УДД Пројекат - 2025/26

Увод у пројекат:
С обзиром на растућу потребу за интегрисаним безбедносним системима који омогућавају ефикасну анализу дигиталних форензичких артефаката и анализу малвера, циљ овог пројекта је имплементација система за мониторинг и претрагу *форензичких извештаја* и *обавештајних података о претњама* (*threat intelligence*).

Сви извештаји биће чувани у MinIO репозиторијуму у сировом облику (PDF). Коришћењем ELK платформе, систем омогућава прецизну претрагу, као и визуелизацију података за потребе digital forensics аналитике, класификације претњи и праћења трендова у домену сајбер претњи.

Бодовање

Целина	**Секција**	**Обавезно**	**Бодови**
Аутентификација	4	**ДА**	**3**
Парсирање и индексирање докумената	**1**	**ДА**	**10**
Основна претрага	**2**	**ДА**	**17**
Геолокацијска претрага	**3.1**	**НЕ**	**10**
ELK stack интеграција	**3.2**	**НЕ**	**10**

Захтеви пројекта:

1. **Парсирање докумената**
    - Сваки форензички извештај чува се у MinIO као PDF и парсира за следеће податке:
        - Име и презиме форензичара који је радио анализу
        - Назив CERT/CSIRT или безбедносне организације која је обрадила узорак
        - Назив малвера/претње која је анализирана
        - Опис понашања малвера/претње
        - Класификација претње (ниска, средња, висока, критична)
        - Хеш вредност анализираног узорка (MD5/SHA256)
    - Приликом индексирања, документ (извештај) се прво парсира на серверу, након чега се кориснику приказује форма где се налази свако од горе наведених поља, са вредношћу које је парсер препознао у њима. Корисник има опцију да промени било које од датих поља, након чега може да потврди индексирање и процес се наставља. Уколико корисник одустане, процес се завршава и документ се не индексира.

2. **Страница за претрагу:**
    - **Претраживање извештаја** према имену и презимену форензичара, хеш вредности и према класификацији претње. **(2 бода)**
    - **Претраживање по називу организације** (CERT/CSIRT) и **називу анализираног малвера/претње**. **(1 бод)**
    - ***Approximate KNN* претрага** извештаја на основу *free text* уноса из поља за претрагу. **(2 бода)**
    - **Претрага описа** из извештаја (full-text претрага у ПДФ-у). **(1 бод)**
    - **Комбинована *boolean semi-structured* претрага** уз подршку за AND, OR и NOT оператора (операције морају да се ланчају неограничено а редослед извршавања треба да је исти као у неком *C-like* програмском језику). **(4 бода)**
    - **Подршка за PhrazeQuery** у свим пољима *semi-structured* претраге. **(3 бода)**
    - **Предпроцесирање упита** користећи **SerbianAnalyzer** ради лингвистичке нормализације (Није довољно само укључити дефаултни анализатор већ је потребно од 0 креирати своју конфигурацију по узору на ону дату на вјежбама, минимално је потребно користити ICU токенизатор као и српски *stemmer* као и уклонити српске стоп ријечи). **(2 бода)**
    - **Приказ резултата:** динамички приказ сажетка са истакнутим кључним појмовима (highlighter). **(2 бода)**
3. **Напредна претрага и визуелизације**
    - **1) Геолокациона претрага** инцидената према локацији CERT/CSIRT организације, користећи име града или адресу унутар задатог кружног радијуса. **(10 бодова)**
    - **2) Визуелизације у Kibana-и:**
        - **Као предуслов за ову функционалност, потребно је увезати и исконфигурисати *ELK stack*. Логове апликације је потребно у произвољном формату прво уписивати у фајл који ће *Logstash* покупити, парсирати и филтрирати помоћу *GROK* филтера, и коначно их уписивати у *Elasticsearch*. Након тога, потребно је имплементирати доле наведене визуелизације. (4 бода)**
        - Град са највише пријављених узорака малвера. **(2 бода)**
        - Топ 3 форензичара са највећим бројем обрађених инцидената. **(2 бода)**
        - Удео појављивања малвера/претња у односу на све анализиране инциденте. **(2 бода)**
4. **Механизам за аутентификацију и ауторизацију**
    - Потребно је додати основну аутентификацију путем форме за пријаву. Неулогован корисник нема приступ ни једној страници сем страници за пријаву док улогован корисник има приступ свему. **(3 бода)**

Овај систем омогућава ефикасну претрагу, анализу и визуализацију докумената и логова повезаних са безбедносним догађајима и организацијама.

Прва контролна тачка

За предмет Управљање Дигиталним Документима биће организована једна контролна тачка. Контролна тачка мора бити предата у задатом року и позитивно оцењена (*pass/fail* принцип, ако напишете нешто што нема смисла, **документ неће бити читан даље**) од стране асистента да би студент могао да ради додатне задатке за највише оцене (поред ове оцене, добићете и бодове од 0-20 које ћете моћи да освојите имплементацијом додатних задатака, имејте у виду да ове бодове морате да потврдите на одбрани и да их не можете надокнадити чак и да урадите више него што сте навели у КТ).

Контролна тачка није обавезна, али је предуслов за добијање највиших оцена. За контролну тачку неопходно је предати PDF документ у коме је описана архитектура апликације и конфигурација *ELK Stack*-а и *MinIO*-а. Из овог *PDF* документа мора бити јасно дефинисана архитектура апликације (стил архитектуре, слојевитност, комуникација компоненти…), како ће се комуницирати са *Elasticsearch*-ом и како ће *Elasticsearch* комуницирати са *Kibana*-ом и *Logstash*-ом, где ће бити складиштена колекција извештаја, који ће подаци бити складиштени у бази података. Такође, из овог *PDF* документа мора бити јасно како ће бити подигнут и конфигурисан *Elasticsearch*, како ће бити обављено претпроцесирање српских текстова, како ће изгледати *indexing unit* (*JSON/POJO* који се индексира), и како ће бити реализована геопросторна претрага. У овом моменту се не очекује да су функционалности апликације повезане са корисничким интерфејсом, односно да се позивају као реакција на неку корисникову акцију. Нема ограничења у броју страна, а рок за предају је 25.01.2026. до 23:59 часова, слањем *e-mail*-а на адресу **ivan.mrsulja@uns.ac.rs**.