{\max} + \lambda_3 \times l_i + \lambda_4 \times m_i + \lambda_5 \times r_i \quad (2)$$

Where t and \max represent the maximum time the user expects to spend on the page, with all previous factors in the estimation. The user is flexible to change and set different values of all five factors in the feedback model. An important factor is the

time the user spends viewing the documents. In this way, the user defines the importance of the documents for him.

Research results

In the continuation of the paper, practical, empirical results of web mining that can improve online marketing activities and results will be presented.

The practical approach consists of extracting information about the behavior of users who visit the website and processing the data with the help of SI (Swarm Intelligence) to find the optimal solution and results. Google Analytics has been implemented on the website of the Faculty of Computer Sciences in Belgrade. The site is the main source of information about the institution and all activities related to it. This site contains news, academic events, graduate and postgraduate study information, as well as information about teaching and administrative staff. It also includes platforms for several sub-sites belonging to the Faculty of Computer Science and Research Centers.

The collection of user search data and the validation period are set from October to November 2021, which is shown in Figure 1. Google Analytics is an important and useful tool for diagnosing pages and for seeing which page content users visit the most in order to improve the site for new users. The review of the content of the site and pages was monitored and collected in the period from October 23 to November 23, which can be seen in the diagram in Figure 1. The number of page views increased in November to 829 views per day. Total pageviews are 8,308, unique pageviews 6,159, along with an average time on page of up to 1 minute and 24 seconds. The maximum number of pages on a website is 143, but for optimization purposes, the first 10 are collected and analyzed as they are the most visited as illustrated in Figure 3.

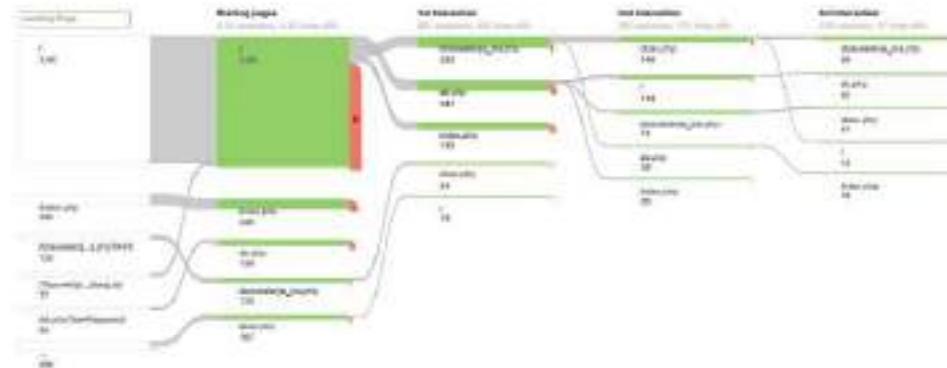
Figure 1. Overview of site content and search



Source: Author's research

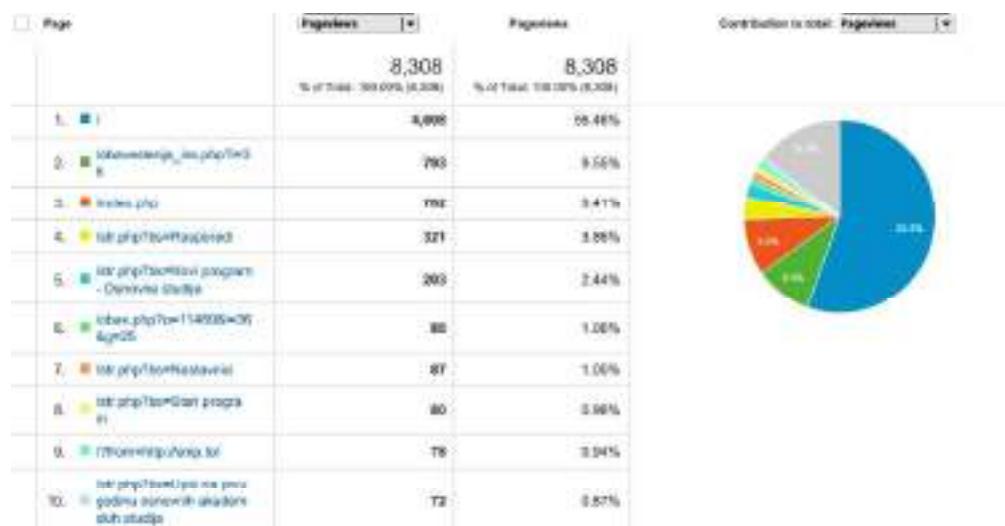
The most important part of the collection of user data is the behavior of users when they browse the site. This part is important because it can display results and recommend to users which page suits their needs. This entire process is used to improve the pages and adapt them to new users and their preferences. This is an essential part in the optimization goal, in order to improve the existing model and provide the best solutions and results as can be seen in Figure 2. The home page is the index page represented in green, and the flow shows the iteration of the view per user. With this behavior, data is collected and implemented into an algorithm to display results for the website and future users.

Figure 2. Flow of user behavior



Source: Author's research

Figure 3. The most visited pages



Source: Research author

Implementation and experimental results of the svitac algorithm for Web Usage Mining

A software framework has been developed for Web Usage Mining using the svetac algorithm. The framework includes methods for parameters used as input, firefly optimized optimization code and methods for sorting web pages defined by weight (most visited pages). The method was tested on 10 web pages from the website of the Faculty of Computer Science, collected by Google Analytics. The GUI (Graphical

User Interface) that was created for this frame consists of parameters: number of scrolls; max cycle, run time, seed, alpha, gamma, betaMin, InputFiles as data source.

Figure 4: GUI Firefly algorithm For Web Usage Mining



Source: Research author

Home page population fireflies are first divided into SP equals subpopulations using next expression:

$$SP = \frac{SN}{PN} (3)$$

where SP is the subpopulation number, SN is the solution number (the number of firefly agents in the population) and PN is the number of web pages in the data sample. So, in this case, the values for SP, SN and PN are 6, 60 and 10, respectively.

Figure 5: Results Firefly optimization



Source: Research author

Creating effective communication with users with the help of web data

Every company, as a business entity, whether large or small, consists of many elements. The only constant in all companies are consumers, and without them, companies do not exist. That is why maintaining constant communication with consumers is not only an element that must exist in every business, but also the foundation of effective communication with the public.

Improving user experience is becoming imperative for businesses to stay ahead of the competition. To improve user experiences, businesses need to gain critical insights into the behavior of customers/users visiting a website, including what they need and where they are in the purchase journey. When real-time data is available, these insights provide a comprehensive view of users that goes beyond their name, age, gender, or other simple "signals" of purchase intent or other targeted behavior.

The implementation of the svitac algorithm for Web Usage Mining, from the attached research in the paper, shows that it can enable companies to more effectively manage their interactions with current and potential users. Based on the data obtained by extracting information about the behavior of users who visit the website and processing data using swarm intelligence, companies can prepare more relevant, timely and personalized campaigns that will attract future customers and retain existing ones. Some other benefits include the following:

- It increases the efficiency and effectiveness of sales, while improving the transparency of marketing return on investment (ROI - Return on investment).

- Companies can strengthen customer loyalty by better understanding customers and thereby improving the quality of their services.
- It provides actionable data-driven insights that support and help optimize key product decisions.
- Companies can predict campaign effectiveness and adjust their marketing strategies across channels to drive cross-sell and up-sell strategies.

These insights help improve customer service workflows and guide future marketing campaigns to improve customer satisfaction and customer experience.

Legal regulation - protection of personal data and privacy on the Internet

Data protection is a set of strategies and processes that Internet users can use to ensure the privacy, availability, and integrity of personal data. A data protection strategy is of great importance to any organization that collects, handles or stores sensitive data. With the rapid development of digital technology and the Internet, this right has been seriously challenged. In the Law on the Protection of Personal Data of Serbia, it is established that the commissioner is obliged to compile and publicly publish on his website a list of the types of actions of processing the collected data. The Charter of Fundamental Rights of the European Union guarantees that everyone has the right to the protection of personal data in all aspects of life: at home, at work, when shopping, when receiving treatment, at the police station or on the Internet (European Commission, 2016). In the digital age, the collection and storage of personal data is essential. Data is used by all companies that exist on the Internet and have their official websites. Behavior on the Internet is subject to national law regulation. However, due to the characteristics of the Internet and its global spread, the domains of protection are also being redefined, because electronic communication is international in nature, and therefore protection must be internationally oriented (Dimitrijević, 2010). One way users can protect themselves is to familiarize themselves with the terms of use and privacy policy before they start browsing the web.

Conclusion

In this paper, a meta-heuristic optimization of the svitac algorithm is proposed in order to determine the optimal solution for web usage mining and with the help of these data to create effective communication with future and existing users. Web mining is a challenging technique, but it has been successfully solved by metaheuristics of swarm intelligence, which facilitates the grouping or segmentation of the audience based on similar patterns in the collected data. The use of the web has recently grown to a great extent. The web is not only a source of information, but also an excellent platform for successful business management. Some of the main uses of the web in the company's business are in the areas of e-commerce and digital marketing. One of the technologies that serves as a tool in these areas is web mining,

which can be used to extract the required patterns of information in a large database (Yadao et al., 2022). Based on this data, the market can be segmented more successfully, which will greatly contribute to the creation of effective communication with users. Given that the Internet is an indispensable tool for the business of every company, efforts are directed towards achieving a high degree of privacy protection for Internet users, which should enable their Internet browsing and business to be safe. In future research, other intelligence swarm algorithms could be adapted to solve problems in the original and modified implementation of marketing strategy creation.

Literature

1. A. Brun, S. Castagnos, and A. Boyer, "From Community Detection to Mentor Selection in Rating-Free Collaborative Filtering", *Advances in Multimedia*, 2017, pp.1 -19.
2. Achar, C., So, J., Agrawal, N., Duhachek, A., What we feel and why we buy: the influence of emotions on consumer decision-making, *Current Opinion in Psychology*, Vol.10, 2016, pp. 166 -170.
3. A. Grbić, M. Tankosić, " Social and legal aspects and ethics parenthood in the republic Serbia ", *Megatrend Review: international magazine For applied the economy*. - ISSN 1820-3159 (Year 19, No. 1, 2022, pp. 67-84)
4. D. Dong, "Exploration on Web Usage Mining and its Application", *International Workshop on Intelligent Systems and Applications*, 2013, pp.1 – 4.
5. Dimitrijević, P. Law of information technology, 2010, Sven, Niš
6. European Commission (2016): Reform of EU data protection rules Data Protection Directive 95/46/ EC. Available at: <http://ec.europa.eu/justice/data-protection/reform/indexen.htm> (17.8.2022)
7. F. Bonchi, "Web log data warehousing and mining for intelligent Web caching", *Data and knowledge engineering*, 2018, Vol. 39, Issue 2, pp. 165–189.
8. H. Qahri Saremi, B. Abedin, and A. Meimand Kermani, "Website structure improvement: quadratic assignment problem approach and ant colony meta-heuristic technique," *Applied Mathematics and Computation*, vol. 195, no. 1, pp. 285–298, 2008.
9. H. Ishikawa, "An intelligent Web recommendation system: A Web usage mining approach", Proceeding ISMIS '02 Proceedings of the 1 3th International Symposium on Foundations of Intelligent Systems, 2016, Vol. 13, pp. 342-350.
10. I. Dzitac, Advanced AI Techniques for Web Mining, Mathematical methods, COMPUTATIONAL TECHNIQUES, NON-LINEAR SYSTEMS, INTELLIGENT SYSTEMS, 2017.
11. I. Brajovic and M. Tuba, "An upgraded artificial bee colony algorithm (ABC) for constrained optimization problems", *Journal of Intelligent Manufacturing*, vol. 24, no. 4, pp. 729-740, Aug 2013.
12. L. Dai, W. Wang, W. Shu, An Efficient Web Usage Mining Approach Using Chaos Optimization and Particle Swarm Optimization Algorithm Based on Optimal Feedback Model, Hindawi Publishing Corporation Mathematica Problems in Engineering Volume 2 013, Article ID 3 4 0 4 8 0, August 2017.

13. M. Jafari, Farzad Soleimani Sabzchi, Applying Web Usage Mining Techniques to Design Effective Web Recommendation Systems: A Case Study, ACSIJ Advances in Computer Science: an International Journal, Vol. 3, Issue 2, No.8, March 2018.
14. M. Dorigo and T. Stützle. *Ant Colony Optimization*. MIT Press, Cambridge, MA, 2004.
15. M. Sajwan, K. Acharya, S. Bhargava, Swarm Intelligence Based Optimization for Web Usage Mining in Recommender System, International Journal of Computer Applications Technology and Research Volume 3– Issue 2, 119 - 124, 2018, ISSN: 2319 -865.
16. MR Morris, J. Teevan, and S. Bush, "Enhancing collaborative web search with personalization: groupization, smart splitting, and group hit-highlighting," in Proceedings of the ACM Conference on Computer Supported Cooperative Work (CSCW '08), pp. 481–484, San Diego, Calif, USA, November 2008
17. M. Tuba, and N. Bacanin, "Artificial bee colony algorithm hybridized with firefly metaheuristic for cardinality constrained mean-variance" The Scientific World Journal, Volume 2014.
18. M. Tuba and R. Jovanovic, "Improved ant colony optimization algorithm with pheromone correction strategy for the traveling salesman problem", International Journal of Computers, Communications & Control, vol. 8, issue 3, pp. 477–485, June 2013.
19. M. Subotic, and M. Tuba, "Parallelized Multiple Swarm Artificial Bee Colony Algorithm (MS-ABC) for Global Optimization," Studies in Informatics and Control, vol. 23, no. 1, pp. 117-126, 2014.
20. N. Bacanin, M. Tuba,I. Strumberger, " RFID Network Planning by ABC Algorithm Hybridized with Heuristic for Initial Number and Locations of Readers", 2015 17th UKSIM-AMSS International Conference on Modeling and Simulation, IEEE DOI 10.1109/UKSim, 2015.
21. N. Bacanin, M. Tuba, "Firefly Algorithm for Cardinality Constrained Mean-Variance Portfolio Optimization Problem with Entropy Diversity Constraint", The Scientific World Journal, Volume 2014, Article ID 721521, Hindawi Publishing Corporation, May 2014.
22. P. Shoval, V. Maidel, B. Shapira, and M. Taieb-Maimon,"Ontological content-based filtering for personalized newspapers: A method and its evaluation", Online Information Review, 2018, Vol. 34, No. 5, pp. 729-756.
23. Plassmann H., Ramsøy TZ., Milosavljevic M., Branding the brain: A critical review and outlook, Journal of Consumer Psychology, Vol. 22, 2012, pp. 18-36.
24. R. Jovanovic and M. Tuba, "An ant colony optimization algorithm with improved pheromone correction strategy for the minimum weight vertex cover problem", Applied Soft Computing, vol. 11, issue 8, pp. 5360–5366, Dec 2011.
25. R. Jovanovic and M. Tuba, "Ant colony optimization algorithm with pheromone correction strategy for the minimum connected dominating set problem", Computer Science and Information Systems (ComSIS), vol. 10, no. 1, pp. 133–149, Jan 2013..

26. Sharma, N., Patterson, PG., The impact of communication effectiveness and service quality on relationship commitment in consumer, professional services, *Journal of Services Marketing*, Vol. 13, 1999, pp. 151-170.
27. Shivani Yadao, A. Vinaya Babu, Midhunchakkavarthy Janarthanan, and Amiya Bhaumik, "A Combined Approach of Web Content Mining and Neural Networks for Predicting the societal impact of covid -19 through twitter", *Journal of Pharmaceutical Negative Results*, pp. 58–74, Oct. in 2022
28. T. Nguyen, D. Phung, B. Adams, T. Tran, and S. Venkatesh, "Classification and Pattern Discovery of Mood in Weblogs ", *Advances in Knowledge Discovery and Data Mining*, 2018, Vol. 6119, pp. 283-290.
29. Tankosic, M., Influence of Media to Social Behavior and the Changes that have been Caused by New Media, *Proceedings of the 5th International Conference on Applied Economics, Business and Development (AEBD '13)*, Chania, Crete Island, Greece, 2013, pp. 143-148.
30. Tankosic, M., Grbic A., Krivokapic Z., "The Marginalization and Exploitation of Women in Media Industry", Globalization and Its Impact on Violence Against Vulnerable Groups, IGI Global publisher of timely knowledge, 2020, pp.75 - 94
31. Tankosic, M., Trifunovic, D., Development of Cultural Institutions through Implementation of Service Innovation, *Proceedings of the 5th International Conference on Applied Economics, Business and Development (AEBD '13)*, Chania, Crete Island, Greece, 2013, pp. 161-166.
32. Tankosic, M., Ivetic, P., Mikelic, K., Managing internal and external Communication in a Competitive Climate via EDI concept, *International Journal of Communications, COPE Ethics Guidelines*, Vol. 2, 2017, pp. 1 - 6, 2367-8887.
33. X.-S. Yang, *Nature-Inspired Metaheuristic Algorithms*, Luniver Press, 2008.
34. XS Yang, " Multiobjective firefly algorithm for continuous optimization", *Engineering with Computers*, Vol. 29, No. 2, pp. 175-184, March 2017.
35. Y. Shang and L. Li, "Precision evaluation of search engines," *World Wide Web: Internet and Web Information Systems*, vol. 5, pp. 159–173, 2013
36. Yan-fei Zhu Overview of swarm intelligence Computer Application and System Modeling (ICCASM), 2010 International Conference on Taiyuan

Datum prijema (Date received): 23.11.2022.

Datum prihvatanja (Date accepted): 29.12.2022.