SVEUČILIŠTE U ZAGREBU

FAKULTET ORGANIZACIJE I INFORMATIKE

V A R A Ž D I N

**Hrvoje Vujasinović**

ANALIZA I VIZUALIZACIJA DNEVNIČKIH ZAPISA U PYTHONU

ZAVRŠNI rad

Varaždin, 2020.

SVEUČILIŠTE U ZAGREBU

FAKULTET ORGANIZACIJE I INFORMATIKE

V A R A Ž D I N

Hrvoje Vujasinović

Matični broj: 46252/17–R

Studij: Informacijski sustavi

ANALIZA I VIZUALIZACIJA DNEVNIČKIH ZAPISA U PYTHONU

ZAVRŠNI RAD

Mentor/Mentorica:

Dr. sc. Miran Zlatović

Varaždin, kolovoz 2020.

*Hrvoje Vujasinović*

Izjava o izvornosti

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

*Autor/Autorica potvrdio/potvrdila prihvaćanjem odredbi u sustavu FOI-radovi*

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Sažetak

Tema ovog završnog rada su dnevnički zapisi, njihova svrha i uloga na osobnim računalima i web poslužiteljima te analiza i vizualizacija podataka dobivenih iz dnevničkih zapisa. U radu ću najprije objasniti što su dnevnički zapisi i čemu služe, a zatim ću obraditi temu web poslužitelja. Nakon toga bit će opisana organizacija i struktura dnevničkih zapisa na osobnim računalima i to na Linux platformama za osobna računala, a za Windows platforme obradit ću Windows Event Viewer. Nakon toga obradit ću dnevničke zapise na web poslužiteljima i nekoliko formata dnevničkih zapisa koji se koriste na web poslužiteljima. U praktičnom dijelu rada izradit ću Python aplikaciju s grafičkim sučeljem pomoću modula Tkinter za parsiranje jednog od formata dnevničkih zapisa. Aplikacija će omogućavati parsiranje i pretraživanje podataka, ekstrakciju podskupa podataka i vizualizaciju ekstrahiranih podataka pomoću dostupnih Python biblioteka.

Ključne riječi: logs; event logs; server logs; Pyhon; web poslužitelji; parsiranje podataka; analiza podataka; vizualizacija podataka

Sadržaj

[Sadržaj iii](#_Toc49452386)

[1. Uvod 1](#_Toc49452387)

[2. Dnevnički zapisi i njihova svrha 2](#_Toc49452388)

[2.1. Dnevnički zapisi na osobnim računalima 3](#_Toc49452389)

[2.1.1. Windows Event Viewer 3](#_Toc49452390)

[2.1.2. Dnevnički zapisi na Linux platformama 5](#_Toc49452391)

[2.2. Dnevnički zapisi na web poslužiteljima 7](#_Toc49452392)

[2.2.1. Web poslužitelji 7](#_Toc49452393)

[3. Python aplikacija za analizu i vizualizaciju 9](#_Toc49452394)

[4. Zaključak 10](#_Toc49452395)

[Popis literature 11](#_Toc49452396)

[Popis slika 12](#_Toc49452397)

[Popis tablica 13](#_Toc49452398)

[Prilozi (1, 2, …) 14](#_Toc49452399)

1. Uvod

Osobna računala, web poslužitelji i informacijski sustavi generiraju dnevničke zapise (eng. *Logs*) koji dokumentiraju aktivnosti sustava. Takvi se zapisi sastoje od niza kronološki poredanih poruka koje sadrže informacije o aktivnostima i operacijama unutar operacijskog sustava, neke aplikacije, web poslužitelja ili nekog drugog uređaja. Analiza podataka dobivenih iz dnevničkih zapisa omogućuje praćenje ponašanja sustava i otkrivanje problema kao što su greške u radu ili napadi na sustav. Danas postoje i brojna programska rješenja za analizu dnevničkih zapisa koja omogućuju bolju ekstrakciju podataka iz dnevničkih zapisa i na taj način olakšavaju pronalaženje trendova i uzoraka u podacima pomoću kojih administratori sustava mogu donositi bolje poslovne odluke ili nadzirati sigurnost sustava. (E. Zhang, 2018)

U ovom radu bit će opisana organizacija dnevničkih zapisa na osobnim računalima i web poslužiteljima te izrađena vlastita Python aplikacija za analizu jednog formata dnevničkog zapisa.

1. Dnevnički zapisi i njihova svrha

Dnevnički zapisi su datoteke automatski generirane od strane računala koje sadrže popis događaja na tom računalu u obliku strukturiranih, kronološki poredanih poruka. Većina dnevničkih zapisa spremljeni su u formatu običnog teksta što omogućuje pregled u bilo kojem programu za uređivanje teksta i osigurava malu veličinu datoteke. (Christensson, 2010)

Podaci, struktura, vrste i format poruka u dnevničkom zapisu nisu propisani pa tako ovise o odlukama i implementaciji razvojnih programera koji rade na razvoju određene aplikacije ili sustava. Iako se implementacije i formati poruka razlikuju sve su sastavljene od nekoliko zajedničkih komponenti. Svaka poruka ima datum i vrijeme kada je nastala, razinu zapisa, na primjer informativna poruka, poruka upozorenja, poruka greške ili neka druga definirana razina, i informacije o kontekstu koje nam pružaju dodatne informacije o stanju sustava ili aplikacije i okruženju u kojem je poruka nastala. (Lee, 2019)

Prema (Lee, 2019) dnevničke zapise možemo podijeliti na nekoliko osnovnih skupina, a to su aplikacijski dnevnički zapisi (eng. *Application log*), sustavski dnevnički zapisi (eng. *System log*) i poslužiteljski dnevnički zapisi (eng. *Server log*). Aplikacijski dnevnički zapisi služe razvojnim programerima aplikacije kako bi lakše otkrili i otklonili probleme u radu aplikacije budući da se neki problemi ne mogu otkriti sve dok se aplikacija ne nađe u produkcijskoj okolini i bude dana na korištenje krajnjim korisnicima. Sustavski dnevnički zapisi sadrže informacije o radu sustava i sustavskim procesima dok poslužiteljski dnevnički zapisi bilježe informacije o korisnicima koji ga koriste.

Poslužiteljski dnevnički zapisi su od posebne važnosti za velike organizacije čije poslovanje ovisi o njihovoj mrežnoj infrastrukturi. Analiza i praćenje poslužiteljskih dnevničkih zapisa omogućuje im da povećaju pouzdanost svojih sustava ispravljanjem grešaka, sporih upita ili netočnih odgovora poslužitelja i na taj način poboljšaju iskustvo krajnjeg korisnika. Isto tako praćenjem dnevničkih zapisa osiguravaju sigurnost sustava i podataka od vanjskih napadača budući da ovi zapisi vode detalje o događajima vezanim za sigurnost kao što su uspiješne i neuspiješne prijave. Još jedan razlog za analizu poslužiteljskih dnevničkih zapisa je praćenje ponašanja korisnika na temelju kojega možemo optimizirati sustav da bolje odgovara njihovim potrebama ili donositi bolje poslovne odluke. Također organizacije mogu na vrijeme odgovoriti na povećanje opterećenja sustava zbog rasta broja korisnika uvođenjem dodatnih kapaciteta. (*What is a Log File? | Sumo Logic*, 2019)

Možemo reći da je glavna svrha svih dnevničkih zapisa dati korisnicima ili administratorima informacije o ponašanju sustava ili aplikacije kroz vrijeme kako bi pravovremeno mogli reagirati na probleme, greške i sigurnosne prijetnje u radu sustava ili mogli donositi bolje poslovne odluke. U narednim poglavljima ovog rada detaljnije će biti opisana organizacija dnevničkih zapisa na osobnim računalima i poslužiteljima

* 1. Dnevnički zapisi na osobnim računalima

Na modernim operacijskim sustavima za osobna računala ugrađeni su mehanizmi i protokoli za snimanje događaja i vođenje dnevničkih zapisa o događajima u sustavu. Takvi zapisi daju korisnicima uvid u rad sustava i prikupljaju informacije o radu sklopovlja i programa, problemima u sustavu i sigurnosnim događajima. U ovom poglavlju bit će obrađena organizacija sustavskih dnevničkih zapisa na Windows platformi i alat Windows Event Viewer te organizacija sustavskih dnevničkih zapisa na Linux platformi.

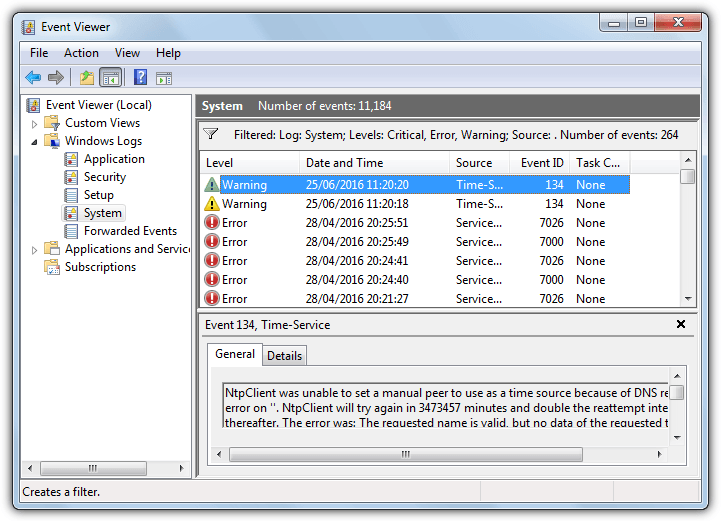
* + 1. Windows Event Viewer

Na Windows operacijskim sustavima dnevnički zapisi spremaju se u XML formatu na lokaciju *C:\WINDOWS\system32\config*. Svaki zapis sastoji se od (*Event Log Monitoring Tool - A Tutorial*, bez dat.):

1. Datum - datum na koji je događaj nastao
2. Vrijeme - vrijeme u koje je događaj nastao
3. Korisnik - prijavljeni korisnik u trenutku nastanka događaja
4. Računalo - naziv računala na kojem se događaj dogodio
5. ID događaja - identifikator događaja koji specificira tip događaja
6. Izvor - aplikacija ili komponenta koja je uzrokovala događaj
7. Tip ili razina - tip događaja prema važnosti

Na Windows platformi tip događaja ili razina zapisa može biti informacija (eng. *Information*), upozorenje (eng. *Warning*), greška (eng. *Error*), uspješna autorizacija (eng. *Success Audit*) ili neuspješna autorizacija (eng. *Failure Audit*). Informacija je događaj koji opisuje uspješnu operaciju neke radnje u aplikaciji, upravljačkom programu ili pokrenutom servisu. Upozorenje je događaj koji nije od prevelike važnosti, ali može ukazivati na neke buduće probleme koji bi mogli nastati. Greška je događaj koji ukazuje na problem koji je nastao u radu neke aplikacije ili komponente i pri tome uzrokovao gubitak podataka ili funkcionalnosti. Uspješna i neuspješna autorizacija su događaji koji opisuju uspješan ili neuspješan sigurnosni događaj, na primjer kada se korisnik uspješno ili neuspješno prijavi. (*Event Log Monitoring Tool - A Tutorial*, bez dat.)

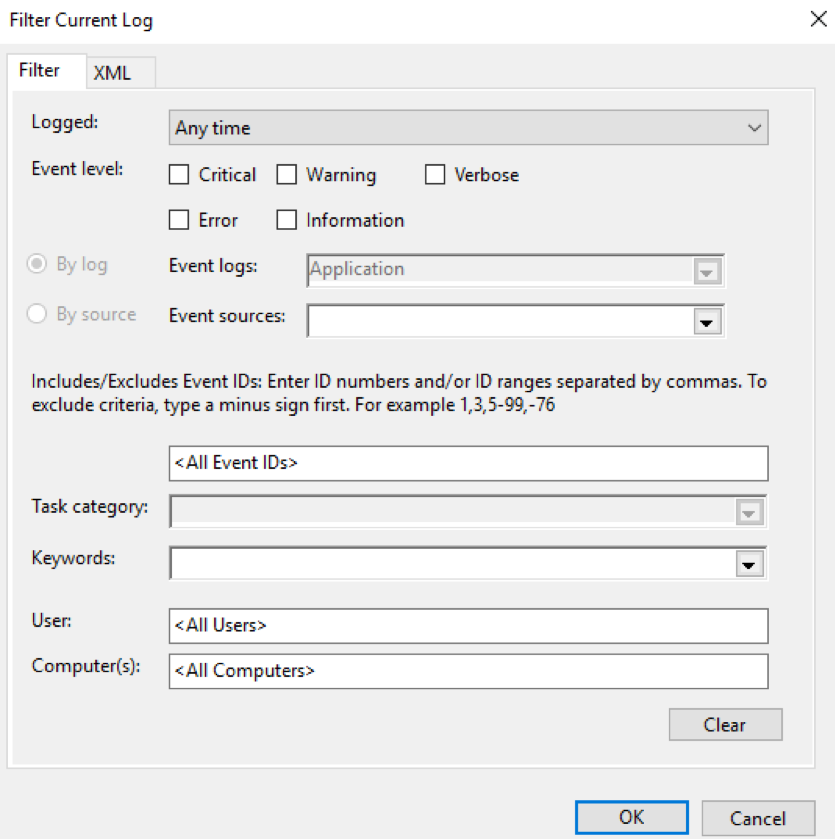
Event Viewer je sustavski alat koji dolazi u sklopu Windows operacijskog sustava i omogućuje pregled dnevničkih zapisa koje taj operacijski sustav generira. Alat omogućuje i dodatne opcije kao što su pretraživanje zapisa, filtriranje zapisa, izvoz zapisa u nekom drugom formatu pogodnom za daljnju analizu ili definiranje korisničkih vrsta događaja i dnevničkih zapisa.



Slika 1: Izgled sučelja Windows Event Viewer-a (Izvor: raymond.cc, 2020)

S lijeve strane sučelja nalazi se okvir za navigaciju u kojem možemo izabrati dnevnički zapis koji želimo pregledati. Postoji pet zadanih kategorija Windows dnevničkih zapisa i to aplikacije (eng. *Application*), sigurnost (eng. *Security*), instalacija (eng. *Setup*), sustav (eng. *System*) i proslijeđeni događaji (eng. *Forwarded Events*). U dnevničkom zapisu aplikacije nalaze se informacije o događajima koje su generirale aplikacije instalirane na računalu. Dnevnički zapis sigurnost sadrži informacije o uspješnim i neuspješnim sigurnosnim događajima na računalu. Dnevnički zapis instalacija sadrži poruke o događajima vezanim uz instalaciju i nadogradnju operacijskog sustava. Dnevnički zapis sustav sadrži poruke koje je generirao operacijski sustav dok dnevnički zapis proslijeđeni događaji sadrži poruke o događajima s drugih računala na mreži u slučaju da je računalo konfigurirano da prikuplja te podatke. (*Windows Logging Basics - The Ultimate Guide To Logging*, 2015)

Na desnoj strani sučelja nalazi se popis poruka u odabranom dnevničkom zapisu s detaljima o vremenu i okruženju u kojem je poruka nastala. Poruke su poredane kronološki od najstarijih na dnu prema najnovijim na vrhu. Odabirom neke od poruka možemo vidjeti opis događaja u donjem desnom okviru pod kraticom *General* i dodatne detalje pod kraticom *Details*. Alat omogućuje i napredno filtriranje dnevničkih zapisa prema različitim kriterijima kao što su vrijeme nastanka poruke, razina ili tip poruke, ključnim riječima, izvoru događaja, ID događaja, računalu ili korisniku.

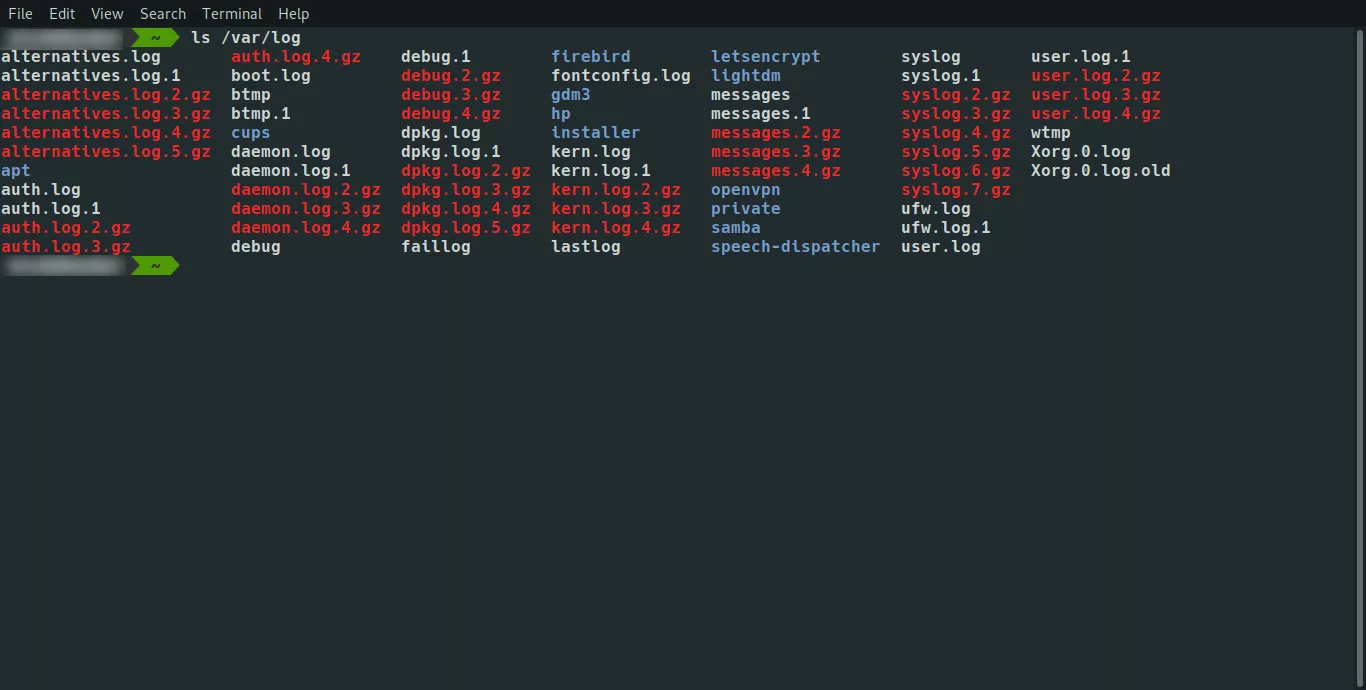


Slika 2: Opcije filtriranja u Event Viwer-u (Izvor: loggly.com, 2020)

* + 1. Dnevnički zapisi na Linux platformama

Program koji obavlja prikupljanje i sastavljanje dnevničkih zapisa na Linux platformama je *Sysklogd* i on se sastoji od dva alata, *syslogd* za prikupljanje sustavskih poruka i *klogd* za prikupljanje poruka jezgre sustava. (*sysklogd(8): system logging utilities - Linux man page*, bez dat.)

Na Linux operacijskim sustavima dnevnički zapisi spremljeni su u tekstualnom obliku na lokaciji */var/log* ili u nekom od poddirektorija vezanim uz pojedinu aplikaciju. Neki dnevnički zapisi komprimiraju se zbog svoje veličine pa imaju *.gz* nastavak. Svaki zapis definiran je *Syslog* formatom i sastoji se od standardiziranog zaglavlja. Zaglavlje sadrži datum i vrijeme nastanka događaja, naziv aplikacije koja je generirala događaj, mjesto na računalu koje je izvor poruke i prioritet poruke. Pri tome se prioritet poruke sastoji od dva broja od kojih prvi opisuje tip procesa koji je generirao događaj i može biti u rasponu od 0 do 23 gdje 0 označava poruku jezgre sustava, a 23 poruku aplikacije pokrenute na sustavu. Drugi broj odnosi se na važnost poruke i može biti između 0 i 7 gdje 0 znači da je poruka hitna i sustav se ne može koristiti, a poruka važnosti 7 opisuje pogrešku. Format poruke može se modificirati pomoću datoteke za konfiguraciju. (*Linux Logging Basics - The Ultimate Guide To Logging*, 2015)



Slika 3: Primjer sadržaja */var/log* direktorija (Izvor: lifewire.com, 2020)

Prema (*Linux Logging Basics - The Ultimate Guide To Logging*, 2015) neki od važnijih dnevničkih zapisa na Linux platformama su:

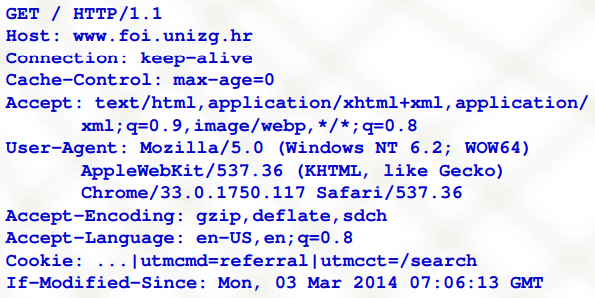
1. */var/log/syslog* ili */var/log/messages*
2. */var/log/auth.log* ili */var/log/secure*
3. */var/log/kern.log*
4. */var/log/cron*

Točan naziv datoteke ovisi o verziji operacijskog sustava, na primjer verzije bazirane na *Debian*-u koriste */var/log/syslog* i */var/log/auth.log* dok verzije bazirane na *Red Hat*-u koriste */var/log/messages* i */var/log/secure.* Dnevnički zapis *syslog* ili *messages* sadrži opće poruke sustava i informacije vezane uz rad sustava. Dnevnički zapis *auth.log* ili *secure* sadrži informacije o događajima vezanim za sigurnost kao što su prijava korisnika i slično. *Kern.log*  sadrži zapise o događajima, greškama i upozorenjima koje je generirala jezgra operacijskog sustava. *Cron* sadrži zapise o izvođenju korisnički zakazanih zadataka.

Dnevnički zapisi na Linux platformama također podržavaju rotiranje kako bi se izbjegla prevelika veličina datoteka. Alat *log rotate* obavlja ovu funkciju na način da briše dnevničke zapise starije od postavljenog vremenskog intervala i stvara nove umjesto njih. Namještanje vremeskog intervala vrši se u */etc/logrotate.conf* datoteci te se za svaku aplikaciju može postaviti vlastiti interval rotacije dnevničkog zapisa, hoće li se zapis komprimirati, slanje zapisa na e-mail prije brisanja i brojne druge opcije. (Haas, 2020)

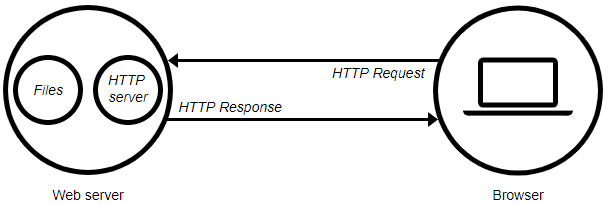
* 1. Dnevnički zapisi na web poslužiteljima
     1. Web poslužitelji

Web poslužitelji (eng. *Web server*)su računala s pripadnom programskom podrškom koja mogu zaprimati, obrađivati i odgovarati na zahtjeve klijenta pomoću HTTP ili drugih protokola. Klijent pomoću web preglednika (eng. *Browser*) šalje HTTP zahtjev koji sadrži informacije o klijentu, vrsti zahtjeva i traženom sadržaju prema web poslužitelju koji zatim prihvaća zahtjev i pronalazi traženi sadržaj te ga vraća klijentu u obliku HTTP odgovora. Ako traženi sadržaj nije pronađen ili je došlo do greške web poslužitelj vraća jedan od definiranih HTTP statusnih kodova. (*What is a web server? - Learn web development | MDN*, 2020)



Slika 4: Primjer HTTP zahtjeva (Izvor: Kermek D., 2020)

Sadržaj koji se vraća klijentu može biti statička web stranica ili dinamički generirana web stranica. Statička web stranica ima predefiniranu strukturu i sadržaj, najčešće pomoću HTML-a i CSS-a, pa je ista za svakog klijenta kojem se šalje. Svaka stranica spremljena je na web poslužitelju kao zaseban dokument i ukoliko ga je potrebno izmijeniti to može napraviti samo administrator. S druge strane, dinamički generirane web stranice generiraju se u stvarnom vremenu uzimajući u obzir klijenta koji je poslao zahtjev. Na web poslužitelju spremljeni su samo predlošci i dizajn za web stranicu dok se sadržaj popunjava iz baze podataka ili nekog drugog izvora pomoću nekog od programskih jezika za programiranje na strani poslužitelja kao što su PHP, ASP i drugi. (Robben, 2019)



Slika 5: Princip rada web poslužitelja (Izvor: MDN, 2020)

1. Python aplikacija za analizu i vizualizaciju
2. Zaključak

Ovdje treba sažeto rezimirati najvažnije rezultate razrade teme rada. Potrebno je sažeto opisati što je predmet rada, koje su metode, tehnike, programski alati ili aplikacije korištene u razradi rada te koje su pretpostavke dokazane, a koje opovrgnute. Sadržajno, ono što se u uvodu rada najavljuje i kasnije je obuhvaćeno u samom radu, moralo bi biti opisano u zaključnom dijelu kroz rezultate rada.

Popis literature

Christensson, P. (2010, travanj 14). *Log File Definition*. https://techterms.com/definition/logfile

E. Zhang. (2018). *What is Log Analysis? Use Cases, Best Practices, and More | Digital Guardian*. https://digitalguardian.com/blog/what-log-analysis-use-cases-best-practices-and-more

*Event Log Monitoring Tool - A Tutorial*. (bez dat.). Preuzeto 24. kolovoz 2020., od https://www.manageengine.com/network-monitoring/Eventlog\_Tutorial\_Part\_I.html

Haas, J. (2020, veljača 14). *What Are Linux Log Files and How Can You Read Them?* https://www.lifewire.com/introduction-to-linux-log-files-2192233

Lee, D. (2019). *Application Logs: What They Are and How to Use Them - XpoLog*. https://www.xplg.com/application-logs-what-how/

*Linux Logging Basics - The Ultimate Guide To Logging*. (2015, rujan 15). https://www.loggly.com/ultimate-guide/linux-logging-basics/

Robben. (2019, studeni). *How do servers work? A Detailed Guide into Web Servers*. https://host4geeks.com/blog/how-do-servers-work/?

*sysklogd(8): system logging utilities - Linux man page*. (bez dat.). Preuzeto 26. kolovoz 2020., od https://linux.die.net/man/8/sysklogd

*What is a Log File? | Sumo Logic*. (2019). https://www.sumologic.com/glossary/log-file/

*What is a web server? - Learn web development | MDN*. (2020). https://developer.mozilla.org/en-US/docs/Learn/Common\_questions/What\_is\_a\_web\_server

*Windows Logging Basics - The Ultimate Guide To Logging*. (2015, rujan 15). https://www.loggly.com/ultimate-guide/windows-logging-basics/

Popis slika

[Slika 1: Izgled sučelja Windows Event Viewer-a 4](#_Toc49452422)

[Slika 2: Opcije filtriranja u Event Viwer-u 5](#_Toc49452423)

[Slika 3: Primjer sadržaja */var/log* direktorija 6](#_Toc49452424)

[Slika 4: Primjer HTTP zahtjeva 7](#_Toc49452425)

[Slika 5: Princip rada web poslužitelja 8](#_Toc49452426)

Popis tablica

Popis tablica treba biti izrađen po uzoru na indeksirani sadržaj, te upućivati na broj stranice na kojoj se tablica može pronaći.

[Tablica 1: Prikaz podataka o učestalosti pojavljivanja objekta 7](#_Toc496692416)

Prilozi (1, 2, …)