# Univerzitet Sinigidunum

Tehnički fakultet

Seminarski rad iz osnova digitalne forenzike na temu:

# ,,Analiza Alata za Identifikaciju Tipova Datoteka u Digitalnim Istragama''

Mentor: dr Jelena Gavrilović

Student: Luka Mutić

*Beograd, jun 2024. godine*

# Uvod

Digitalna forenzika je disciplina koja se bavi identifikacijom, očuvanjem, analizom i dokumentovanjem digitalnih dokaza u svrhu podrške pravnim i istražnim postupcima. Kako tehnologija napreduje i sve više informacija postaje digitalizovano, tako raste i potreba za pouzdanim i efikasnim alatima koji mogu pomoći forenzičkim ekspertima da obrade velike količine podataka brzo i tačno.

U svetu digitalne forenzike, alati za identifikaciju tipova datoteka igraju ključnu ulogu. Ovi alati omogućavaju istražiteljima da brzo i precizno identifikuju tipove datoteka prisutnih na digitalnim medijima, što je neophodno za efikasno sortiranje i analizu velikih količina podataka. Identifikacija tipova datoteka može biti izazovna zbog različitih strategija koje kriminalci koriste za prikrivanje pravih tipova datoteka, kao što su promene ekstenzija ili korišćenje steganografskih tehnika. Alati za identifikaciju tipova datoteka koriste različite metode, uključujući prepoznavanje "magičnih brojeva" u hederima datoteka, analizu sadržaja datoteka i druge tehnike.

S obzirom na važnost ovih alata, važno je proceniti njihovu efikasnost i pouzdanost. Rad "Comparative Study of File Type Identification Tools for Digital Investigations" istražuje performanse deset različitih alata za identifikaciju tipova datoteka, uključujući komercijalne i open-source rešenja. Cilj rada je bio da se utvrdi koji alati pružaju najbolju kombinaciju tačnosti i brzine obrade, kao i da li zadovoljavaju potrebe forenzičkih istraga u stvarnim uslovima.

Ovaj seminarski rad pruža pregled ključnih aspekata rada, uključujući metodologiju evaluacije, rezultate testiranja, diskusiju o prednostima i manama različitih alata, kao i zaključke o njihovoj praktičnoj primeni u digitalnim forenzičkim istragama. Analiza je sprovedena na dva skupa podataka: jedan sa 17,500 datoteka sa 110 različitih tipova i drugi sa skoro milion datoteka sa 63 različita tipa. Testirane su strategije prikrivanja, kao što su promena ili uklanjanje ekstenzija datoteka, kako bi se procenila otpornost alata na ove metode obmanjivanja.

Rezultati ovog istraživanja pružaju dragocene uvide u performanse i pouzdanost različitih alata za identifikaciju tipova datoteka, što može pomoći forenzičkim stručnjacima u odabiru najpogodnijih alata za njihove potrebe.

## Cilj Rada

Cilj rada je da se proceni relevantnost alata za forenzičke svrhe kroz analizu tačnosti i vremena obrade na dva različita skupa podataka. Glavno pitanje je da li alati zadovoljavaju očekivanja u stvarnim forenzičkim istragama.

### Pregled Alata

Rad obuhvata deset alata, uključujući komercijalne i open-source alate:

1. **Unix command File**: Osnovno Unix rešenje zasnovano na prepoznavanju "magičnih brojeva" u datotekama.
2. **filetype**: Python modul za identifikaciju tipova datoteka koristeći magične brojeve, prepoznaje 64 tipa datoteka.
3. **file-type**: JavaScript paket koji detektuje binarne formate datoteka, podržava 135 tipova datoteka.
4. **detect-file-type**: JavaScript API koji prepoznaje binarne i tekstualne formate datoteka, podržava 94 tipa datoteka.
5. **guess-file-type**: JavaScript API koji koristi različite testove za identifikaciju tipa datoteke, podržava 34 tipa datoteka.
6. **Fidentify**: Alat zasnovan na bazi podataka Photorec potpisima, podržava 440 tipova datoteka.
7. **TrID**: Alat koji koristi bazu podataka definicija obrazaca za prepoznavanje tipova datoteka, podržava 14,374 tipa datoteka.
8. **EnCase**: Komercijalni softver za digitalne forenzičke istrage, koristi potpisivanje datoteka za identifikaciju.
9. **Autopsy**: Open-source GUI alat za analizu disk imidža, koristi The Sleuth Kit za prepoznavanje tipova datoteka.
10. **ForENSIque**: Akademski alat razvijen u Rust programskom jeziku, koristi magične brojeve za prepoznavanje 155 tipova datoteka.

### Metodologija

Evaluacija alata je sprovedena na dva skupa podataka:

1. **Dataset 17500**: Sadrži 17,500 datoteka sa 110 različitih tipova.
2. **Dataset 1M**: Sadrži skoro milion datoteka sa 63 različita tipa.

Eksperimenti su obuhvatali procenu tačnosti i vremena obrade za svaki alat. Takođe su testirane strategije prikrivanja, kao što su promena ili uklanjanje ekstenzija datoteka.

### Rezultati

1. **Tačnost Alata**:
    - **Fidentify**: Najpreciznij alat sa tačnošću od 94.4% na Dataset 17500 i 98.1% na Dataset 1M.
    - **Autopsy i Unix command File**: Pokazali visoku tačnost.
    - **guess-file-type**: Značajan pad tačnosti pri promeni ekstenzija.
2. **Vreme Obrade**:
    - **Unix command File**: Najbrži alat, ali sa manjom tačnošću u poređenju sa Fidentify.
    - **Fidentify**: Dobar balans između brzine i tačnosti.
3. **Detalji Performansi**:
    - Alati kao što su TrID i EnCase pokazali su dobru tačnost ali su spori.
    - Kombinovanjem alata može se poboljšati ukupna tačnost, kao što je demonstrirano sa kombinacijom Fidentify i ForENSIque, što je povećalo tačnost na 96.2% na Dataset 17500.

## Diskusija

Rezultati pokazuju značajnu razliku u performansama alata. Većina open-source alata nije dovoljno efikasna za forenzičke svrhe. Profesionalni alati, kao što su Fidentify, Autopsy i EnCase, pokazali su se kao pouzdana rešenja sa različitim kompromisima između brzine i tačnosti.

## Zaključak

Fidenti se izdvaja kao najefikasnije rešenje među testiranim alatima, kombinujući visoku tačnost sa prihvatljivim vremenom obrade. Profesionalni alati kao što su EnCase i Autopsy pokazali su se kao pouzdana rešenja za kompleksnije istrage. Kombinacija različitih alata može dodatno poboljšati ukupnu tačnost identifikacije tipova datoteka.

Ovaj rad značajno doprinosi profesionalizaciji digitalne forenzike pružajući rigorozne evaluacione protokole za procenu forenzičkih alata. Pravilnim odabirom i kombinacijom alata, forenzički eksperti mogu poboljšati efikasnost i tačnost svojih istraga, čime se povećava ukupna pouzdanost digitalne forenzike. Buduća istraživanja mogu se fokusirati na integraciju mašinskog učenja i drugih naprednih tehnologija kako bi se dodatno unapredile performanse forenzičkih alata.

---

## Reference


- Dubettiera, A., Gernota, T., Gigueta, E., & Rosenbergera, C. (2023). Comparative Study of File Type Identification Tools for Digital Investigations. *Forensic Science International: Digital Investigation*.
- https://www.sciencedirect.com/science/article/abs/pii/S2666281723000835