

# Optimizacija web mjesta za potrebe CRM-a

---

Mudri, Lorena

Undergraduate thesis / Završni rad

2021

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **University of Zagreb, Faculty of Organization and Informatics / Sveučilište u Zagrebu, Fakultet organizacije i informatike**

Permanent link / Trajna poveznica: <https://um.nsk.hr/um:nbn:hr:211:377782>

Rights / Prava: [Attribution-NonCommercial-NoDerivs 3.0 Unported / Imenovanje-Nekomercijalno-Bez prerada 3.0](#)

Download date / Datum preuzimanja: **2024-07-03**



Repository / Repozitorij:

[Faculty of Organization and Informatics - Digital Repository](#)



**SVEUČILIŠTE U ZAGREBU  
FAKULTET ORGANIZACIJE I INFORMATIKE  
VARAŽDIN**

**Lorena Mudri**

**OPTIMIZACIJA WEB MJESTA ZA  
POTREBE CRM-a**

**ZAVRŠNI RAD**

**Varaždin, 2021.**

**SVEUČILIŠTE U ZAGREBU**  
**FAKULTET ORGANIZACIJE I INFORMATIKE**  
**V A R A Ž D I N**

**Lorena Mudri**

**Matični broj: 0016135907**

**Studij: *Ekonomika poduzetništva***

**OPTIMIZACIJA WEB MJESTA ZA POTREBE CRM-a**

**ZAVRŠNI RAD**

**Mentor:**

Izv. prof. dr. sc. Sandro Gerić

**Varaždin, rujan 2021.**

*Lorena Mudri*

### **Izjava o izvornosti**

Izjavljujem da je moj završni rad izvorni rezultat mojeg rada te da se u izradi istoga nisam koristila drugim izvorima osim onima koji su u njemu navedeni. Za izradu rada su korištene etički prikladne i prihvatljive metode i tehnike rada.

*Autorica potvrdila prihvaćanjem odredbi u sustavu FOI-radovi*

---

## **Predgovor**

Zahvaljujem svom mentoru, izv. prof. dr. sc. Sandru Geriću na ukazanom povjerenju, savjetima i pruženoj pomoći prilikom izrade mog završnog rada.

Također se zahvaljujem svim profesorima Fakulteta organizacije i informatike na prenesenom znanju. Veliko hvala svim kolegama na pomoći i potpori tijekom studiranja. Od srca se zahvaljujem svojoj obitelji i dečku na podršci koju su mi pružali tijekom studija.

## Sažetak

Tema ovog završnog rada je optimizacija web mjesta za potrebe CRM-a. Gotovo je nemoguće zamisliti svakidašnji život bez pristupa Internetu. On nam omogućuje pravovremeno informiranje, istraživanje, obavljanje kupnje, kao i komunikaciju s ljudima. S ubrzanim razvojem interneta kao i same tehnologije, mijenjali su se i marketing i načini upravljanja odnosima s kupcima u svrhu prilagođavanja proizvoda istima te poboljšanju poslovanja. Web mjesto u tom aspektu igra veliku ulogu s obzirom na to da mogućnost pristupa web mjestu organizacije toj organizaciji daje prednost i to joj može pružiti priliku da bude u bliskom kontaktu s kupcima uočavajući njihove potrebe, a krajnji je cilj zadovoljavanje istih. Poznato je da, primjerice, kupnji putem interneta prethodi upisivanje željenog pojma u tražilicu web preglednika, pa će u radu biti objašnjen princip funkcioniranja web mjesta te web tražilica. Također je naglasak na promociji koja se postiže oglašavanjem na web tražilicama te optimizacijom web mjesta. U ovom će se radu objasniti i koncept Velikih podataka, čija analiza ima veliku ulogu u marketingu za tražilice i općenito optimizaciji web mjesta.

Središnji dio rada predstavljaju načini optimizacije web mjesta, a to su optimizacija za Internet pretraživače, marketing za Internet pretraživače te tzv. *Trusted feed*. Također će se navesti i alati koji pomažu kod optimizacije web mjesta, koja je ostvariva tek kad se prepoznaju kupci, uoče njihove želje i potrebe temeljene na njihovom razmišljanju. Tada su potrebne analize i praćenja web mjesta te promišljeno i pravovremeno donošenje ispravnih odluka i zaključaka o daljnjim postupcima vezanim uz cilj, a to je optimizirati, odnosno poboljšati web mjesto kako bi se pružili optimalni proizvodi ili usluge na zadovoljstvo obje strane. Zaključku ovog rada prethode problemi koji se javljaju prilikom optimizacije web mjesta.

**Ključne riječi:** optimizacija, web mjesto, web tražilice, algoritmi, indeksiranje, rangiranje, *PageRank*, *CT-rank*, *Big data*, CRISP-DM, SEO, *On-page SEO*, *Off-page SEO*, *SERP*, *SEM*, *Trusted feed*, *Google Analytics*, *Google Search Console*, *Google Tag Manager*, ključne riječi, *black hat SEO*, *Sandbox*

# Sadržaj

<b>1. Uvod</b> .....	1
<b>2. Web mjesto</b> .....	2
2.1. Vrste web mjesta .....	2
2.2. Dimenzije web mjesta.....	3
2.3. Razlike između web mjesta i web stranice.....	3
<b>3. Web tražilice</b> .....	5
3.1. Funkcioniranje web tražilica.....	6
3.1.1. Pretraživanje web mjesta.....	6
3.1.2. Indeksiranje sadržaja.....	6
3.1.3. Rangiranje web mjesta.....	9
3.1.3.1. Tradicionalni algoritmi rangiranja.....	9
3.1.3.2. Algoritam rangiranja povezan s vremenom.....	14
<b>4. Koncept Velikih podataka</b> .....	19
4.1. Karakteristike Velikih podataka.....	19
4.2. Tehnike i proces analize Velikih podataka.....	21
4.3. Alati za analizu Velikih podataka.....	24
4.4. Prednosti i nedostaci koncepta Velikih podataka.....	27
<b>5. Načini optimizacije web mjesta</b> .....	28
5.1. Optimizacija za Internet pretraživače.....	28
5.1.1. Vrste optimizacije za Internet pretraživače.....	29
5.1.2. Važnost plana za optimizaciju za Internet pretraživače.....	32
5.1.3. Postupak optimizacije za Internet pretraživače.....	33
5.1.4. Odabir ključnih riječi.....	34
5.2. Marketing za Internet pretraživače.....	36
5.2.1. Uvod u marketing za Internet pretraživače.....	36
5.2.2. Način funkcioniranja marketinga za Internet pretraživače.....	37
5.2.3. Isticanje oglasa.....	39
5.3. <i>Trusted feed</i> .....	41

<b>6. Alati za optimizaciju web mjesta za potrebe CRM-a</b> .....	42
6.1. <i>Google Analytics</i> .....	42
6.2. <i>Google Search Console</i> .....	45
6.3. <i>Google Tag Manager</i> .....	49
6.4. <i>Free Monitor for Google</i> .....	51
6.5. <i>Word Tracker</i> .....	52
6.6. <i>Keyword Discovery</i> .....	53
<b>7. Problemi kod optimizacije</b> .....	54
7.1. Konkurencija.....	54
7.2. <i>Black hat SEO</i> .....	54
7.3. <i>Sandbox</i> .....	55
<b>8. Zaključak</b> .....	57
<b>Popis literature</b> .....	58
<b>Popis slika</b> .....	64
<b>Popis tablica</b> .....	66



# 1. Uvod

Tema ovog završnog rada je Optimizacija web mjesta za potrebe CRM-a. Opće je poznato da Internet posljednjih godina ima veliku važnost kao komunikacijski kanal te alat koji svoju svrhu nalazi u ostvarivanju marketinških ciljeva. U okviru ostvarivanja tih ciljeva potrebno je uspješno upravljati odnosima s kupcima, odnosno prepoznati njihove želje i potrebe, prikupljati podatke o istima te ih analizirati u svrhu poboljšanja poslovanja. Gotovo je nemoguće zamisliti organizaciju koja nije dostupna javnosti putem Interneta. Ta dostupnost zapravo predstavlja prednost kao i konkurentnost na tržištu jer daje mogućnost organizaciji da poboljša svoje poslovanje i približi se kupcima. Web mjesto, koje će biti objašnjeno na početku ovog završnog rada, kupcima daje uvid u organizaciju kao i organizaciji uvid u interese kupaca, što se ostvaruje upravo u okviru CRM-a. Također će biti objašnjeno pretraživanje web mjesta koje se obavlja putem tražilica, koje na temelju korisničkih upita prikazuju web mjesta koja odgovaraju tim željama. Kako bi web mjesto neke organizacije bilo relevantno, važno je da se ono prilagodi tražilicama (eng. *SEO*). Često se pod *SEO*-om percipira tehnika dobrog pozicioniranja na *Google*-u, no u principu to nije tako, jer se pod time smatra prikupljanje te obrada prikupljenih informacija, a naposljetku i primjena tehnika za postizanje željene pozicije na tražilicama, a to se sve provodi u svrhu CRM-a. Biti će objašnjeni načini optimizacije, odnosno optimizacija za Internet pretraživače, eng. *SEO*, marketing za Internet pretraživače, eng. *SEM*, i *Trusted feed*, kao i koncept Velikih podataka koji omogućuje poboljšanje optimizacije korištenjem tehnika analize podataka. Također će biti navedeni i alati koji pružaju pomoć kod optimizacije web mjesta, ali i problemi s kojima se prilikom optimizacije web mjesta stručnjaci često susreću.

S obzirom na ubrzani tempo života današnjice, sve se više ljudi okreće kupnji putem Interneta iz razloga što je brza, jednostavna i može se obavljati iz udobnosti vlastitog doma, dakle nije potrebno putovati i hodati po trgovinama te trošiti energiju kao i gorivo, a kao što je već navedeno, pruža i uštedu vremena. Svakim upisivanjem željenog pojma u tražilicu i otvaranjem određenog web mjesta neke organizacije čiji smo potencijalni kupci, dajemo toj određenoj organizaciji mogućnost da nas privuče sadržajem na tom web mjestu. Upravo je zato bitno da organizacija uloži trud i napore usmjerene optimizaciji web mjesta i privuče čim više kupaca te ih učini zadovoljnim, lojalnim kupcima koji će podizati kredibilitet i ugled organizacije.

Optimizacija web mjesta za potrebe CRM-a zanimljiva je tema jer daje uvid u načine poboljšanja samog poslovanja organizacija, neovisno o njihovoj veličini, ako se uloži dovoljno truda i marljivog rada na tehnikama navedenim u ovom završnom radu.

## 2. Web mjesto

„Web mjesto ili web sjedište (eng. *website*) zbirka je međusobno povezanih web stranica različitog sadržaja (tekst, slike, video, ...) dostupna putem Internet ili lokalne mreže na određenoj Internet adresi ili URL-u.“ (Dobrinić D. i sur., 2011., str. 278.)

Dobrinić D. i sur. (2011) s obzirom na funkciju smatraju da web mjesta mogu biti osobna, komercijalna, web mjesta neprofitnih organizacija, vladina web mjesta itd.

S obzirom na način njihova ostvarivanja web mjesta mogu biti statička, odnosno temeljena na HTML ili statičkim web stranicama i dinamička, odnosno temeljena na skriptnim jezicima poput ASP, PHP koji generiraju dinamičke web stranice. (Dobrinić D. i sur., 2011., str. 278.)

**Statička web mjesta** sadržavaju predefinirane, statičke informacije, tj. informacije koje se ne mijenjaju često. Primjer takvih web mjesta su razne brošure, odnosno web mjesta „od pet stranica“, koja su pogodna za manja poduzeća zbog njihovog jednostavnog i jeftinog postavljanja, no ažuriranje zahtijeva veće napore. (Dobrinić D. i sur., 2011., str. 278.)

**Dinamička web mjesta** su dinamičkog sadržaja i kod njihove izrade potrebno je uložiti veće napore, kao i vrijeme i financijska sredstva, u odnosu na statička web mjesta, no kasnije su lakša za održavanje i ažuriranje. (Dobrinić D. i sur., 2011., str. 278.)

### 2.1. Vrste web mjesta

Prema Chafney i sur. (kao što citiraju Dobrinić i sur.) postoje četiri tipa web mjesta s obzirom na naglašeniju ulogu marketinga.

Prvi tip je **transakcijsko web mjesto** za elektroničku trgovinu, odnosno web mjesto koje omogućuje naručivanje proizvoda *online* i to web mjesto pridonosi poslovanju prodajom proizvoda, a u isto vrijeme je i izvor informacija o kupcima koji preferiraju naručivanje proizvoda putem Interneta.

Drugi tip je **web mjesto orijentirano uslugama i izgradnji odnosa** i ono pruža informacije o poticanju na kupnju te izgradnju odnosa s kupcima.

Treći tip je **web mjesto za izgradnju *branda*** i ono osigurava/pruža iskustva kako bi se potpomognuo *brand* proizvoda.

Četvrti i posljednji tip web mjesta je **portal ili medijsko web mjesto** koje pruža informacije ili novosti o nizu tema na samom web mjestu ili putem poveznice na druga web

mjesta te ima mogućnost generiranja prihoda od oglašavanja, prodaje korisničkih podataka i slično. (Dobrinić D. i sur., 2011., str. 278.)

## 2.2. Dimenzije web kvalitete

Prema Aladwani i Palvia (kao što citiraju Dobrinić i sur. ) kriteriji kvalitete svrstani su u tri glavne dimenzije.

Prva dimenzija tiče se **tehničke prikladnosti** (eng. *technical adequacy*). Primjeri stavki koje spadaju pod ovu dimenziju su sigurnost, lakoća navigacije po web mjestu, usluge prijenosa, ograničeno korištenje specijalnih dodataka (eng. *plug-inova*), mogućnost pretraživanja, raspoloživost, važeće poveznice, brzo učitavanje web stranice, višejezična podrška i slično. (Dobrinić D. i sur., 2011., str. 280.)

Druga dimenzija je **dimenzija web sadržaja** i ona je usko vezana uz korisnost sadržaja, cjelokupnost sadržaja, jasnoću navedenog sadržaja, njegovu jedinstvenost, širinu, točnost, izvornost i vrijednost, kao i pronalaženje održavatelja web mjesta te pronalaženje poveznica na relevantna web mjesta i slično. (Dobrinić D. i sur., 2011., str. 280.)

Treća dimenzija vezana je uz **izgled web mjesta**. Ona naglašava važnost atraktivnog izgleda, funkcionalnosti i organiziranosti samog web mjesta. Važno je da se prikladno koristi font, kao i boje i slike. Naglasak je i na uravnoteženosti količine slika i teksta, prikladnosti korištenja multimedije te dosljednosti stila. (Dobrinić D. i sur., 2011., str. 280.)

## 2.3. Razlike između web mjesta i web stranice

Kad se dođe na temu pretraživanja Interneta, pojmovi web mjesto i web stranica su u širokoj upotrebi i važno je navesti ključne karakteristike po kojima se razlikuju.

Iako je u ovom poglavlju rada već navedena definicija web mjesta, za lakše razumijevanje će biti navedena još jednom. Web mjesto (eng. *website*) je skup povezanih web stranica obično serviranih s jedne domene (Internetske adrese). S druge strane, web stranica (eng. *webpage*) predstavlja dokument koji je napisan u obliku običnog teksta s instrukcijama formatiranja koda zvanog *Hypertext Markup Language* ili *Extended HTML*. („Web stranica: medij za promociju i kredibilnost brenda“, 2016).

Ključne razlike web mjesta i web stranica su navedene u nastavku.

Web stranica je manji dio web mjesta koji sadrži određen sadržaj i poveznice na druge web stranice, dok je web mjesto platforma s mnogo web stranica adresiranih određenim URL-om (eng. *Uniform Resource Locators*).

Svako web mjesto uključuje jedinstvenu adresu domene i sve su web stranice povezane pod tom domenom, što znači da su one ovisne o web mjestu, dok web mjesto funkcionira neovisno o web stranicama i ako se iste izbrišu.

Web stranice uobičajeno sadrže dodatak u svojim URL adresama, primjerice .html, .php., .aspx., .htm i slično, dok web mjesto takve dodatke ne sadrži.

Razvoj, dizajn i održavanje web mjesta složenije je i iziskuje više vremena od istih aktivnosti za web stranice, sukladno odnosu web mjesta i web stranice.

Strukturiranje i programiranje web mjesta također je kompleksnije od strukturiranja i programiranja web stranica, jer je razvoj web stranica relativno lak proces kad je web mjesto već strukturirano.

Pristup web mjestu moguć je direktno putem URL linka ili putem web mjesta, dok se web mjestu može pristupiti putem adrese domene. („Gadget-info.com“, bez dat.)

Primjeri web mjesta su Amazon.com, JavaTpoint.com, a primjeri web stranica na tim web mjestima su Kontakt, Registracija i Prijava. („Gadget-info.com“, bez dat.)

Kako bi se razumio princip funkcioniranja i uloge web mjesta i web stranica na Internetu i generalno u marketingu, važno je znati njihova obilježja.

### 3. Web tražilice

S obzirom na to da Internet predstavlja velik dio naše svakodnevice, u svakom trenutku može se u vrlo kratkom roku doći do željene informacije tako da se u web tražilicu upiše odgovarajući pojam.

S korisničkog gledišta, web tražilica (eng. *Search engine*) je web stranica koja omogućuje upisivanje fraza ili ključnih riječi i potom daje rezultate ovisno o relevantnosti dokumenata koje sadrži u svom indeksu, a ta je relevantnost određena automatiziranim programima, odnosno algoritmima. (George, 2005., str.2)

Ružić i sur. (2014.) smatraju da su tražilice specijalizirani sustavi koji prikupljaju informacije o internetskim resursima te omogućuju pretraživanje prikupljenih informacija.

Individualni su pretraživači baze podataka web mjesta koje sastavljaju roboti, odnosno programi koji pretražuju i indeksiraju web mjesta, a razlikuju se s obzirom na brzinu, veličinu baze podataka, opcijama pretraživanja, prikazu rezultata i dr. Najvažnija je njihova osobina algoritam rangiranja rezultata pretraživanja, što predstavlja svojevrsnu poslovnu tajnu i preduvjet uspješnosti pretraživača. (Dobrinić i sur., 2011., str. 292.)

Kako bi se razumjela svrha web tražilica, važno je razumjeti način na koji iste funkcioniraju. Svaka od njih koristi vlastiti softver, no sve provode tri osnovna zadatka. Prvo prouče sadržaj pretraživanjem radi indeksiranja, potom kategoriziraju dijelove tog sadržaja, što predstavlja indeksiranje i potom odlučuju koji je sadržaj najkorisniji i najrelevantniji za korisnika, što predstavlja rangiranje. (Google Digitalna Garaža, 2021.)

Kako bi se tražilice probile do web mjesta, koriste alate za indeksiranje, tzv. *botove* (skraćeno od robota), koji prelaze s jednog web mjesta na drugo slijedeći poveznice (eng. *links*) web mjesta. Potom slijedi kategoriziranje sadržaja, odnosno indeksiranje, a kako bi se razumio princip indeksiranja, važno je razumjeti značenje samog pojma indeks, a to je popis web mjesta i sadržaja koje *botovi* nađu. Cilj rangiranja, koje nastupa nakon indeksiranja, jest pružiti korisniku najrelevantnije rezultate koji se podudaraju s upitom koji je korisnik unio u web tražilicu. (Google Digitalna Garaža, 2021.)

U nastavku rada detaljnije će se objasniti princip funkcioniranja web tražilica, jer je razumijevanje tog postupka vrlo korisno za poslovanje, obzirom na to da je za poboljšanje poslovanja važno imati što bolju rangiranost, kako bi se doprlo do čim većeg broja klijenata.

## 3.1. Funkcioniranje web tražilica

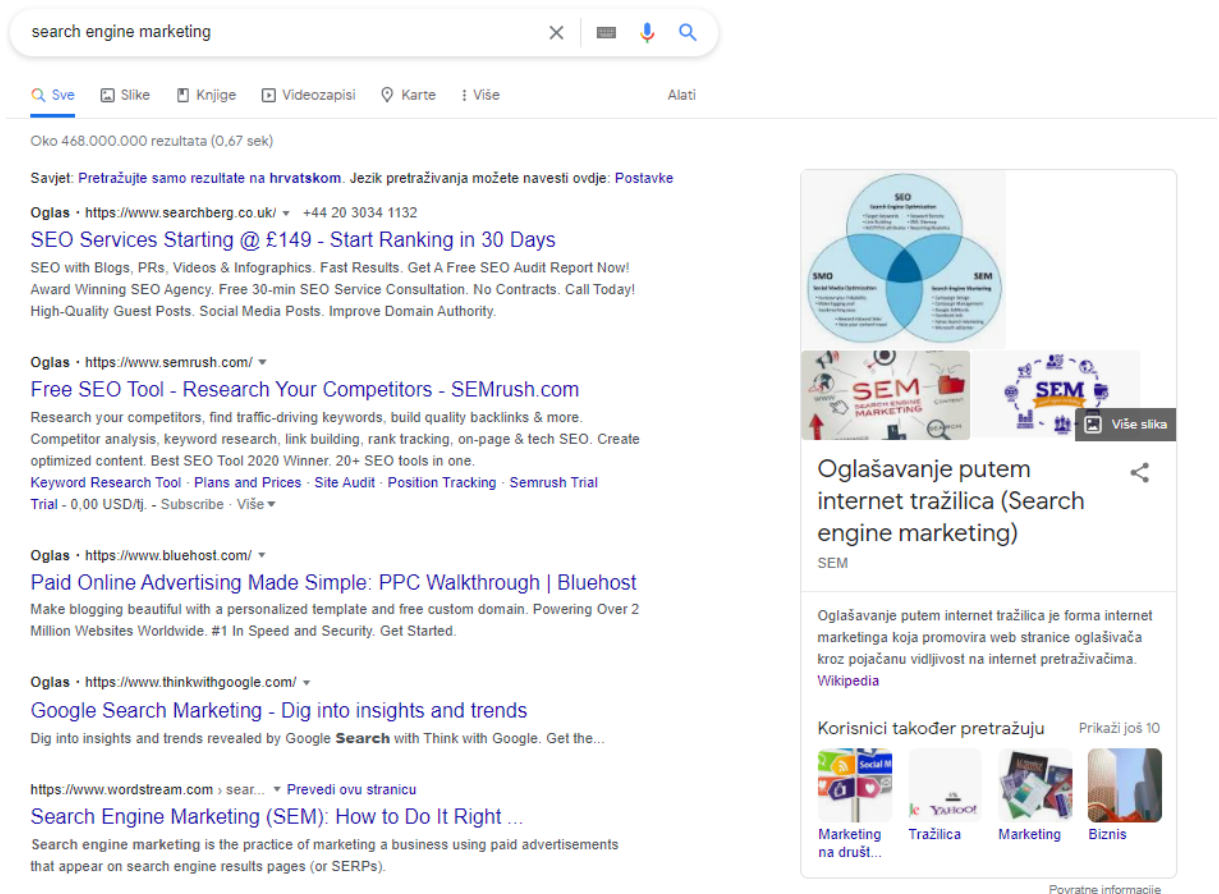
Web tražilice predstavljaju vrata Interneta, s obzirom na to da gotovo svakom posjetu određenom web mjestu prethodi upisivanje određenog pojma u neku od web tražilica. Kako bi web tražilice prikupile podatke koje korisnici traže i prikazale ih, one trebaju razumjeti web mjesta koja te podatke pružaju.

### 3.1.1. Pretraživanje web mjesta

Za pretraživanje web mjesta, web tražilice koriste posebne softvere, već navedene *botove*. Primjerice, *Google* koristi *Googlebotove*, koji slijede poveznice i pretražuju informacije na web mjestima, počevši na mapi weba koju administratori uobičajeno pošalju web tražilicama. Pomoću tih informacija, *Google* određuje koliko se često web mjesta ažuriraju, pa sukladno tome određuje i učestalost posjećivanja i ponovnog pretraživanja tih web mjesta. Na taj se način osigurava da *Google* indeksira nova web mjesta i ažurira već postojeća. Učestalost pretraživanja nije strogo određena i ona varira od nekoliko puta dnevno pa sve do jednom u nekoliko mjeseci godišnje. (Krušec, 2017)

### 3.1.2. Indeksiranje sadržaja

Korištenjem web pretraživača ili tražilice, korisnici ne pretražuju cijeli web, već samo bazu, koja se sastoji od web mjesta koja je web pretraživač kategorizirao i pohranio, s obzirom na to da bi pretraživanje cijelog weba predugo trajalo zbog prevelikog broja postojećih web mjesta. Za primjer će u *Google* tražilicu biti upisan pojam „*search engine marketing*“ i na slici 1 prikazani su rezultati tog pretraživanja. (Slika 1.)



Slika 1: Prva stranica rezultata pretraživanja pojma „search engine marketing“, prikazana na prijenosnom računalu

Na slici iznad vidljivo je da rezultat pretraživanja (eng. *Search Engine Result Page*) pojma „search engine“ ne predstavlja rezultat pretrage cijelog weba, već trenutne baze tražilice. U vremenu od 0,67 sekundi pronađeno je oko 468.000.000 rezultata.

U okviru rezultata pretrage, razlikuju se dvije vrste rezultata: organski i plaćeni rezultati. Organski rezultat je rezultat koji je web tražilica pronašla na temelju algoritama pretraživanja, dok je plaćeni rezultat onaj kojeg je određena tvrtka platila kako bi se on prikazivao u rezultatima pretraživanja, a najlakše ga je prepoznati po pojmu „Oglas“ koji stoji iznad određenog web mjesta.

Svaki od rezultata pretrage predstavlja jedno web mjesto, odnosno jedan SERP isječak (eng. *SERP Snippet*). Svaki od tih rezultata sastoji se od 3 dijela. To su naslov (1), URL (2) i isječak teksta (3). (Slika 2.)

Oglas · <https://www.semrush.com/> ▾ 2

1 **Free SEO Tool - Research Your Competitors - SEMrush.com**

3 Research your competitors, find traffic-driving keywords, build quality backlinks & more.  
Competitor analysis, keyword research, link building, rank tracking, on-page & tech SEO. Create optimized content. Best SEO Tool 2020 Winner. 20+ SEO tools in one.  
[Keyword Research Tool](#) · [Plans and Prices](#) · [Site Audit](#) · [Position Tracking](#) · [Semrush Trial](#)  
[Trial - 0,00 USD/tj.](#) - [Subscribe](#) · [Više](#) ▾

Slika 2: Jedan SERP isječak, prikazan na prijenosnom računalu

Sljedeće pitanje koje se postavlja je na koji se način web mjesta pronadu i prikazuju u rezultatima pretraživanja.

Web tražilice sastoje se od modula neovisnih o korisnicima i njihovim upitima, a to su šunjalo, spremište stranica, modul za indeksiranje, indeksi i moduli ovisni o upitima, a to su modul upita i modul rangiranja.

Prethodno u radu već spomenuta šunjala, (eng. *crawlers*) kreiraju virtualne robote koji su kontinuirano prisutni na webu s ciljem prikupljanja novih informacija i web mjesta, koje potom privremeno pohranjuju u centralnom spremištu (eng. *page repository*).

Web mjesta ostaju u centralnom spremištu sve do trenutka njihova slanja modulu za indeksiranje (eng. *indexing module*). Taj modul generira ključne informacije kako bi se kreirala komprimirana, odnosno sažeta verzija određenog web mjesta. Prilikom ažuriranja web mjesta, modul za indeksiranje ponovo preuzima cjelovito web mjesto i ažurira sažetak istog, izdvajajući ključne deskriptore i pohranjuje ga u različite indekse.

Indeksi (eng. *indexes*) se sastoje od sažetih informacija o svakom web mjestu, a postoje tri tipa.

Indeks sadržaja (eng. *content index*) omogućuje pohranu sadržaja u sažetom obliku u obrnuti indeks koji djeluje kao kazalo. S obzirom na veličinu Weba, broj web mjesta koja sadrže popularne termine je velik, pa je sukladno tome i datoteka obrnutog indeksa velika.

Indeks strukture (eng. *structure index*), sadrži strukturu *hiperlinkova* stranica, dok indeks specijalne namjene (eng. *special-purpose index*) sadrži specifični kontekst, primjerice slike, pdf datoteke i slično. (Dobrinić i sur., 2011., str. 293-294.)



### 3.1.3. Rangiranje web mjesta

Svaka web tražilica koristi svoj algoritam za rangiranje web mjesta koji predstavlja poslovnu tajnu, no poznato je koji faktori utječu na rangiranje.

Prema Ružić i sur. (2014.) većina tražilica prilikom rangiranja u obzir uzima sljedeće elemente, odnosno faktore:

- Naziv web mjesta
- Glavni naslov
- Cijeli tekst na web mjestu
- Prvi redak svih odlomaka
- Naslovi slika i grafičkih prikaza
- Podnaslovi
- Ključni elementi u meta oznakama (eng. *meta tags*)
- Broj ponavljanja ključnih riječi

#### 3.1.3.1. Tradicionalni algoritmi rangiranja

Horvat i Mundžar (2017) navode da su prve tražilice devedesetih godina funkcionirale na način da su prolaskom kroz indeksirani popis web mjesta brojale koliko se puta uneseni pojam pojavljuje na svakom od web mjesta pa su se, sukladno tome, korisniku prikazivala ona web mjesta koja su sadržavala najveći broj pojavljivanja određenog pojma koji je korisnik unio u web tražilicu. Međutim, pojavio se problem jer se na web mjestu moglo nalaziti bezbroj ponavljanja istog pojma i nijedna korisna informacija o tom pojmu, koja bi mogla pomoći korisniku u potrazi za određenim informacijama. No, vremena su se promijenila i danas web tražilice koriste drugačiji pristup, uzimajući u obzir relevantnost, važnost ili popularnost web mjesta. Slijedeći taj princip, web tražilica stavlja na vrh popisa rezultata pretraživanja relevantnija web mjesta, uz pomoć algoritma rangiranja.

Poznati algoritmi rangiranja koji se temelje na analizi linkova s jednog web mjesta na drugo su **PageRank algoritam** koji rangira web mjesta po povezanosti neovisno od upita i **HITS algoritam**, koji rangira web mjesta po povezanosti ovisno o upitu, a koji će biti opisani u nastavku rada.

**PageRank algoritam** za rangiranje web stranica razvijen je od strane Larryja Page i Sergeja Brina na sveučilištu Stanford. Taj algoritam predstavlja funkciju količine i snage dolaznih poveznica te procjenjuje vjerojatnost da će određenom web mjestu pristupiti korisnik koji pretražuje Internet i prati poveznice s jednog web mjesta na drugo. Njegovo funkcioniranje se temelji na relativnoj mjeri popularnosti poveznica neovisno o upitu korisnika u pretraživaču, što znači da se, neovisno o upitu svakoj stranici pridružuje neka vrijednost radi mjerenja kvalitete same stranice te se u trenutku upita ta vrijednost koristi sa ili bez kriterija ovisnog o upitu, s ciljem rangiranja dokumenata ili web mjesta koja odgovaraju upitu. Vrijednost algoritma svojevrsno predstavlja i ugled web stranice proporcionalan sumi pokazatelja ugleda web stranica koje *linkaju*, odnosno ukazuju na određenu web stranicu. Vrijednost *PageRank*-a ( $PR(D_i)$ ) se za promatranu web stranicu  $D_i$  izračunava na sljedeći način:

$$PR(D_i) = (1-d) + d [(PR(D_1)/C(D_1) + \dots + (PR(D_m) / C(D_m))],$$

Gdje su  $D_1, D_2, \dots, D_m$  stranice koje ukazuju na određenu web stranicu,  $d$  označava parametar, a  $C(D_m)$  broj vanjskih poveznica na stranicu  $D_m$ . (Dobrinić D. i sur., 2011., str. 295.)

To znači da su neki linkovi „jači“ od drugih jer stranica s većim *PageRankom* ima više šansi da bude posjećena. Drugim riječima, taj se algoritam bazira na pretpostavci da je relevantno web mjesto svako na koje se referira velik broj ostalih, više ili manjih, relevantnih web mjesta, odnosno, da su neki linkovi „jači“ od drugih jer web mjesta s većim *PageRankom* imaju veće vjerojatnosti da budu posjećena, što je sasvim logično, s obzirom na to da nijedno ozbiljno web mjesto neće postaviti poveznicu na drugo nekvalitetno ili neprovjereno web mjesto. (Horvat i Mundžar, 2017.)

**Kleinbergov HITS algoritam** otkriven je krajem 90-ih godina, a funkcionira na način da procjenjuje kvalitetu web mjesta temeljeno na dva numerička faktora, vrijednosti autoriteta (eng. *authority*) i vrijednosti središta (eng. *hub score*) za svaki čvor iz grada susjeda, što predstavlja set stranica koje poveznicama pokazuju na web mjesto koje se prikazuje kao rezultat upita. Autoritet sadržaja nekog web mjesta određen je brojem referenci, odnosno poveznica s drugih web mjesta, dok je autoritet poveznica, odnosno vrijednost središta određen kvalitetom poveznica, odnosno, visokom autoritetu sadržaja doprinosi često referenciranje na drugim web mjestima, dok visokom autoritetu poveznica doprinosi visoka kvaliteta drugih web mjesta koja referenciraju određeno web mjesto. (Hong, Lihua, Xiaowen, Peiquan, 2010., str. 100)

Po tom se principu dokumenti rangiraju prema obje vrijednosti, pa se za dokumente s visokim autoritetom smatra da imaju relevantan sadržaj, a za dokumente središta da imaju poveznice na relevantne dokumente. Drugim riječima, za one se dokumente ili web mjesta koja pokazuju na mnoge druge smatra da su dobra središta, dok se za one koje pokazuju mnogi drugi dokumenti ili web mjesta, smatra da su dobri autoriteti. (Macarol, 2017.)

Na sljedećem je primjeru prikazan proces po kojem *HITS* algoritam računa vrijednosti središta i autoritete, kako navodi Macarol.

Algoritam za izračun vrijednost središta i vrijednosti autoriteta:

1. Svaki čvor počinje s obje vrijednosti postavljene na 1, odnosno,  
 $Auth(p) = 1$   
 $Hub(p) = 1$
2. Autoritet svakog čvora ažurira se tako da je jednak zbroju svih vrijednosti kao središta čvorova koji poveznicama šalju korisnika na njega.

$$Auth(p) = \sum_{i=1}^n hub(i)$$

3. Vrijednost središta ažurira se tako da je jednaka zbroju autoriteta svih čvorova na koje šalje poveznicama.

$$Hub(p) = \sum_{i=1}^n auth(i)$$

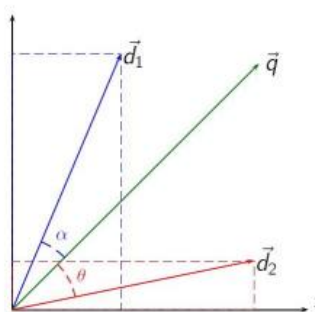
4. Vrijednosti je potrebno normalizirati
5. Proces se ponavlja od drugog koraka prema potrebi  $n$  puta. (Macarol, 2017.)

Drugi tip tradicionalnih algoritama rangiranja određen je u okviru *online* modela rangiranja, odnosno rangiranja koje se događa prilikom obrade upita kojeg je korisnik uputio web tražilici.

**Algoritam po Booleovom modelu** predstavlja jednu od najranijih i najjednostavnijih tehnika za pretraživanje informacija, a ime je dobio po Booleovoj algebri. George Boole bio je britanski matematičar koji je svojevremeno stvarao modernu simboličku logiku. Booleova algebra, odnosno, matematička logika omogućuje donošenje zaključka logičkim izvođenjem, tako da zaključak odgovara početnoj premisi ili tvrdnji. (Encyclopaedia Britannica, 2021.)

Kako navodi Macarol, algoritam funkcionira na način da provjerava podudarnost ključnih riječi u korisničkom upitu (unosu u web tražilicu, npr. u *Google*) s web mjestima, pomoću stanja istina (eng. *true*), ako su ključne riječi iz upita prisutne na web mjestu i laž (eng. *false*), ako ključne riječi iz upita nisu prisutne na web mjestu. Algoritam koristi logičke veznike I, ILI i NE, koji logički spajaju riječi i oblikuju upit. Kod naprednih opcija pretraživanja polje „sve ove riječi“ odgovara operatoru I, polje „bilo koja od ovih riječi“ odgovara operatoru ILI i polje „nijedna od ovih riječi“ odgovara operatoru „NE“. Svrhu pronalazi zbog jednostavnosti kreiranja i programiranja pretraživača, brzine procesuiranja upita i primjenjivosti na veće kolekcije dokumenata. (Macarol, 2017.)

Algoritam koji prikazuje upite i web mjesta kao zasebne vektorske veličine jednakih duljina naziva se **algoritam vektorskog prostora** (eng. *Vector Space Model*). Sličnost vektora upita i vektora web mjesta, kod ovog algoritma mjerilo je relevantnosti web mjesta za korisnički upit. Ta se sličnost računa kao kosinus kuta između vektora, odnosno, kosinusna sličnost, što rezultira poretkom dohvaćenih dokumenata (web mjesta) po vjerojatnosti da su relevantni korisničkom upitu. Na slici 3 prikazan je graf vektorskog prostora, na kojem je vidljiv upit  $q$  i dva dokumenta iz kolekcije  $d_1$  i  $d_2$ . Uvažavajući kosinusnu sličnost, može se zaključiti kako dokument  $d_1$  ima veću sličnost s upitom  $q$  od dokumenta  $d_2$ , stoga se  $d_1$  smatra relevantnijim od  $d_2$ . (Slika 3.)



Slika 3: Graf vektorskog prostora (*Vector Space Model: Examples, bez dat.*)

Nedostatak ovog algoritma je njegova stroga orijentiranost odnosu upita i web mjesta, koja se ne može preusmjeriti na ispitivanje međusobnih odnosa između web mjesta, kao i velika količina računalne snage koju iziskuje. (Hong, Lihua, Xiaowen, Peiquan, 2010.,str. 100)

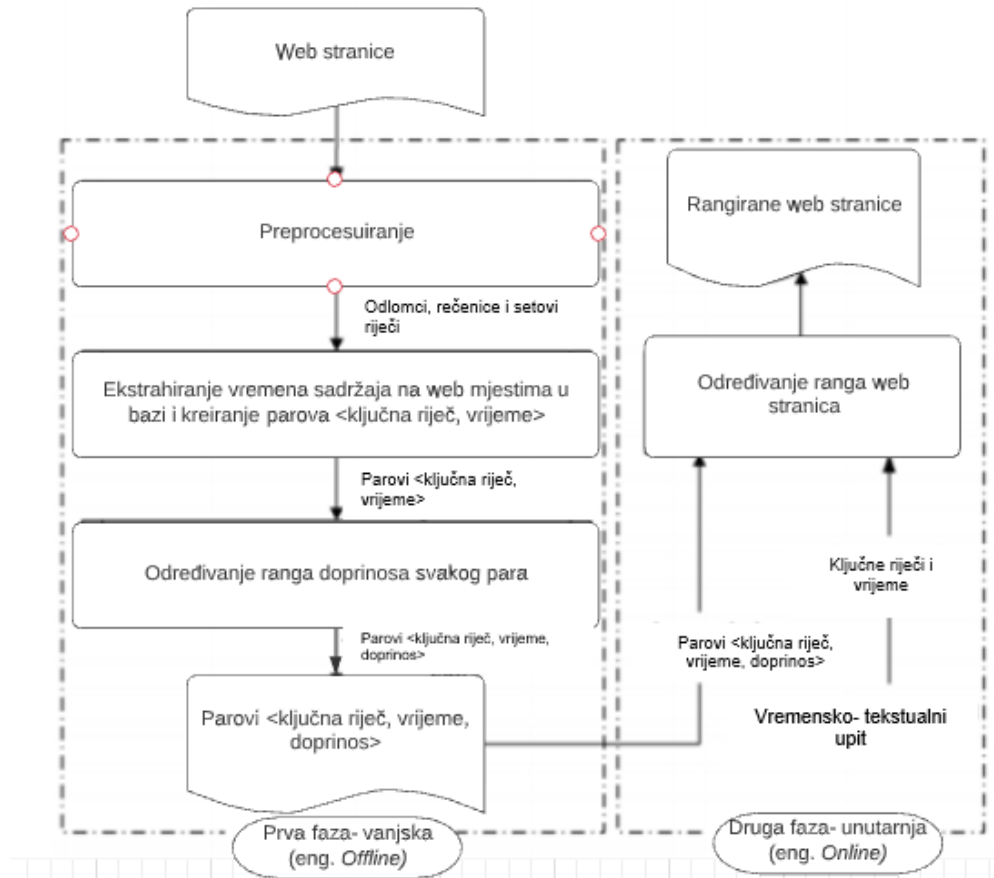
Algoritam koji se nadovezuje na prethodni je **algoritam modela vjerojatnosti** (eng. *Probability Model*) koji rangira web mjesta prema vjerojatnosti relevantnosti između upita i web mjesta. Samim time, mora pretpostaviti početne vrijednosti i onda ih iterativno pokušati poboljšati sve dok ne dobije konačnu rang listu vjerojatnosti da je dokument, odnosno, web mjesto relevantno za korisnički upit. Jedini nedostatak je taj što je algoritam iznimno kompleksan i težak za programiranje. Međutim, omogućuje učenje na temelju korisnikovog ponašanja, što je preduvjet nastanka trećeg tipa algoritama, opisanih u nastavku rada. (Hong, Lihua, Xiaowen, Peiquan, 2010.,str. 100)

Treći tip tradicionalnih algoritama određen je relevantnošću povratne informacije, koji funkcioniraju po principu samoučenja, odnosno, automatskom prilagođavanju upitu korisnika koje se temelji na prethodno dohvaćenim rezultatima. Na taj se način mogu filtrirati rezultati prema korisnosti koju su imali za korisnika, odnosno mogu se izdvojiti oni koji su korisniku bili korisni od onih koji mu nisu bili korisni, no tu je tehniku potrebno koristiti u kombinaciji s drugim algoritmima, jer se time povećava efikasnost pružanja rezultata pretraživanja. (Hong, Lihua, Xiaowen, Peiquan, 2010.,str. 100)

### 3.1.3.2. Algoritam rangiranja povezan s vremenom

Algoritam koji je također zastupljen kod rangiranja web mjesta na web tražilicama je **algoritam rangiranja temeljen na sadržaju i vremenu** (eng. *Content-Time-based Ranking - CT-Rank*). Vrijeme je novac i igra veliku ulogu u ljudskim životima, pa je iz tog razloga vrijeme i uvršteno u rad navedenog algoritma. Na mnogim web mjestima korisnicima je omogućeno da vide vrijeme posljednjeg ažuriranja web mjesta kao i sadržaja web mjesta, odnosno informacija koje web mjesto korisnicima nudi. Oslanjajući se na to, u prikazu rezultata pretraživanja web tražilice će biti uvažena i dimenzija vremena, a ne samo ključne riječi. Na taj se način rezultati pretraživanja ne prikazuju samo kao relevantni po pitanju ključnih riječi, već i relevantni prema vremenu, što povećava mogućnost da će informacije koje web tražilica vraća korisniku biti korisne, a pretraživanje učinkovito. Na taj način algoritam istovremeno u obzir uzima u relevantnost sadržaja i vremena. (Hong, Lihua, Xiaowen, Peiquan, 2010.,str. 102.)

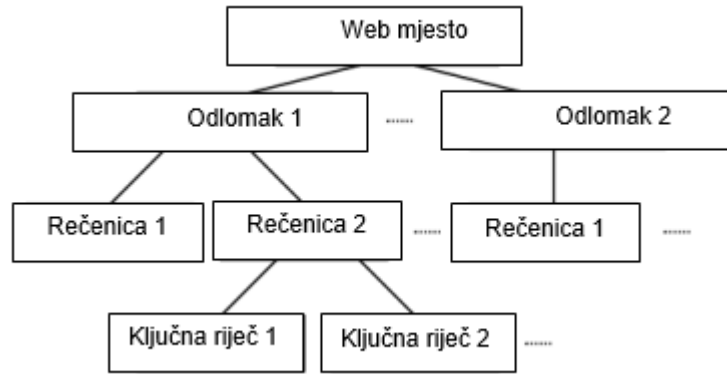
Strategija implementacije ovog algoritma ima dvije faze. Prva faza je vanjska (eng. *offline*), a tiče se izdvajanja i kreiranja parova <ključna riječ, vrijeme, rang> za web mjesto, a druga je faza unutarnja, (eng. *online*) i u okviru te faze izračunavaju se i prikazuju konačni rezultati rangiranja web mjesta na temelju *PageRank* algoritma, koji se oslanja na broj poveznica s drugih web mjesta, relevantnosti naslova web mjesta i vremenske ograničenosti svake ključne riječi. Slika 4 prikazuje dvofazni proces ovog algoritma. (Hong, Lihua, Xiaowen, Peiquan, 2010.,str. 102.)



Slika 4: Proces CT-Rank algoritma u dvije faze (prema Hong, Lihua, Peiquan, Xiaowen, 2011., str. 102).

Prvi korak je preprocesuiranje (eng. *pre-processing*), u sklopu kojeg se sa svakog web mjesta u bazi izdvajaju odlomci i njihov sadržaj, pa su rezultati ove faze odlomci, rečenice i setovi riječi. U ovom se koraku do podataka o vremenu ažuriranja web mjesta dolazi pomoću meta oznaka (eng. *meta tags*), a do naslova odlomaka u web mjestima pomoću oznaka naslova (eng. *title tags*). (Hong, Lihua, Xiaowen, Peiquan, 2010.,str. 103.)

Konačni rezultat ove faze prikazan je na slici 5.



Slika 5: Rezultat preprocesuiranja (prema Hong, Lihua, Peiquan, Xiaowen, 2011., str 103.)

Drugi je korak ekstrahiranje vremena sadržaja na web mjestima u bazi i kreiranje parova <ključna riječ, vrijeme>. Kako bi se to provelo, slijede se *TIMEX2* i *TimeML* standardi. Prilikom ekstrahiranja vremena određuje se broj rečenica i njihovi položaji u tekstu na pojedinom web mjestu. Primjerice, upit postavljen web tražilici „Stanovništvo Republike Hrvatske 2015. godine“ sastoji se od ključnih riječi „Stanovništvo“ i „Republika Hrvatska“ te vremena „2015. godina“, što zajedno predstavlja prethodno navedeni par koji se ekstrahira u sklopu ovog koraka. Algoritam koji kreira te parove prikazan je na slici 6. Ulaz ili input algoritma predstavlja set odlomaka s web mjesta, a izlaz ili output parovi <ključna riječ, vrijeme>. (Hong, Lihua, Xiaowen, Peiquan, 2010.,str. 103.)

**Algoritam** Mapiranje vremena

**Ulaz:**  $P = \langle P_1, P_2, \dots, P_n \rangle$ , gdje je  $P_i$  i-ti odlomak na stranici.

$P_i = \langle S_1, S_2, \dots, S_m \rangle$ , gdje je  $S_j$  j-ta rečenica u  $P_i$  \*  $S_j = \langle W_1, W_2, \dots, W_k \rangle$ , gdje je  $W_t$  t-a ključna riječ u  $S_j$ .

$ut$ : vrijeme ažuriranja web stranice

**Izlaz:**  $M$ : popis parova za mapiranje <ključna riječ, vrijeme>

**Započni:**

**Ako**  $S_1 \in P_1$  i  $S_1$  ne sadrži vrijeme **Onda**

//inicijaliziranje prve rečenice u prvom odlomku

Postavljanje vremena naslova kao vrijeme od  $S_1$

**Završi Ako:**

**Za** svaki  $P_i$  u  $P$ ,  $i=1,2,\dots$  **Učini**

**Za** svaki  $S_j$  u  $P_i$ ,  $j=1,2,\dots$  **Učini**

**Ako**  $S_j$  sadrži vrijeme  $t$  **Onda**

**Za** svaki  $W_i$  u  $S_j$  **Dodaj**  $\langle W_i, t \rangle$  u  $M$ ;

**Inače**

$t \leftarrow$  NađiSličnoVrijeme  $(S_j, P_i)$ ;

**Za** svaki  $W_i$  u  $S_j$  **Dodaj**  $\langle W_i, t \rangle$  u  $M$ ;

**Završi Ako**

**Završi za**

**Završi za**

**Završi** Mapiranje vremena

Slika 6: Algoritam koji kreira parove <ključna riječ, vrijeme> (prema Hong, Lihua, Peiquan, Xiaowen, 2011., str. 104.)



Korak koji slijedi je određivanje ranga doprinosa svakog para generiranog u prethodnom koraku. Pritom se koristi metoda učestalosti pojma- inverzne učestalosti (eng. *tf-idf* (*term frequency- inverse document frequency*)). Izrazi koji se koriste u okviru ove metode navedeni su u tablici 1.

Tablica 1: Izrazi u formuli/metodi *tf-idf*

IZRAZ	ZNAČENJE
$num_{time}$	Konačan broj parova <ključna riječ, vrijeme> na web mjestu, ako se ključna riječ i vrijeme potpuno podudaraju s unesenim upitom u web tražilicu
$num_{total}$	Konačan broj svih ključnih riječi na web mjestu
$pages_{total}$	Konačan broj web mjesta
$pages_{keyword}$	Konačan broj web mjesta koja sadržavaju unesenu ključnu riječ (upit)

U okviru tradicionalne *tf-idf* metode izračunavala se učestalost dane ključne riječi na cijelom web mjestu, dok se u okviru poboljšane *tf-idf* metode izračunava učestalost samo onih ključnih riječi koje su povezane vremenom u paru <ključna riječ, vrijeme>. Ta je metoda definirana sljedećom formulom:

$$doprinis (ključna riječ) = \frac{broj\ parova_{vrijeme}}{sve\ ključne\ riječi\ na\ web\ mjestu\ zbroj} \times \log\left(\frac{web\ mjesta_{zbroj}}{web\ mjesta_{ključna\ riječ}}\right)$$

U drugoj, unutarnjoj fazi, računa se konačan rang web mjesta temeljem 3 faktora: prethodno navedenom metodom *tf-idf* dobiva se vrijednost parova <ključna riječ, vrijeme>, uz pomoć algoritma *PageRank* generira se rang web mjesta koji se uzima u obzir te doprinos naslova web mjesta. Konačan rang dobiva se pomoću sljedeće formule:

$$rang(stranice) = \sum_{i=1}^n doprinos(w_i) + rang\ web\ mjesta + doprinos(naslov)$$

Doprinos ( $w_i$ ) dobiva se pomoću *tf-idf* metode, rang web mjesta je rezultat *PageRank* algoritma, a doprinos naslova računa se s obzirom na ključnu riječ u upitu i vrijeme, pa

sukladno tome, ako se u naslovu web mjesta nalaze ključna riječ i vrijeme iz upita, tada se doprinosu naslova dodaje vrijednost 1. (Hong, Lihua, Xiaowen, Peiquan, 2010.,str. 106.)

Glavni faktor po kojem se razlikuje ovaj algoritam od prethodno navedenih je kombiniranje vremenske dimenzije sadržaja i vremenske dimenzije ažuriranja web mjesta te ključnih riječi iz korisničkog upita web tražilici. (Hong, Lihua, Xiaowen, Peiquan, 2010.,str. 102.)

## 4. Koncept Velikih podataka

U posljednjih nekoliko godina, ono što je do nedavno bila tema isključivo akademskih radova i časopisa, postaje sve konkretniji i opipljiviji dio ljudskih života. Da je koncept Velikih podataka (eng. *Big Data*) važan dio naših života, govori činjenica da se smatra da smo na rubu revolucije koja će potpuno transformirati ljudske živote u bližoj budućnosti. Naziv Veliki podaci često navodi na pogrešno shvaćanje pojma, jer doslovni prijevod navodi na razmišljanje o ogromnoj količini podataka koju imamo na raspolaganju. Međutim, koncept na koji se odnosi pojam Velikih podataka, odnosi se na sposobnost za obradu, analizu i generiranje aplikacija iz svih podataka koji su na raspolaganju te na kraju krajeva ekstrakciju vrijednih informacija i znanja iz raspoloživih podataka („Što učiniti sa svim tim podacima?“, 2020).

### 4.1. Karakteristike Velikih podataka

Kako bi se određeni skup podataka smatrao Velikim podacima, potrebno je da posjeduje karakteristike popularno nazvane „3V“. Te su karakteristike nazvane prema njihovim početnim slovima na engleskom jeziku- eng. *volume* ili volumen, eng. *velocity* ili brzina te eng. *variety* ili raznolikost. . (Informatička tehnologija ITpedia, bez dat.).

**Volumen** se odnosi na dostupnost velike količine podataka koja je pohranjena na Internetu i kojoj se može pristupiti. Količina tih podataka obično se kreće od nekoliko gigabajta, pa sve do zetabajta. Volumen ili obujam podataka određuje možemo li određene podatke smatrati Velikim podacima, što govori o važnosti ove dimenzije

**Brzina** se odnosi na dinamiku kojom se podaci stvaraju, odnosno dohvaćaju iz izvora kao što su poslovni procesi, zapisnici aplikacija, web mjesta društvenih mreža, mobilni uređaji i slično. Primjerice, kod mobilnih aplikacija, velika količina podataka neprestano odlazi u sustav, od kojeg mi zahtijevamo smisao tih istih podataka u realnom vremenu. Brzina određivanja takvih zadataka određuje stvarni potencijal podataka, a ujedno i izazov cijelom sustavu, čiji dijelovi moraju biti koordinirani kako bi brzina bila zadovoljavajuća. (Informatička tehnologija ITpedia, bez dat.).

Posljednja karakteristika, **raznolikost**, povezana je s izvorima i prirodom podataka, koji mogu biti strukturirani i nestrukturirani.

Strukturirani podaci su vrsta podataka koji su mogu organizirati, primjerice, u proračunske tablice, a predstavljaju određenu, organiziranu shemu poput datuma, dok su nestrukturirani podaci oni koji se ne mogu organizirati u tablice, kao što su tekstualne datoteke, statusi na društvenim mrežama te audio ili video sadržaji.

Posljednjih godina pojavila su se još „2V“, kao dodatne dvije karakteristike Velikih podataka. To su **vrijednost** (eng. *value*) i **istinitost** (eng. *veracity*).

Što se tiče novih karakteristika, vrijednost je za svaki podatak različita, a važan segment je i istinitost, jer je bitno razaznati koliko su skupljeni podaci istiniti, pa s obzirom a to nova „V“ upotpunjuju opis tehnologije *Big Data*-e. („Big Data tehnologija- što je i gdje se primjenjuje?“, 2019).

Iako tehnički nepotpuno, ali Veliki podaci mogu se predstaviti kao sve ono što ne stane u MS Excel (Marjanović, 2015).

Ljudi koji svoje poslovanje temelje na digitalnoj tehnologiji za prodaju, Velike podatke dijele po vlastitom iskustvu u četiri kategorije:

**Podaci o kupcu**, odnosno, svi kontakt podaci trenutnih kupaca, podaci o potencijalnim kupcima i podaci iz službe za korisnike, mail liste i slično.

**Podaci o kupovini**, odnosno, svi transakcijski podaci kupaca koji su ostvarili kupnju proizvoda ili usluga (dostupni putem ERP sustava poduzeća ili sustava plaćanja),

**Društveni podaci**, odnosno, podaci s društvenih mreža, blogova i foruma.

**Podaci spajanja**, odnosno, svi podaci s bilo kojih spojenih uređaja na Internetu, uobičajeno nazvani Internet objekti (eng. *Internet of Things*), a to mogu biti automobili, četkice za zube, hladnjaci i slično. (Connell-Waite, 2014.)

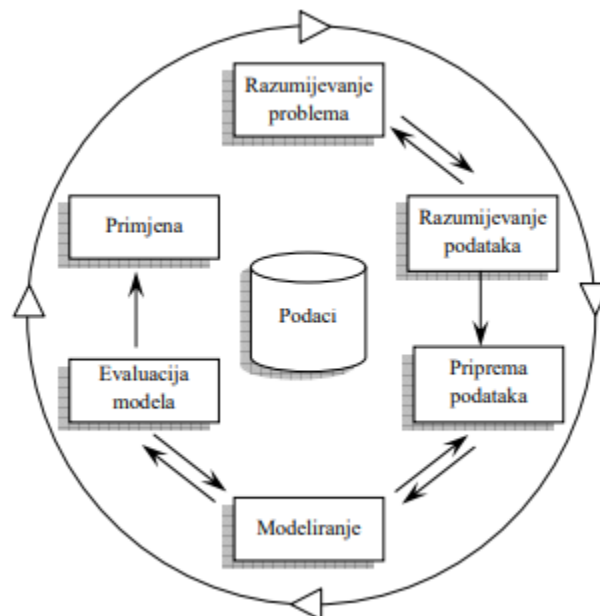
## 4.2. Tehnike i proces analize velikih podataka

Srž i temelj analize podataka čine tehnike modeliranja podataka, a koje su zasnovane na postupcima strojnog učenja. Međutim, nije dovoljno podatke samo modelirati, već ih je potrebno pripremiti i upoznati se s njima, s obujmom, raspršenošću te šumom u podacima kako bi im se na ispravan način pristupilo. (Ujević, bez dat.)

Proces dubinske analize Velikih podataka (eng. *Data Mining*) definiran je na osnovu iskustava u praktičnoj primjeri metodologije koja se naziva CRISP-DM (eng. *CRoss Industry Standard Process for Data Mining*). Osnovne su faze tog procesa:

- Razumijevanje problema,
- Razumijevanje podataka,
- Priprema podataka,
- Modeliranje,
- Evaluacija modela i
- Primjena

Navedene su faze prikazane na slici 7.



Slika 7: Osnovne faze procesa dubinske analize podataka (CRISP-DM) (Ujević, bez dat.)

Prva je faza u procesu dubinske analize podataka **razumijevanje problema**. Kako je u svakom segmentu života potrebno sagledati problem sa svih aspekata, kako bi se isti riješio, tako je i u okviru analize podataka potrebno razumjeti problem i ograničenja koja utječu na proces dubinske analize podataka. U ovoj se fazi određuju ciljevi i kriteriji uspješnosti procesa analize. Primjer definiranog cilja i kriterija je sljedeći: cilj poboljšanja poslovanja neke tvrtke može biti povećanje konverzije u sklopu marketinške kampanje novog proizvoda, a kriterij kvantitativno izražen kao postotak željenog povećanja konverzija. U sklopu ove faze potrebno je i izraditi okvirni plan odvijanja procesa analize, kao i plan trajanja i cijene cjelokupnog postupka. (Ujević, bez dat.)

Osnovni resurs ovog procesa su podaci. Sukladno tome, druga faza analize Velikih podataka je **razumijevanje podataka**. Kako bi se podaci razumjeli, potrebno ih je prikupiti. Istražujući izvore podataka, prikupljaju se potrebni podaci i pohranjuju za daljnju obradu. Nakon što se prikupe, određuju se karakteristike podataka, koristeći deskriptivnu statistiku i vizualizaciju podataka. U tom se segmentu određuju srednje vrijednosti, mjere disperzije i oblici distribucije skupova podataka, ovisno o tipovima atributa, koji mogu biti numerički (intervalni i omjerni) ili kategorijski (ordinalni i nominalni). Također se određuju i stršila, vrijednosti koje odstupaju od ostalih, koeficijent korelacije ili povezanosti između varijabli, kao i kauzalnost. (Ujević, bez dat.)

Faza koja oduzima najviše vremena cjelokupnog procesa dubinske analize podataka je **priprema podataka**, a sastoji se od 4 koraka. Prvi je korak odabir podataka, a on se može vršiti na razini primjera i na razini atributa. Prilikom odabira podataka na razini primjera, kriterij kvalitete nalaže da se odabiru kompletni i ispravni primjeri podataka za analizu, dok je prilikom odabira podataka na razini atributa moguće isključiti attribute s nekvalitetnim vrijednostima (npr. šum, neodređeni podaci i sl.) i provesti redukciju i transformaciju podataka. Kako bi se osigurala maksimalna kvaliteta podataka koji se analiziraju, potrebno ih je pročistiti u smislu normalizacije podataka, odnosno svođenja vrijednosti numeričkih atributa na interval  $[0, 1]$ , zaglađivanja podataka te zamjene neodređenih vrijednosti statistički opravdanim vrijednostima. Uz svojstva odabrane tehnike modeliranja podataka, vezan je korak formiranja novih podataka, a to je izvedivo stvaranjem novog atributa na osnovu jednog ili više postojećih, stvaranjem novih primjera i atributa agregiranjem informacija iz više primjera i slično. Na kraju ove faze podaci se mogu i formatirati, odnosno prilagoditi određenoj tehnici modeliranja. (Ujević, bez dat.)

**Modeliranje podataka** središnji je dio procesa dubinske analize Velikih podataka. Potrebno je pomno i detaljno razmotriti tehnike modeliranja, budući da kvaliteta i oblik rezultata cjelokupnog procesa bitno ovise o odabranoj tehnici. Prilikom odabira tehnike, u obzir treba uzeti odnos temeljnih karakteristika problema te raspoloživih podataka. Zatim je važno odijeliti podatke koji služe učenju od podataka za testiranje pomoću statističkih metoda. Nadalje, nad podacima za učenje konstruiraju se modeli pravilnosti u podacima. (Ujević, bez dat.)

Modeliranje je iterativni proces, pa je različit za nadzirano i nenadzirano strojno učenje. Vrste nadziranog strojnog učenja su stablo odlučivanja, k-najbližih susjeda, neuronska mreža i Bayesova mreža, a nenadziranog strojnog učenja klasteriranje i faktorska analiza. Nadzirano učenje može se objasniti na način da je poznata ulazna varijabla i potrebno je predvidjeti točnu izlaznu varijablu te je u najjednostavnijim slučajevima odgovor oblika „da“ ili „ne“. S druge strane, nenadzirano učenje nema točnih oznaka, odnosno, ne zna se koje su varijable ulazne, a koje izlazne, već mu se prepušta da samostalno otkrije strukturu podataka.

Sukladno tome, određuje se cilj modeliranja, što može biti deskripcija (opis) ili predikcija (procjena). (Ujević, bez dat.)

Faza koja slijedi nakon modeliranja je faza **evaluacije modela**. U okviru ove faze interpretiraju se dobiveni rezultati i ocjenjuje pouzdanost i smislenost konstruiranih modela. Revizijom svih obavljenih faza procesa naglašavaju se eventualni nedostaci u provođenju procesa analize, moguća poboljšanja i alternativna rješenja za pojedine faze u procesu analize Velikih podataka. Prije prelaska na finalnu fazu procesa analize, potrebno je utvrditi je li potrebno ponoviti određene faze s ciljem poboljšanja modela. (Ujević, bez dat.)

Posljednja je faza procesa analize Velikih podataka faza **primjene rezultata**. Kako bi se proces analize uspješno priveo kraju, potrebno je specificirati plan primjene dobivenog modela, kao i praćenja daljnjih rezultata i održavanja modela. Ukoliko je na početku procesa cilj bio povećati broj posjeta web mjestu poduzeća, taj se cilj ne smije zanemariti u ovoj fazi procesa analize. U model je potrebno uvrstiti sve moguće načine ostvarenja cilja i na njima kontinuirano raditi i pratiti povratne informacije, jer one mogu ukazati na nepravilnu upotrebu modela ili neplanirane manjkavosti, koje tek prilikom upotrebe dolaze do izražaja. Na samom završetku procesa potrebno je rezimirati rezultate procesa u završnom izvještaju, navodeći pretpostavke koje su prethodile samom procesu, ograničenja na koja se naišlo prilikom procesa analize i podatke koji su bili dostupni u procesu. Neizostavan su dio izvještaja iskustva stečena tijekom procesa analize podataka i postignuti rezultati (Ujević, bez dat.).

### 4.3. Alati za analizu Velikih podataka

U okviru koncepta Velikih podataka i alata za analizu istih, najčešće se spominje Hadoop. To je programsko okruženje otvorenog koda koje koristi jednostavne programske modele kako bi pohranilo i obradilo velike skupove podataka na računalnim nakupinama (eng. *cluster*). 2005. godine su ga kreirali Doug Cutting i Mike Cafarella uz pomoć programskog jezika Java. („Apache Hadoop“, bez dat.)

Međutim, kako bi se objasnio ovaj alat, prethodno je potrebno objasniti samo značenje programskog okruženja i programa otvorenog koda.

**Programsko okruženje** (eng. *framework*) pojam je koji označava skup alata koji se nalaze na jednom mjestu, a potrebni su za razvoj i pokretanje programskih aplikacija, primjerice, programa, veza i sl. Hadoop sadrži velik broj alata koji služe različitim potrebama, a što ga čini idealnim partnerom u radu s Velikim podacima, a to odrađuje po principu razbijanja velikih skupova podataka u blokove, njihovom multiplikacijom i spremanjem na poslužitelje. (SAS Institute, bez dat.)

**Program otvorenog koda** (eng. *open-source software*) program je kojeg kreiraju i održavaju programeri iz cijelog svijeta putem mreže. Kako je njegov dizajn javno dostupan, taj se pojam odnosi na program koji se može modificirati i koji pruža zajednički pristup (Opensource, bez dat.).

Hadoop se prilikom svog rada ne oslanja na sklopovlje već na same biblioteke koje su dizajnirane s ciljem otkrivanja i otklanjanja kvarova na aplikacijskom sloju te isporuke visokoraspodivih usluga na vrhu nakupine računala. Jezgra Hadoop-a sastoji se od dva dijela: dijela za pohranu- **HDFS** (eng. *Hadoop Distributed File System*) i dijela za obradu- **MapReduce**, koji prenoseći zapakirane kodove čvorova omogućuje njihovu paralelnu obradu. (The Apache Software Foundation, bez dat.).

4 glavne komponente od kojih se Hadoop sastoji navedene su u nastavku. Hadoop Common paket sadrži biblioteke i uslužne programe za druge module, Hadoop raspodijeljeni sustav datoteka (HDFS) sustav je koji pohranjuje podatke na klasteru računala generalne namjene, Hadoop YARN je komponenta odgovorna za upravljanje računalnim resursima u nakupinama računala, koja ih koristi kod raspoređivanja korisničkih aplikacija te Hadoop MapReduce, sustav za paralelnu obradu velikih skupova podataka. (The Apache Software Foundation, bez dat.).



Kako okosnicu Hadoopa čine Hadoop distribuirani datotečni sustav (HDFS), zadužen za pohranu podataka i MapReduce, zadužen za obradu, o njima će biti više rečeno u nastavku ovog rada. (The Apache Software Foundation, bez dat.).

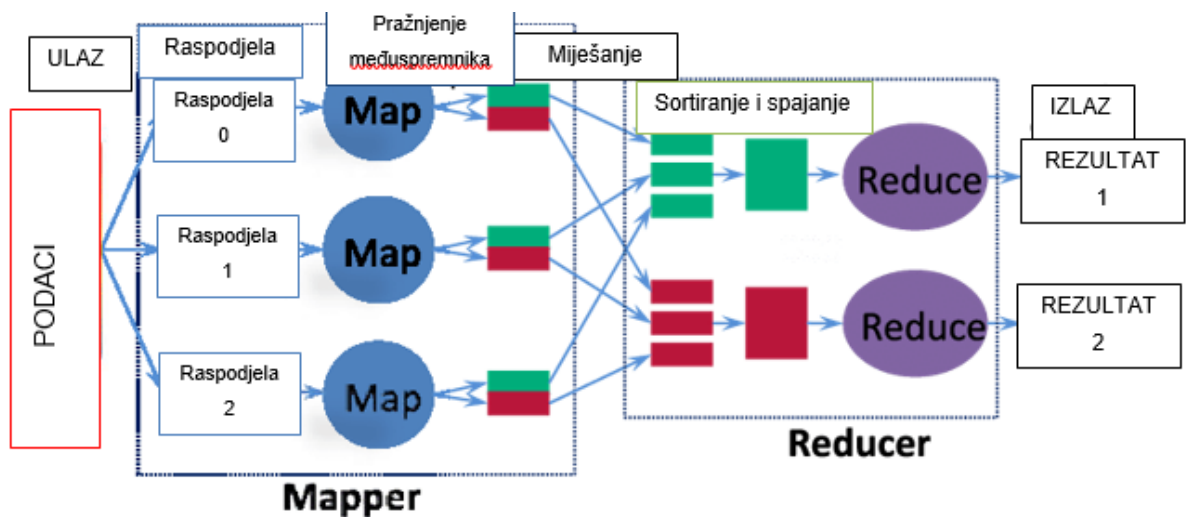
Grozdek navodi da je HDFS dizajniran s pretpostavkom da je pojava greške i kvara hardvera sasvim uobičajena situacija, pa sukladno tome pojava greške ili kvara ne izaziva dugo čekanje ni prekid rada, iz razloga što HDFS sprema podatke multiplicitirajući ih na više računala u jednom klasteru. Uobičajena veličina datoteka koje se pohranjuju u HDFS je 1GB ili više, što je optimalna veličina, jer HDFS nije prikladan u slučaju velikog broja podataka male veličine, npr. nekoliko MB ili manje. Kako bi se osigurala maksimalna točnost i efikasnost ovog sustava, podaci koji se pohranjuju kasnije se čitaju nekoliko puta, a izračunavanje se odvija u neposrednoj blizini tih podataka. Pohrana datoteka odvija se u memorijskim blokovima, čija veličina iznosi 128 MB, a moguće ju je i povećati. Kako bi se utjecalo na moguć gubitak podataka u slučaju pojave greške, HDFS pohranjuje tri kopije svakog bloka na različitim čvorovima. Klaster sadrži dvije vrste čvorova, glavni čvor i podatkovni čvor. Glavni čvor nalazi se u ulozi gospodara i zadužen je za imenički prostor datotečnog sustava, odnosno odražava datotečno stablo i meta podatke u stablu te sprema informacije o svim podatkovnim čvorovima na kojima se nalaze blokovi te datoteke. Podatkovni čvorovi imaju ulogu robova i oni pohranjuju i dohvaćaju blokove podataka. (Grozdek, 2017).

MapReduce je model programiranja i pripadne implementacije za obradu i generiranje Velikih podataka na klasteru računala. Taj se model sastoji od ulaznih podataka, MapReduce programa i konfiguracijskih informacija. Zadaci koji se odvijaju u okviru ove komponente Hadoop-a su *map* zadatak (eng. *map task*) i *reduce* zadatak (eng. *reduce task*). *Map* zadaci imaju veći prioritet od *reduce* zadataka, budući da se svi *map* zadaci trebaju završiti kako bi se *reduce* zadaci mogli početi izvršavati. *Map* zadatak služi za sortiranje i filtriranje podataka, a *reduce* zadatak kombinira podatke koje je obradio *map*. (Grozdek, 2017).

MapReduce svoj rad započinje na način da se ulazni podaci raspodijele na manje dijelove (eng. *input splits*) i za svaki manji dio se kreira *map* zadatak kojem ti podaci predstavljaju ulazne podatke. Ti se podaci najprije zapisuju u međuspremnik (eng. *buffer*). (Grozdek, 2017).

Kad podaci popune 80% međuspremnika, podaci se počinju zapisivati iz međuspremnika na disk i kreira se nova datoteka za konkretnu akciju pražnjenja međuspremnika (eng. *disk spill*). Neposredno prije završetka *map* zadatka, sve se datoteke spajaju u jednu datoteku ili particiju koja predstavlja izlazni podatak. Izlaznu je datoteku moguće komprimirati prije zapisivanja na disk, ako je to potrebno. Nakon toga slijedi faza

kopiranja. U toj je fazi prvo potrebno prikupiti sve ulazne podatke, spojiti ih i zapisati na disk i ukoliko se u *map* fazi koristila kompresija podataka, u ovoj je fazi potrebno provesti dekompresiju podataka. Nakon faze kopiranja slijedi faza sortiranja, u kojoj se ti podaci spajaju s ciljem kreiranja sortiranog niza ulaznih podataka za *reduce* zadatak. Takvi se sortirani podaci prosljeđuju funkciji smanjivanja, (eng. *reduce*). Izlaz (eng. *output*) ove faze predstavljaju manji skupovi podataka čime se završava *reduce* zadatak. (Slika 8.) (Grozdek, 2017).



Slika 8: Princip rada *MapReduce*-a (prema Hadoop MapReduce framework, bez dat.)

## 4.4. Prednosti i nedostaci koncepta Velikih podataka

Analiza Velikih podataka predstavlja velik utjecaj na poslovanje organizacije, od razvoja marketinške strategije, zadržavanja postojećih kupaca pa sve do stjecanja novih i poboljšanja prodajnih aktivnosti. Prilikom lansiranja novih linija proizvoda i usluga, analiza Velikih podataka na društvenim mrežama može biti od veće pomoći od ispitivanja putem upitnika ili anketa, jer se na društvenim mrežama uobičajeno mogu naći spontana, neformalna i iskrena mišljenja korisnika. Kako bi marketinški stručnjaci u organizacijama mogli predvidjeti ponašanje i želje korisnika, analiza informacija o mišljenjima i iskustvima korisnika rezultira podacima koji stručnjacima u odjelu marketinga mogu u velikoj mjeri pomoći prilikom kreiranja budućih marketinških kampanja i strategija. (Russom, 2011, str. 11)

Budući da koncept Velikih podataka predstavlja relativno novu tehnologiju, najveći nedostatak koncepta predstavlja nedostatak vještina i osoblja potrebnog za analitiku Velikih podataka, pa se organizacije odlučuju na prekvalifikacije i edukacije postojećih i novih zaposlenika, što rezultira dodatnim troškovima, pod kojima se podrazumijevaju troškovi edukacija i troškovi hardvera i softvera. Prema istraživanju agencije TDWI (*Transforming Data With Intelligence*) 70% ispitanika u ovoj tehnologiji vidi priliku za napredak, dok 30% ispitanika smatra da tehnologija Velikih podataka predstavlja svojevrsni problem za organizaciju na čijem je rješavanju potrebno kontinuirano i marljivo raditi za dobrobit organizacije, a kojoj će analiza Velikih podataka u velikoj mjeri pomoći u poslovanju. (Russom, 2011, str. 12)

Ukoliko se u odnos stave nedostaci i prednosti ove tehnologije, prednosti, čije su koristi višestruke i važne za organizaciju, se ističu usprkos nedostacima koji su premostivi. Od organizacija će se u budućnosti očekivati edukacije osoblja i kontrola troškova kako bi se iskoristile sve prednosti ove tehnologije, poglavito u pogledu napretka i rasta poslovanja te poboljšanja odnosa s kupcima.

## 5. Načini optimizacije web mjesta

Tražilice poput *Google*-a danas uistinu predstavljaju vrlo relevantan i između ostalog, ključan izvor informacija o uslugama, proizvodima i *brandovima* i definitivno predstavljaju zamjenu za nekadašnje telefonske imenike, časopise i knjige. Upravo to je rezultat promjena u vremenu i tehnologiji, a kako bismo se prilagodili promjenama u pretraživanju i stjecanju korisničkog iskustva, važno je posvetiti posebnu pažnju tehnikama optimizacije web stranica.

Kako bi se istovremeno utjecalo i na relevantnost i na popularnost web stranice, web stranice je potrebno optimizirati za Internetske pretraživače. Web mjesto može se optimizirati i na način da se posveti marketingu za Internet pretraživače, što podrazumijeva usluge plaćenih oglasa te uz pomoć tzv. *Trusted feed*-a.

U nastavku će biti detaljnije opisan svaki od načina optimizacije web mjesta.

### 5.1. Optimizacija za Internet pretraživače

Optimizacija za Internet pretraživače (eng. *Search Engine Optimization- SEO*) predstavlja niz tehnika, metoda i strategija za visoko rangiranje web stranica na tražilicama kao što su *Google*, *Bing*, *Yahoo* i druge, s ciljem doseganja većeg broja korisnika i na kraju krajeva, većeg broja posjeta web stranici. („Softwise d.o.o.“, bez dat.)

Kako bismo se upoznali s postupkom optimiziranja tražilica, najprije je važno shvatiti na koji način iste funkcioniraju. Iako je u trećem poglavlju ovog rada funkcioniranje web tražilica već razrađeno, u nastavku će isto biti ponovo ukratko objašnjeno. Dvije ključne funkcije tražilica su tzv. *crawlanje* (eng. *crawling*) i indeksiranje (eng. *indexing*). Pojam *crawlanje* dolazi od engleskih pojmova *crawlers* ili *spiders*, koji predstavljaju softverske robote. Oni su zaduženi za pronalaženje specifične web stranice kao odgovor na korisnički upit. Da bi se to izvršilo, *crawleri* ili *spideri* najprije skeniraju i identificiraju web stranice, slike, dokumente i PDF-ove koji se nalaze na Internetu i potom daju njihove podatke tražilici. Zatim tražilica prikazuje samo one rezultate koji su relevantni upitu korisnika i rangira ih prema popularnosti web stranice i ti se rezultati nazivaju SERP (eng. *Search Engine Result Pages*). Međutim, kako bi se rezultati rangirali prema popularnosti, potrebno je istu i odrediti. Za to su zaduženi kompleksni matematički modeli, odnosno algoritmi., koji se sastoje od stotinu varijabli i oni predstavljaju srž optimizacije za tražilice, a objašnjeni su u trećem poglavlju ovog rada. („Softwise d.o.o.“, bez dat.)

U počecima upotrebe Interneta, tražilice su svoje funkcioniranje temeljile na tek nekoliko varijabli algoritama, koje su se mogle podijeliti na dvije široke kategorije: na varijable

na stranici (eng. *On-page SEO*) i one izvan stranice (eng. *Off-page SEO*). („Softwise d.o.o.“, bez dat.)

### 5.1.1. Vrste optimizacije za Internet pretraživače

Optimizacija web stranice može se postići na dva načina, na stranici (eng. *On-page*) i izvan stranice (eng. *Off-page*). Korištenje jednog načina ne isključuje važnost korištenja drugog, jer je vrlo bitno ukomponirati oba u optimizaciju web mjesta, kako bi ista bila potpuna i učinkovita. („Što je SEO i zašto je optimizacija bitna?“, 2018.)

**Optimizacija na stranici** (eng. *On-page SEO*) temelji se na mišljenju da se treba staviti u korisnikove cipele i prepoznati kako i što on uistinu pretražuje, svodeći se na odabir najboljih ključnih riječi i dokazivanje tražilici da se web stranica temelji na tim ključnim riječima. Drugim riječima, ova se optimizacija fokusira se na uređenje komponenata web mjesta kako bi ono rezultiralo prepoznatljivošću web tražilicama. Samim time, zadaća unutarnje optimizacije je kvalitetno prikazati algoritmima tražilica i posjetitelju o čemu se radi na web stranici koju posjećuju. („Što je SEO i zašto je optimizacija bitna?“, 2018.)

Kako bi optimizacija bila kvalitetno izvedena potrebno je slijediti 3 smjernice: iz sadržaja web stranice mora biti lako zaključivo o čemu je riječ i taj sadržaj mora biti kvalitetan, a u isto vrijeme povezan uz samu temu relevantnim ključnim riječima te koristan kako bi bio visoko pozicioniran u rezultatima pretraživanja. („Što je SEO i zašto je optimizacija bitna?“, 2018.)

Ključni elementi unutarnje optimizacije su meta oznake (eng. *meta tags*), oznake naslova (eng. *title tags*), sadržaj i lokalizacija, a u nastavku rada svaki od tih elemenata bit će detaljnije objašnjen. („Ultimativni vodič: Optimizacija za tražilice za početnike“, bez dat.)

**Meta oznake** predstavljaju službene opise svake web stranice i uloga im je obavještavati tražilice o ključnim informacijama o sadržaju stranice, što im naposljetku pomaže pri odluci je li stranica prikladna za prikaz u rezultatima pretraživanja. („Ultimativni vodič: Optimizacija za tražilice za početnike“, bez dat.)

U suštini, tzv. *meta description* oznake služe da se u jednoj ili dvije rečenice ukratko opiše sadržaj web stranice, a tzv. *meta keywords*, odnosno, meta ključne riječi služe istoj svrsi, samo to postižu nabranjem ključnih riječi. („SEO.hr“, bez dat.)

**Oznake naslova** su HTML oznake koje sadrže tekst koji opisuje sadržaj pripadajućeg web mjesta, a njihova je dužina ograničena i preporučena dužina iznosi 77 znakova. („Ultimativni vodič: Optimizacija za tražilice za početnike“, bez dat.)

Ako web mjesto nema **sadržaj**, korisnici nemaju razlog navedeno posjetiti, jer neće pronaći informaciju za kojom tragaju. Prilikom već prethodno navedenog skeniranja web stranice, *crawleri* se bave povezivanjem sadržaja s ključnim riječima koje se pojavljuju u sadržaju. Iz tog se razloga velik značaj ponovo stavlja na odabir relevantnih i popularnih ključnih riječi u kontekstu unutarnje optimizacije, ali i na ideji da se opisi proizvoda i usluga na web mjestu predstave na jednostavan i jedinstven način, bez besmislenog ponavljanja ključnih riječi. („Ultimativni vodič: Optimizacija za tražilice za početnike“, bez dat.)

Poduzećima koja svoje poslovanje temelje na određenom geografskom području (kvartu, gradu ili državi), **lokalizacija** je vrlo bitan čimbenik unutarnje optimizacije. Takvim poduzećima se kod pisanja sadržaja na web mjestu preporuča korištenje lokaliziranih ključnih riječi, što potom utječe na tražilicu da prepozna da su proizvodi ili usluge povezani s određenom lokacijom. Najpoznatiji alati koji mogu biti od pomoći za optimizaciju lokacije su *Google My Business*, *Yahoo Local Listings* i *Bing Places for Business*. („Ultimativni vodič: Optimizacija za tražilice za početnike“, bez dat.)

S druge strane, **Vanjska optimizacija** (eng. *Off-page SEO*) odnosi se na tehnike koje se koriste za poboljšanje pozicije web stranice u rezultatima pretraživanja. („Ultimativni vodič: Optimizacija za tražilice za početnike“, bez dat.)

Razlika između te dvije kategorije je ta da vanjska optimizacija uključuje i aktivnosti koje nisu povezane s web mjestom, a to su izgradnja poveznica (eng. *link building*), koja označava utjecaj na rang web mjesta uzimajući u obzir kvalitetu i kvantitetu poveznica koje vode prema web stranici te marketing preko društvenih mreža (eng. *social media marketing*). („Ultimativni vodič: Optimizacija za tražilice za početnike“, bez dat.)

Izgradnja poveznica (eng. *link building*) je područje optimizacije za Internet pretraživače koje se bavi generiranjem i privlačenjem kvalitetnih ulaznih poveznica na web mjesto. Što više takvih, kvalitetnih ulaznih poveznica neko web mjesto ima, to će ono biti bolje rangirano u rezultatima pretraživanja. Cilj ovog područja optimizacije je osigurati da se na web mjestu nalazi kvalitetan, pouzdan, vrijedan i relevantan sadržaj i da takvo web mjesto bude rangirano u vrhu rezultata pretraživanja. Praćenjem specifičnih uzoraka, algoritmi današnjice mogu identificirati one poveznice koje upućuju prema web mjestu koje želi skupiti čim više poveznica, a koje su posljedica dogovora marketera i vlasnika nekog web mjesta. Time se sprječava manipulacija izgradnjom poveznica i smanjuje nekvalitetan i nerelevantan sadržaj na web mjestima. („Ultimativni vodič: Optimizacija za tražilice za početnike“, bez dat.)

Najpoznatiji alati za izgradnju poveznica su: *Majestic SEO*, koji funkcionira koristeći dva indeksa, tzv. *Fresh Index* i *Historic Index*. Prvi indeks sadrži popis web mjesta i poveznica unazad 120 dana, kao i popis „izgubljenih“ poveznica u tom razdoblju. Drugi indeks predstavlja

bazu podataka koja može poslužiti kao mjerilo kvalitete i uspješnosti *SEO* optimizacije budući da obuhvaća veliku količinu URL adresa od unazad 5 godina. Sljedeći alat je *MozBar* koji se fokusira na analiziranje ključnih riječi i funkcionira kao ekstenzija preglednika *Chrome*. Jednostavan je za korištenje budući da je potrebno samo upisati ključnu riječ i alat potom analizira koliko je ista kompetitivna, odnosno kakav utjecaj ima na optimizaciju uzevši u obzir konkurenciju. *Hunter.io* je alat koji olakšava pronalaženje e-mail adresa povezanih s određenim web mjestom, a funkcionira na način da korisnik upiše URL adresu web mjesta čije povezane e-mail adrese ga interesiraju. Zatim alat generira e-mail adrese, a uz to pruža i mogućnost verifikacije istih, kako bi bilo vidljivo je li e-mail adresa aktivna i u upotrebi ili ne. Navedeni alati dostupni su u besplatnim verzijama, no moguće su i određene opcije plaćanja kako bi se moglo pristupiti svim funkcijama alata. (Camarce, 2020)

Važan faktor vanjske optimizacije predstavljaju društvene mreže. Taktike optimizacije važno je primijeniti na društvenim mrežama na način da se kreiraju gumbi za dijeljenje članaka na društvenim mrežama kao pozivi na akciju, a uz to se može ponuditi korisnicima nešto za uzvrat, npr. za jedno dijeljenje članka korisnik može dobiti besplatnu online knjigu (eng. *e-book*). Influenceri također imaju velik utjecaj na spoj marketinga preko društvenih mreža i optimizacije za Internet pretraživače, jer imaju pristup i utjecaj na veliku količinu ljudi, pa samim time i mogućnost dugoročne suradnje s poduzećima čije proizvode promoviraju. („Ultimativni vodič: Optimizacija za tražilice za početnike“, bez dat.)

### 5.1.2. Važnost plana za optimizaciju za Internet pretraživače

Važnost planiranja u svakodnevnom životu neopisiva je. Ritam današnjice ponekad je teško pratiti ukoliko nemamo dobar plan za ciljeve koje želimo ostvariti. Dobar plan optimizacije omogućuje da se trud usmjerava na precizniji i djelotvorniji način. Prilikom pohađanja tečaja „Uvod u digitalni marketing“ u sklopu *Google digitalne garaže*, naišla sam na primjer poduzetnika Mikea, koji je svoje djelovanje usmjerio u tvrtku s balonima. Navodi kako je važno koristiti karte radi ciljanja lokalnih tvrtki koje pokušava privući, kao i detaljno istražiti koje ključne riječi njegovi klijenti pretražuju. Nadalje, bitna je prezentacija u javnosti, koja se može postići postavljanjem veza na web stranicu, primjerice, ako neki od klijenata pišu blogove, bilo bi poželjno da, ako imaju ikakvo iskustvo u vezi sadržaja na web stranici, isto i podijele u svom blogu i spomenu tu web stranicu. Poželjno je da se u okviru plana, odnosno, strategije postave rokovi i odrede prioritete. Normalno je da se najprije teži napraviti nešto što donosi zaradu, no treba biti realističan i imati strpljenja. Plan će se s vremenom mijenjati i bit će ga potrebno ažurirati i konstantno poboljšavati pozive na radnju, kao i redovito pregledavati rezultate, o čemu će biti riječ u nastavku ovog rada. (Google digitalna garaža, 2021.)

Strategija optimizacije za Internet pretraživače proces je organizacije sadržaja web mjesta na način da se ne kreira sadržaj za kojeg se smatra da ga korisnici traže, već da se kreira sadržaj kojeg korisnici uistinu traže. (Fortin, 2021.)



### 5.1.3. Postupak optimizacije za Internet pretraživače

Optimizacija tražilice neprekidan je postupak i ne postoji jedinstveni prečac za ostvarenje iste.

Prvi korak je istraživanje ključnih riječi, odnosno istraživanje koje fraze ili riječi korisnici pretražuju prilikom traženja proizvoda i/ili usluga povezanih sa tvrtkom. Primjerice, osoba koja posjeduje farmu i pokrenula je uslugu dostave svježeg voća i povrća može pretpostaviti da njeni korisnici pretražuju održive plodove ili vegetarijanske recepte, pa sukladno tome optimizirati sadržaj i ponudu na web stranici. To se može ostvariti objavama tjednih recepata ili pisanjem bloga o uzgoju voća i povrća na farmi. Vrlo je važno biti aktivan na web stranici, jer više će korisnika posjetiti stranicu na koju se svaki tjedan dodaju novi sadržaji, nego onu koja je posljednje bila ažurirana prije, primjerice, godinu dana. (Google digitalna garaža, 2021.)

Pažnju je potrebno posvetiti i brzini učitavanja web mjesta, kao i odabiru njegove pravilne strukture, koja je također vrlo važna samom korisniku, a naposljetku i vlasniku web mjesta. To su zadaci tzv. tehničke optimizacije. Web mjesto koje nije tehnički optimizirano ili optimizacija nije kvalitetno izvedena, u budućnosti će zasigurno imati pad korisnika i malu stopu konverzije. Što se tiče brzine učitavanja web mjesta, potrebno je obratiti pozornost na sadržaj istog u smislu veličine, odnosno, poželjno je da fotografije na web mjestu sadrže manje MB, *plug-inova* i redirekcija, web mjesto će se brže učitati. To se može postići korištenjem *Photoshopa* za optimizaciju fotografija, *CDN service-a* i *browser cache-a*, koji će, prilikom prvog učitavanja web mjesta na koje se korisnik vraća, spremiti određene statične dijelove istog i pri svakom sljedećem učitavanju upamtiti *cached* verziju. Za provjeru postignutih rezultata, preporučljivo je koristiti *Googleov* alat, *PageSpeed Insights* ili *GT metrix* koji pružaju uvid u performanse i brzinu učitavanja web mjesta, a uz to daju i prijedloge za poboljšanje iste. („Detaljan vodič za tehnički SEO“, 2019.)

Također je važno prilagoditi web stranicu mobilnim uređajima, odnosno u plan optimizacije uvrstiti responzivni dizajn. Web stranica je responzivna ako se prilagođava uređajima, odnosno različitim veličinama ekrana, što naposljetku omogućuje dobro korisničko iskustvo, neovisno o uređaju koji se koristi za pregledavanje. Važnost ovog faktora optimizacije potvrđuje činjenica da čak 945 korisnika pametnih telefona traži informacije putem mobitela i 84% njih kontaktira tvrtku ili obavi kupnju, odnosno napravi neku akciju. (Papratović, 2018.)

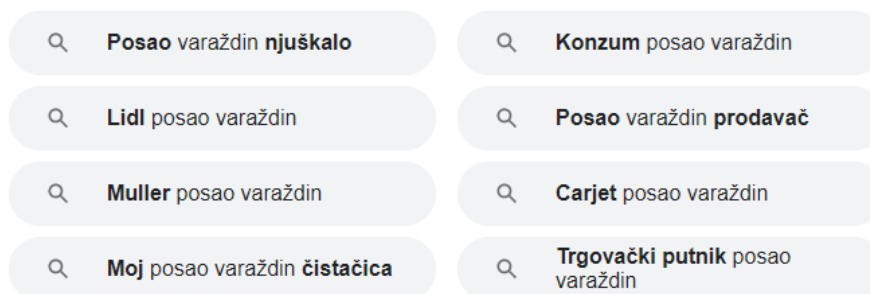
### 5.1.4. Odabir ključnih riječi

Ključne riječi u digitalnom svijetu su pojmovi ili izrazi koji opisuju proizvod ili uslugu, a pomoću njih tražilice određuju koje će fotografije i članke prikazati osobi koja pretražuje Internet. (Što su ključne riječi i gdje se koriste na webu?, 2017.)

Šunjala (eng. *Crawler*), ranije spomenuta u radu, prilikom skeniranja stranice bilježe pojmove koji se na njoj pojavljuju. Ukoliko pojmovi na web stranici odgovaraju u velikoj mjeri ključnoj riječi koju je korisnik koji pretražuje sadržaj Interneta upisao, algoritmi će web stranicu rangirati bliže prvoj poziciji u rezultatima pretraživanja. U tom slučaju, važno je odrediti koje ključne riječi valja koristiti, iz razloga što one predstavljaju osnove *SEO*-a. Prvi korak u analizi ključnih riječi je izrada popisa od desetak ključnih riječi koje su povezane s proizvodima ili uslugama koji se nude javnosti na web stranici. Potom se sastavljen popis analizira u jednom od alata za istraživanje ključnih riječi, primjerice *Google Keyword Planner*. Rezultat ovog koraka je prikaz varijacija grupiranih ključnih riječi uzevši u obzir konkurenciju i obujam pretraživanja te povezane ključne riječi koje generira korišteni alat. Ukratko, alat koji će biti objašnjen u šestom poglavlju ovog rada pruža uvid u to koliko se često određene riječi pretražuju i kako su se pretraživanja mijenjala u vremenu, što naposljetku može pomoći kod sužavanja popisa ključnih riječi. („Ultimativni vodič: Optimizacija za tražilice za početnike“, bez dat.)

Međutim, ako je riječ o malom poduzeću, koje još nije postiglo velik tržišni udio, tada se fokus stavlja na proširene fraze koje korisnici upisuju u tražilice (eng. *long-tail keywords*). Iako će one privući manje prometa, imat će veću vrijednost konverzije. Brz i učinkovit način pronalaska proširenih fraza je pregled *Googleovih* povezanih pretraživanja na dnu stranice rezultata pretraživanja. Primjer koji se može vidjeti na slici ispod vezan je uz traženje posla na području Varaždina u rezultatima povezanih istraživanja. (Slika 9.) („Ultimativni vodič: Optimizacija za tražilice za početnike“, bez dat.)

#### Povezana pretraživanja



Slika 9. Povezana pretraživanja („posao Varaždin“) prikazana na prijenosnom računaru

Kako bi određivanje ključnih riječi imalo maksimalan utjecaj na optimizaciju web mjesta, potrebno je pokrenuti *AdWords* oglas koji je optimiziran oko točno jedne ključne riječi. Važnu ulogu u tom procesu ima i *Google Analytics* koji prikazuje rezultate prometa po ključnoj riječi, a pruža i mogućnost izračuna točnog broja ljudi koji su izvršili određenu akciju, od kupnje, *downloada*, prijave na *newsletter* i slično, što naposljetku daje povratnu informaciju isplati li se optimizacija za određenu ključnu riječ. („Ultimativni vodič: Optimizacija za tražilice za početnike“, bez dat.)

O alatima za optimizaciju biti će riječi u nastavku ovog rada.

## 5.2. Marketing za Internet pretraživače

Marketing za Internet pretraživače (eng. *Search engine marketing- SEM*) kao način optimizacije web mjesta označava korištenje plaćenih pozicija u rezultatima pretraživanja na web pretraživačima, s ciljem postizanja veće vidljivosti web mjesta prilikom pretraživanja. (Dobrinić D. i sur., 2011., str. 232.)

Ovaj način optimizacije provodi se korištenjem plaćenih oglasa, oglašavanjem s *banner* oglasima na web stranicama, e-mail marketingom i ostalim vrstama oglašavanja koji u suštini povećavaju posjećenost i popularnost određenog web mjesta, ali i olakšavaju prikupljanje podataka o posjetiteljima, odnosno korisnicima. („Softwise d.o.o.“, bez dat.)

Kako bi taj marketing bio učinkovit, potrebno je provesti analizu podataka na web mjestima, uzimajući u obzir koncept Velikih podataka (eng. *Big data* koncept). U tom se kontekstu provodi otkrivanje znanja u podacima, najčešće pomoću procesa rudarenja podataka (eng. *Data mining*), o čemu je već bila riječ u 4. poglavlju ovog rada.

### 5.2.1. Uvod u marketing za Internet pretraživače

Tradicionalno oglašavanje dopire do cijelog svijeta, no ciljano oglašavanje dopire do točno određene grupe ljudi, koja se zove ciljna grupa korisnika. Moć marketinga za Internet pretraživače objasniti će se u nastavku rada na primjeru tvrtke Rachel's kitchen, objašnjenom u tečaju *Google* digitalne garaže. (Google Digitalna garaža, 2021.)

Vlasnica tvrtke Rachel's kitchen opisala je koliko joj je *SEM* pomogao u rastu i razvoju poslovanja. Da je željela oglašavati svoju tvrtku za vrijeme blagdana ili praznika, kao što su Majčin dan ili Valentinovo, trebala je izdvojiti popriličan iznos za višemjesečni marketing. Međutim, prve godine prilikom pokretanja *Ads* kampanje za Majčin dan, udvostručio se promet u odnosu na prethodnu godinu, a uzrok tome bili su plaćeni oglasi. Veze koje se pojavljuju na glavnom dijelu u rezultatima, prilikom pretraživanja web tražilice, organski su rezultati i njihovo prikazivanje je besplatno. No, na vrhu stranice i najčešće u desnom stupcu prikaza rezultata, nalaze se plaćeni oglasi. Upravo *SEM* omogućuje upotrebu navedenog prostora radi oglašavanja potencijalnim klijentima prilikom pretraživanja određenih izraza ili riječi koji su važni za tvrtku, i kao što je u prethodnom dijelu rada navedeno, to su ključne riječi, koje imaju veliku ulogu u kontekstu optimizacije web mjesta. Ovaj se način oglašavanja zove još plaćanje po kliku, (eng. *Pay-per-click*), jer oglašivač plaća samo ako netko klikne na oglas, a ako se oglas prikaže i nitko na njega ne klikne, oglašivač ne plaća ništa. Upravo to je razlika između *SEM*-a i tradicionalnog načina oglašavanja. Kod tradicionalnog oglašavanja, oglas se prikazuje

velikom broju ljudi bez obzira na to treba li im usluga ili proizvod koji se oglašava ili ne, dok se kod marketinga za tražilice zna da su korisnici koji su kliknuli zapravo zainteresirani za proizvod ili uslugu koja se oglašava, jer su to sami pretraživali. (Google Digitalna garaža, 2021.)

### 5.2.2. Način funkcioniranja marketinga za Internet pretraživače

*SEM* koristi najmanje pet metoda i mjernih podataka u svrhu optimizacije web mjesta. („Web Identitet“, bez dat.)

Prva metoda je istraživanje i analiza ključnih riječi, a njom se postiže uvid u percepciju pretraživanja, odnosno utjecaj pretraživanja robne marke ili *branda* na samu percepciju potrošača, a vodeći se idejom da je online pretraživanje najčešće prvi korak potencijalnim potrošačima, tj. kupcima, takvo pretraživanje stvara dojam robne marke ili *branda* svakom individualnom potrošaču. („Web Identitet“, bez dat.)

Druga metoda je proučavanje prisutnosti web mjesta na tražilicama, a to se analizira kroz broj stranica neke web lokacije koje su indeksirane od strane tražilica i popularnost iste. („Web Identitet“, bez dat.)

Treća metoda su tzv. *back-end* alati, koji pružaju podatke na web mjestu posjetiteljima istog, što omogućuje mjerenje uspjeha web mjesta. Ti alati prisutni su u obliku jednostavnog brojača prometa, primjerice, broj posjeta web mjestu u jednom danu, pa sve do rada s log datotekama i stavljanja *JavaScripta* ili slike na web mjesto za praćenje radnji. EBSCO, baza podataka koja sadrži velik broj znanstvenih radova, članaka i disertacija, koristi tri alata. Prvi je alat za analizu dnevnika, *WebTrends by NetiQ*, potom analitički alat temeljen na oznakama, *WebSideStory Hitbox* i alat temeljen na transakcijama *TeaLeaf RealITea*. Pomoću tih alata provjeravaju se nevidljivi dijelovi web mjesta i analiziraju se potencijalni problemi koji utječu na opći dojam korisnika prilikom posjete stranici, ali i na njenu popularnost. („Web Identitet“, bez dat.)

Četvrta metoda su *Whois* alati, koji se odnose na zaštitu autentičnosti web mjesta, a uloga im je otkrivanje vlasnika različitih web mjesta i pružanje informacija koje se tiču autorskih prava i zaštitnih znakova. („Web Identitet“, bez dat.)

Peta metoda vezana je uz, prethodno naveden, responzivni dizajn, a zove se *Google Mobile-Friendly* i temelji se na analizi URL adrese web mjesta, koja zatim daje povratnu informaciju o tome je li dizajn web mjesta pogodan za mobilne uređaje ili nije. Potrebno je istaknuti važnost responzivnog dizajna, s obzirom na velik broj korisnika koji web mjestu pristupaju upravo putem svojih mobilnih uređaja. („Web Identitet“, bez dat.)

S obzirom na to da ovaj način optimizacije omogućuje oglašivačima licitiranje za određene ključne riječi ili fraze prije samog plaćenog oglašavanja, može se reći da se *SEM* temelji na sustavu dražbe. Ukoliko se za primjer uzme traženje usluge fotografiranja vjenčanja i ako se ta usluga unese u tražilicu, na desnoj strani i na vrhu rezultata pretraživanja, vidljivi su plaćeni oglasi, za koje se oglašivači nadmeću. Na *SEM* dražbi, ključna je relevantnost oglasa i konkurentna licitacija. Naposljetku, ta dva faktora ključna su za uspjeh u *SEM*-u. Maksimalni *CPC* ili maksimalna cijena po kliku predstavlja iznos koji je oglašivač spreman platiti za jedan klik na njegov oglas. Ako je jedan oglašivač spreman platiti do 2 eura kad korisnici kliknu na njegov oglas nakon određenog pretraživanja, a njegov konkurent je spreman platiti tek iznos od 1 euro za isto, prvi će oglašivač imati prednost zbog veće licitacije. (Google Digitalna garaža, 2021.)

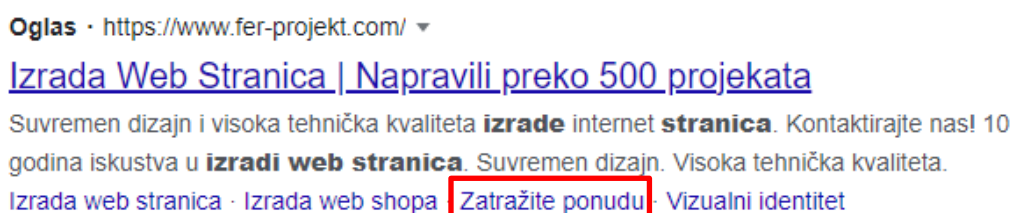
Drugi čimbenik, relevantnost oglasa, povezan je s ocjenom na ljestvici od 1 do 10 i pokazuje koliko je oglas usko povezan s onim što korisnik pretražuje i na tražilicama poput *Google*-a ili *Binga* to se zove ocjena kvalitete. Relevantnost oglasa može se objasniti na primjeru usluge fotografiranja vjenčanja. Ukoliko oglašivač u svom oglasu navede da nudi fotografiranje vjenčanja u Varaždinu, njegov će oglas biti relevantniji od oglasa drugog oglašivača koji nudi usluge za vjenčanja u Varaždinu, iako se oba oglasa odnose na vjenčanja. Kad su takva dva oglašivača prisutna na dražbi i ukoliko oba imaju isti iznos licitacije, oglasi oglašivača s višom ocjenom kvalitete prikazat će se više na stranici rezultata pretraživanja. (Google Digitalna garaža, 2021.)

### 5.2.3. Isticanje oglasa

Ono što je lako za primijetiti jest da, prilikom pretraživanja i generiranja rezultata pretraživanja, većina oglasa izgleda slično. Tajna uspješnog marketinga za tražilice leži u istaknutim, prepoznatljivim i uočljivim oglasima s jasnim pozivima na radnju. (Google Digitalna garaža, 2021.)

Ako se ponovo uzme primjer usluge fotografiranja vjenčanja i pretraži pojam „najam fotografa vjenčanja“, može se primijetiti da je dostupno mnogo oglasa i svi su u velikoj mjeri slični i to je zadatak uspješnog *SEM*-a, stvoriti jedinstven oglas koji se izdvaja od konkurencije. U ovom kontekstu ponovo se uzima u obzir relevantnost, koja je glavna smjernica u marketingu za tražilice. Iz tog razloga, oglas treba sadržavati sve ključne riječi koje korisnik pretražuje, a rezultat će savršenim podudaranjem s pretraživanjem. Korisnici također vole jedinstvene ponude i promocije, pa je poželjno i da se u oglas uvrste rasprodaje i akcije, na primjer, u slučaju fotografiranja vjenčanja, „Besplatna dostava za kupnju iznad 150 kuna.“ Također je poželjno da je u oglasu istaknut poziv na konkretnu radnju, a to se može realizirati na način da se istaknu riječi poput „Pregledajte naš album“, „Rezervirajte na vrijeme“, ili „Kupite“, jer istraživanja pokazuju da korisnici reagiraju kad ih se usmjeri na određenu radnju. Konkretni primjer ispravnog poziva na akciju na web mjestu pomaže korisniku da nađe upravo ono što traži, pa tako primjerice, ako korisnik traži uslugu izrade web stranice, ono što treba dobiti je ponuda za tu uslugu kao rezultat unosa upita u tražilicu. (Slika 10.) (Google Digitalna garaža, 2021.)

Takav poziv poželjno je uvrstiti u drugi redak u oglasu, tako da već u rezultatima pretraživanja to bude vidljivo i da korisnici znaju već unaprijed što im se na web mjestu nudi. (Google Digitalna garaža, 2021.)



Slika 10. Poziv na akciju na oglasu tvrtke Fer projekti (<https://www.fer-projekt.com/hr>) prikazan na prijenosnom računaru

Kako bi se istaknula specifičnost i temeljitost na web mjestu, poželjno je da se generira više grupa oglasa, odnosno da se napravi nekoliko skupova ključnih riječi, koje će potom utjecati na vizualni dojam samih oglasa, ali i lakše razvrstavanje i kontroliranje istih. Kako bi

se istaknula jedinstvenost proizvoda ili usluga potrebno je prije kreiranja samog oglasa razmisliti što proizvod ili uslugu čini jedinstvenom ili jedinstvenim, kako se razlikuju od konkurencije na tržištu i kakvu korist kupci imaju od usluge ili proizvoda. („Kako napisati učinkoviti *Google AdWords* oglas“, 2017.)

Odgovori na ta pitanja predstavljaju *USP* (eng. *Unique selling proposition*), ključ efikasne prodaje koji se sastoji od isticanja originalnosti i unikatnosti proizvoda i/ili usluga koje određena tvrtka nudi na tržištu. *USP* se može temeljiti na karakteristikama proizvoda, na strategiji određivanja cijena ili promociji, uzevši u obzir elemente marketinškog miksa, 4P, odnosno proizvod, promociju, distribuciju i cijenu. Primjerice, Charles Revlon, osnivač tvrtke *Revlon*, tvrdi da prodaje nadu, a ne kozmetičke proizvode. Na taj način ističe da proizvodi tvrtke *Revlon* nisu isključivo kozmetičke prirode, već da je u njima utkana nada, trud i ljubav prema kozmetici i to bi trebao biti putokaz za isticanje oglasa, kako u tradicionalnom marketingu, tako i u internetskom. („Entrepreneur EUROPE“, bez dat.)



### 5.3. Trusted feed

Nešto manje zastupljeni oblik oglašavanja i optimiziranja web mjesta je tzv. *trusted feed*, odnosno *XML feed*, budući da se temelji na HTML protokolu. Kod *XML feeda* je bitno da se poveznice, odnosno URL-ovi poduzeća dostave izravno u indekse pretraživača putem ključnih riječi, meta opisa i ostalih meta podataka, bez posredstva *crawlera*. Korištenjem *trusted feeda*, web mjesto se pojavljuje u organskim rezultatima na pretraživačima. (Dobrinić D. i sur., 2011., str. 298.)

Još jedna karakteristika je ta da *trusted feed* omogućuje tzv. *cloaking*, tehniku optimizacije web mjesta koja se sastoji od prikaza drugačijeg sadržaja ili informacija alatima pretraživača od onog koji je vidljiv korisnicima. Jedan primjer implementacije navedene tehnike je dodavanje dodatnih ključnih riječi, koje su najčešće u istoj boji kao i boja pozadine, pa ih ljudi ne vide, no *botovi* tražilice ih mogu pratiti. („What is cloaking in SEO?“, 2020)

Slične tehnike su i plaćanje za plasman (PFP- eng. *Pay for Placement*) i plaćanje za uključivanje (PFI- eng. *Paid for Inclusion*), kod kojih se plaćene poveznice ugrađuju u organske rezultate pretraživanja. Zajedničko tim tehnikama i *trusted feedu* je oslanjanje na model cijene po kliku (PPC- eng. *Pay-per-click*), no kod *trusted feeda* ta je cijena fiksna, dok kod tehnika plaćanja za plasman cijene variraju ovisno o ponudama oglašivača. („Just what exactly are Trusted Feed programs?“, 2004.)

Ukoliko proizvođač ili distributer nemaju određene proizvode na zalihi, uz pomoć *trusted feeda* oni su u mogućnosti ukloniti privremeno nedostupne proizvode s web mjesta uz pomoć već navedenih meta opisa. („Just what exactly are Trusted Feed programs?“, 2004.)

## 6. Alati za optimizaciju web mjesta za potrebe CRM-a

U ovom će poglavlju biti detaljnije prikazani alati koji su korisni kod optimizacije web mjesta te način na koji isti funkcioniraju.

### 6.1. *Google Analytics*

*Google Analytics* je alat koji je dostupan u okviru *Google-a*, a pruža uvid u statističke podatke i ponašanje na web mjestu. Navedeni alat također pruža uvida u organski izvor prometa i praćenje korisnika, čime se može doći do informacija o tome koliko je konverzija izvršeno na temelju jednog organskog posjeta, kao i do informacija o broju posjeta web mjestu, o pregledniku pomoću kojeg je korisnik posjetio web mjesto, o vremenu zadržavanja na web mjestu te mnogo drugih važnih informacija koje kasnije mogu koristiti u unaprjeđenju web mjesta. (Google Support, 2021.)

U okviru marketinga i upravljanja odnosima s kupcima, postoji koncept lijevka za kupnju, koji se dijeli u tri koraka:

- Akvizicija- uključenje svijesti i stjecanja interesa korisnika
- Ponašanje – odnos korisnika s tvrtkom
- Konverzija- korisnik postaje kupac (Slika 11.) (Google Support, 2021.)



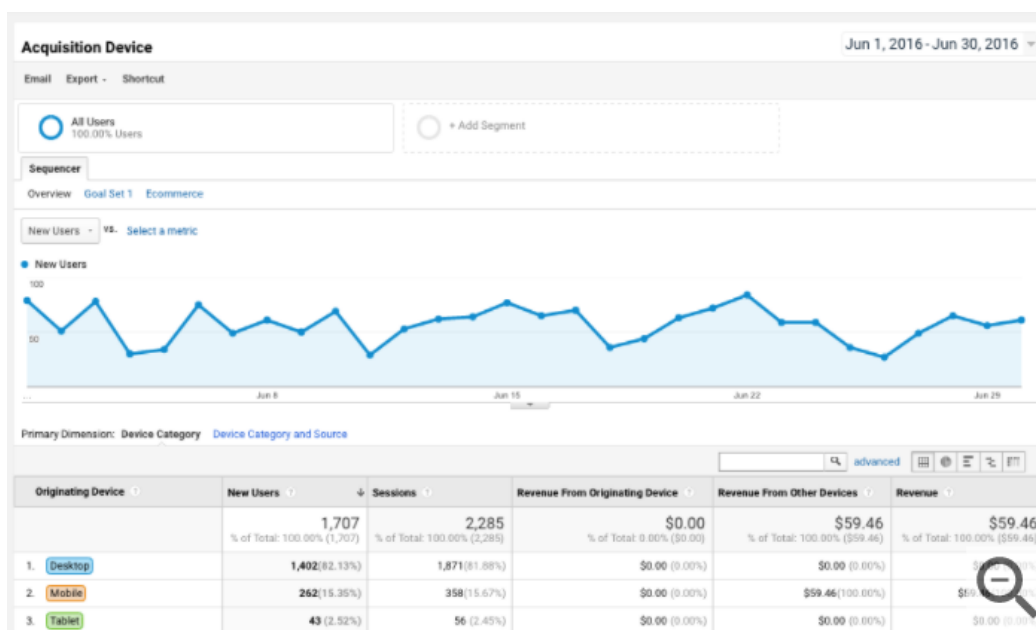
Slika 11:Koncept lijevka za kupnju (prema Google Support, bez dat.)

Pomoću ovog alata moguće je praćenje ponašanja koje je dovelo do kupnje, što potom ima utjecaj na privlačenje novih, potencijalnih kupaca. (Google Support, 2021.)

Kako je u Republici Hrvatskoj zbog pandemije uzrokovane koronavirusom online trgovina porasla za 13,9%, odnosno u periodu „lockdowna“ su se potrošačke navike kupaca promijenile pa se sve više ljudi odlučilo na online kupovinu, e- trgovine su se suočile s novim izazovom. (Majic, 2020.)

Webshop neke trgovine može postaviti cilj povećanja prodaje određenih artikala, a put do ostvarenja tog cilja krije se u digitalnoj analitici kojom se prikupljaju i analiziraju podaci iz reklamnih kampanja kako bi se usporedila njihova učinkovitost. Također je uz pomoć digitalne analitike stvoriti lojalnu publiku i uskladiti oglašavanja za svakog individualnog potrošača. („Horizont Solutions“, bez dat.)

Kako bi se ciljevi ostvarili, potrebno je prikupljati podatke o ponašanju korisnika i na temelju njih kreirati izvješća. Primjer mogućnost koju pruža ovaj alat je mjerenje akvizicije. Izvješće naziva „Uređaj akvizicije“ koje se generira u ovom alatu daje uvid u doprinos različitih uređaja privlačenju novih korisnika. U donjem dijelu slike vidljiv je redak za svaku kategoriju uređaja (stolno računalo, tablet i mobilni uređaj). (Slika 12.) („Horizont Solutions“, bez dat.)



Slika 12: Izvještaj Uređaj akvizicije, dostupan na: <https://support.google.com/analytics/answer/3123672?hl=hr>, bez dat.)

Sada se postavlja pitanje što je sve *Google Analyticsu* potrebno za pružanje informacija pomoću izvješća. To su HTTP zahtjev korisnika, informacije o pregledniku ili sustavu te kolačići.

HTTP zahtjev sadrži detalje o pregledniku i računalu koje podnosi zahtjev, a to su vrsta računala, preglednik, preporuka i jezik. Ti detalji korisni su *Google Analyticsu* prilikom kreiranja izvještaja karte, preglednika i preporučenih web lokacija.

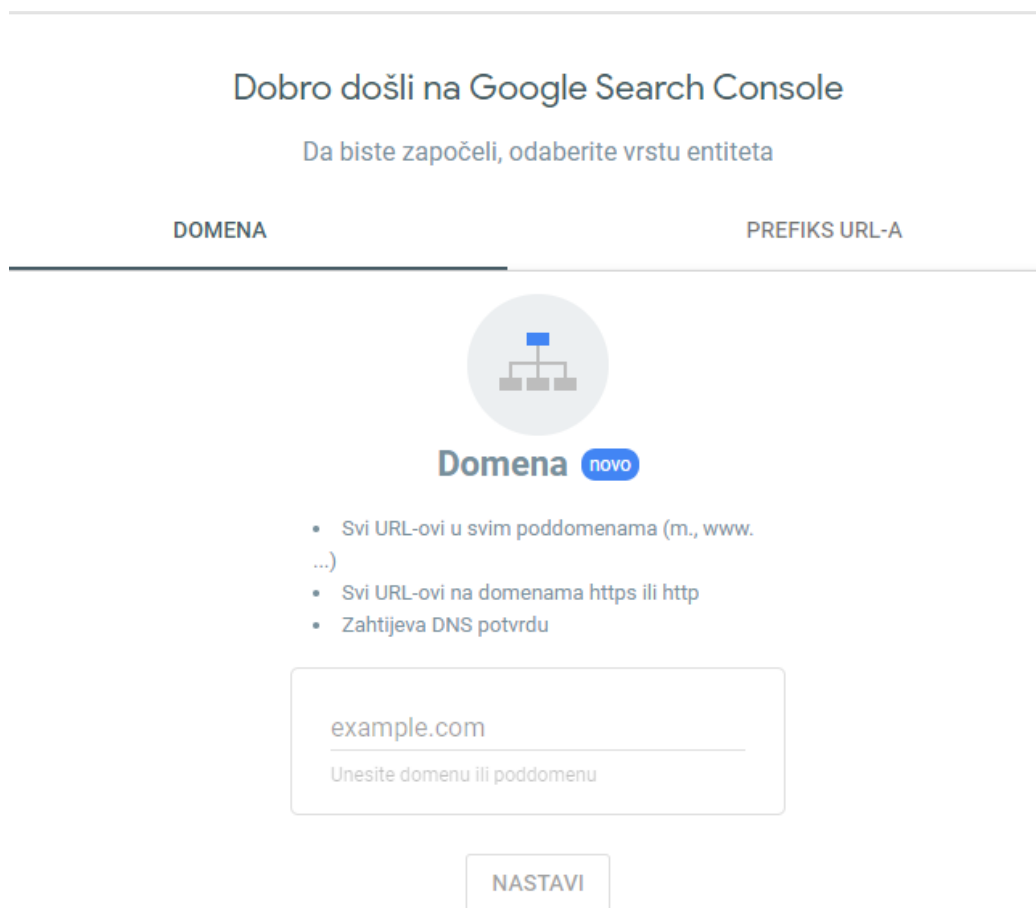
Kolačići (eng. *cookies*) koriste *Google Analyticsu* dobivanje korisničke sesije i svih informacija o oglasnoj kampanji iz zahtjeva na web stranici.

Kad se svi potrebni podaci prikupe, šalju se analitičkim poslužiteljima u popisu parametara priloženih zahtjevu, a potom se obrađuju i generiraju u izvješća. (Akhtar, 2020.)

## 6.2. Google Search Console

Alat koji predstavlja „nadopunu“ *Google Analyticsu* je alat *Google Search Console*, koji se do 20.5.2015. zvao *Google Webmaster Tools*, a koji daje uvid u način na koji *Google* vidi web mjesto i eventualne greške i poteškoće („*Google Webmaster Tools* postaje *Google Search Console*“, 2015).

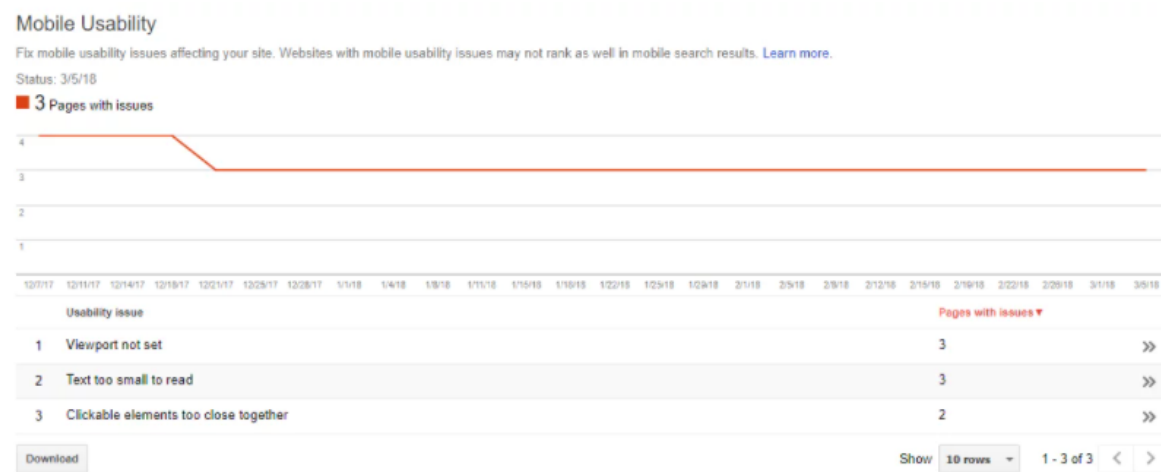
Kako bi se pokrenula analiza web mjesta pomoću ovog alata, prvo je potrebno posjetiti sljedeću web stranicu: <https://search.google.com/search-console/about>, potom se prijaviti s *Google* računom, po mogućnosti onim koji se koristi i za pristup *Google Analyticsu* i sljedeći korak je odabir vrste entiteta. (Slika 13.) („Neil Patel Digital“, bez dat.).



Slika 13: Odabir vrste entiteta u *Google Search Console*-u; prikazan na prijenosnom računalu

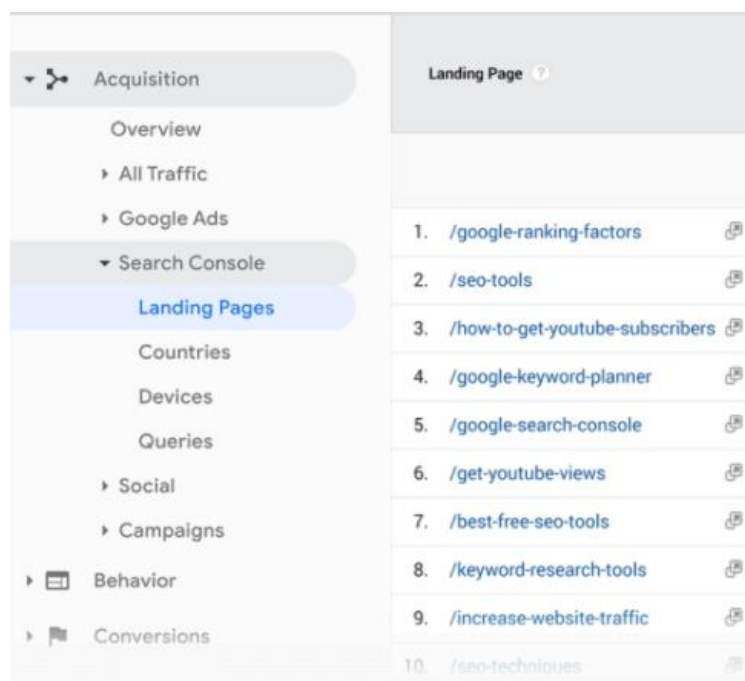
Sljedeći korak je verifikacija za pristup web mjestu, koju je moguće učiniti na čak sedam načina, a kad se to učini, uz pomoć ovog relativno jednostavnog, a efikasnog alata dobije se uvid u područja web mjesta kao što su pogreške indeksiranja, ukupni broj klikova, pojavljivanja, poziciju, stopu klicanja i upite korisnika na koje se web mjesto prikazuje u rezultatima pretraživanja, brzinu učitavanja, vizualnu stabilnost, popis povratnih i internih poveznica itd. Što se tiče već prethodno spomenutog responzivnog dizajna, u ovaj je alat uključena i opcija koja pokazuje kvalitetu upotrebe i učitavanja web mjesta na mobilnim uređajima, u isto vrijeme ukazujući i na eventualne poteškoće, kao što je veličina teksta ili blizina elemenata web mjesta. (Slika 14.) („Neil Patel Digital“, bez dat.).

Ukoliko postoji razlog za to, web mjesto se pomoću ovog alata može i privremeno, otprilike na 90 dana sakriti iz pretraživanja („Neil Patel Digital“, bez dat.).



Slika 14: Prikaz u *Google Search Console*-u, dostupan na: <https://neilpatel.com/blog/google-search-console-guide/> (bez dat.).

Kad su *Google Analytics* i *Google Search Console* povezani, moguće je vidjeti prikaz određinih web stranica, koje su navele korisnike na pristup web mjestu koje se želi ovim alatom optimizirati. (Slika 15.) („Neil Patel Digital“, bez dat.).



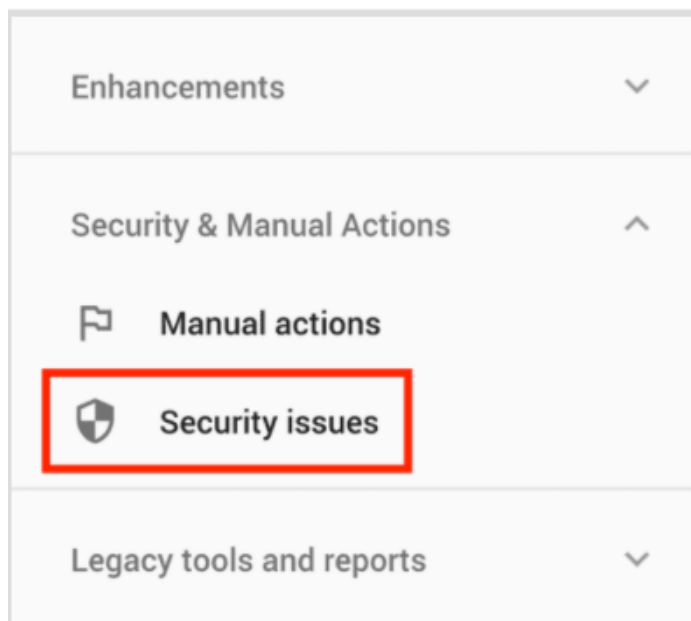
Slika 15: Prikaz u *Google Search Console-u*, dostupno na: <https://backlinko.com/google-search-console>, (bez dat.)

U ovom se alatu također mogu vidjeti podaci o ključnim riječima, primjerice, koliko je klikova bilo po određenoj ključnoj riječi na web mjestu. (Slika 16.) („Neil Patel Digital“, bez dat.).

Search Query ?	Clicks ?
	<b>38,149</b> <small>% of Total: 60.57% (62,975)</small>
1. [redacted]	<b>2,113</b> (5.54%)
2. [redacted]	<b>824</b> (2.16%)
3. [redacted]	<b>795</b> (2.08%)
4. [redacted]	<b>553</b> (1.45%)
5. [redacted]	<b>528</b> (1.38%)
6. [redacted]	<b>402</b> (1.05%)
7. [redacted]	<b>372</b> (0.98%)
8. [redacted]	<b>353</b> (0.93%)
9. [redacted]	<b>311</b> (0.82%)
10. [redacted]	<b>308</b> (0.81%)

Slika 16: Prikaz u *Google Search Console-u*, dostupno na: <https://backlinko.com/google-search-console>, prikazan na prijenosnom računaru (bez dat.)

Mogućnost koju alat nudi je i provjera sigurnosti, odnosno provjera čimbenika koji bi mogli naštetiti optimizaciji pomoću opcije Sigurnosni problemi, eng. *Security issues*. (Slika 17.) („Neil Patel Digital“, bez dat.).



Slika 17: Prikaz u *Google Search Console-u*, dostupno na: <https://backlinko.com/google-search-console>, prikazan na prijenosnom računalu (bez dat.)



## 6.3. Google Tag Manager

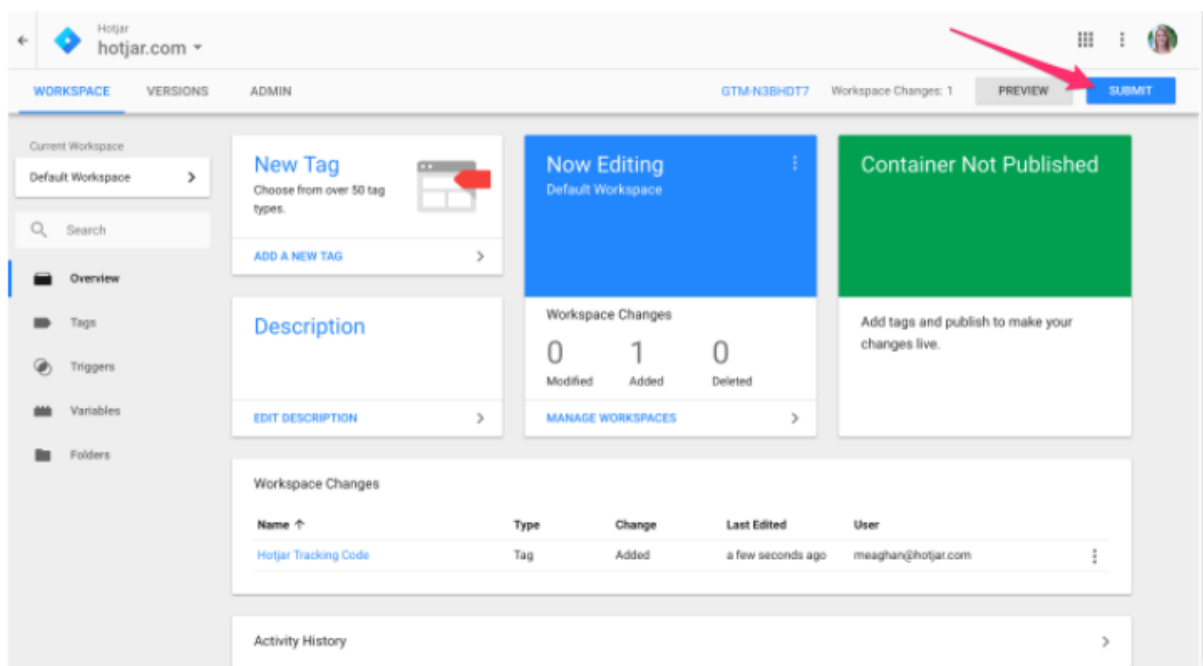
Google Tag Manager alat je koji funkcionira kao dodatak pretraživaču Chrome, a koji testira implementaciju Google Analyticsa i u realnom vremenu prikazuje greške s tracking kodovima ukoliko je implementacija neispravna. („39 alata za SEO i content marketing bez kojih u Arboni ne možemo“, 2017)

Ovaj alat omogućuje brzo i jednostavno ažuriranje oznaka i isječaka koda na web mjestu ili mobilnoj aplikaciji koji su namijenjeni analizi prometa te optimizaciji. Međutim, prvo je potrebno objasniti značenje oznake (eng. tag). Oznaka je isječak koda koji šalje podatke trećoj strani, primjerice, Google Analyticsu ili Google AdWords-u. (Arbona, bez dat.)

Google Tag Manager sastoji se od 3 dijela:

- Tagova- dijelovi koda koji se dodaju na web stranicu
- Triggera- definiraju kada se pojedina oznaka uključuje
- Varijabli- primaju i spremaju informacije koje ostala dva dijela Google Tag Managera koriste (Arbona, bez dat.)

Korisničko sučelje alata prikazano je na sljedećoj slici. (Slika 18.)



Slika 18: Korisničko sučelje alata Google Tag Manager, dostupno na: <https://www.arbona.hr/ostalo/cesta-pitanja/google-analytics-400/sto-je-google-tag-manager/673> (bez dat.)

Razloga za korištenje *Google Tag Managera* je mnogo, a u nastavku su neki od njih. Sve su njegove značajke besplatne.

Svako unošenje oznaka, promjene i provjere rada može se napraviti samostalno, bez potrebe za suradnjom s developerima, što dodatno štedi vrijeme.

Kako bi se dobio bolji uvid u analitiku, moguće je povezivanje s *Google Analyticsom*.

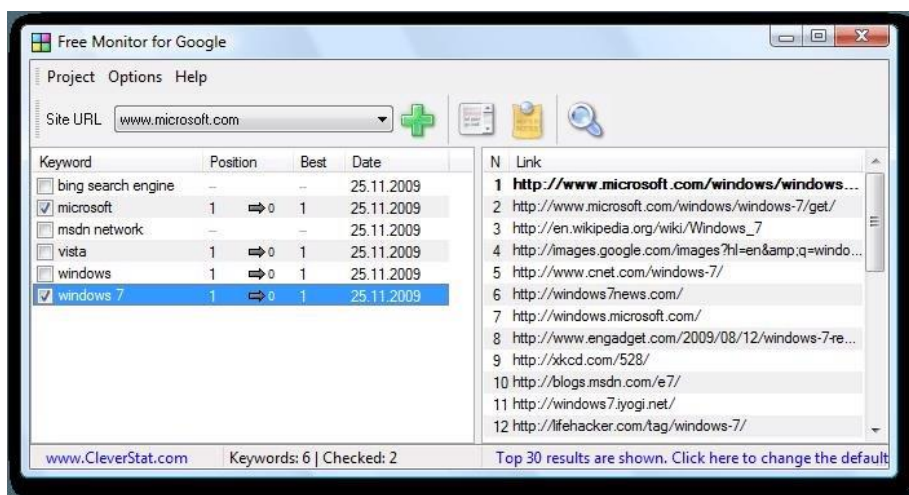
Sadrži sigurnosne značajke, kao što je dvostruka autentikacija preko lozinke i numeričkog koda putem poruke ili mobilne aplikacije.

Moguć je pristup *Google Tag Manageru* putem aplikacija na platformama Android i iOS. (Arbona, bez dat.)

## 6.4. Free Monitor for Google

Ovaj alat nije proizveo sam *Google*, no veoma je koristan kod praćenja pozicija web mjesta za koje se radi optimizacija.

Taj alat omogućuje prikaz pozicije web mjesta u određenoj zemlji po određenim ključnim riječima za samo nekoliko sekundi. Ovaj brz, jednostavan i kvalitetan alat pruža besplatan pregled pozicije na *Google*-u te olakšava odabir ključnih riječi, jer daje informacije da ista ključna riječ ne daje jednake rezultate u svakoj zemlji, pa time pomaže i prilikom izbora ključnih riječi. (Slika 19.) (Cleverstat, bez dat.).



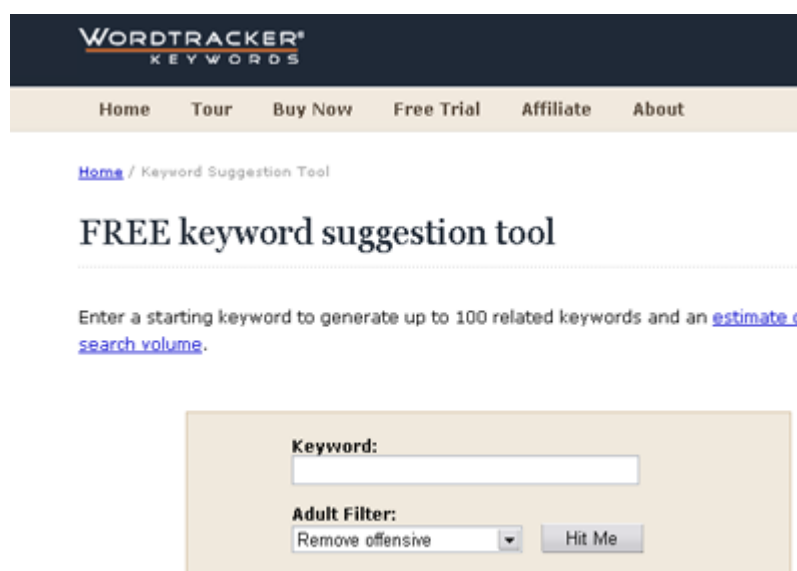
Slika 19: Prikaz alata Free Monitor for Google, dostupan na: <https://alternativeto.net/software/free-monitor-for-google/about/> (bez dat.)

Rezultati koje generira ovaj alat mogu biti prebačeni u HTML, a ono što ga čini bitnim alatom je to što odmah daje uvid koja se područja web mjesta trebaju razvijati uz prikaz ključnih riječi koje stvaraju promet i koje ga ne stvaraju. Uz navedeno, daje uvid u položaj u odnosu na konkurenciju te preporuke za poboljšanja. (Cleverstat, bez dat.).

## 6.5. *WordTracker*

*WordTracker* je alat koji služi kod odabira ključnih riječi, odnosno, pomaže kod identificiranja ključnih riječi i fraza koje su relevantne za određena web mjesta i koje će korisnici prilikom formiranja upita u web tražilici najvjerojatnije koristiti.

Taj alat pruža uvid u broj konkurentskih web mjesta koja koriste identične ključne riječi, kao i uvid u fraze i ključne riječi koje imaju najveći potencijal u aspektu ostvarivanja prometa na web mjestu. Prilikom korištenja prikazuje prvih 100 fraza ili ključnih riječi u tablici te pored svake broj puta koliko su bile pretražene u jednom danu. (Slika 20.) (Francescevic, 2007.)



The screenshot shows the top navigation bar of the WordTracker website with the logo 'WORDTRACKER KEYWORDS' and menu items: Home, Tour, Buy Now, Free Trial, Affiliate, and About. Below the navigation bar is a breadcrumb trail: Home / Keyword Suggestion Tool. The main heading is 'FREE keyword suggestion tool'. Below the heading is a text prompt: 'Enter a starting keyword to generate up to 100 related keywords and an estimate of search volume.' The form contains a 'Keyword:' input field, an 'Adult Filter:' dropdown menu set to 'Remove offensive', and a 'Hit Me' button.

Slika 20: Pregled alata *WordTracker*, dostupno na: <https://seekandhit.com/hr/oglasavanje/alati-za-generiranje-i-predlaganje-kljucnih-rijeci/> (2007.)

Postoji besplatna verzija i standardna verzija koja se plaća, no u svakom je slučaju alat od velike pomoći prilikom izdvajanja grupa ključnih riječi prilikom optimizacije web mjesta. (Francescevic, 2007.)

## 6.6. *Keyword Discovery*

*Keyword Discovery* je alat koji prikuplja podatke o ključnim riječima i frazama iz mnogo tražilica diljem svijeta te na temelju tih podataka sastavlja niz statistika pretraživanja. Kako bi omogućio ostvarenje konkretnog cilja, primjerice, povećanja prodaje, pronalazi točne ključne riječi ili fraze koje kupci koriste ne bi li pronašli određene proizvode ili usluge. (Slika 21.) (Francescevic, 2007.)

Na taj se način web mjesta koja ne sadrže te ključne riječi ili fraze niti ne prikazuju u rezultatima pretraživanja. Taj alat sadrži bazu od preko 36 milijardi pretraživanja sa preko 200 svjetskih pretraživača i uz sve navedene mogućnosti, pruža i mogućnost identificiranja pravopisnih pogrešaka i sezonskih trendova pretraživanja koja imaju utjecaj na ključne riječi i fraze. (Francescevic, 2007.)

Search Term	Total
seo	21861
seo tools	1947
seo taiji	617
seo services	487
seo software	422
professional seo	395
internet marketing seo	355
seo tool	330
tips on seo	265

Slika 21: Prikaz alata *Keyword Discovery*, dostupno na: <https://seekandhit.com/hr/oglasavanje/alati-za-generiranje-i-predlaganje-kljucnih-rijeci/> (2007.)

## 7. Problemi kod optimizacije

Prilikom optimizacije web mjesta, često se nailazi na probleme koji tu optimizaciju usporavaju, a najčešći su konkurencija, crni šešir (eng. *Black hat SEO*) i kutija pijeska (eng. *Sandbox*), o kojima će biti riječi u nastavku ovog rada.

### 7.1. Konkurencija

Ključ je uspjeha poslovanja neke organizacije postići konkurentnost na tržištu. Upravo se zbog postizanja što boljeg rejtinga i nadvladavanja konkurencije na tržištu koristi optimizacija web mjesta, kojom se nastoji postići bolji položaj web mjesta u rezultatima pretrage na web tražilici te veći broj posjetitelja web mjesta.

Kako bi bili korak ispred konkurencije, važno je kontinuirano raditi na napretku, prihvaćati inovacije, istraživati i težiti konstantnom poboljšanju.

### 7.2. *Black hat SEO*

Erceg navodi, da je crni šešir (eng. *Black hat*) metodologija koja uključuje tehnike i radnje koje nisu u skladu sa smjernicama pretraživača za webmastere. Drugim riječima, pod tim pojmom podrazumijevaju se ne-etične i amoralne radnje, koje su najčešće izvan okvira zakona ili osobe koje takve radnje izvode, odnosno hakeri koji infiltriraju određeni računalni sustav sa zlim namjerama, nazivaju se hakeri crnog šešira (eng. *black hat hackers*).

Svoje ne-etično djelovanje provode na način da prevare web tražilice kako bi rangirale određeno web mjesto više nego što bi isto trebalo biti rangirano, a to se postiže na način da skrivaju tekst od korisnika i omogućuju jedino robotima da taj tekst vide, kao i prikazivanje jedne web stranice korisnicima, a druge robotima ili neovlašteno i namjerno plagiranje sadržaja s drugih web mjesta na svoje i slično. (Erceg, 2018.)

## 7.3. Sandbox

*Google Sandbox* je *Googleov* algoritam čija je uloga onemogućavanje visokog rejtinga novim domenama, unatoč opsegu optimizacije.

Ideja koja stoji iza tog algoritma jest da nova web mjesta možda nisu toliko relevantna kao stara, pa je iz tog razloga prisutno ovo ograničenje. (Munjaj, 2019.)

Još uvijek nije poznata informacija što točno utječe na *Sandbox*, odnosno koji je razlog „upada“ u isti, no poznato je da je moguće preko noći nestati s tražilica i da je najčešći uzrok tome nova domena koja je upala u *Googleov* filtar i neko će vrijeme provesti u filtru. Pretpostavlja se, no to nije službeno potvrđeno od strane *Googlea*, da unutarnja optimizacija web mjesta i izgradnja poveznica mogu imati utjecaj na *Sandbox*.

Međutim, ukoliko postoji sumnja, moguće je provjeriti nalazi se web mjesto u *Sandboxu*. (Munjaj, 2019.)

Šarić navodi primjer na kojem će se objasniti način provjere za stranicu seo.hr (<https://www.seo.hr/>), koja se pokušava dovesti u top 10 za malo jače ključne riječi: web dizajn i izrada web stranica, za malo slabije: optimizacija, optimizacija web stranica i za generičnu ključnu riječ seo.hr.

To se može provjeriti uz pomoć *allinanchor* operatora u *Googlu* u kojeg se unesu sljedeći unosi:

- Allinanchor: web dizajn
- Allinanchor: izrada web stranica
- Allinanchor: optimizacija
- Allinanchor: optimizacija web stranica
- Allinanchor: seo.hr

Ista se stvar može provjeriti s još jednim operatorom, *allintitle*, uz pomoć sljedećih unosa:

- Allintitle: web dizajn
- Allintitle: izrada web stranica
- Allintitle: optimizacija
- Allintitle: optimizacija web stranica
- Allintitle: seo.hr (Šarić, bez dat.)

Nakon toga se preporuča provjeriti rejting na web tražilicama. Ukoliko se u ovim provjerama web mjesto nije pronašlo u top 50 ili niže, a očekivani ili željeni rejting je biti u top 50, riječ je samo o lošoj optimizaciji, no ako se web mjesto našlo prilikom provjera u top 50, a normalnim pretraživanjem svih tih ključnih riječi se nalazi iza top 100 ili se uopće ne može naći, tada je riječ o *Sandboxu*, koji zahvaća cijelu domenu u cijelosti, odnosno kompletno

web mjesto, a ne pojedinačnu web stranicu. Još jedan način za provjeru web mjesta, ako postoji sumnja na *Sandbox* je statistika posjeta. Ukoliko je web mjesto imalo po 10,20 ili 30 posjeta dnevno i ako se kontinuiranim praćenjem posjeta uočilo da je broj posjeta pao na 2 ili 5 posjeta na mjesec, taj se pad može pripisati *Sandboxu*. (Šarić, bez dat.)

Iako utječe na pad ranga web mjesta u rezultatima pretraživanja, ne smije utjecati na pad samopouzdanja i motivacije za optimizacijom web mjesta, već treba nastaviti raditi na poboljšanju stranica i izgradnji poveznica.



## 8. Zaključak

Optimizacija web mjesta neizostavan je dio poslovanja koje teži poboljšanju i modernizaciji. U tom procesu optimizacija za Internet pretraživače igra ključnu ulogu prilikom stvaranja prisutnosti web mjesta na Internetskim tražilicama. Sukladno tome, web mjesto koje je prisutno i visoko rangirano na web tražilicama, posjetit će veća količina ljudi, odnosno, potencijalnih kupaca, što utječe na rast i razvoj poslovanja, rast profita te mogućnost širenja poslovanja na međunarodna tržišta. Kako bi web mjesto bilo optimizirano, važno je kontinuirano raditi na prikupljanju, obradi i analizi podataka koji svoju važnost pronalaze u marketingu za tražilice, još jednom načinu optimizacije web mjesta s ciljem poboljšanja upravljanja odnosima s klijentima.

Nijedna od tehnika za optimizaciju navedenih u ovom radu ne pokazuje rezultate preko noći. Isto tako, teško je procijeniti u kolikom udjelu neka od tehnika naposljetku poveća zaradu organizacije. Tehnike se svakodnevno mijenjaju, nadopunjuju i razvijaju pa je stoga važno biti u koraku s vremenom i pratiti trendove, kako bi optimizacija bila uspješna. U svakom je pogledu važno pratiti i razumjeti ponašanje klijenata, pratiti i razumjeti ponašanje tehnologije, pratiti i razumjeti načine funkcioniranja web mjesta i web tražilica te na kraju krajeva, imati motivacije i strpljenja. Ponekad je potrebno vrlo malo vremena, dodataka i novca kako bi se određeno web mjesto optimiziralo do prihvatljive razine, ali i za to je potrebno imati odgovarajuća znanja i vještine.

Optimizacija je dugotrajan proces koji zahtijeva određena znanja i vještine opisane u ovom radu i proces u kojeg je nužno ulagati, jer se njime naposljetku povećava promet i prodaja, a samim time i zarada, krajnji cilj koji uvjetuje opstanak, rast i razvoj organizacija.

# Popis literature

1. Akhtar, A. (2020). *How does Google Analytics Work?* Preuzeto 26.7.2021. s <https://www.monsterinsights.com/how-does-google-analytics-work-beginners-guide/>
2. Apache Hadoop. (bez dat.). U Wikipedia. Preuzeto 24.7.2021. s [https://en.wikipedia.org/wiki/Apache\\_Hadoop](https://en.wikipedia.org/wiki/Apache_Hadoop)
3. Arbona (bez dat.) *Što je Google Tag Manager?* Preuzeto 26.7.2021. s <https://www.arbona.hr/ostalo/cesta-pitanja/google-analytics-400/sto-je-google-tag-manager/673>
4. „*Big Data tehnologija- što je i gdje se primjenjuje?* (2019). Preuzeto 24.7.2021. s <https://www.ekonomska-klinika.hr/2019/11/07/big-data-tehnologija-sto-je-i-gdje-se-primjenjuje/>
5. Camarce, J. *10 Best Free Link Building Tools to Help You Build Backlinks* [Blog post]. Preuzeto 13.7.2021. s <https://userp.io/free-link-building-tools/>
6. Chen H., Lihua Y., Peiquan J., Xiaowen L. (2010). *CT-Rank: A Time-aware Ranking Algorithm for Web Search*. Journal of Convergence Information Technology, 5(6), DOI:10.4156/jcit.vol5.issue6.10
7. Chen H., Lihua Y., Peiquan J., Xiaowen L. (2010). [Slika] *CT-Rank: A Time-aware Ranking Algorithm for Web Search*. Journal of Convergence Information Technology, 5(6), DOI:10.4156/jcit.vol5.issue6.10
8. Cleverstat (bez dat.) *Free monitor for Google*. Preuzeto 26.7.2021. s <https://cleverstat.com/en/google-monitor-query.htm>
9. Connell-Waite J. (2014). *A Simple Guide to Big Data*. Preuzeto 25.7.2021. s <https://www.linkedin.com/pulse/20140912112657-9245190-a-simple-guide-to-big-data-2014/>

10. David George, *The ABC of SEO, Search Engine Optimization Strategies*. Lulu Press, 2005. Dostupno 21.7.2021. na <http://books.google.hr>
11. *Detaljan vodič za tehnički SEO* [Brošura] (2019). Preuzeto 12.7.2021. s <https://www.markething.hr/vodic-za-tehnicki-seo/>
12. Dobrinić D. i ostali (2011) *Marketing i baze podataka*, Varaždin: Fakultet organizacije i informatike, CRODMA - Hrvatska udruga za direktni i interaktivni marketing, 278-299.
13. Entrepreneur EUROPE (bez dat.) *Unique Selling Proposition (USP)*. Preuzeto 12.7.2021. s <https://www.entrepreneur.com/encyclopedia/unique-selling-proposition-usp>
14. Erceg, M. (2018.) Optimizacija web mjesta kao element promotivnog spleta internetskog marketinga (Završni rad, Zagreb). Preuzeto s <https://repozitorij.foi.unizg.hr/islandora/object/foi:4049/datastream/PDF/download>
15. Fortin D. (2021.) *How to Create an SEO Strategy for 2021* [Template Included] [Blog post]. Preuzeto 13.7.2021. s <https://blog.hubspot.com/marketing/seo-strategy>
16. Francescevic, M. (2007.) *Alati za generiranje i predlaganje ključnih riječi* (dio drugi) [Blog post]. Preuzeto 26.7.2021. s <https://seekandhit.com/hr/oqlasavanje/alati-za-generiranje-i-predlaganje-kljucnih-rijeci/>
17. Gadget-info.com (bez dat.) *Razlika između web stranice i web mjesta*. Preuzeto 14.7.2021. s <https://hr.gadget-info.com/difference-between-web-page>
18. George Boole. (2021) U Encyclopaedia Britannica. Preuzeto 24.7.2021. s <https://www.britannica.com/biography/George-Boole>
19. Google Search Console [Slika] (bez dat.) Preuzeto 14.7.2021. sa <https://neilpatel.com/blog/google-search-console-guide/>
20. Google Support (bez dat.) *Welcome to Google Analytics for Beginners*. Preuzeto 25.7.2021. s <https://support.google.com/analytics/answer/6383002?hl=en>

21. *Google Webmaster Tools postaje Google Search Console* (2015). Preuzeto 14.7.2021. s <https://www.arbona.hr/blog/seo/google-webmaster-tools-postaje-google-search-console/203>
22. Google, Digitalna garaža, tečaj „Uvod u digitalni marketing“, dostupno na <https://learndigital.withgoogle.com/digitalnagaraza/course/become-searchable-online/lesson/41>
23. Google, Digitalna garaža, tečaj „Uvod u digitalni marketing“, dostupno na <https://learndigital.withgoogle.com/digitalnagaraza/course/become-searchable-online/lesson/47#!#%2F>
24. Google, Digitalna garaža, tečaj „Uvod u digitalni marketing“, dostupno na <https://learndigital.withgoogle.com/digitalnagaraza/course/promote-business-online/lesson/54#!#%2F>
25. Google, Digitalna garaža, tečaj „Uvod u digitalni marketing“, dostupno na <https://learndigital.withgoogle.com/digitalnagaraza/course/promote-business-online/lesson/55#!#%2F>
26. Grozdek, A. (2017). *Tehnologije za rad s Velikim podacima* (Diplomski rad, Zagreb). Preuzeto s <https://repositorij.pmf.unizg.hr/islandora/object/pmf%3A767/datastream/PDF/view>
27. Hadoop MapReduce framework [Slika] (bez dat.) Preuzeto 25.7.2021. s [https://www.researchgate.net/figure/Hadoop-MapReduce-framework\\_fig3\\_305716828](https://www.researchgate.net/figure/Hadoop-MapReduce-framework_fig3_305716828)
28. Horizont Solutions (bez dat.) *Alati za SEO analizu*. Preuzeto 14.7.2021. s <https://www.horizont.com.hr/alati-za-seo-analizu-10-blog>
29. Horvat, D. i Mundar, D. (2017.) *Rangiranje web stranica*. Osječki matematički list 17(2017), 51.-62.
30. Informatička tehnologija ITpedia (bez dat.) *BIG DATA: Vrste, karakteristike i prednosti*. Preuzeto 24.7.2021. s <https://hr.itpedia.nl/2017/08/29/big-data-soorten-kenmerken-en-voordelen/>

31. Izvješće Uređaj akvizicije [Slika] (bez dat.) Preuzeto 14.07.2021. sa <https://support.google.com/analytics/answer/3123672?hl=hr>
32. *Just what exactly are Trusted Feed programs?* (2004). Preuzeto 13.07.2021. s <https://www.webmoves.net/blog/news/just-what-exactly-are-trusted-feed-programs-383/>
33. *Kako napisati učinkoviti Google AdWords oglas* (2017). Preuzeto 12.07.2021. s <https://logobox.agency/kako-napisati-dobar-google-adwords-oglas/>
34. Koncept lijevka za kupnju [Slika] (bez dat.) Preuzeto 25.7.2021. s <https://support.google.com/analytics/answer/6383002?hl=en>
35. Krušec, L. (2017). *Kako radi Google i ostale tražilice?* [Blog post]. Preuzeto 21.7.2021. s <https://www.info-kon.hr/2017/02/11/kako-radi-google-i-ostale-trazilice/>
36. Macarol V. (2017). *Algoritmi za rangiranje web stranica* (Završni rad, Zagreb). Preuzeto 24.7.2021. s [https://www.bib.irb.hr/896435/download/896435.MacarolVedran\\_BA.pdf](https://www.bib.irb.hr/896435/download/896435.MacarolVedran_BA.pdf)
37. Majic, S. (2020, Rujan 14). *Hrvatska: Zbog pandemije online trgovina porasla za 13,9 posto*. Anadolu Agency. Preuzeto 25.7.2021. s <https://www.aa.com.tr/ba/ekonomija/hrvatska-zbog-pandemije-online-trgovina-porasla-za-13-9-posto/1972920#>
38. Marjanović D. (2015). *Podaci su zlato- šta je big data, šta su Hadoop i Spark, i kako krenuti sa njima?* [Blog post]. Preuzeto 25.7.2021. s <https://startit.rs/big-data-hadoop-apache-spark/>
39. Munjal, H. (2019). *Google Sandbox: Is it still affecting new sites in 2019?* Preuzeto 26.7.2021. s <https://www.searchenginewatch.com/2019/07/01/google-sandbox-2019/>
40. Neil Patel Digital (bez dat.) *A Beginner's Guide to Google Search Console* Preuzeto 14.7.2021. s <https://neilpatel.com/blog/google-search-console-guide/>
41. Opensource (bez dat.) *What is open source?* Preuzeto 25.7.2021. s <https://opensource.com/resources/what-open-source>

42. Papratović, N. (6.3.2018.) *Zašto je važno da vaše web stranice imaju responzivni dizajn?* [Blog post]. Preuzeto 4.7.2021. s <https://neomedia.hr/zasto-responsivni-dizajn-web-stranice-potreban/>
43. Russom, P. (2011). *TDWI Best Practices Report: Big Data Analytics*, 11-12.
44. Ružić, D., Biloš, A., Turkalj, D. (2014.) *E-marketing*, Osijek, Factum d.o.o.
45. SAS Institute (bez dat.) *Hadoop- What it is and why it matters*. Preuzeto 25.7.2021. s [https://www.sas.com/en\\_my/insights/big-data/hadoop.html](https://www.sas.com/en_my/insights/big-data/hadoop.html)
46. SEO.hr (bez dat.) *Meta tags- Meta tagovi*. Preuzeto 13.07.2021. s <https://www.seo.hr/blog/29/meta-tags-meta-tagovi/>
47. Softwise d.o.o. (bez dat.) *SEO i SEM*. Preuzeto 7.7.2021. s [https://softwise.hr/hr\\_HR/usluge/seo-i-sem](https://softwise.hr/hr_HR/usluge/seo-i-sem)
48. Šarić, B. (bez dat.) *Što je Google Sandbox?* Preuzeto 26.7.2021. s <https://www.seo.hr/blog/16/sto-je-google-sandbox/>
49. *Što je SEO i zašto je optimizacija bitna?* (2018). Preuzeto 13.07.2021. s <https://directdesign.hr/web-trendovi/sto-je-seo-i-zasto-je-optimizacija-bitna.html>
50. *Što su ključne riječi i gdje se koriste na webu?* (2017). Preuzeto 05.07.2021. s <https://www.womeninadria.com/sto-su-kljucne-rijeci-i-gdje-se-koriste-na-webu/>
51. *Što učiniti sa svim tim podacima?* (2020). Preuzeto 25.7.2021. s <https://www.fluentis.com/hr/blog/tehnoloski-trendovi/big-data-i-data-analytics/>
52. The Apache Software Foundation (bez dat.) *What is Apache Hadoop?* Preuzeto 25.7.2021. s <http://hadoop.apache.org/>
53. *Trideset devet alata za SEO i content marketing bez koji u Arboni ne možemo* (2017). Preuzeto s <https://www.arbona.hr/blog/seo/39-alata-za-seo-i-content-marketing-bez-kjih-u-arboni-ne-mozemo/667>

54. Ujević F. (bez dat.) *Postupci i tehnike dubinske analize podataka* (Kvalifikacijski rad).  
Preuzeto s  
[https://inf.uniri.hr/images/studiji/poslijediplomski/kvalifikacijski/Filip\\_Ujevic -  
\\_Postupci i tehnike dubinske analize podataka.pdf](https://inf.uniri.hr/images/studiji/poslijediplomski/kvalifikacijski/Filip_Ujevic_-_Postupci_i_tehnike_dubinske_analize_podataka.pdf)
55. *Ultimativni vodič Optimizacija za tražilice za početnike*. [Brošura] (bez dat.) Arbona  
d.o.o. Preuzeto 5.7.2021. s <http://arbona.info/ebook/ultimativni-vodic-seo.pdf>
56. Vector Space Model: Examples [Slika] (bez dat.) Preuzeto 24.7.2021. s  
[https://study.com/academy/lesson/vector-space-model-examples.html#quiz-course-  
links](https://study.com/academy/lesson/vector-space-model-examples.html#quiz-course-links)
57. Web Identitet (bez dat.) *Što je Search Engine Marketing (SEM)?* Preuzeto  
12.07.2021. s <https://www.webidentitet.com/sem-marketing/>
58. *Web stranica: medij za promociju i kredibilnost brenda* (2016). Preuzeto 24.7.2021. s  
<https://marketingfancier.com/web-stranica-domena-seo-hosting/>
59. *What is cloaking in SEO?* (2020). Preuzeto 13.07.2021. s  
<https://www.gsmresults.com/what-is-cloaking-in-seo/>

# Popis slika

Slika 1: Prva stranica rezultata pretraživanja pojma „search engine marketing“, prikazana na prijenosnom računalu .....	7
Slika 2: Jedan SERP isječak, prikazan na prijenosnom računalu .....	8
Slika 3: Graf vektorskog prostora (Vector Space Model: Examples, bez dat.) .....	12
Slika 4: Proces CT-Rank algoritma u dvije faze (prema Hong, Lihua, Peiquan, Xiaowen, 2011., str. 102). .....	15
Slika 5: Rezultat preprocesuiranja (prema Hong, Lihua, Peiquan, Xiaowen, 2011., str 103.).....	16
Slika 6: Algoritam koji kreira parove <ključna riječ, vrijeme> (prema Hong, Lihua, Peiquan, Xiaowen, 2011., str. 104.).....	16
Slika 7: Osnovne faze procesa dubinske analize podataka (CRISP-DM) (Ujević, bez dat.).....	21
Slika 8: Princip rada MapReduce-a (prema Hadoop MapReduce framework, bez dat.) .....	26
Slika 9. Povezana pretraživanja („posao Varaždin“) prikazana na prijenosnom računalu .....	34
Slika 10. Poziv na akciju na oglasu tvrtke Fer projekti ( <a href="https://www.fer-projekt.com/hr">https://www.fer-projekt.com/hr</a> ) prikazan na prijenosnom računalu .....	39
Slika 11:Koncept lijevka za kupnju (prema Google Support, bez dat.) .....	42
Slika 12: Izvještaj Uređaj akvizicije, dostupan na: <a href="https://support.google.com/analytics/answer/3123672?hl=hr">https://support.google.com/analytics/answer/3123672?hl=hr</a> , bez dat.) .....	43
Slika 13: Odabir vrste entiteta u Google Search Console-u; prikazan na prijenosnom računalu .....	45
Slika 14:Prikaz u Google Search Console-u, dostupan na: <a href="https://neilpatel.com/blog/google-search-console-guide/">https://neilpatel.com/blog/google-search-console-guide/</a> (bez dat.).....	46
Slika 15: Prikaz u Google Search Console-u, dostupno na: <a href="https://backlinko.com/google-search-console">https://backlinko.com/google-search-console</a> , (bez dat.) .....	47
Slika 16: Prikaz u Google Search Console-u, dostupno na: <a href="https://backlinko.com/google-search-console">https://backlinko.com/google-search-console</a> , prikazan na prijenosnom računalu (bez dat.)	47
Slika 17: Prikaz u Google Search Console-u, dostupno na: <a href="https://backlinko.com/google-search-console">https://backlinko.com/google-search-console</a> , prikazan na prijenosnom računalu (bez dat.)	48
Slika 18: Korisničko sučelje alata Google Tag Manager, dostupno na: <a href="https://www.arbona.hr/ostalo/cesta-pitanja/google-analytics-400/sto-je-google-tag-manager/673">https://www.arbona.hr/ostalo/cesta-pitanja/google-analytics-400/sto-je-google-tag-manager/673</a> (bez dat.).....	49



Slika 19: Prikaz alata Free Monitor for Google, dostupan na: <a href="https://alternativeto.net/software/free-monitor-for-google/about/">https://alternativeto.net/software/free-monitor-for-google/about/</a> (bez dat.).....	51
Slika 20: Pregled alata WordTracker, dostupno na: <a href="https://seekandhit.com/hr/oglasavanje/alati-za-generiranje-i-predlaganje-kljucnih-rijeci/">https://seekandhit.com/hr/oglasavanje/alati-za-generiranje-i-predlaganje-kljucnih-rijeci/</a> (2007.) .....	52
Slika 21: Prikaz alata Keyword Discovery, dostupno na: <a href="https://seekandhit.com/hr/oglasavanje/alati-za-generiranje-i-predlaganje-kljucnih-rijeci/">https://seekandhit.com/hr/oglasavanje/alati-za-generiranje-i-predlaganje-kljucnih-rijeci/</a> (2007.) .....	53

## Popis tablica

Tablica 1: Izrazi u formuli/metodi tf-idf .....	17
---	----