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](#_Toc48747678)

[1. Uvod 1](#_Toc48747679)

[2. Dnevnički zapisi i njihova svrha 2](#_Toc48747680)

[3. Web poslužitelji 3](#_Toc48747681)

[4. Organizacija dnevničkih zapisa na osobnim računalima 4](#_Toc48747682)

[4.1. Windows Event Viewer 4](#_Toc48747683)

[4.2. Dnevnički zapisi na Linux platformama 4](#_Toc48747684)

[5. Dnevnički zapisi na web poslužiteljima 4](#_Toc48747685)

[6. Python aplikacija za analizu i vizualizaciju 4](#_Toc48747686)

[7. Zaključak 5](#_Toc48747687)

[Popis literature 6](#_Toc48747688)

[Popis slika 7](#_Toc48747689)

[Popis tablica 8](#_Toc48747690)

[Prilozi (1, 2, …) 9](#_Toc48747691)

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)

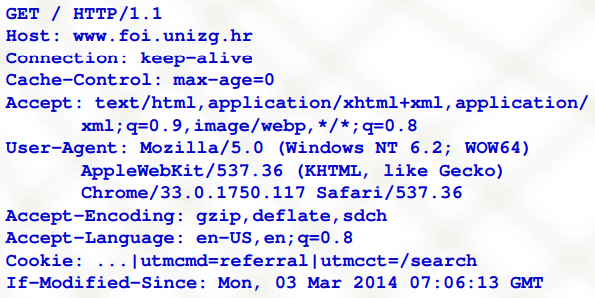
Podaci, struktura, vrste i format poruka u dnevničkom zapisu nisu propisani pa tako ovise o odlukama i implementaciji razvojnih programera koji rade na razvoju određene aplikacije ili sustava. Iako se implementacije i formati poruka razlikuju sve su sastavljene od nekoliko zajedničkih komponenti. Svaka poruka ima datum i vrijeme kada je nastala, razinu zapisa koja može biti 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. Sustavski i poslužiteljski dnevnički zapisi bit će detaljnije opisani u narednim poglavljima ovog rada.

Možemo reći da je glavna svrha svih dnevničkih zapisa dati korisnicima ili administratorima informacije o ponašanju sustava kroz vrijeme kako bi pravovremeno mogli reagirati na probleme, greške i sigurnosne prijetnje u radu sustava ili donositi bolje poslovne odluke.

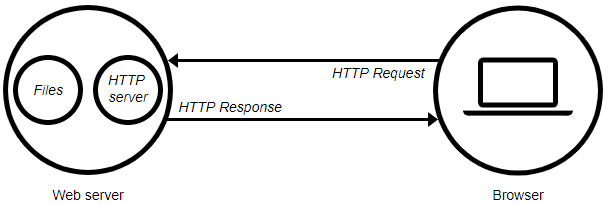
1. Web poslužitelji

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 1: Primjer HTTP zahtjeva (Izvor: Kermek, 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 2: Princip rada web poslužitelja (Izvor: MDN, 2020)

1. Organizacija dnevničkih zapisa na osobnim računalima
   1. Windows Event Viewer
   2. Dnevnički zapisi na Linux platformama
2. Dnevnički zapisi na web poslužiteljima
3. Python aplikacija za analizu i vizualizaciju
4. 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.

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Nam cursus. Morbi ut mi. Nullam enim leo, egestas id, condimentum at, laoreet mattis, massa. Sed eleifend nonummy diam. Praesent mauris ante, elementum et, bibendum at, posuere sit amet, nibh. Duis tincidunt lectus quis dui viverra vestibulum. Suspendisse vulputate aliquam dui. Nulla elementum dui ut augue. Aliquam vehicula mi at mauris. Maecenas placerat, nisl at consequat rhoncus, sem nunc gravida justo, quis eleifend arcu velit quis lacus. Morbi magna magna, tincidunt a, mattis non, imperdiet vitae, tellus. Sed odio est, auctor ac, sollicitudin in, consequat vitae, orci. Fusce id felis. Vivamus sollicitudin metus eget eros.

Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. In posuere felis nec tortor. Pellentesque faucibus. Ut accumsan ultricies elit. Maecenas at justo id velit placerat molestie. Donec dictum lectus non odio. Cras a ante vitae enim iaculis aliquam. Mauris nunc quam, venenatis nec, euismod sit amet, egestas placerat, est. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Cras id elit. Integer quis urna. Ut ante enim, dapibus malesuada, fringilla eu, condimentum quis, tellus. Aenean porttitor eros vel dolor. Donec convallis pede venenatis nibh. Duis quam. Nam eget lacus. Aliquam erat volutpat. Quisque dignissim congue leo.

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

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

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

*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

Popis slika

[Slika 1: Primjer HTTP zahtjeva 3](#_Toc48747735)

[Slika 2: Princip rada web poslužitelja 4](#_Toc48747736)

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, …)