Szegedi Tudományegyetem

Informatikai Intézet

**SQL injection detektálás természetes nyelvi feldolgozó rendszerek és gépi tanulás által**

**SQL injection detection using natural language processing systems and machine learning**

Szakdolgozat

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Készítette: |  | Témavezető: |  |
|  | Stiller Marianna |  | Dr. Vidács László |  |
|  | programtervező informatika szakos hallgató |  | egyetemi oktató |  |

Szeged

2021

## Feladatkiírás

Az SQL Injection napjaink leggyakoribb és legveszélyesebb sérülékenysége, amely a felhasználói input elégtelen mértékű validációjából ered, következménye pedig SQL lekérdezések futtatása érzékeny adatokon.

A rendszer természetes nyelvi feldolgozó rendszerek és gépi tanulás által képes lesz forrásódban SQL Injection sérülékenységeket keresni. Működésének alapja, hogy valós PHP alapú forráskódokból reprezentál absztrakt szintaxis fákat, melyekből mintákat nyer ki, amit később megtanul.

## Tartalmi összefoglaló

A téma megnevezése:

SQL injection detektálás természetes nyelvi feldolgozó rendszerek és gépi tanulás által.

A megadott feladat megfogalmazása:

Minták által tanított program képes reális becslést adni egy kód SQL injectionbeli sebezhetőségével kapcsolatban.

A megoldási mód:

PHP kódokból generálok absztrakt szintaxis fákat, amelyeket felcímkézek, mint sérülékeny és nem sérülékeny példák. Ezek átesnek egy vektorizáláson, majd egy kifejezett modell segítségével – amely a legkevesebb hibás döntéseket hozza - a program megtanulja felismerni a sérülékenység jeleit. Ennek hatására egy nem tanuló kódról képes kiértékelést készíteni.

Alkalmazott eszközök, módszerek:

A IntelliJ fejlesztői környezetet, emellett a gépi tanulás során Python, a grafikus asztali felület megvalósításához Java programozási nyelvet alkalmaztam.

Elért eredmények:

A program a bemenetként megadott PHP kódra és classifier típusra kimenetként annak hatékonyságát, valamint egy valószínűségi számot ad, amely a kód sérülékenységének mértéke.

Kulcsszavak:

Felügyelt tanulás, train/test készlet, K-Fold, confusion mátrix, recall.

## Tartalomjegyzék

[Feladatkiírás 2](#_Toc69730693)

[Tartalmi összefoglaló 3](#_Toc69730694)

[Tartalomjegyzék 4](#_Toc69730695)

[BEVEZETÉS 5](#_Toc69730696)

[1. SQL INJECTION 5](#_Toc69730697)

[2. GÉPI TANULÁS 5](#_Toc69730698)

[3. PROGRAM FELÉPÍTÉSE 5](#_Toc69730699)

[4. EREDMÉNY ÉS ANALÍZIS 5](#_Toc69730700)

[Irodalomjegyzék 6](#_Toc69730701)

[Nyilatkozat 7](#_Toc69730702)

[Köszönetnyilvánítás 8](#_Toc69730703)

## BEVEZETÉS

1. SQL INJECTION
2. GÉPI TANULÁS
3. PROGRAM FELÉPÍTÉSE
4. EREDMÉNY ÉS ANALÍZIS

## Irodalomjegyzék

## Nyilatkozat

Alulírott Stiller Marianna programtervező informatikus BSc szakos hallgató, kijelentem, hogy a dolgozatomat a Szegedi Tudományegyetem, Informatikai Intézet Szoftverfejlesztés Tanszékén készítettem, programtervező informatikus BSc diploma megszerzése érdekében. Kijelentem, hogy a dolgozatot más szakon korábban nem védtem meg, saját munkám eredménye, és csak a hivatkozott forrásokat (szakirodalom, eszközök, stb.) használtam fel. Tudomásul veszem, hogy szakdolgozatomat / diplomamunkámat a Szegedi Tudományegyetem Informatikai Intézet könyvtárában, a helyben olvasható könyvek között helyezik el.

2021.

## Köszönetnyilvánítás