

Alati za pretraživanje sekundarnih internetskih informacija

Pfeifer, Bruno

Master's thesis / Diplomski rad

2024

Degree Grantor / Ustanova koja je dodijelila akademski / stručni stupanj: **Josip Juraj Strossmayer University of Osijek, Faculty of Economics and Business in Osijek / Sveučilište Josipa Jurja Strossmayera u Osijeku, Ekonomski fakultet u Osijeku**

Permanent link / Trajna poveznica: <https://urn.nsk.hr/urn:nbn:hr:145:562871>

Rights / Prava: [In copyright](#)/[Zaštićeno autorskim pravom.](#)

Download date / Datum preuzimanja: **2025-01-31**



Repository / Repozitorij:

[EFOS REPOSITORY - Repository of the Faculty of Economics in Osijek](#)



Sveučilište Josipa Jurja Strossmayera u Osijeku
Ekonomski fakultet u Osijeku
Sveučilišni diplomski studij (Poslovna informatika)

Bruno Pfeifer

Alati za pretraživanje sekundarnih internetskih informacija

Diplomski rad

Osijek, 2024.

Sveučilište Josipa Jurja Strossmayera u Osijeku
Ekonomski fakultet u Osijeku
Sveučilišni diplomski studij (Poslovna informatika)

Bruno Pfeifer

Alati za pretraživanje sekundarnih internetskih informacija

Diplomski rad

Kolegij: Marketing – informacijski sustav

JMBAG: 00102227169

e-mail: bruno.fajfa@gmail.com

Mentor: Prof. dr. sc. Antun Biloš

Osijek, 2024.

Josip Juraj Strossmayer University of Osijek
Faculty of Economics and Business in Osijek
Graduate study (Business informatics)

Bruno Pfeifer

Secondary Data Research Tools on the Internet

Graduate paper

Osijek, 2024.

IZJAVA
O AKADEMSKOJ ČESTITOSTI,
PRAVU PRIJENOSA INTELEKTUALNOG VLASNIŠTVA,
SUGLASNOSTI ZA OBJAVU U INSTITUCIJSKIM REPOZITORIJIMA
I ISTOVJETNOSTI DIGITALNE I TISKANE VERZIJE RADA

1. Kojom izjavljujem i svojim potpisom potvrđujem da je (navesti vrstu rada: završni / diplomski / specijalistički / doktorski) rad isključivo rezultat osobnoga rada koji se temelji na mojim istraživanjima i oslanja se na objavljenu literaturu. Potvrđujem poštivanje nepovredivosti autorstva te točno citiranje radova drugih autora i referiranje na njih.
2. Kojom izjavljujem da je Ekonomski fakultet u Osijeku, bez naknade u vremenski i teritorijalno neograničenom opsegu, nositelj svih prava intelektualnoga vlasništva u odnosu na navedeni rad pod licencom Creative Commons Imenovanje – Nekomercijalno – Dijeli pod istim uvjetima 3.0 Hrvatska.
3. Kojom izjavljujem da sam suglasan/suglasna da se trajno pohrani i objavi moj rad u institucijskom digitalnom repozitoriju Ekonomskoga fakulteta u Osijeku, repozitoriju Sveučilišta Josipa Jurja Strossmayera u Osijeku te javno dostupnom repozitoriju Nacionalne i sveučilišne knjižnice u Zagrebu (u skladu s odredbama Zakona o visokom obrazovanju i znanstvenoj djelatnosti, NN 119/2022).
4. izjavljujem da sam autor/autorica predanog rada i da je sadržaj predane elektroničke datoteke u potpunosti istovjetan sa dovršenom tiskanom verzijom rada predanom u svrhu obrane istog.

Ime i prezime: Bruno Pfeifer

JMBAG: 0010222716

OIB: 00767942044

e-mail za kontakt: bruno.fajfa@gmail.com

Naziv studija: Poslovna informatika

Naslov rada: Alati za pretraživanje sekundarnih internetskih informacija

Mentor/mentorica rada: Antun Biloš

U Osijeku, 2024. godine.

Potpis 

Sažetak

Internetski informacijski prostor, je virtualna knjižnica koja pohranjuje tekstove, slike, videozapise, animacije i glazbene dokumente povezane putem poveznica. Korisnici mogu pristupiti tim resursima s bilo kojeg mjesta putem internetskih preglednika. Internet je globalna mreža koja prenosi podatke pomoću različitih medija. Putem interneta moguće je slati elektroničku poštu i razmjenjivati poruke putem aplikacija. Pretraživanje na internetu je popularna aktivnost koja uključuje odabir tražilice, postavljanje upita i pregledavanje rezultata. Google je najpoznatija tražilica, ali postoje i druge popularne tražilice poput Yahooa, Binga, Navera i Yandexa. Sekundarni podaci su prikupljeni podaci iz već postojećih istraživanja, dok se primarni podaci stvaraju tijekom istraživanja. Tražilice koriste programe koji pretražuju web-prostor i prikupljaju relevantne informacije. Iako je Google najrasprostranjenija tražilica na svijetu, postoje alternativne tražilice poput onih u Rusiji i Kini. U Rusiji je popularan Yandex, koji nudi pretragu na ruskom jeziku i ima 55% udjela u korištenju tražilica u toj zemlji. Također, Yandex je i tehnološka firma i medijska kompanija. U Kini je Google zabranjen, pa se koristi tražilica Baidu. Baidu se bavi razvojem umjetne inteligencije i samovozećih automobila. Kina ima vlastiti internetski prostor s vlastitim jezikom, kartama, medijima i enciklopedijom. Prema dubini internetskog prostora, postoji površinski (engl. *surface*) web-prostor što bi bila strana interneta koja se svakodnevno koristi u različite svrhe pretraga informacija. Na kraju, postoje tamni (engl. *dark*) web-prostor i duboki (engl. *deep*) web-prostor kao različiti dijelovi interneta. *dark web* je namjerno sakriveni dio interneta koji sadrži ilegalne aktivnosti. Pristup *dark webu* moguć je putem posebnih preglednika i tražilica kao što je Tor. S druge strane, *deep web* obuhvaća sve što nije indeksirano i nedostupno putem uobičajenih tražilica koje nisu u mogućnosti otvoriti podatke kao što su osobni podaci, medicinski nalazi, bankovne financije, OIBi.

Ključne riječi: Sekundarna informacija, internetski informacijski prostor, tražilice, indeksiranje, alati za pretraživanje informacija

Abstract

The Internet information space is a virtual library that stores texts, images, videos, animations, and music documents linked through hyperlinks. Users can access these resources from anywhere via web browsers. The Internet is a global network that transmits data through various media. It allows for sending emails and exchanging messages through applications. Internet search is a popular activity involving choosing a search engine, entering queries, and reviewing results. Google is the most well-known search engine, but there are other popular search engines like Yahoo, Bing, Naver, and Yandex. Secondary data are collected from existing research, while primary data are generated during research. Search engines use programs to crawl the web and gather relevant information. Although Google is the most widely used search engine globally, there are alternatives like the ones from Russia and China. Yandex is popular in Russia, offering searches in the Russian language and holding a 55% market share. It is also a technology and media company. In China, where Google is banned, Baidu is the most popular search engine. Baidu focuses on artificial intelligence development and autonomous vehicles. China has its own internet space with its language, maps, media, and encyclopedia. In terms of internet depth, there is the surface web, which is the everyday part of the internet used for various information searches. Lastly, there are the Dark web and Deep web as different parts of the internet. The Dark web intentionally hides illegal activities and can be accessed through specialized browsers and search engines like Tor. On the other hand, the Deep web encompasses unindexed and unavailable content for regular search engines, such as personal data, medical records, financial information, and identification numbers.

Keywords: Secondary data, world wide web, search engine, indexing, data search tools

SADRŽAJ

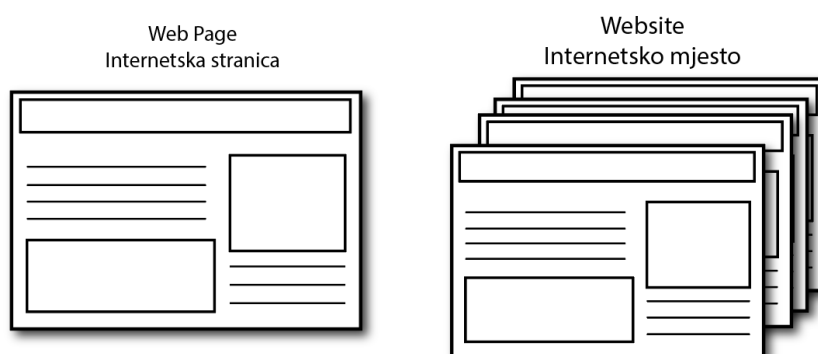
1. Uvod	1
2. Internetski informacijski prostor	2
2.1. Kako pristupiti informacijskom prostoru – <i>Webu</i>	3
2.2. Sekundarna informacija.....	4
2.3. Primarna informacija.....	4
3. Alati za pretraživanje	6
3.1. Tražilice	6
3.2. Metatražilice	7
3.3. Specijalizirane tražilice	8
3.4. Agregatori vijesti	9
3.5. Web scraping	9
4. Metode pretraživanja internetskog prostora	11
4.1. Indeksiranje	11
4.2. Ključne riječi	12
4.3. Operatori pretraživanja.....	14
4.4. Napredne tehnike pretraživanja	16
4.5. Algoritam pretraživanja.....	16
4.6. Horizontalno i vertikalno pretraživanje.....	17
5. Testiranje i analitika alata za pretraživanje internetskog prostora	19
5.1. Svjetski poznati pojmovi	20
5.2. Geografska područja	24
5.3. Edukacijske ustanove	28
5.4. Krićka analiza	30
6. Druge velike tražilice i internetski sustavi	31
6.1. Yandex.....	31
6.2. Baidu	31
7. Tamna strana interneta	33
7.1. Dark web	33
7.2. <i>Tor</i> preglednik.....	34
8. Zaključak	36
Literatura	37
Popis slika	42
Popis tablica	43

1. Uvod

Cilj diplomskog rada je analizirati opširnu temu alata za pretraživanje sekundarnih internetskih informacija, stvoriti kontekst razumijevanja i približiti čitatelju kako i na koji način te informacije može prikupljati. Prvi dio rada biti će postavljanje temelja što je internetski informacijski prostor, u odnosu na internet, te razlika između internetske stranice i internetskog mjesta ili sjedišta. Kako u naslovu stoji pojam sekundarne informacije, povući će se jasna linija između primarne i sekundarne informacije, kako se prikupljaju i što su njihove prednosti i nedostaci. Doticanje samih internetskih tražilica je glavni dio, gdje će biti obrađene najpoznatije tražilice interneta. U radu se prikazuje kako te tražilice funkcioniraju prilikom korištenja u svrhe svakodnevnog pretraživanja informacija. Njihova količina korištenja i koje tražilice se koriste najviše u kojim dijelovima svijeta. Obrađuju se osnovni koncepti kao što su ključne riječi, indeksiranje, i *Web crawling*, što je interakcija korisnika koji upisuje te riječi i efikasnost tražilice i njenih botova koji pronalaze te ključne riječi, zatim indeksiraju, te šalju rezultate natrag korisniku. Također analizirati će se 3 najpoznatije i korištenije tražilice 2023. godine odnosno Google, Bing i Yahoo. Detaljniji statistički podaci biti će izvučeni iz konteksta rezultata te tri tražilice, gdje će biti uspoređene u tablicama radi lakšeg čitanja i razumijevanja dobivenih resursa analize. Odabranih 10 ključnih riječi koristiti će se u testiranju i analizi Googlea, Binga i Yahooa, za prikaz razlike u količini informacija koje te tri tražilice mogu generirati. U radu se pozornost posvećuje i specifičnostima tržišta na kojima dominiraju drugačije tražilice i tehnološke kompanije kao što su Yandex u Rusiji ili Baidu u Kini. Naposljetku rad se fokusira na specifične dijelove internetskog informacijskog prostora odnosno *deep web* i *dark web*. Rad se sastoji od 8 poglavlja. Nakon prvog, uvodnog poglavlja, u drugom se opisuje internetski informacijski prostor i specifičnosti dostupnih informacija. Alati za pretraživanje internetskog informacijskog prostora se obrađuju u trećem poglavlju rada, a povezane metode i tehnike pretraživanja u četvrtom. U petom poglavlju prikazuje se testiranje i usporedna analiza najpopularnijih internetskih tražilica i njihovih mogućnosti. Šesto poglavlje fokus stavlja na dodatne popularne tražilice i specifična tržišta, a tamna strana internetskog informacijskog prostora analizira se u sedmom poglavlju. U završnom osmom poglavlju prikazuju se osnovni zaključci rada.

2. Internetski informacijski prostor

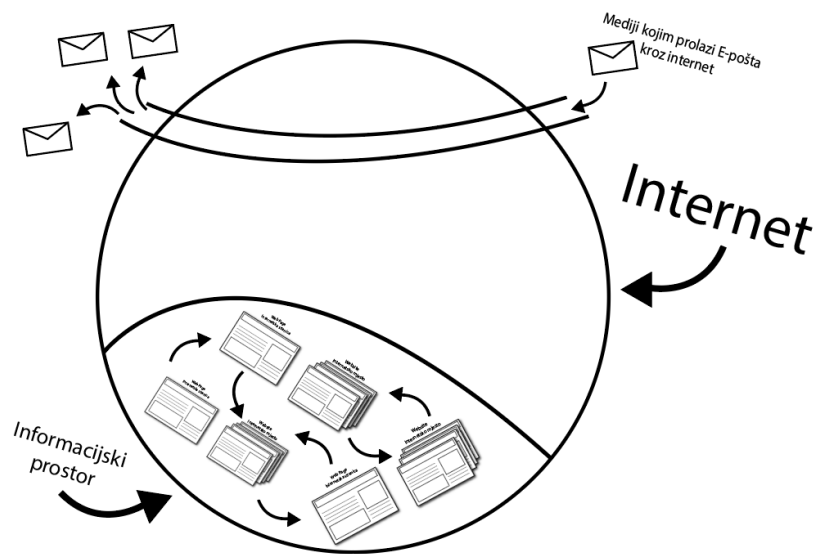
Internetski informacijski prostor je skup svih informacija koje su publicirane internetskim putem. To je virtualna knjižnica ili skladište dokumenata i web-resursa kojima se pristupa putem interneta. Kako je internet sačinjen od brojnih servisa, najznačajniji dio internetskih dostupnih informacija čini *world wide web* ili WWW, odnosno web-prostor. Informacijski prostor je ustvari kolekcija tekstualnih stranica, digitalnih fotografija, videa, animacija, digitalnih komunikacija, glazbenih dokumenata. Posebnost toga stoji u načinu na koji su međusobno povezani. Temelj internetskog informacijskog prostora je *web-stranica* (*Web page*), skupina *web-stranica* čini internetsko mjesto (*Website*). Svako internetsko mjesto unutar sebe ima poveznice, klikajući na poveznice korisnik se kreće unutar internetskog mjesta na druge *web-stranice* (*Web pages*). Osim kretanja korisnika po stranicama unutar internetskog mjesta, korisnik može naići na poveznicu koji vodi na drugo internetsko mjesto, a koje opet ima svoje poveznice sa kojima se kreće na tom internetskom mjestu. Terminologija *Web* tu je sa razlogom, internetski prostor dizajniran je da se lako pristupi različitim internetskim resursima gdje god se u internetskom prostoru korisnik nalazi ili sa bilo kojeg internetskog mjesta polazi (Vučina, 2006).



Slika 1: Vizualni prikaz web-stranice i internetskog prostora
Izvor: Autorski rad

„Prije 20 do 30 godina, postojalo je nešto što se moglo uzeti zdravo za gotovo; moglo se ući u javnu knjižnicu, otvoriti referentnu knjigu i pronaći bilo koje informacije o gotovo bilo kojoj temi koju osoba želi. Što danas uzimamo zdravo za gotovo je da možemo sjediti kući i za bilo kojim računalom na svijetu i pristupiti internetskoj knjižnici informacija, koja je daleko moćnija od bilo koje javne knjižnice u fizičkom obliku i nosi veću količinu informacijskih resursa. Život 21. stoljeća ovisan je o *Webu*, jedan od najvećih izuma svih vremena, no također još uvijek nov.“ (Woodford, 2023).

Internet s druge strane je globalno povezani mrežni sustav koji funkcionira na način da podatke prenosi putem različitih vrsta medija. Povezivanjem na mrežni sustav, korisnik dolazi do željenih informacija koje se nalaze unutar tog sustava, te se putem poveznica kreće kroz internetski prostor. Slanje elektroničke pošte ne zahtjeva *Web* već samo internet. „Internet šalje napisane riječi kroz telefonsku liniju do drugih korisnika. Kada razgovarate s nekim online, vrlo vjerojatno koristite i internet - jer je to *Net* koji razmjenjuje vaše poruke naprijed i nazad. Ali kada ažurirate blog ili tražite informacije putem *Googlea* kako bi se pronašla neka tražena informacija tada se koristi Web putem Interneta.“ (Woodford, 2023).



Slika 2: Jednostavan prikaz informacijskog prostora i medija za komuniciranje
Izvor: Autorski rad

2.1. Kako pristupiti informacijskom prostoru – *Webu*

Informacijskom prostoru i internetu općenito pristupamo preko internetskih preglednika. Preglednici su posebne aplikacije koje nakon instalacije na računalo omogućuju pristup virtualnom moru informacijskih i medijskih resursa. Putem preglednika, povlače se hipertekstualni dokumenti koji se strukturirano slažu na korisnikov uređaj u trenutku kada je zatražio pristup nekoj stranici ili direktnim klikom na poveznicu. Preglednik dešifrira pozadinu traženog dokumenta, te ga pretvara iz strojnog jezika u ispisani dizajn, i sadržaj traženog dokumenta. Najviše korišteni preglednici su Chrome, Safari, Edge, Opera, Firefox i Samsung Internet (Oberlo, 2023). Prema statističkim prezentaciji rezultata Oberloa i statističkom istraživanju Statcountera, u 2023. godini 62,85% korisnika sa pristupom na internet koristi Google Chrome preglednik, zatim Safari preglednik je na drugom mjestu, koristi ga 20,72% internetske populacije, što je znatna razlika od 42,13% između Chroma i Safarija. Na treće

mjesto dolazi Microsoft Edge sa 5,31% korištenja, Opera 2,82%, te prije korišteniji no i dalje jedan od najpoznatijih preglednika, Firefox se nalazi na 5. mjestu sa visinom korištenja od 2,77% populacije i Samsung internet, iako tematski dobro odrađen za Samsungove korisnike, nije toliko zaživio, pa ima stopu od 2,5% korištenja od populacije (Oberlo, 2023).

2.2. Sekundarna informacija

„U jednostavnom obliku, sekundarna informacija je svaki podatak koji nije prikupljen od strane autora ili analiza podataka prikupljena od nekog drugog“ (Boslaugh, 2007). Prikupljene informacije koje nisu prvi put u korištenju, odnosno informacije koje je netko već prikupio u prijašnjem razdoblju čine sekundarnu informaciju. Prikupljena na temelju neke druge potrebe, no može imati korist ukoliko korisnik ima relevantnu potrebu za njegovo trenutno istraživanje. Sekundarna informacija ne znači ništa loše, ne mora značiti da su podatci istekli ili da uopće više nisu od koristi. Primjer; pisanje završnih i diplomskih radova je istraživanje svih dostupnih sekundarnih informacija, kako bi se složila željena struktura i smisao tog rada, no ukoliko rad ima primarno istraživanje, te je korisnik provodio to prikupljanje podataka prvi put, ta informacija je primarna, jer se koristi prvi puta i autor rada ju je sam provodio i bilježio. Također, moguće je provoditi identično istraživanje u dva različita vremenska razdoblja, primjerice u 2007. i 2018. godini. Dobiveni podaci ne moraju biti identični. Moglo je doći do promjena u rasponu između te dvije godine. Stoga korištenje sekundarnih informacija i dalje znači pronaći najbolju i ažurnu informaciju. „Sekundarna informacija može uključivati prijašnje prikupljene podatke koji se smatraju da će biti korisni jednog dana, kada bi se ponovo mogli koristiti za neko drugo pitanje.“ (Vartanian, 2010). Prikupljanje sekundarnih informacija treba biti u okviru provjerenih podataka, znanstvenih radova, pisanih javnih članaka, knjiga, statističkih državnih institucija. Treba pažljivo birati i spajati sekundarne podatke sa određenom temom koja se obrađuje, kako se ti podatci ne bi krivo interpretirali, nakon čega nastaje problem gdje informacija postepeno gubi svoje prvobitno značenje što dulje kruži.

2.3. Primarna informacija

Primarna informacija za razliku od sekundarne se prikuplja u trenutku istraživanja koje kreira sami autor istog. Tehnike prikupljanja primarnih informacija uključuju ankete, intervjue, eksperimente, posebno napravljeno za razumijevanje i rješavanje problema unutar aktualnog istraživanja. Primarno istraživanje uz sebe kreira i veće troškove u poslovnom okruženju.

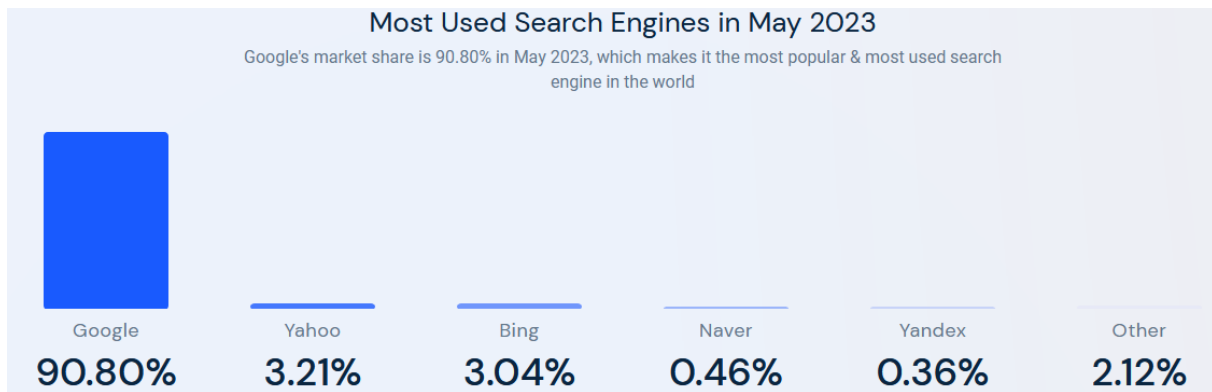
Istraživanje tržišta može zahtijevati cijeli tim ljudi koji će raditi isključivo na istraživanju primarnih potrebnih informacija za proširenje. Dok istraživanje sekundarnih podataka, temeljeni na prijašnjim istraživanjima određenog tržišta može rudarenjem podataka i organizacijom istih napraviti i jedan radnik. „Dok su primarni podaci izrazito dobri, budući da pružaju pouzdanost i transparentnost o fenomenima na koje istraživači usmjeravaju pažnju, također je vjerojatno da su rezultat napornog rada, skupi za prikupljanje i u konačnici ograničeni.“ (Cunha, Martins i Serra, 2018).

3. Alati za pretraživanje

Pretraživanje na internetu je druga najpopularnija aktivnost, odmah je iza aktivnosti koje su vezane za slanje i primanje e-pošte, odnosno direktne komunikacije populacije na internetu (Central Statistics Office, 2020). Pretraživanje se sastoji od 3 koraka ili procesa. Prvo treba odabrati tražilicu, koju korisnik odabire po svom ukusu, vještinama korištenja i potrebi pretraživanja, zatim slijedi sastavljanje upita, što znači, upisuje željenu informaciju ili temu u tražilicu. Te naposljetku dolazi pregledavanje rezultata. Rezultate koje pretraživač vraća sadrže dovoljno informacija za korisnika da odluči jesu li stranice relevantne za njegovo pretraživanje. (Meng i T.Yu, 2011.)

3.1. Tražilice

Tražilica je alat koji se koristi na svakodnevnoj bazi. To su servisi za pretraživanje sadržaja koji se nalazi na internetskom prostoru. Funkcioniraju na način da prikupljaju podatke sa mrežnih stranica, u to se uključuju programi koji automatizirano posjećuju pojedine stranice, pregledavaju dostupne sadržaje i upisuju pronađeno u vlastitu bazu podataka. Za skupljanje informacija s mrežnih stranica koriste se takozvani „pauci“, (engl. *spider* i dodatni nazivi *crawler*, *worm*) (Google Search Central, 2023). To su računalni programi kojima tražilica pretražuje dostupan internetski informacijski prostor. Ti računalni programi ne pretražuju internet u cijelosti, već samo dio koji su vlasnici internetskih mjesta dozvolili da se pretražuje. Kada korisnik upiše traženi pojam u tražilicu, ona pretražuje svoju bazu podataka i nudi relevantne rezultate. Takav proces se odvija u manje od jedne sekunde, te je vrijeme potrebno do pronalaženja rezultata prikazano zajedno sa količinom dobivenih rezultata. Najpopularnija tražilica na svijetu je Google, osnovala su je dva studenta, Larry Page i Sergey Brin 1997. godine, i njihov moto glasi „Googleova misija je, kako kažu, organizirati informacije svijeta te ih učiniti korisnima“ (Periera, 2023). Vrijedni spomena su Yahoo i Bing, koji će zajedno sa Googleom biti testirani u analitičkom dijelu rada, radi istraživanja koliko informacija i rezultata pridonose svaka od pojedine tražiice. Zatim slijedi tražilica Naver i na posljetku, Yandex. Yandex koristi isti princip tražilice kao ostale, no striktno povezano sa ruskim internetskim informacijskim prostorom.



Slika 3: Statistički prikaz razine korištenja tražilica na internetu 2023. godine
Izvor: SimilarWeb (2023).

3.2. Metatražilice

Metatražilice ili pod drugim nazivom agregatori su tražilice koje šalju zahtjeve prema nekoliko klasičnih tražilica i agregiraju rezultate u jednu veliku glavnu listu ili kategoriziraju rezultate po tražilici koju su obradili, odnosno tražili preko nje (Wordstream, 2023). Glavna prednost korištenja metatražilice je količina informacija koju korisnik može zadobiti u vrlo kratkom roku, sa time i veći izbor rezultata među kojima može birati koji mu se čini ili je najrelevantniji. Loša strana je u tome što su ovakve tražilice i dalje vrlo općenite. Pretraživanje treba biti jednostavno, s jednostavnim i ciljanim ključnim riječima. Neka veća i detaljnija pretraživanja se i dalje treba obavljati sa običnim tražilicama. S obzirom da metatražilica vuče rezultate od više običnih, nekada se događa preklapanje, odnosno više identičnih rezultata (duplikata) se korisniku može pojaviti. Kako bi se to izbjeglo, metatražilicu treba koristiti na način da pronađemo koja tražilica daje najviše i najrelevantnije rezultate, te zatim se prebaciti na tu tražilicu i nastaviti detaljnije pretraživanje. Također više vrsta metatražilica postoje koje i obavljaju malo drugačije algoritme jedna od druge. „Broj pronađenih dokumenata dobivenih kao rezultat pretraživanja uz pomoć metatražilice uvijek je manji od zbroja dokumenata koje biste dobili kao rezultat pretraživanja pretražujući putem nekoliko tražilica. Ipak, metatražilice su korisne za usporedbu tražilica. Tako možete vidjeti da, npr. Google daje kvalitetnije rezultate pretraživanja od Lycosa, pa ćete pretraživanje nastaviti pretražujući putem Googlea.“ (Vučina, 2006).

Tablica 1: Primjeri metatražilica za pretraživanje internetskog prostora i njena obilježja

Dogpile – kada korisnik želi skup rezultata od svih glavnih običnih tražilica
Ithaki – kada se želi agregirati rezultat koji će biti rangiran unutarnjim algoritmom
Polymeta – vrhunski inteligentan meta-pretraživač koji prepoznaje kolokvijalni jezik i koristi algoritam za organiziranje rezultata
SearchSalad – kada korisnik želi kombinaciju rezultata tražilice i informacije od svih glavnih web-stranica za recenzije
Seekz – kada korisnik želi koristiti metatražilicu koja daje najrelevantniji rezultat i miče duplikate rezultata.

Izvor: WordStream (2023).

3.3. Specijalizirane tražilice

Specijalizirane tražilice su posebne platforme koji omogućuju korisnicima da pretražuju digitalne resurse unutar njihove željene tematike (Yext, 2021). Iako je preko tražilice Google dostupna iznimno velika količina informacija, te informacije mogu često biti općenite ili korisnik za traženu informaciju dobije malu količinu poveznica koji su vezani za njegovu željenu pretragu. Kako bi se ta količina povećala i resursi bili relevantniji, koriste se specijalizirane tražilice. Primjer takve tražilice, subjektima je najpoznatija i najraširenija *Google Scholar*, odnosno Google znalac tražilica. Kako sama riječ „*Scholar*“ ili „*Znalac*“ predstavlja nešto što bi bilo usko povezano sa školovanjem, edukacijama, studiranjem može se pretpostaviti da se u takvoj tražilici nalazi isključivo znanstvena i stručna literatura. Kada bi se upisale ključne riječi „Sekundarna internetska informacija“, tražilica Google bi korisnika zasula sa poveznicama blogova, običnim informacijama što je to, pa čak i nekim nevezanim informacijama po pitanju definicije „Sekundarne internetske informacije“. Dok upisujući te ključne riječi u Google znalac, korisnik bi dobio isključivo poveznice na završne radove, članke, znanstvene ili stručne radove vezano uz temu „Sekundarne internetske informacije“. Koristiti tražilicu Google za dolazak do takvih resursa je u redu, no ukoliko korisnik želi dublje pročitati i naučiti nešto o tome, može se koristiti specijaliziranim tražilicama kako bi ušao dublje u temu koje ga zanima. „Specijalizirane tražilice, također poznate kao tražilice za specifični niše ili vertikalne tražilice, funkcioniraju drugačije od popularnih tražilica poput Googlea. Specijalizirane tražilice prikazuju samo web-stranice koje su kategorizirane unutar njihove specifične teme. Ova vrsta pretraživanja često prikazuje brojne stranice koje se ne bi pojavile u Google pretraživanju - ili barem ne u prvih pet stranica.“ (Yext, 2021). Osim Google

Znalca, treba spomenuti Hrčak, hrvatsku otvorenu platformu za pristup znanstvenim i stručnim časopisima u različitim disciplinama, Elephind tražilicu za povijesne primjere tiskanih novina i Science za pristup Američkim istraživanjima vlade.

3.4. Agregatori vijesti

Kako sam naziv glasi, to su posebne platforme u kojima *botovi* prikupljaju vijesti sa različitih dijelova interneta i unose na stranicu koja je akumulator tih istih. Grupiranje informacija i vijesti se događa putem ključnih riječi koji se nalazi na tim vijestima, tako da recimo pomorska katastrofa neće biti u rubrici gdje je i turistička ekspedicija uspješno popela *Mount Everest*, već će biti nešto bliže temi kao što su na primjer; čišćenje oceana, najveći turistički kruzer pristize u Španjolsku, ili bilo kakva tema što je vezana za mora, oceane i plovidbe. Isto tako na već uzetom primjeru, s obzirom da postoje dvije ključne riječi „pomorska“ i „katastrofa“ ta ista vijest može se svrstati u rubriku sa ostalim „katastrofama“. To su stranice koje olakšavaju pristup i pretraživanje vijesti koje su globalno dostupne. Objavljuju se na način da uz pisani kontekst i kratki sadržaj događaja uvijek dolaze sa poveznicama koji vodi direktno na izvor te vijesti, odnosno citirana je putem poveznice na izvor. „Neke stranice za agregiranje vijesti također omogućuju praćenje omiljenih publikacija. Na taj način se osigurava da se dobije više sadržaja s web-stranica kojima korisnik vjeruje u njegovom prikazu. Međutim, da se pronađe određena publikacija na stranici za agregiranje vijesti, ta publikacija mora imati vlastiti profil na toj stranici.“ (Fitzgerald, 2023). Primjer takvog internetskog mjesta bi bila platforma Feedly, koji se može koristiti uz plaćanje ili besplatno uz ograničenje na 100 izvora po korisniku, zatim tražilica Panda koja je za web-dizajnere i programere. Prati stranice poput Dribblea, Behancea, Product Hunta i GitHuba (Timeisle, 2023).

3.5. Web scraping

Web scraping se može definirati kao vađenje informacija sa internetskog mjesta. Program prikuplja i procesira podatke kako bi pojednostavio i prikazao korisniku samo određene i željene informacije. *Web scraping* se može raditi ručno u kodovima ili u već postojećim programima. U program se prvo učitaju poveznice stranice odnosno *URL* (engl. *uniform resource locator*) koji se procesira unutar programa za *Web scraping*, napredni programi učitaju HTML, CSS i Javascript elemente stranice kako bi se korisnik lakše kretao i odabirao željeni sadržaj koji želi da mu se prikaže. Korisnik zatim može složiti listu željenih resursa koji da mu

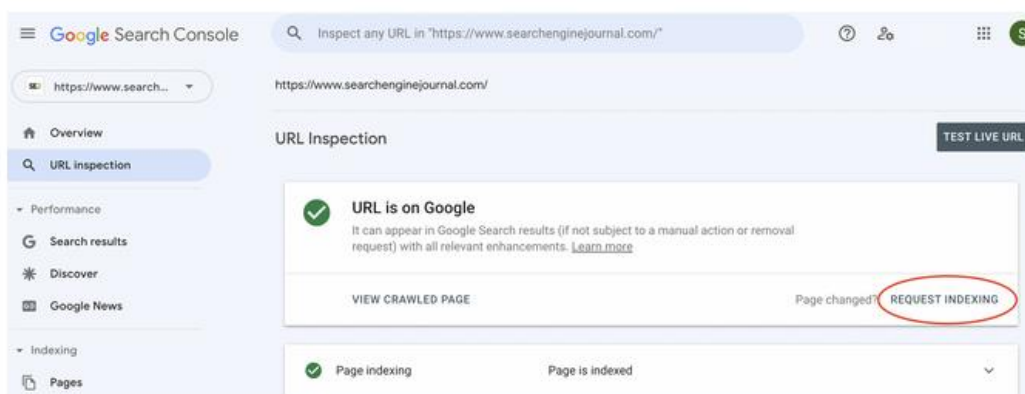
se prikazuju. Na primjer; korisnik želi vidjeti na internet trgovini samo cijenu i model proizvoda da mu se prikazuju, no komentare, vrijeme dostave i način plaćanja ga ne zanimaju, pa to neće označiti. Program zatim najčešće prikupljene informacijske resurse konvertira u korisniku pristupačniji način isčitavanja u *Excel* listu. *Web scraperi* mogu biti u obliku dodatka internetskom pregledu ili desktop aplikacije, no isto tako može se koristiti i putem *back-end* programiranja poznatije pod *PhP Web scraping*. (Perez, 2023; Freecodecamp, 2023).

4. Metode pretraživanja internetskog prostora

Prošli naslovi su se dotakli programa i aplikacija kojima ulazimo i pretražujemo internetski informacijski prostor, no u ovom dijelu diplomskog rada cilj će biti približiti način kojom metodom korisnik uopće dolazi do željenog sadržaja putem tih programa i kako ti programi očitavaju, uspoređuju i šalju točne resurse natrag istom korisniku.

4.1. Indeksiranje

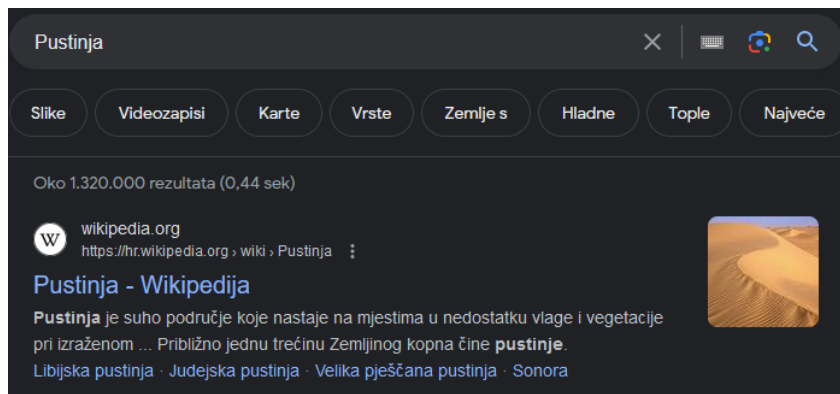
Indeksiranje je proces koji se odvija nakon što je internetsko mjesto (engl. *website*) pretraženo. Tražilica koja se koristi zatim pokušava razumijeti o čemu se radi na toj stranici, tu spada obrada i analiziranje tekstualnog sadržaja, te ključnih oznaka i atributa koji se nalaze unutar HTML strukture. Zatim atributi koji su upisani kao mali opis koji se nalazi ispod poveznice kada Google izbaci rezultate, te procesira slike, sadržaje i video zapise koji se nalaze unutar stranice. Tijekom tog procesa tražilica provjerava je li stranica duplicirana, relevantna, legalna ili uopće dostupna. Nakon toga stranica bude indeksirana odnosno stavljena na listu kao odobrena stranica sa relevantnim informacijama, što znači biti će jedan od rezultata kada tražilica ponudi sve poveznice koji i nakon filtriranja mogu biti u ogromnim brojevima (Google Search Center, 2023). Preko upita korisnika tražilica može također svrstavati stranice koje će nam ponuditi po geografskoj lokaciji, jeziku i jednostavnosti korištenja odnosno korisničkom iskustvu. Korisnik u slučaju da posjeduje internetsko mjesto može ga predložiti na indeksiranje, ali ne može samostalno indeksirati odnosno može dozvoliti da tražilica pretraži i indeksira povezane web-stranice (Davies, 2023). Zajedno s tim potrebno je i obratiti pažnju da ključne riječi na vlastitoj stranici što će biti više spomenuto u slijedećem podnaslovu.



Slika 4: Primjer zahtjeva za indeksiranje određene stranice
Izvor: Davies (2023).

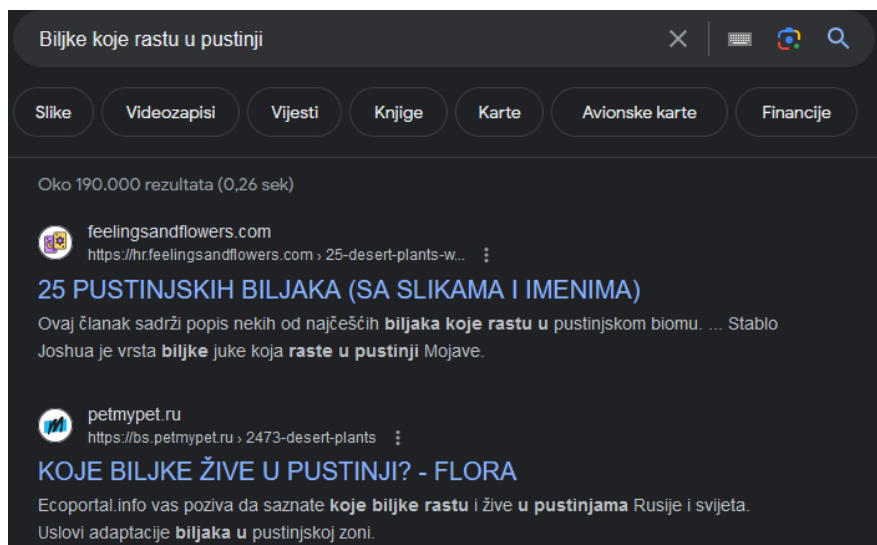
4.2. Ključne riječi

Prije svega treba razlikovati ključne riječi iz pozicije korisnika i ključne riječi iz pozicije vlasnika. Ključne riječi korisniku su fraze kojima dohvaća željeni sadržaj. Može biti jedna riječ ili više riječi, opisnog oblika kako bi korisnik pokušao pronaći točno željeni rezultat (Dennis, 2022). Na strani vlasnika web-sjedišta, ključne riječi imaju ulogu olakšati korisnicima da lakše pronađu njihove sadržaje na web-stranicama. To je organski način dolaska do posjetitelja web-sjedišta. Pažljivo se istražuje odabrana grupa korisnika, te se ciljano unose na stranici ključne riječi relevantne za proizvode ili usluge tog web-sjedišta (Southwick, 2022). Ključne riječi trebaju biti relevantne sadržaju stranice, i trebaju sadržavati ono što korisnik traži. Prije svega treba istražiti koje su ključne riječi popularne za određenu temu, kakva je statistika pretraga i na kojim se ključnim riječima korisnik najčešće zadržava ili pridržava prilikom njegove pretrage. Svaka stranica i tema neke stranice zahtjeva i ima svoje određene ključne riječi koje treba istražiti što više, no isto tako ne treba stranicu pretrpati nepotrebnom količinom tih ključnih riječi. Kako ne postoji definiran i formalan broj koliko ključnih riječi smije ili treba biti, mora se uzeti u obzir da ukoliko se stavlja prevelika količina tih riječi, stranica postaje općenitija, više nije usko ciljana, te zbog takve situacije ta ista stranica može završiti niže u rezultatima pa čak i nekoliko strana u tražilici dalje. Sa razlogom se zapošljavaju ljudi sa znanjem optimiziranja web-stranica za tražilice (engl. SEO ili *search engine optimization*), jer napisati ključne riječi koje se vežu za stranicu može svatko, no napisati prave riječi i odabrati pravu količinu nije lagano. Statističkim analizama, i detaljnom procjenom, programeri i SEO stručnjaci moraju predvidjeti nekada laičke upite koje korisnik upisuje na tražilicu, te putem toga stvoriti strukturu koju kada korisnik upiše jednu, dvije ili više riječi u upit tražilice da mu se ta ista prikaže barem unutar prve strane tražilice. Korisnik sa druge strane je taj koji ne vidi koje su ključne riječi u pozadini, na njemu je da željeni sadržaj koji traži upotpuni sa konkretnim upitom. Kao primjer uzet će se prikaz pretrage od Semrush Bloga na temelju pretraživanja biljaka (Harsel, 2022). Za čitatelje bit će prikazano na hrvatskom jeziku za lakše praćenje.



Slika 5: Pretraživanje ključne riječi "Pustinja" + rezultati
Izvor: slika zasolna tražilice Google

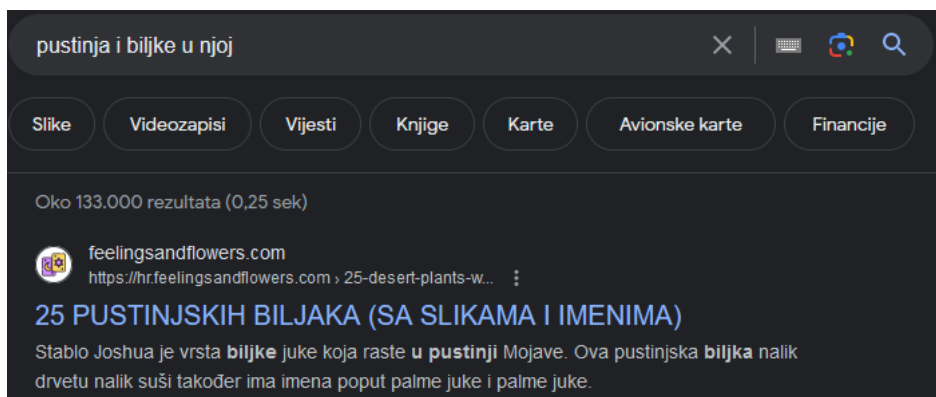
Prilikom upisa ključne riječi „Pustinja“ već se može pretpostaviti kako će rezultati biti velik, te će prikazati rezultate svih stranica koje u sebi imaju ključnu riječ pustinja, to čini 1,320,000 pronađenih rezultata unutar 0,44 sekunde. Kako se nije koristio dodatni opis u upitu, odnosno dodatne i popratne ključne riječi prvi rezultat je općenita Wikipedija, definicija pustinje i pustinje kao prirodnog fenomena. No ukoliko se nadoda još kratak detalj ili kratak opis tog riječi, rezultat postaje automatski drugačiji i ciljaniji. Slijedi slika dodatne pretrage.



Slika 6: Pretraživanje iste ključne riječi uz dodatne opisne ključne riječi + rezultati
Izvor: slika zaslona tražilice Google

Odmah se može primjetiti značajan pad u rezultatima i vrijeme pronalaska i ponude rezultata se smanjio skoro za pola. Rezultata sada ima 190,000 koji su pronađeni unutar 0,26 sekundi. Sada kada su se dodale još ključnih riječi, u ovom slučaju su to biljke koje rastu u pustinji, umjesto općenite definicije pustinje, i općenitog sadržaja Wikipedije, prvi rezultat glasi „25 pustinjskih biljaka (sa slikama i imenima)“. S obzirom da je bio cilj pronaći biljke koje rastu u pustinji, korisnik može početi pregledavati poveznice koji su mu ponuđeni, te sam zaključiti je

li zadovoljan sadržajem ili će se upustiti u još dodatnu pretragu poveznica, ili čak dodavanjem još ključnih riječi. No za ovaj primjer, uočljiva je promjena i ciljani sadržaj. Može se reći kako unutar prvih 1.320.000 rezultata općenitog pojma „Pustinja“ nalazi i poveznica sa biljkama iz pustinje, samo što tražilica nije očitala riječ „biljka“ te nije imala potrebu kao prve rezultate staviti linkove sa biljkama. Kako ne bi došlo do zabune, riječ „Pustinja“ i „Pustinji“ nisu iste, iako se zna o čemu se priča tražilica to prepoznaje i daje iste rezultate. Za primjer slijedi slika:



Slika 7: Primjer rezultata sa drugačijim sufiksom (zadnjim nastavkom)
Izvor: slika zaslona tražilice Google

Iako su riječi vrlo slične, što ne bi trebalo praviti preveliku razliku, ovaj primjer je uzet samo kao kontrola, za dokaz da je prva poveznica ostala ista, iako su ključne riječi poredane na malo drugačiji način.

4.3. Operatori pretraživanja

Operatori pretraživanja su posebne oznake i simboli koji olakšavaju korisniku lakše usmjeravanje pretraživanja za neku određenu informaciju. Funkcionira na način dodavanja oznaka prije, unutar ili poslje pojma koji se traži. Može specificirati informacije, oduzimati kontekst informacija ili zbrajati odnosno grupno pretražiti dva ili više pojmova koji korisnika zanimaju, koji ukoliko ne napisan ili grupiran, ne bi bio pri vrhu traženih rezultata (Hardwick, 2023).

Tablica 2: Primjeri operatora tražilica

Operator	Što radi	Primjer
„ „	Traži isključivo rezultate sa riječi u navodnicima	„Jedrilica“

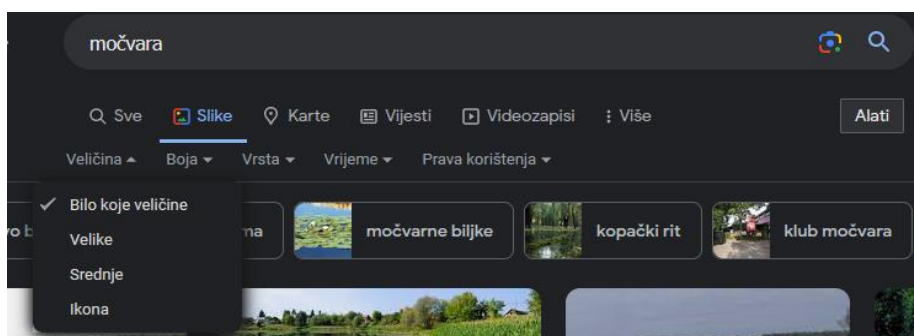
OR	Traži rezultate koji su X ili Y	„Jedrilica ili jahta“
!	Traži rezultate koji su X ili Y isto	„Jedrilica ! jahta“
AND	Traži rezultate koji su povezani sa X i Y	„Jedrilica i jahta“
-	Traži rezultate koji ne sadrži određenu riječ	„Jedrilica -brod“
*	Algoritam podudara bilo koje riječi i fraze	„Jedrilica * brod* jahta“
()	Grupira više pretraga	(Jedrilica ili jahta) brodovi
Define:	Traži rezultate u kojima je definicija ključne riječi	Definiraj:brod
Cache:	Traži najnovije informacije na određenoj stranici	Cache:Jadranska obala
Filetype:	Traži određene vrste dokumenta (PDF, Word)	Pomorstvo filetype:pdf
Ext:	Isto kao i „Filetype:“	Pomorstvo ext:pdf
Site:	Traži rezultate sa određene stranice	Site:pomorstvo.hr
Related:	Traži stranice relevantne danj domeni	Related:pomorstvo.hr
Intitle:	Traži stranice sa određenim riječima u računalskom simbolu <title></title>	Intitle:pomorstvo
Allintitle:	Traži stranice sa više ključnih riječi u računalskom simbolu <title></title>	Allintitle:pomorstvo jahta
Inurl:	Traži stranice sa određenom riječu unutar usklađenog lokatora sadržaja (engl. URL)	Inurl:pomorstvo
Allinurl:	Traži stranice sa više određenih riječi unutar usklađenog lokatora sadržaja (engl. URL)	Allinurl:pomorstvo jadrans
Intext:	Traži stranice sa određenim riječima u sadržaju stranice	Intext:pomorstvo jadrans
Allintext:	Traži stranice sa više određenih riječi u sadržaju stranice	Allintext:pomorstvo jadrans
Weather:	Traži vremensku prognozu na određenoj lokaciji	Weather:Dalmacija
Stocks:	Traži informacije o dionicama	Stocks:tesla
Map:	Upit za google da prikaže mapu	Map:Split
Movie:	Traži informacije o određenom filmu	Movie:Avatar
In	Pretvara jednu valutu u drugu	750kn u EUR

Source:	Traži rezultate sa određenih izvora o Google vjestima	Jadran source:turizam
After:	Traži rezultate nakon zadanog datuma	Jadran after:20.06.20018.

Izvor: Hardwick (2023).

4.4. Napredne tehnike pretraživanja

Napredne tehnike pretraživanja omogućuju korisniku da skрати i olakša proces pretraživanja željenog rezultata. To su dodatni alati ili implementirane sekcije već unutar tražilice koje se mogu odabrati prilikom pretraživanja. Primjeri takvog pretraživanja bi bile Google slike, karte, videozapisi, veličine resursa odnosno u ovom slučaju datoteka sa informacijama (Fisher, 2023). Google slike korisniku omogućava da prilikom upisivanja ključnih riječi u tražilicu, neće biti opterećen općenitim poveznicama kao rezultatima tražilice, već će biti automatski prebačen u posebni prozor gdje se prikazuju isključivo i samo slike na temu njegove pretrage koje su onda potkrijepljene poveznicama ispod njih, kako bi korisnik mogao pristupiti stranici na kojoj se slika nalazi, ako to uopće želi. Unutar sekcije pretrage samo slika, postoje dodatni alati kojima se može regulirati veličina i rezolucija slike, boja, vrijeme i vrsta. Isto tako vrijedi i za geografsku pretragu gdje upisivanjem pojma nekog mjesta, korisnika se šalje na geografsku kartu koja označava pojam koji predstavlja neko određeno mjesto, sa dodatnim linkovima i slikama ukoliko je to mjesto trgovina prikazana na kartama, da uz geografski položaj može i pristupiti stranici ukoliko imaju dodatne ponude. (Search Engine Land, 2022.)



Slika 8: Primjer napredne tehnike pretraživanja
Izvor: Slika zaslona tražilice Google

4.5. Algoritam pretraživanja

Algoritam pretraživanja funkcionira putem matematičkih i kontekstualnih filtriranja rezultata kako bi korisniku koji je oblikovao upit tražilici na određene ključne riječi dobije što

relevantnije, točnije i bitne informacije. Algoritam može pozicionirati stranicu pri vrhu stranice rezultata pretrage ukoliko je stranica često posjećena i dijeljena poveznicom od strane vlasnika stranice, visoko ocijenjena i na posljetku koliko ključnih riječi sadrži u sebi koje je korisnik upisao na tražilicu (Widmer i Hardwick, 2022). Svaka tražilica ima svoj implementirani oblik algoritma, koji nije javno objavljen kako bi se zadržao jedinstven pristup sustava svake tražilice. Drugim riječima, svaka tražilica ima svoj algoritam koji se sastoji od većeg broja varijabli, te koji je u pravilu poslovna tajna, iako tražilice sugeriraju smjernice kako oblikovati web-stranice za dobro pozicioniranje. Mogu se razlikovati dvije generalne vrste algoritama. Linearni algoritam, koji se podrazumijeva kao najobičniji oblik svih algoritama tražilica, koji zahtijevaju vrlo jednostavnu implementaciju, te je najbolji za pretrage kratkih lista poveznica koje su obično neorganizirane. Prilikom pretrage, algoritam pretražuje jedinice s liste, te kad ih pronade, pretraga je završena. S druge strane, binarni algoritam za razliku od linearnog, organizira liste rezultata, što je pogodno ukoliko će ti rezultati količinski rasti. Algoritam kreće od sredine liste, ukoliko je ciljana jedinica u drugoj polovici liste, algoritam miče gornji dio liste i pretraga se fokusira na ostatak gdje se tražena jedinica nalazi. Petlja se ponavlja kako bi se filtrirali što točniji rezultati, odnosno algoritam ponovo počinje u sredini liste koja je ostala, te je dijeli na dva dijela i miče dio liste koji je manje relevantan (Teran, 2018). Nadogradnje možemo svrstati u nekoliko skupina: važne, opsežne i manje nadogradnje (Cox, 2022). Važne nadogradnje su rijetke i često se bave određenim problemima u pretraživačkom algoritmu. Na primjer, nedavna nadogradnja Core Web Vitals bavi se problemima vezanim uz korisničko iskustvo na web-stranicama. Tražilice ih obično objavljuju jednom ili dva puta godišnje. Opsežne nadogradnje sredotočuju se na ciljanje niskokvalitetnih stranica. Obično prilagođavaju važnost nekoliko faktora rangiranja. Na primjer, mogu odlučiti da je brzina učitavanja stranice sada važnija od ukupnog broja povratnih veza. Ove nadogradnje obično se javljaju jednom svakih 4-5 mjeseci. Manje nadogradnje obično ne stvaraju značajne vidljive promjene u performansama i analitici internetskog mjesta. Često su to manje izmjene koje poboljšavaju iskustvo tražilice i ne utječu na rangiranje visokokvalitetnih web-stranica. Manje nadogradnje mogu se provoditi svakodnevno ili tjedno.

4.6. Horizontalno i vertikalno pretraživanje

Horizontalno pretraživanje se odnosi na generalno i općenito pretraživanje koje tražilica obavlja. Na horizontalno pretraživanje bi se odnosile tražilice poput Googlea, Yahooa i Binga.

Pretražuju internetski prostor, te pokrivaju velik broj rezultata i medijskih resursa. Svaki korisnik koji upiše željenu informaciju u tražilicu, stisne *Enter* i dobije rezultate je obavio horizontalno pretraživanje. Vertikalno pretraživanje sa druge strane je specificirano pretraživanje određene teme. Već spominjane specijalizirane tražilice su oblik verikalnog pretraživanja. Funkcioniraju na način da unutar te tražilice korisnik ne može tražiti sve postojeće rezultate već samo rezultate koji su vezani uz tu nišu. To su Elephind, Science, Hrčak, Google znalac. Korisnik se kreće „Gore“ i „Dolje“ po određenoj temi, ovisno koliko duboko ju želi istražiti. Dok horizontalnim traženjem ide „Naprijed“ i „Nazad“ dotičajući se željenih rezultata, no oni su općeniti (Hillborn, 2017).

Tablica 3: Horizontalno i vertikalno pretraživanje

	Pretraživanje	Tražilica
Horizontalno	Horizontalno pretraživanje, je generalan oblik pretraživanja koji se proteže kroz cijeli internet. Primjer traženja može biti „Mayanski hramovi“ što će rezultirati u velikom broju rezultata rangiranim po algoritmu za najbolji rezultat.	Velike tražilice poput Google, Yahoo i Binga su sve primjeri horizontalnih tražilica
Vertikalno	Vertikalno pretraživanje se može predstaviti kao tematsko ili specijalno pretraživanje. Pretražuje se samo određeni dio interneta sa određenom medijskom temom. Primjer takvog traženja može biti „Mayanski hramovi“ + „vijesti“, „slike“, „video zapisi“	Pretraživanje unutar drugih sekcija na velikim tražilicama poput Google slike ili tražilica medijskih sadržaja poput Youtubea su primjeri vertikalnog pretraživanja.

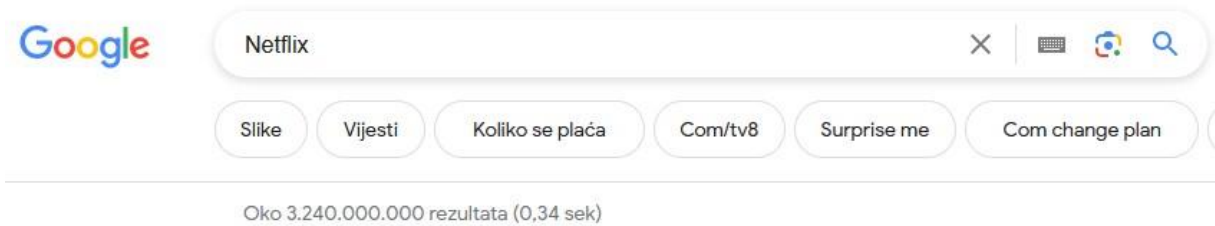
Izvor: Hillborn (2017).

5. Testiranje i analitika alata za pretraživanje internetskog prostora

Cilj ovog dijela diplomskog rada je utvrditi rezultate tri najpoznatije tražilice na internetu. Količina njihovih rezultata, koji će biti pretvoreni u postotak odstupanja jedna od druge svrstati će se unutar tablice uz komentar radi lakšeg razumijevanja dobivenih rezultata. Premda je iz prethodne analize razvidno koja je tražilica dominantno najpopularnija, svakako treba uspostaviti formalan pregled koliko su ostale tražilice relevantne odnosno može li se i iz njihove količine i kvalitete rezultata doći do željenih informacija. U ovom istraživanju biti će obrađene tražilice Google, Bing i Yahoo kao tri najkorištenije tražilice na internetu. Od ključnih riječi biti će postavljena 4 svjetski poznata pojma, 4 geografska područja i 2 edukacijske ustanove. Odabrane ključne riječi su autorski odabir. No isto tako odabrani su općeniti pojmovi poznati većini populacije i većini korisnika interneta. Te kako ključne riječi ne bi bile samo internetske stranice i aplikacije za narudžbe, dodana su geografska područja i edukacijske ustanove, sa pokušajem pronalaska veće razlike u dobivenim rezultatima. Datum provođenja analize je 27.06.2023. na lokaciji grada Osijeka, u Osiječko-Baranjskoj županiji. Te računalo korišteno za analizu je prijenosno računalo Asus TUF gaming F15. Ključne riječi glase: **Svjetski poznati pojmovi:** Netflix, Amazon, WHO (World Health Organisation) i Ikea. **Geografska područja:** Alpe, Jadran, Sahara i Skandinavija. **Edukacijske ustanove:** Harvard, Oxford. S obzirom da je Google dominantno vodeća tražilica, pretpostavlja se da će dobiveni rezultati biti znatno brojniji na Google tražilici nego na Bing i Yahoo tražilici (Forsey 2023). Google će biti bazni standard unutar analiziranja rezultata dobivenih između te tri tražilice. Napraviti će se također omjeri koliko brojnost rezultata tražilica Yahoo i Bing postotno odstupa od bazne tražilice Googlea. **Formula** za izračunavanje odstupanja rezultata jedne tražilice od druge, biti će:

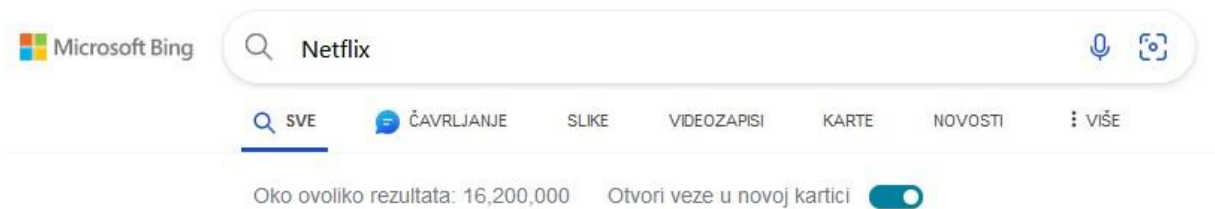
$$\frac{\textit{tražilica A}}{\textit{tražilica B}} * 100$$

5.1. Svjetski poznati pojmovi



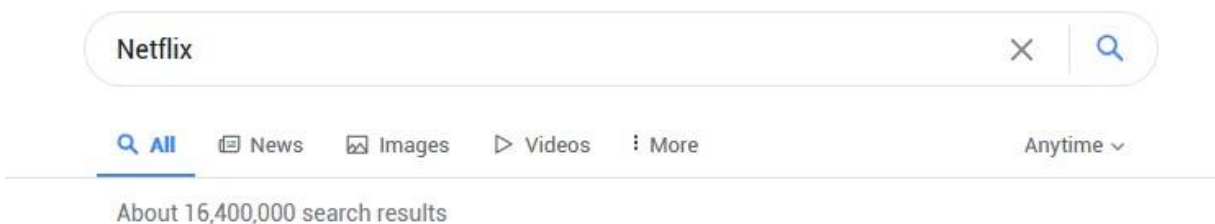
Slika 9: Analitika dobivenih rezultata na Google-u, pod ključnom riječi „Netflix“

Izvor: Slika zaslona tražilice Google



Slika 10: Analitika dobivenih rezultata na Bing-u, pod ključnom riječi „Netflix“

Izvor: Slika zaslona tražilice Bing



Slika 11: Analitika dobivenih rezultata na Yahoo, pod ključnom riječi „Netflix“

Izvor: Slika zaslona tražilice Yahoo

Tablica 4: Analiza rezultata za ključnu riječ "Netflix"

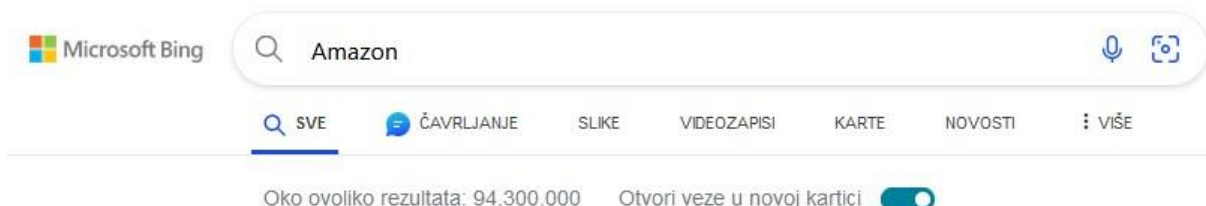
NETFLIX – Statistički podaci			
Tražilice	Google	Bing	Yahoo
Rezultati	3,240,000,000	16,200,000	16,400,000
Omjer	100%	0,5%	0,51%

Rezultati prikazani na tablici ukazuju na sljedeće; Google ima 3,240,000,000 rezultata na ključnu riječ Netflix, Bing ima 16,200,000 rezultata i Yahoo ima 16,400,000 rezultata. Uzimajući u obzir brojnost rezultata pretrage između Binga i Googlea, može se zaključiti da Bing nudi 0,5% rezultata u odnosu na broj koji nudi Google. Brojnost rezultata između Yahoo i Googlea, može se zaključiti da Yahoo nudi 0,51% rezultata u odnosu na broj koji nudi Google.



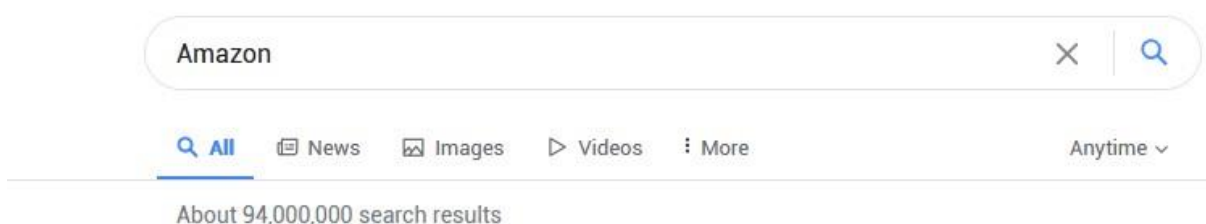
Slika 12: Dobiveni rezultati pretrage tražilicom Google, na ključnu riječ "Amazon"

Izvor: Slika zaslona tražilice Google



Slika 13: Dobiveni rezultati pretrage tražilicom Bing, na ključnu riječ "Amazon"

Izvor: Slika zaslona tražilice Bing



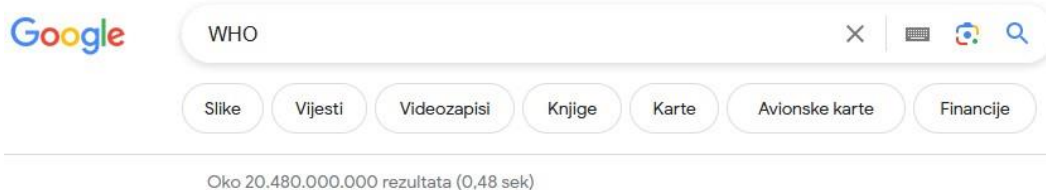
Slika 14: Dobiveni rezultati pretrage tražilicom Yahoo, na ključnu riječ "Amazon"

Izvor: Slika zaslona tražilice Yahoo

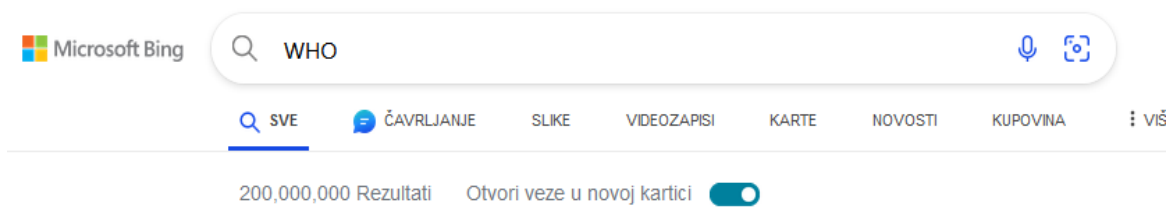
Tablica 5: Analiza rezultata na ključnu riječ "Amazon"

AMAZON – Statistički podaci			
Tražilice	Google	Bing	Yahoo
Rezultati	5,040,000,000	94,300,000	94,000,000
Omjer	100%	1,870%	1,865%

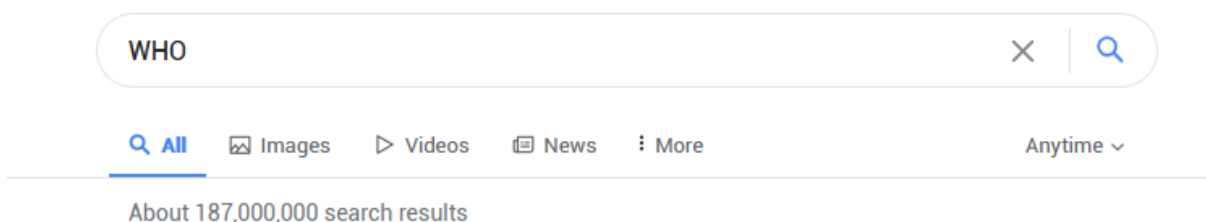
Rezultati prikazani na tablici ukazuje na sljedeće; Google ima 5,040,000,000 rezultata na ključnu riječ Amazon, Bing ima 94,300,000 rezultata i Yahoo ima 94,000,000 rezultata. Uzimajući u obzir brojnost rezultata pretrage između Binga i Googlea, može se zaključiti da Bing nudi 1,87% rezultata u odnosu na broj koji nudi Google. Brojnost rezultata između Yahoo i Googlea, može se zaključiti da Yahoo nudi 1,86% rezultata u odnosu na broj koji nudi Google.



Slika 15: Dobiveni rezultati pretrage tražilice Google, na ključnu riječ „WHO“ (World health organisation)
Izvor: Slika zaslona tražilice Google



Slika 16: Dobiveni rezultati pretrage tražilice Bing, na ključnu riječ „WHO“ (World health organisation)
Izvor: Slika zaslona tražilice Google



Slika 17: Dobiveni rezultati pretrage tražilice Yahoo, na ključnu riječ „WHO“ (World Health Organisation)
Izvor: Slika zaslona tražilice Google

Tablica 6: Analiza rezultata na ključnu riječ „WHO“

WHO – Statistički podaci			
Tražilice	Google	Bing	Yahoo
Rezultati	20,480,000,000	200,000,000	187,000,000
Omjer	100%	0,97%	0,91%

Rezultati prikazani na tablici ukazuju na sljedeće; Google ima 20,480,000,000 rezultata na ključnu riječ WHO, Bing ima 200,000,000 rezultata i Yahoo ima 187,000,000 rezultata. Uzimajući u obzir brojnost rezultata pretrage između Binga i Googlea, može se zaključiti da Bing nudi 0,97% rezultata u odnosu na broj koji nudi Google. Brojnost rezultata između Yahoo i Googlea, može se zaključiti da Yahoo nudi 0,91% rezultata u odnosu na broj koji nudi Google.



Oko 443.000.000 rezultata (0,49 sek)

Slika 18: Dobiveni rezultati pretrage tražilice Google, na ključnu riječ "Ikea"

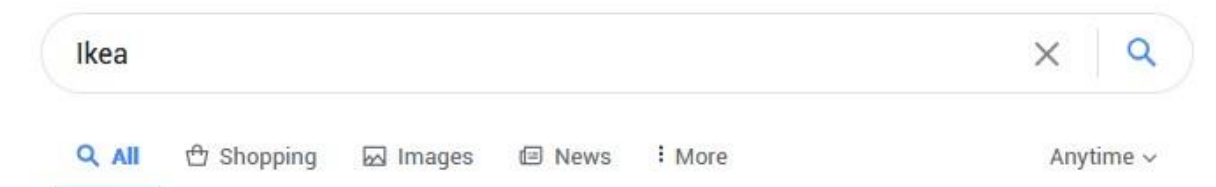
Izvor: Slika zaslona tražilice Google



Oko ovoliko rezultata: 7,960,000 Rezultati u blizini Osijek, Osječko-baranjska županija · Promij

Slika 19: Dobiveni rezultati pretrage tražilice Bing, na ključnu riječ "Ikea"

Izvor: Slika zaslona tražilice Bing



About 7,980,000 search results

Slika 20: Dobiveni rezultati pretrage tražilicom Yahoo, na ključnu riječ "Ikea"

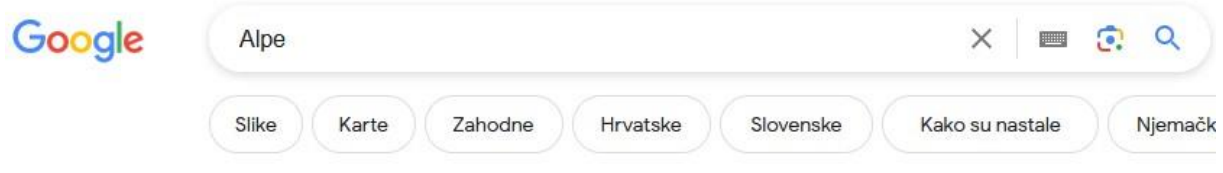
Izvor: Slika zaslona tražilice Yahoo

Tablica 7: Analiza rezultata na ključnu riječ "Ikea"

IKEA – Statistički podaci			
Tražilice	Google	Bing	Yahoo
Rezultati	443,000,000	7,960,000	7,980,000
Omjer	100%	1,79%	1,80%

Rezultati prikazani na tablici ukazuje na sljedeće; Google ima 443,000,000 rezultata na ključnu riječ Ikea, Bing ima 7,960,000 rezultata i Yahoo ima 7,980,000 rezultata. Uzimajući u obzir brojnost rezultata pretrage između Binga i Googlea, može se zaključiti da Bing nudi 1,79% rezultata u odnosu na broj koji nudi Google. Brojnost rezultata između Yahoo i Googlea, može se zaključiti da Yahoo nudi 1,80% rezultata u odnosu na broj koji nudi Google.

5.2. Geografska područja



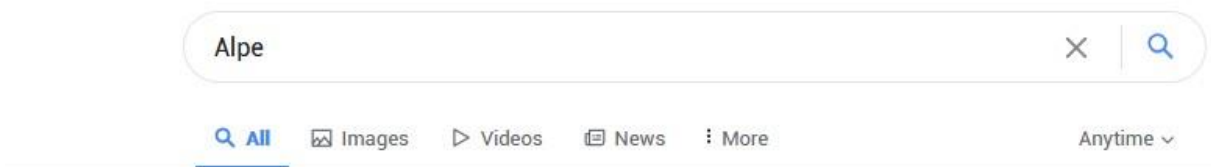
Slika 21: Dobiveni rezultati pretrage tražilice Google, na ključnu riječ „Alpe“

Izvor: Slika zaslona tražilice Google



Slika 22: Dobiveni rezultati pretrage tražilice Bing, na ključnu riječ „Alpe“

Izvor: Slika zaslona tražilice Bing



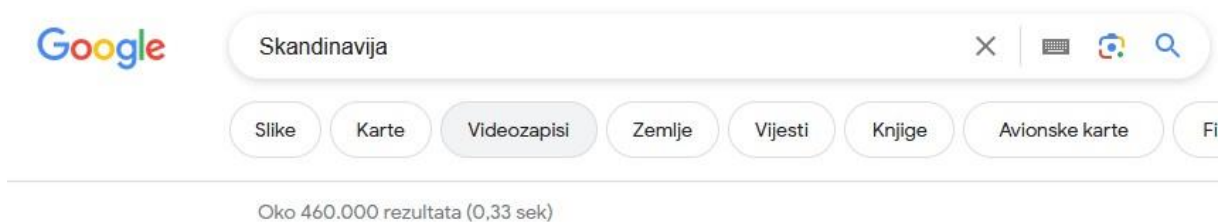
Slika 23: Dobiveni rezultati pretrage tražilice Yahoo, na ključnu riječ „Alpe“

Izvor: Slika zaslona tražilice Yahoo

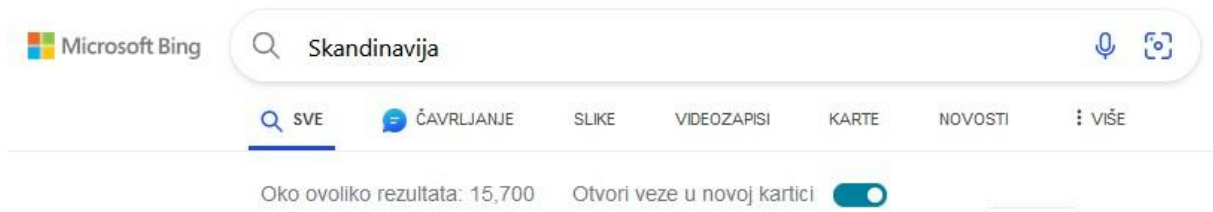
Tablica 8: Analiza rezultata na ključnu riječ „Alpe“

ALPE – Statistički podaci			
Tražilice	Google	Bing	Yahoo
Rezultati	35,600,000	367,000	367,000
Omjer	100%	1,03%	1,03%

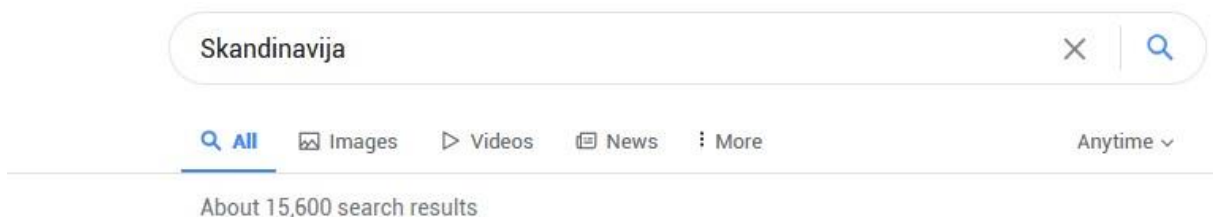
Rezultati prikazani na tablici ukazuje na sljedeće; Google ima 35,600,000 rezultata na ključnu riječ Alpe, Bing ima 367,000 rezultata i Yahoo ima 367,000 rezultata. Uzimajući u obzir brojnost rezultata pretrage između Binga i Googlea, može se zaključiti da Bing nudi 1,03% rezultata u odnosu na broj koji nudi Google. Brojnost rezultata između Yahoo i Googlea, može se zaključiti da Yahoo nudi 1,03% rezultata u odnosu na broj koji nudi Google.



Slika 24: Dobiveni rezultati pretrage tražilice Google, na ključnu riječ "Skandinavija"
Izvor: Slika zaslona tražilice Google



Slika 25: Dobiveni rezultati pretrage tražilice Bing, na ključnu riječ "Skandinavija"
Izvor: Slika zaslona tražilice Bing

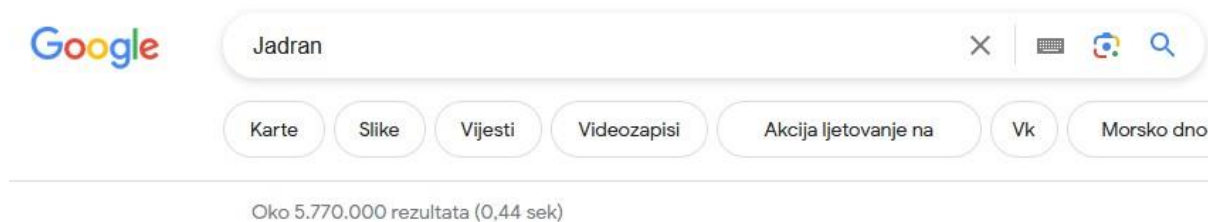


Slika 26: Dobiveni rezultati pretrage tražilice Yahoo, na ključnu riječ "Skandinavija"
Izvor: Slika zaslona tražilice Yahoo

Tablica 9: Analiza rezultata na ključnu riječ "Skandinavija"

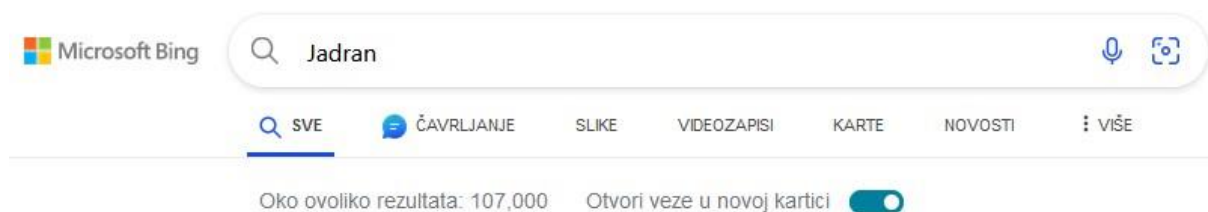
SKANDINAVIJA – Statistički podaci			
Tražilice	Google	Bing	Yahoo
Rezultati	460,000	15,700	15,600
Omjer	100%	3,41%	3,39%

Rezultati prikazani na tablici ukazuje na sljedeće; Google ima 460,000 rezultata na ključnu riječ Skandinavija, Bing ima 15,700 rezultata i Yahoo ima 15,600 rezultata. Uzimajući u obzir brojnost rezultata pretrage između Binga i Googlea, može se zaključiti da Bing nudi 3,41% rezultata u odnosu na broj koji nudi Google. Brojnost rezultata između Yahoo i Googlea, može se zaključiti da Yahoo nudi 3,39% rezultata u odnosu na broj koji nudi Google.



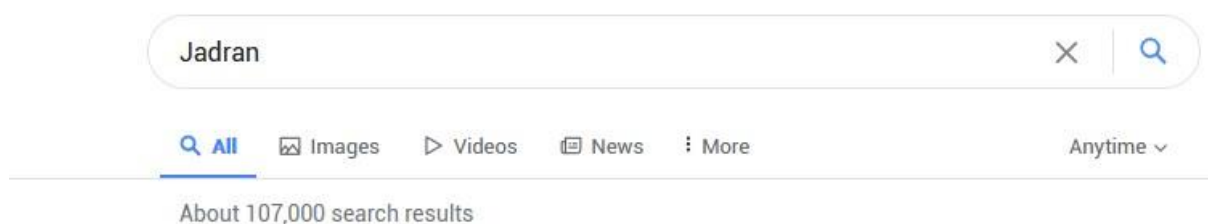
Slika 27: Dobiveni rezultati pretrage tražilice Google, na ključnu riječ „Jadran“

Izvor: Slika zaslona tražilice Google



Slika 28: Dobiveni rezultati pretrage tražilice Bing, na ključnu riječ „jadran“

Izvor: Slika zaslona tražilice Bing



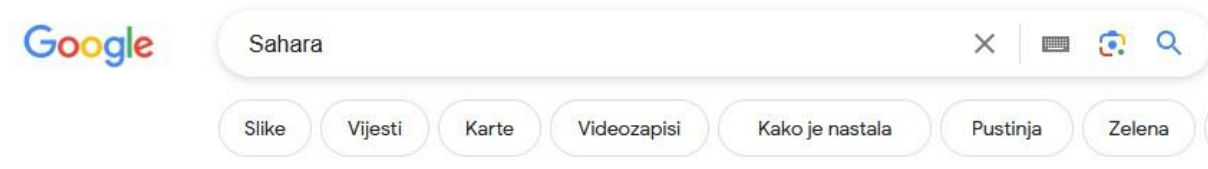
Slika 29: Dobiveni rezultati pretrage tražilice Yahoo, na ključnu riječ „Jadran“

Izvor: Slika zaslona tražilice Yahoo

Tablica 10: Analiza rezultata na ključnu riječ „Jadran“

JADRAN – Statistički podaci			
Tražilice	Google	Bing	Yahoo
Rezultati	5,770,000	107,000	107,000
Omjer	100%	1,85%	1,85%

Rezultati prikazani na tablici ukazuju na sljedeće; google ima 5,770,000 rezultata na ključnu riječ Jadran, Bing ima 107,000 rezultata i Yahoo ima 107,000 rezultata. Uzimajući u obzir brojnost rezultata pretrage između Binga i Googlea, može se zaključiti da Bing nudi 1,85% rezultata u odnosu na broj koji nudi Google. Brojnost rezultata između Yahoo i Googlea, može se zaključiti da Yahoo nudi 1,85% rezultata u odnosu na broj koji nudi Google.



Oko 615.000.000 rezultata (0,33 sek)

Slika 30: Dobiveni rezultati pretrage tražilice Google, na ključnu riječ "Sahara"

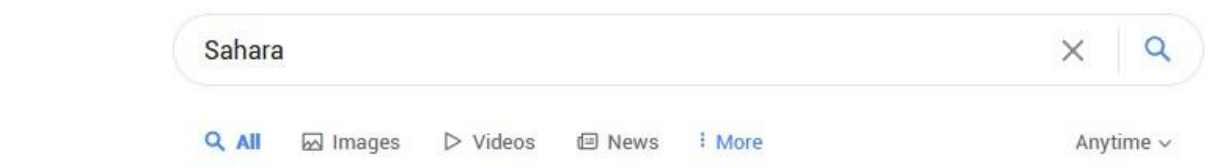
Izvor: Slika zaslona tražilice Google



Oko ovoliko rezultata: 1,480,000 Otvori veze u novoj kartici

Slika 31: Dobiveni rezultati pretrage tražilice Bing, na ključnu riječ "Sahara"

Izvor: Slika zaslona tražilice Bing



About 1,480,000 search results

Slika 32: Dobiveni rezultati pretrage tražilice Yahoo, na ključnu riječ "Sahara"

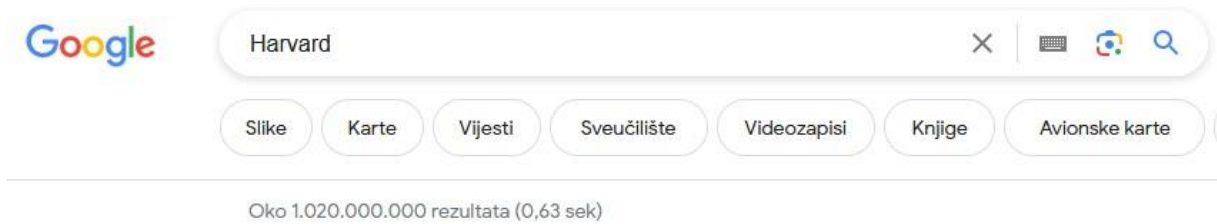
Izvor: Slika zaslona tražilice Yahoo

Tablica 11: Analiza rezultata na ključnu riječ "Sahara"

SAHARA – Statistički podaci			
Tražilice	Google	Bing	Yahoo
Rezultati	615,000,000	1,480,000	1,480,000
Omjer	100%	0,24%	0,24%

Rezultati prikazani na tablici ukazuju na sljedeće; google ima 615,000,000 rezultata na ključnu riječ Sahara, Bing ima 1,480,000 rezultata i Yahoo ima 1,480,000 rezultata. Uzimajući u obzir brojnost rezultata pretrage između Binga i Googlea, može se zaključiti da Bing nudi 0,24% rezultata u odnosu na broj koji nudi Google. Brojnost rezultata između Yahoo i Googlea, može se zaključiti da Yahoo nudi 0,24% rezultata u odnosu na broj koji nudi Google.

5.3. Edukacijske ustanove



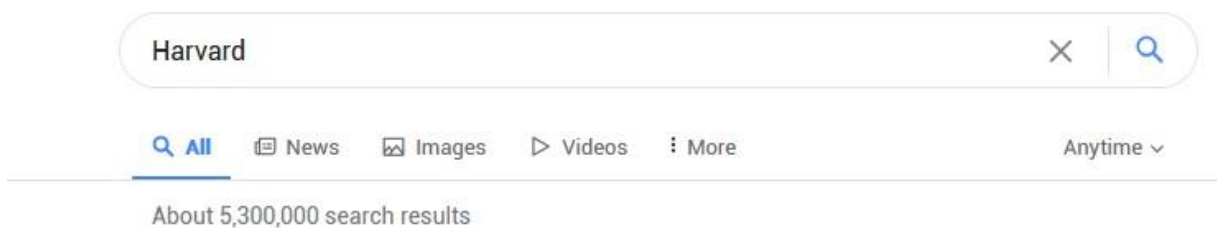
Slika 33: Dobiveni rezultati pretrage tražilice Google, na ključnu riječ „Harvard“

Izvor: Slika zaslona tražilice Google



Slika 34: Dobiveni rezultati pretrage tražilice Bing, na ključnu riječ „Harvard“

Izvor: Slika zaslona tražilice Bing



Slika 35: Dobiveni rezultati pretrage tražilice Yahoo, na ključnu riječ „Harvard“

Izvor: Slika zaslona tražilice Yahoo

Tablica 12: Analiza rezultata na ključnu riječ „Harvard“

HARVARD – Statistički podaci			
Tražilice	Google	Bing	Yahoo
Rezultati	1,020,000,000	5,300,000	5,300,000
Omjer	100%	0,51%	0,51%

Rezultati prikazani na tablici ukazuju na sljedeće; Google ima 1,020,000,000 rezultata na ključnu riječ Harvard, Bing ima 5,300,000 rezultata i Yahoo ima 5,300,000 rezultata. Uzimajući u obzir brojnost rezultata pretrage između Binga i Googlea, može se zaključiti da Bing nudi 0,51% rezultata u odnosu na broj koji nudi Google. Brojnost rezultata između Yahoo i Googlea, može se zaključiti da Yahoo nudi 0,51% rezultata u odnosu na broj koji nudi Google.



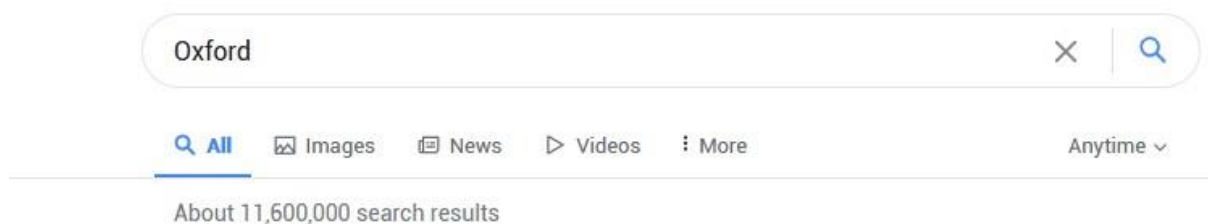
Slika 36: Dobiveni rezultati pretrage tražilice Google, na ključnu riječ "Oxford"

Izvor: Slika zaslona tražilice Google



Slika 37: Dobiveni rezultati pretrage tražilice Bing, na ključnu riječ "Oxford"

Izvor: Slika zaslona tražilice Bing



Slika 38: Dobiveni rezultati pretrage tražilice Yahoo, na ključnu riječ "Oxford"

Izvor: Slika zaslona tražilice Yahoo

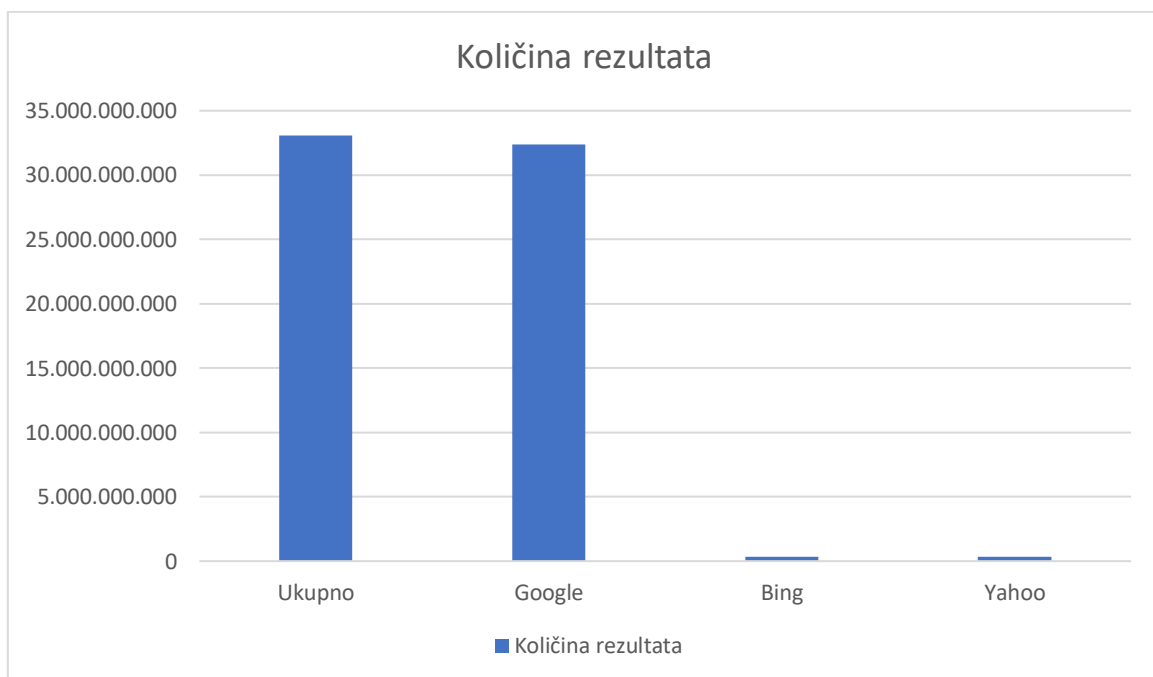
Tablica 13: Analiza rezultata na ključnu riječ "Oxford"

OXFORD – Statistički podaci			
Tražilice	Google	Bing	Yahoo
Rezultati	1,490,000,000	11,500,000	11,600,000
Omjer	100%	0,771%	0,778%

Rezultati prikazani na tablici ukazuju na sljedeće; Google ima 1,490,000,000 rezultata na ključnu riječ Oxford, Bing ima 11,500,000 rezultata i Yahoo ima 11,600,000 rezultata. Uzimajući u obzir brojnost rezultata pretrage između Binga i Googlea, može se zaključiti da Bing nudi 0,771% rezultata u odnosu na broj koji nudi Google. Brojnost rezultata između Yahoo i Googlea, može se zaključiti da Yahoo nudi 0,778% rezultata u odnosu na broj koji nudi Google.

5.4. Kritička analiza

Nakon dobivenih rezultata istraživanja, može se primjetiti kako razlika vodeće i najpoznatije tražilice Google jako odstupa od njegovih alternativnih tražilica. No to ne treba obeshrabriti korisnike Binga i Yahooa. Dobiveni rezultati su brojčano različiti, no njihove prve dobivene informacije i rezultati poveznica su isti. Prve glavne stranice kod Googlea i prve glavne stranice kod Binga i Yahooa su isti linkovi koji vode do istih informacija koji se odvijaju u svijetu, u prijašnjem stanju ili trenutnom. Ostali rezultati su podskupine glavnih poveznica i ključnih riječi, s obzirom kako je google najrasprostranjeniji, očigledno ima više ljudi koji informacije stavljaju na internet u obliku stranica. Ta pojava je i očekivana. Od ukupnog zbroja dobivenih rezultata preko svih tražilica iznosi 33,058,219,300. Od te količine, google zauzima 97.92% tih rezultata, što čini 32.269.760.000 rezultata. Bing i Yahoo čine samo ostalih 1,041283% i 1,041282% od ukupne količine sa svojih 344,229,700 i 344,229,600 rezultata, redosljedno. Zanimljiva činjenica je da se ovi rezultati i njihovi postotci ne razlikuju puno u odnosu na kompletan *market share* svih tražilica 2023. godine. Što znači da što više korisnika koristi neku tražilicu to će generacija informacija i stranica svakodnevno biti veća, događa se agregacija rezultata. Više ljudi, veća potreba za raznolikosti, otprilike dovodi ova dva rezultata da se razlikuju u samo nekoliko postotaka.



Grafikon 1: Histogram ukupnih rezultata preko 3 najpoznatije tražilice na internetu

Izvor: Autorski rad

6. Druge velike tražilice i internetski sustavi

Iako je Google najrasprostranjenija tražilica na svijetu sa svojih 92,63% korištenja populacije, malo drugačija priča se događa u Rusiji i Kini (Marsden, 2023). Naime Rusija i Kina imaju specifičan internetski prostor i svoje internetske tražilice. Ruski poslovni subjekt Yandex, svojim stanovnicima je pružio kompletno pretraživanje informacijskog prostora na ruskom jeziku. No u pitanju nije samo da se Google tražilica prebaci na ruski jezik, već Yandex nudi pretragu ruskog internetskog informacijskog prostora, s druge strane u Kini, Google je kompletno zabranjen, te oni koriste svoj internetski sustav i tražilicu pod nazivom Baidu (Bleichert, 2023).

6.1. Yandex

Najpoznatija ruska tražilica koja dominira svojim područjem. Iako Google nije zabranjen u Rusiji, Yandex tražilica ima postotak korištenja od 55% ostatak su Google i Google.ru po snazi korištenja (Brigham, 2019). Osim što je Yandex na ruskom jeziku, te je tako prilagođen građanima Rusije, začuđujuća informacija nalaže kako je Yandex i stariji od Googla za godinu dana. Te u tom periodu je imao prednost u izlasku na tržište, odnosno ljudi u rusiji su ju počeli prvo koristiti i nisu imali potrebu prijeći na Google. Ljudi koriste ono na što su navikli. Tako da Google nije imao neke šanse probiti se na njihovo tržište, pogotovo jer je tek došao 2004. godine u Rusiju (Brigham, 2019). Osim internetskog pretraživanja, Yandex je i najveća tehnološka firma isto tako i medijska kompanija. U globalu toliko je rasprostranjena i prati događaje, razvoj i napredak populacije i tehnologije, da prelazak na google bi značio pad točnosti i relativnosti podataka i rezultata koji se prikazuju u toj zemlji. Uz to, nakon toliko godina implementiranja i prikupljanja resursa, mogu pretpostaviti njihovo sljedeće nadograđivanje i putanje što bi korisnici htjeli. Na poslijetku tekstovi su na njihovom jeziku i pisani njihovim pismom (CNBC, 2019).

6.2. Baidu

Baidu se naziva Googlom Kine, pokrenut je projekt 2000. godine od strane Robin Li-a i Eric Xu-a, što kasnije postaje kineska najpopularnija internetska tražilica i kineski tehnološki gigant, koji u suvremenom dobu se bavi razvijanjem umjetne inteligencije i samovozećih auta (Kharpal i Chang, 2021). Kao internetska tražilica 70% internetskih pretraživanja prolazi kroz Baidu, te s obzirom na veliku populaciju te države, sama tražilica Baidu dospijeva u top 5 tražilica na

svjetskoj razini (Chris, 2023). Iako taj postotak glasi nešto malo više od 2%, ostale tražilice nisu ni spomenute. Osim čistog pretraživanja informacija, Baidu ima svoje karte, medije, videoe i pretrage o slikama, čak ima i cijelu svoju enciklopediju. Što nije toliko čudno, ali daje malo jasniju sliku zašto se u tako masivnim i visoko populacijskim zemljama koriste vlastite tražilice i generalno vlasiti internetski prostori. To su količine informacija koje zahtijevaju cijelu svoju platformu. Kina sama po sebi koliko god je masivna, uspješna u svojim izvozima i trgovinom. Informacijama je odcjepljena od svijeta, te zbog toga su stvorili svoj vlastiti internetski prostor ili u ovom slučaju internetski svijet. (SEOMandarin 2020).

Tablica 14: Najpopularnije kineske tražilice na svim medijima 2020.

Tražilica	Postotak
Baidu	76,33%
Sogou	11,27%
Google	3,49%
Haosou	3,01%
Shenma	2,87%
Bing	2,84%

Izvor: SEOMandarin (2020).

7. Tamna strana interneta

Na početku ovog poglavlja, treba razlučiti prvo razliku između površinskog web-prostora i dubokog web-prostora (engl. *deep web*), koji se još grana na tamni web-prostor (engl. *dark web*). Površinski web-prostor je dio internetskog informacijskog prostora kojemu se može pristupiti pomoću običnih tražilica kao što je Google, Bing, Yahoo i drugih tražilica koje se koriste u svakodnevnicima (Karr, 2021). Te stranice su indeksirane odnosno upisane u bazu podataka tražilice, što omogućuje tražilici da ih povuče u rangiranu listu rezultata koje zatim prikazuje putem poveznica na korisnikovom ekranu. Sve što korisnik može pronaći putem standardnih tražilica je samo 4% - 10% cijelog interneta (Karr, 2021). Zatim slijede pojmovi duboki i tamni web-prostor. Ta dva pojma se često miješaju. *Deep web* označava sve što se nalazi na internetu, a nije indeksirano, te nedostupno putem svakodnevnih tražilica. Njegov sadržaj uključuje sve što se nalazi iza plaćenog pristupa ili zahtjeva prijavu korisničkih podataka. Također obuhvaća i sadržaj koji vlasnik blokira da ti podaci budu indeksirani od strane tražilica (Guccione, 2021). To su medicinski kartoni, kamatne naplate, internetska članstva, zaštićeni podaci korporacija, šifre, i sakrivene baze podataka kojim se ne može pristupiti bez dozvole i identifikacije vlasnika. Prema procjenama, *deep web* čini oko 96-99% internetskog informacijskog web-prostora, samo mali komad je zapravo internet kojemu se pristupa putem običnih tražilica, što je već ranije spomenuto kao površinski web-prostor (Guccione, 2021).

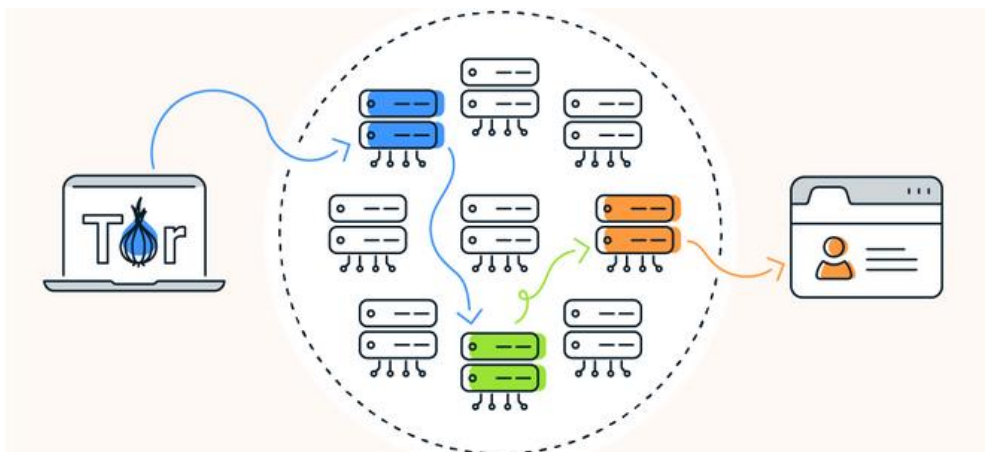
7.1. Dark web

Dark web je podcjelina *deep weba* koji je namjerno sakriven, zahtjeva posebne internetske preglednike i tražilice za bilo kakav pristup toj strani internetskog prostora (Whitmore, 2023). Nikakav podatak nije indeksiran i stranica nije indeksirana, zato ni ne postoje njihove stranice kada bi ih korisnik pokušao naći na primjer, kroz konvencionalne preglednike i tražilice. Prema istraživanjima Daniel Moorea i Thomas Rida, na fakultetu Kings College u Londonu, klasificirali su 2015. godine da postoje 2,723 aktivne stranice *dark weba* unutar 5 tjedana, te da su 57% od ukupnog broja aktivnih stranica na *dark webu* ilegalnog i nezakonitog sadržaja (Guccione, 2021). Godine 2019. provedeno je još jedno istraživanje koje sugerira da je tijekom godina stanje postalo još gore u odnosu na 2015. godinu, te da isključujući krijumčarenje ilegalnih supstanci, 60% ostalih informacija i trgovanja tom stranom interneta mogu imati vrlo negativne posljedice (Guccione, 2021). Pristupom takvom jednom internetskom prostoru, uz

dovoljnu količinu novaca, korisnik može kupiti podatke kreditnih kartica, narkotike i supstance, oružje, krivotvoreni novac, ukradene pretplate, usluge ulaženja u tuđa računala, kreiranje lažnog identiteta i dokumenata (Guccione, 2021). Korisnik se na prvu može iznenaditi kako web-sjedišta na *dark webu* izgledaju. Dizajnom, strukturom i izborom podsjećaju na internetske trgovine na koje su korisnici navikli. No umjesto nekretnina, automobila, mobitela i naprava za zabavu, od sadržaja su ponuđeni najčešće ilegalni proizvodi i usluge. Internetsko mjesto je pozicionirano na više različitih poveznica, gdje iz minute u minutu skače sa jedne na drugu domenu, kako vlasti i antikriminalne organizaciju ukoliko sruše jednu od domena ili stranica, ona nastavlja raditi i prebacivati se na ostale, sve dok se i ta srušena domena ponovo ne uključi. Na taj način web-sjedišta *dark weba* nastavljaju funkcionirati (McNamara, z2023).

7.2. Tor preglednik

Tor je kratica za *The Onion Routing project* što je otvorena mreža za privatnost koja omogućuje anonimno pregledavanje internetskih mjesta (Porutiu, 2023). Svjetska računalna mreža *Tor* koristi sigurne i šifrirane protokole kako bi osigurala zaštitu korisničke digitalne privatnosti. Digitalni podaci i komunikacija korisnika *Tora* su zaštićeni korištenjem slojevite metode koja nalikuje slojevima luka. (Frankenfield, Brown, 2022). Funkcionira na način mreže virtualnih tunela koji omogućuju privatnost kretanja kroz internet. Podatci koji se šalju, prolaze kroz tri glavna nasumična servera koji još prolaze kroz više od 7000 neovisnih mrežnih *relay-eva* u *Torovoj* mreži, te zadnji server ili pod nazivom izlazni server šalje podatke na javni internetski prostor.



Slika 39: Prikaz rada tor servera i enkripcija

Izvor: Ghimiray (2022).

Srednji oblak predstavlja *Torove* servere a ključevi predstavljaju enkripciju koja ide sukladno sa podacima, odnosno traženim informacijama do željene stranice. Svaki server koji dobije promet podataka koji se traži zna samo odakle je taj podatak došao od prijašnjeg servera, na taj način korisnik od prvog unosa ostaje sakriven, s obzirom da prolazi kroz više od 7000 takvih mrežnih sustava, korisnik je u tom vrlo kratkom vremenu promijenio velik broj IP adresa i lokacija, što je možda poznatije kada netko spomene da ulaskom na *dark web* njegova adresa može izgledati kao da je iz Nizozemske, a zapravo živi u Hrvatskoj. Svaki server mijenja enkripciju i svaki idući server samo zna da je zadobio promijenjenu lokaciju sa podatkom, ne znajući odakle je prvobitno slana, zatim opet taj server mijenja lokaciju i enkripciju do sljedećeg, i tako sve dok korisnik ne uđe na željeno mjesto, a sve to se događa u nekoliko sekundi, isto kao i što na običnoj tražilici dobijemo par milijuna rezultata u samo nekoliko milisekundi (Ghimiray, 2022).

8. Zaključak

Internetski informacijski prostor je jedna velika virtualna knjižnica informacija, slika, videozapisa i medija. No doći nekada do točnih i željenih resursa može biti zamarajuće. S time kako bi se korisnik znao kretati efikasno po internetskom prostoru, prvo se mora naučiti kako efikasno i koristiti njegove tražilice. Korisniku je poznato otvoriti preglednik, unijeti željeni upit i početi tražiti rezultate. No mala količina korisnika se zna pravilno i precizno koristiti sa tražilicom i ključnim riječima. Dobro razumijevanje tražilica, ključnih riječi i kako uopće tražilica funkcionira, daje korisniku slobodu da ne istražuje internetski prostor šablonski. Jasna je dominacija Google tražilice na tržištu prema količini korisnika, stoga Google stoji na prvom mjestu kao najkorištenija tražilica 2023. godine, zatim Microsoft Bing je na drugom mjestu i Yahoo na trećem mjestu. Četvrto i peto mjesto zauzele su tražilice Baidu i Yandex, specijalizirane za kinesko odnosno rusko tržište. Za dublje istraživanje nekih tema i potrebe edukacijskih resursa, korisnik bi trebao vertikalno istraživati internet odnosno koristiti specijalizirane tražilice, kako one pohranjuju resurse isključivo za željene teme, može se smatrati da su slojevito obogaćeni informacijama koje korisnik baš treba. Sa razlogom se još nazivaju vertikalne tražilice informacija. Analiza 3 najpoznatije tražilice dovodi do zanimljivih rezultata koji govore kako korištenija tražilica zahtjeva i više informacija, te ujedno ima i više korisnika koji te informacije obogaćuju. Postotni rezultat tog istraživanja se skoro poklapa sa postotkom korištenja tražilica u 2023. godini. Kretanje internetom također može biti i opasno. Površinski web-prostor je pristupačan svim korisnicima putem standardnih tražilica poput Googla. Postoji dio interneta nazvan duboki web-prostor, gdje su skladišteni podaci poput medicinskih dokumenata, bankovnih računa, osobnih podataka. Čine oko 94% ukupnog interneta, i nije pristupačan svakome, niti je indeksiran, pa tražilica ne može tu informaciju svrstati u rezultate sa poveznicama. Kretanje internetom također može biti i opasno s obzirom da se korisnik javno izlaže virtualnom svijetu, u dubokom dijelu web-prostora nalazi se tamni web-prostor. Ovaj dio web-prostora obilježen je s ponudom ilegalnih proizvoda i usluga, te kretanje tim dijelom interneta može rezultirati sa hakiranim osobnim i finansijskim podacima. Korisnikove podatke može se pronaći putem njegove IP adrese, te putem toga doseći sve ostale potrebne informacije o korisniku. Postoje dodatni nastavci i tražilice koje sakrivaju njegovu lokaciju, IP adresu i identitet, treba naglasiti da postoje alternativne tražilice koje odrade takvo sakrivanje umjesto korisnika.

Literatura

1. Avast (2022). Tor Dark web browser. Dostupno na: <https://www.avast.com/c-tor-dark-web-browser> [Pristupljeno: 28.06.2023]
2. Bhar, A. (2021). Secondary Research: Definition, methods and Examples. Question pro. Dostupno na: <https://www.questionpro.com/blog/secondary-research/> [Pristupljeno: 08.06.2023.]
3. Bleichert, J. (2023). Internet censorship. Experte. Dostupno na: <https://www.experte.com/internet-censorship/china> [Pristupljeno: 04.09.2023.]
4. Boslaugh, S. (2007). An introduction to secondary data analysis. Secondary data sources for public health. Cambridge university press publishing.
5. Bouhai, N., Imad, S. (2017). Internet of Things, Evolutions and innovations. John Wiley & sons publishing.
6. Central statistics office (2020). Type of internet activities. CSO. Dostupno na: <https://www.cso.ie/en/releasesandpublications/ep/p-isshh/informationstistics-households2020/typeofinternetactivities/> [Pristupljeno: 05.09.2023.]
7. Chris, A. (2023). Top 10 search engines in the world. Reliablysoft. Dostupno na: <https://www.reliablysoft.net/top-10-search-engines-in-the-world/> [Pristupljeno: 28.09.2023]
8. Cox, C., (2022). Search engine algorithms what to know. Hawksem. Dostupno na: <https://hawksem.com/blog/search-engine-algorithms-what-to-know/> [Pristupljeno: 25.06.2023]
9. Davies, D. (2023). Website indexing. Dostupno na: <https://www.searchenginejournal.com/search-engines/website-indexing/#close> [Pristupljeno: 29.08.2023.]
10. Duo, M. (2023). Google Search Operators: 40 Commands to Know in 2023 (Improve Research, Competitive Analysis, and SEO). Kinsta. Dostupno na: <https://kinsta.com/blog/google-search-operators/> [Pristupljeno: 08.06.2023.]
11. Feta, D. (2022). What are keywords (And why you need to know how to find them) Dostupno na: <https://blog.hubspot.com/marketing/keywords> [Pristupljeno 29.08.2022)
12. Fitzgerald C. (2023). News aggregator websites examples. Theme isle. Dostupno na: <https://themeisle.com/blog/news-aggregator-websites-examples/> [Pristupljeno 22.06.2023.]
13. Forsey, C. (2023). The top 11 search engines, ranked by popularity. Dostupno na: <https://blog.hubspot.com/marketing/top-search-engines> [Pristupljeno: 05.09.2023.]

14. Freecodecamp (2023). PHP *Web scraping*. Freecodecamp. Dostupno na: <https://www.freecodecamp.org/news/web-scraping-with-php-crawl-web-pages> [Pristupljeno: 19. 06.2023.]
15. Ghimiray, D. (2022). What is Tor Browser. Dostupno na: <https://www.avast.com/c-tor-dark-web-browser> [Pristupljeno: 05.09.2023.]
16. Google Search Center (2023). In-depth guide how google search works. Developers. Dostupno na: <https://developers.google.com/search/docs/fundamentals/how-search-works> [Pristupljeno: 05.09.2023.]
17. Guccione, D. (2021). What is the dark web? How to access it and what you will find. Dostupno na: <https://www.csoonline.com/article/564313/what-is-the-dark-web-how-to-access-it-and-what-youll-find.html> [Pristupljeno: 05.09.2023.]
18. Halavais, A. (2018). Search engine society. Polity press publishing.
19. Harsel, L. (2022). What are keywords, simple keyword definition. Semrush blog. Dostupno na: <https://www.semrush.com/blog/what-are-keywords-simple-keyword-definition/> [Pristupljeno: 05.09.2023.]
20. Hardwick, J. (2023). Google Search Operators. Ahrefs Blog. Dostupno na: <https://ahrefs.com/blog/google-advanced-search-operators/> [Pristupljeno: 08.06.2023.]
21. Hardwick, J. (2023). Advanced Search operators. Ahrefs. Dostupno na: <https://ahrefs.com/blog/google-advanced-search-operators/> [Pristupljeno: 26.06.2023.]
22. Hiler, W. (2022). What is Secondary Data?. CF Blog. Dostupno na: <https://careerfoundry.com/en/blog/data-analytics/what-is-secondary-data/> [Pristupljeno: 08.06.2022.]
23. Imperva (2023). Web scrapping. Imperva. Dostupno na: <https://www.imperva.com/learn/application-security/web-scraping-attack/> [Pristupljeno: 08.06.2023.]
24. JotForm (2023). A guide on primary and secondary data-colection methods. Jotform. Dostupno na: <https://www.jotform.com/blog/primary-and-secondary-data-collection-methods/> [Pristupljeno: 08.06.2023.]
25. Khan, Y. J., Yuce, R.M. (2019.) Internet of Things (IoT) Systems and Applications. Jenny Stanford publishing

26. Kharpal, A. (2021). Baidu selfdriving car. Dostupno na: <https://www.cnbc.com/2021/08/18/baidu-launches-robocar-and-ai-chip-in-bid-to-diversify-business.html?&qsearchterm=Baidu> [Pristupljeno: 04.09.2023.]
27. Lewandowski D. (2023). Understanding search engines. Springer international publishing.
28. Martins F. S, da Cunha J. A. C., & Serra, F. A. R (2018). Secondary data in research-uses and opportunities. PODIUM sport, leisure and tourism review, 7(3) I-IV.
29. McNamara, P. (2023). Why is the dark web not shut down ? Dostupno na: <https://mentorcruise.com/blog/why-is-the-dark-web-not-shut-down-3f799/> [Pristupljeno: 04.09.2023.]
30. Muslim, L. (2019). Optimizacija internetskih stranica za web-tražilice kao strategija digitalnog marketinga. Dabar. Dostupno na: <https://urn.nsk.hr/urn:nbn:hr:124:816821> [Pristupljeno: 08.06.2023.]
31. Meng, W., T.Yu, C. (2011). Advanced Metasearch engine Technology. Morgan & Claypool publishing.
32. Noy, N. (2019). Google Dataset Search: Building a search engine for datasets in an open Web ecosystem. Association for Computing Machinery publishing.
33. Oberlo (2023). Browser market share. Dostupno na: <https://www.oberlo.com/statistics/browser-market-share> [Pristupljeno: 22.06.2023.]
34. Patell, N. (2022). What we learned in 2022. and where SEO is going in 2023. NP Digital. Dostupno na: <https://neilpatel.com/blog/search-engine-trends/> [Pristupljeno: 08.06.2023.]
35. Periera, D. (2023). Google mission. Bussiness model analyst. Dostupno na: <https://businessmodelanalyst.com/google-mission-and-vision-statement/> [Pristupljeno: 05.09.2023.]
36. Pillay, N., Qu, R. (2021). Automated Design of Machine Learning and Search Algorithms. Springer international publishing.
37. Search engine land (2022). Advanced google search operators. Dostupno na: <https://searchengineland.com/advanced-google-search-operators-388355> [Pristupljeno 26.06.2023]
38. Seo Mandarin (2020). Top Chinese search engines 2020. Dostupno na: <https://www.seomandarin.com/top-chinese-search-engines-2020> [Pristupljeno: 26.06.2023]

39. Serpanos, D., Wolf, M. (2018). Internet of Things (IoT) Architectures, Algorithms, Methodologies. Springer international publishing.
40. Shieildar, J. (2021). What is Yandex ?. Loganix. Dostupno na: <https://loganix.com/what-is-yandex/> [Pristupljeno: 08.06.2023.]
41. Southwick, K. (2022). Boostblog keywords. Boostability. Dostupno na: <https://www.boostability.com/content/ways-to-implement-your-seo-keyword-strategy-into-your-content> [Pristupljeno: 05.09.2023.]
42. Teran, J. (2018). What is search engine algorithm. Volusion. Dostupno na: <https://www.volusion.com/blog/search-algorithms/> [Pristupljeno: 27.08.2023.]
43. Themeisle (2023). News aggregator websites examples. Dostupno na: <https://themeisle.com/blog/news-aggregator-websites-examples/> [Pristupljeno: 05.09.2023.]
44. Udovičić, L. (2021). Pregled i usporedba web tehnologija. Dabar. Dostupno na: <https://urn.nsk.hr/urn:nbn:hr:195:300432> [Pristupljeno: 08.06.2023.]
45. Vartanian, T. P. (2010). Secondary data analysis. Oxford University Press
46. Vizentaner, D. (2021). Marketinške aktivnosti na tražilicama. Dabar. Dostupno na: <https://urn.nsk.hr/urn:nbn:hr:145:840125> [Pristupljeno: 08.06.2023.]
47. Volle, A. (2022). Search engine. Britannica. Dostupno na: <https://www.britannica.com/technology/search-engine> [Pristupljeno: 08.06.2023.]
48. Vučina Ž. (2006). Pretraživanje i vrednovanje informacija na internetu. Edupoint Publishing.
49. Zahran, O. (2020). The Evolution of search engine. Omarzahran-Medium Dostupno na: <https://omarzahran.medium.com/the-evolution-of-the-search-engine-c9b0bb08bfb0> [Pristupljeno: 08.06.2023.]
50. Žiljak, G. J. (2020). Zbornik radova. Tiskarstvo.net. Dostupno na: <https://www.tiskarstvo.net/printing&design2023/ZbornikRadovaTISKARSTVO2022.>

[pdf#page=65](#)

[Pristupljeno: 08.06.2023.]

51. Yext (2021). What are specialized search engines. Yext. Dostupno na: <https://www.yext.com/blog/2021/10/what-are-specialized-search-engine> [Pristupljeno: 24.06.2023]
52. Whitmore (2023). What is the dark web, and how do you access it ? Dostupno na: <https://nordvpn.com/blog/what-is-the-dark-web/> [Pristupljeno: 05.09.2023.]
53. Widmer, B. (2022). How the google search algorithm. Ahrefs. Dostupno na: <https://ahrefs.com/blog/google-search-algorithm/> [Pristupljeno: 29.08.2023.]
54. Woodford, C. (2023). How the web works. Explain that stuff. Dostupno na: <https://www.explainthatstuff.com/howthewebworks.html> [Pristupljeno: 22.06.2023.]
55. WordStream (2023). Meta search engine, what is meta search engine?. LocalIO Dostupno na: <https://www.wordstream.com/meta-search-engine> [Pristupljeno: 08.06.2023.]

Popis slika

Slika 1: Vizualni prikaz Web page-a i Website-a	2
Slika 2: Jednostavan prikaz informacijskog prostora i medija za komuniciranje.....	3
Slika 3: Statistički prikaz razine korištenja tražilica na internetu 2023. godine.	7
Slika 4: Primjer zahtjeva za indeksiranje određene stranice	11
Slika 5: Pretraživanje ključne riječi "Pustinja" + rezultati	13
Slika 6: Pretraživanje iste ključne riječi uz dodatne opisne ključne riječi + rezultati	13
Slika 7: Primjer rezultata sa drugačijim sufiksom (zadnjim nastavkom).....	14
Slika 9: Primjer napredne tehnike pretraživanja.....	16
Slika 11: Analitika dobivenih rezultata na Google-u, pod ključnom riječi "Netflix"	20
Slika 12: Analitika dobivenih rezultata na Bing-u, pod ključnom riječi "Netflix"	20
Slika 13: Analitika dobivenih rezultata na Yahoo, pod ključnom riječi "Netflix"	20
Slika 14: Dobiveni rezultati pretrage tražilicom Google, na ključnu riječ "Amazon"	21
Slika 15: Dobiveni rezultati pretrage tražilicom Bing, na ključnu riječ "Amazon".....	21
Slika 16: Dobiveni rezultati pretrage tražilicom Yahoo, na ključnu riječ "Amazon"	21
Slika 17: Dobiveni rezultati pretrage tražilice Google, na ključnu riječ "WHO" (World health organisation).....	22
Slika 18: Dobiveni rezultati pretrage tražilice Bing, na ključnu riječ "WHO" (World health organisation).....	22
Slika 19: Dobiveni rezultati pretrage tražilice Yahoo, na ključnu riječ "WHO" (World Health Organisation).....	22
Slika 20: Dobiveni rezultati pretrage tražilice Google, na ključnu riječ "Ikea"	23
Slika 21: Dobiveni rezultati pretrage tražilice Bing, na ključnu riječ "Ikea"	23
Slika 22: Dobiveni rezultati pretrage tražilicom Yahoo, na ključnu riječ "Ikea"	23
Slika 23: Dobiveni rezultati pretrage tražilice Google, na ključnu riječ "Alpe"	24
Slika 24: Dobiveni rezultati pretrage tražilice Bing, na ključnu riječ "Alpe"	24
Slika 25: Dobiveni rezultati pretrage tražilice Yahoo, na ključnu riječ "Alpe"	24
Slika 26: Dobiveni rezultati pretrage tražilice Google, na ključnu riječ "Skandinavija"	25
Slika 27: Dobiveni rezultati pretrage tražilice Bing, na ključnu riječ "Skandinavija"	25
Slika 28: Dobiveni rezultati pretrage tražilice Yahoo, na ključnu riječ "Skandinavija"	25
Slika 29: Dobiveni rezultati pretrage tražilice Google, na ključnu riječ "Jadran"	26
Slika 30: Dobiveni rezultati pretrage tražilice Bing, na ključnu riječ "Jadran"	26
Slika 31: Dobiveni rezultati pretrage tražilice Yahoo, na ključnu riječ "Jadran".....	26
Slika 32: Dobiveni rezultati pretrage tražilice Google, na ključnu riječ "Sahara".....	27
Slika 33: Dobiveni rezultati pretrage tražilice Bing, na ključnu riječ "Sahara"	27
Slika 34: Dobiveni rezultati pretrage tražilice Yahoo, na ključnu riječ "Sahara"	27
Slika 35: Dobiveni rezultati pretrage tražilice Google, na ključnu riječ "Harvard"	28
Slika 36: Dobiveni rezultati pretrage tražilice Bing, na ključnu riječ "Harvard"	28
Slika 37: Dobiveni rezultati pretrage tražilice Yahoo, na ključnu riječ "Harvard"	28
Slika 38: Dobiveni rezultati pretrage tražilice Google, na ključnu riječ "Oxford"	29
Slika 39: Dobiveni rezultati pretrage tražilice Bing, na ključnu riječ "Oxford".....	29
Slika 40: Dobiveni rezultati pretrage tražilice Yahoo, na ključnu riječ "Oxford".....	29
Slika 42: Prikaz rada tor servera i enkripcija.....	34

Popis tablica

Tablica 1: Primjeri metatražilica za pretraživanje internetskog prostora i njena obilježja.....	8
Tablica 2: Analiza rezultata na ključnu riječ "Netflix"	20
Tablica 3: Analiza rezultata na ključnu riječ "Amazon"	21
Tablica 4: Analiza rezultata na ključnu riječ "WHO"	22
Tablica 5: Analiza rezultata na ključnu riječ "Ikea"	23
Tablica 6: Analiza rezultata na ključnu riječ "Alpe"	24
Tablica 7: Analiza rezultata na ključnu riječ "Skandinavija"	25
Tablica 8: Analiza rezultata na ključnu riječ "Jadran"	26
Tablica 9: Analiza rezultata na ključnu riječ "Sahara"	27
Tablica 10: Analiza rezultata na ključnu riječ "Harvard"	28
Tablica 11: Analiza rezultata na ključnu riječ "Oxford"	29