Vilniaus universitetas

LIETUVOS TEISMO EKSPERTIZĖS CENTRAS

Ataskaitinė informatikos krypties doktorantų konferencija 2021-09-30

Andrius Chaževskas (VU DMSTI doktorantas, Išmaniųjų technologijų tyrimų grupė)

Darbo tema.

Teksto semantinės analizės ir mašininio mokymosi algoritmų taikymo slaptažodžių parinkimui tyrimas.

Application of text semantic analysis and machine learning algorithms for passwords guessing.

Darbo vadovas.

Prof. dr. Igoris Belovas.

Doktorantūros studijų laikotarpis.

2020 m. spalio mėn. 1 d. – 2024 m. rugsėjo mėn. 30 d..

Ataskaitinis laikotarpis.

2021 m. kovo mėn. 26 d. – 2021 m. rugsėjo mėn. 30 d..

Visų studijų planas ir jo vykdymo suvestinė

Studijų metai	Egzaminai		Dalyvavimas konferencijose		Publikacijos		
	Planas	Įvykdyta	Planas	Įvykdyta	Planas	Įvykdyta	Būklė
I (2020/2021) Pirmas pusmetis	1	1		1 (L)			
II (2020/2021) Antras pusmetis	1	1	1 (L)	1 (T)		1 (KD/R)	Priimta
II (2021/2022) Pirmas pusmetis	1				1 (KD)		
II (2021/2022) Antras pusmetis	1		1 (L)		1 (KD / R)		
III (2022/2023) Pirmas pusmetis							
III (2022/2023) Antras pusmetis			1 (T)		1 (CA WoS)		
IV (2023/2024) Pirmas pusmetis							
IV (2023/2024) Antras pusmetis			1 (T)		1 (CA WoS)		

Vilniaus universitetas

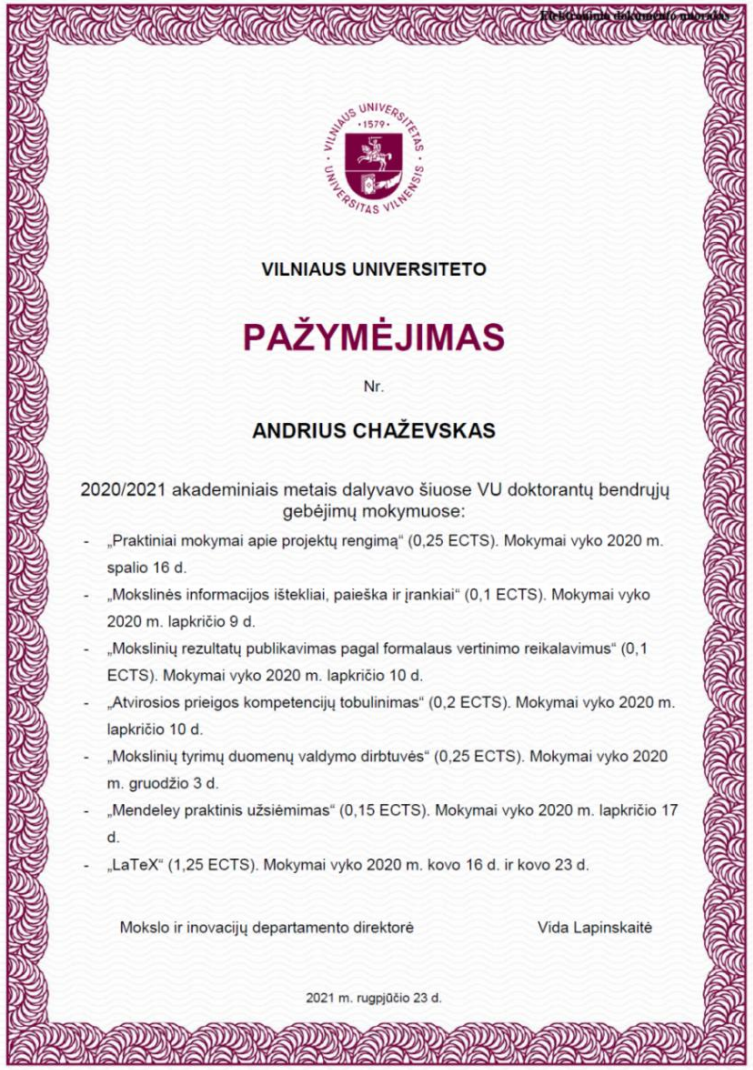
Ataskaitinių metų darbo planas ir jo vykdymo suvestinė

Egzaminai		Dalyvavimas konferencijose		Publikacijos	
Planas	**Įvykdyta**	**Planas**	**Įvykdyta**	**Planas**	**Įvykdyta**
„Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika“	Išlaikyta: Informatikos ir informatikos inžinerijos tyrimo metodai ir metodika“, 2021-06-29	-	Nuotoliniu būdu dalyvauta: 17-ajame tarptautiniame kongrese, Kriminalistika ir teismo ekspertologija: mokslas, studijos, praktika, Bratislava, Slovakija, 2021-09-17 skaitytas pranešimas.	-	Tarptautinei konferencijai, Kriminalistika ir teismo ekspertologija: mokslas, studijos, praktika, Bratislava, Slovakija, įteikta (ir priimta) publikacija: „Forensic password examination in leaked user databases“. Autoriai: Andrius Chaževskas, prof. dr. Igoris Belovas, dr. Virginijus Marcinkevičius.

Vilniaus universitetas

Kvalifikacijos kėlimas

- **Bendrieji gebėjimai surinkta : 2.3 ECTS.**
- **Darbas su VU Duomenų mokslo ir skaitmeninių technologijų instituto Informacinių sistemų inžinerijos bakalauro programos IV k. studentais (jaunesniojo asistento pareigose). Informacinės saugos kurso laboratorinių darbų kuravimas.**



VILNIAUS UNIVERSITETO

PAŽYMĖJIMAS

Nr.

ANDRIUS CHAŽEVSKAS

2020/2021 akademiniais metais dalyvavo šiuose VU doktorantų bendrųjų gebėjimų mokymuose:

- „Praktiniai mokymai apie projektų rengimą“ (0,25 ECTS). Mokymai vyko 2020 m. spalio 16 d.
- „Mokslinės informacijos ištekliai, paieška ir įrankiai“ (0,1 ECTS). Mokymai vyko 2020 m. lapkričio 9 d.
- „Mokslinių rezultatų publikavimas pagal formalaus vertinimo reikalavimus“ (0,1 ECTS). Mokymai vyko 2020 m. lapkričio 10 d.
- „Atvirosios prieigos kompetencijų tobulinimas“ (0,2 ECTS). Mokymai vyko 2020 m. lapkričio 10 d.
- „Mokslinių tyrimų duomenų valdymo dirbtuvės“ (0,25 ECTS). Mokymai vyko 2020 m. gruodžio 3 d.
- „Mendeley praktinis užsiėmimas“ (0,15 ECTS). Mokymai vyko 2020 m. lapkričio 17 d.
- „LaTeX“ (1,25 ECTS). Mokymai vyko 2020 m. kovo 16 d. ir kovo 23 d.

Mokslo ir inovacijų departamento direktorė Vida Lapinskaitė

2021 m. rugpjūčio 23 d.

Visų mokslinių tyrimų ir disertacijos rengimo etapai

	Darbo pavadinimas	Atlikimo terminai	Pastabos
1.	Mokslinių tyrimų disertacijos tema apžvalga ir analizė (Lietuvoje ir užsienyje): 1.1. Analitinės apžvalgos atlikimas. 1.2. Disertacijos tyrimo objekto detalizavimas. 1.3. Mokslinių problemų susietų su tyrimo objektu identifikavimas ir tyrimo tikslo suformavimas.	2020 m. spalio mėn. – 2021 m. rugsėjo mėn.	Atlikta, apibendrinti rezultatai mokslinėje ataskaitoje.
	Mokslinio tyrimo vykdymas:		
2.	**2.1.Tyrimo metodikos sudarymas:** 2.1.1. Uždavinių, skirtų tyrimo tikslui pasiekti, suformulavimas. 2.1.2. Tyrimo metodikos išsikeltiems uždaviniams spręsti parinkimas. 2.1.3. Teorinio ir empirinio tyrimų suplanavimas pagal pasirinktą metodiką.	2021 m. spalio mėn. – 2022 m. sausio mėn.	

Visų mokslinių tyrimų ir disertacijos rengimo etapai

	Darbo pavadinimas	Atlikimo terminai	Pastabos
	2.2. Teorinis tyrimas: 2.2.1. Mašininio mokymosi metodų naudojamų automatizuotame slaptažodžių parinkime tyrimas. 2.2.2. Semantinės slaptažodžių analizės ir šablonų parinkimo metodų tyrimas. 2.2.3.Slaptažodžių parinkimo algoritmų taikant semantinę analizę tyrimas.	2022 m. sausio mėn. – 2022 m. rugsėjo mėn.	
	2.3. Empirinis tyrimas: 2.3.1. Skirtingų algoritmų palyginimas. 2.3.2. Įgyvendintų algoritmų modifikacijos, ar naujų algoritmų kūrimas, sprendžiant apibrėžtus uždavinius. 2.3.3. Sukurtų modifikacijų eksperimentinis tyrimas analizuojant jų efektyvumą	2022 m. spalio mėn. – 2023 m. gegužės mėn.	
	2.4. Gautų rezultatų analizė ir apibendrinimas	2023 m. birželio mėn. – 2023 m. rugsėjo mėn.	

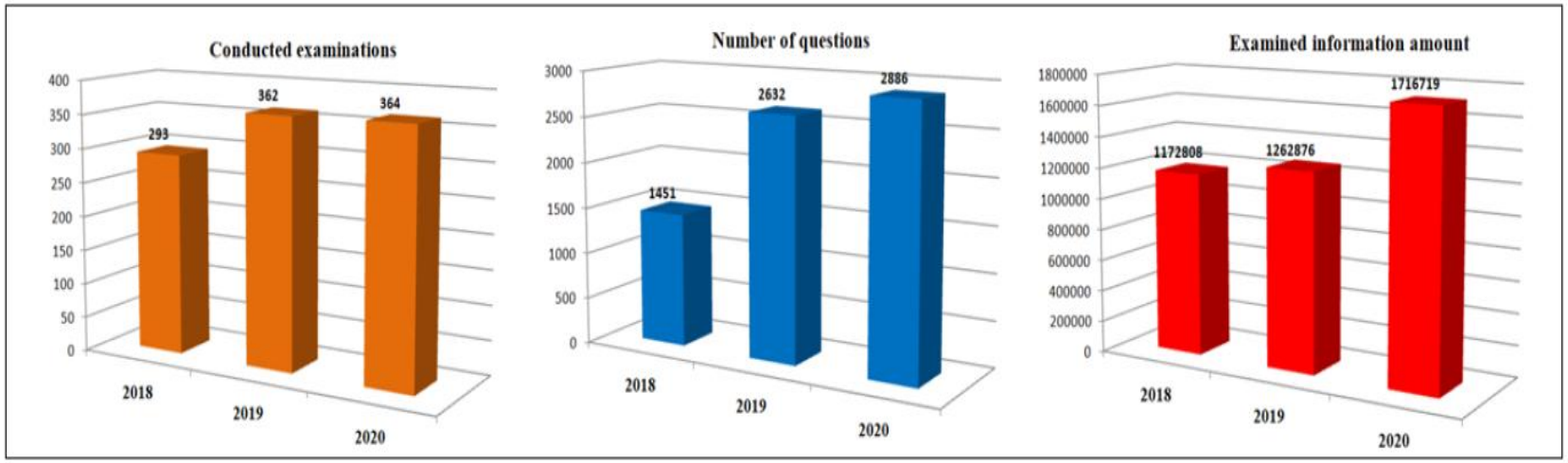
Visų mokslinių tyrimų ir disertacijos rengimo etapai

	Darbo pavadinimas	Atlikimo terminai	Pastabos
3.	Atskirų daktaro disertacijos dalių (tyrimo metodikos, rezultatų, ginamų teiginių, išvadų ir kt.) parengimas: 3.1. Tikslų, uždavinių, tyrimo metodikos, ginamųjų teiginių patikslinimas. 3.2. Analitinės disertacijos dalies parengimas. 3.3. Teorinės disertacijos dalies parengimas. 3.4. Eksperimentinės disertacijos dalies parengimas. 3.5. Bendrųjų išvadų formulavimas.	2023 m. spalio mėn. – 2024 m. gegužės mėn.	
4.	Daktaro disertacijos parengimas ir svarstymas padalinyje	2024 m. birželio mėn.	
5.	Daktaro disertacijos gynimas	2024 m. rugsėjo mėn.	

Tyrimo objektas, tikslas ir uždaviniai

Ekspertiniai tyrimai:

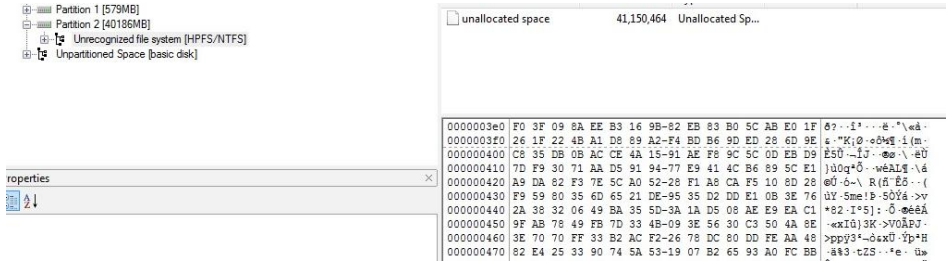
- **Teisminės ekspertizės (susijusios su IT) Lietuvoje.**
- **Pagrindiniai užsakovai.**
- **Tiriamieji objektai.**
- **Tyrimų statistika.**

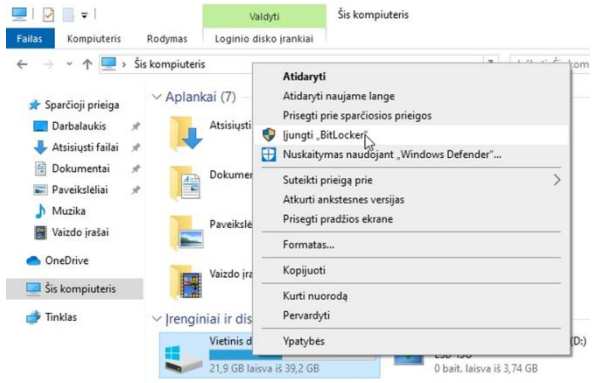


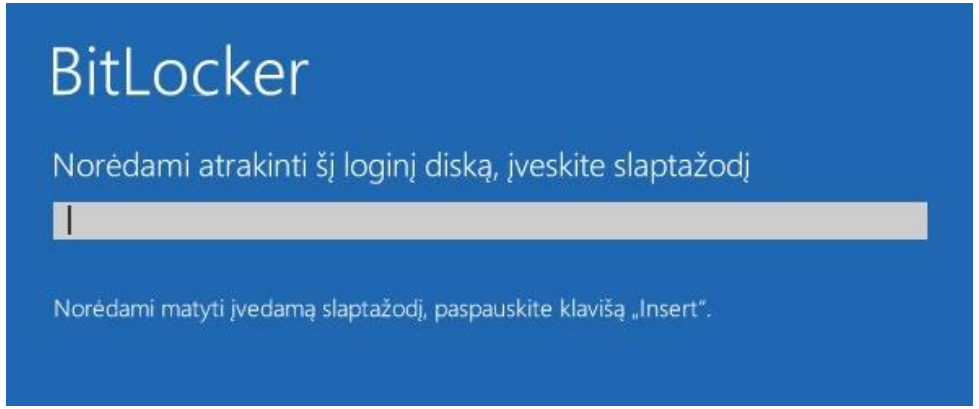
Vilniaus universitetas

Autentifikacija ir duomenų sauga

- **Skaitmeninės informacijos svarba ir saugumas.**
- **Autentifikavimas.**
- **Informacijos šifravimas.**
- **Slaptažodžiai.**
- **Pavyzdžiui** “Bitlocker” apsauga.







Vilniaus universitetas

Problemos

Kaip ištirti šifruotą informaciją?

Slaptažodžių parinkimo metodai:

- **Žodynų taikymas;**
- **Nutekintų slaptažodžių duomenų bazių panaudojimas;**
- **Pilno perrinkimo atakos („brute-force“);**
- **Kombinuotos (mišrios) slaptažodžių parinkimo atakos, skirtinguose etapuose naudojant žodynų ir „brute force“ atakas.**

Slaptažodžių parinkimo priemonės:

- **Laboratorijos aparatūrinė įranga;**
- **Laboratorijos programinė įranga.**

Laikas (kiek galime skirti laiko ir resursų parinkti slaptažodį).

“Brute force” atakos

Slaptažodžio ilgis	Simbolių sekos	Galimų slaptažodžių skaičius
1-10	Skaičiai	11111111110
1-10	Skaičiai, mažosios raidės	3.76 × 1015
1-10	Skaičiai, mažosios ir didžiosios raidės	8.53 × 1017
1-10	Skaičiai, mažosios ir didžiosios raidės, spec. simboliai	1.75 × 1019

Pilno perrinkimo slaptažodžių parinkimo statistika.

Kiekviena laboratorija gali paskaičiuoti savo galimybes:

$$Max\ Time = \frac{(A)^M}{N}$$

Kur: A – simbolių sekos dydis, M - slaptažodžio ilgis, N - bandymai atspėti slaptažodį per sekundę.

Vilniaus universitetas

Vartotojų slaptažodžiai

Nutekintų slaptažodžių duomenų bazių panaudojimas:

- **https://github.com/lexcor/LT-SecList;**
- **http://downloads.skullsecurity.org;**
- **Lietuvos vartotojų duomenų bazės.**

Vartotojų apklausos:

- **https://www.bite.lt/apie/ziniasklaidai/slaptazodi-nulauzti-uztruktu-iki-10-sekundziu;**
- **https://www.bite.lt/apie/ziniasklaidai/paskelbti-10-populiariausiu-2019-uju-slaptazodziu.**

Pirminis darbo tikslas

- **Nutekintų slaptažodžių duomenų bazių analizė rodo, kad žmonės yra linkę naudoti lengvai įsimenamus slaptažodžius, tai reiškia, kad jų pasirinkti slaptažodžiai paprastai turi logišką struktūrą ir nėra atsitiktinių simbolių rinkiniai.**

- **Šiuolaikiniai slaptažodžių parinkimo metodai, remiasi mašininiu mokymusi ir natūralios kalbos apdorojimu, siekiant išnaudoti šią informaciją.**

- **Šio tyrimo pirmas etapas – naujausių slaptažodžių parinkimų metodų ir juose naudojamų strategijų ištyrimas, praktinis jų pritaikymas ir palyginimas. Gauti rezultatai bus panaudoti siekiant galutinio tikslo – naujų slaptažodžių parinkimo, grįstų mašininio mokymusi, metodų (strategijų) kūrimas ir pritaikymas, atsižvelgiant į Lietuvos varototojų naudojamų slaptažodžių ypatybes bei lietuvių kalbos specifiką.**

Atliktas statistinis tyrimas

FORENSIC PASSWORD EXAMINATION IN LEAKED USER DATABASES

17th INTERNATIONAL CONGRESS
CRIMINALISTICS AND FORENSIC EXPERTOLOGY:
SCIENCE, STUDIES, PRACTICE
2021 Bratislava, Slovak Republic

Andrius Chaževskas
internal PhD student at Vilnius University, Institute of Data Science and Digital Technologies,
forensic IT expert at Forensic Science Centre of Lithuania.

Assoc. prof. dr. Igoris Belovas
senior researcher at Vilnius University, Institute of Data Science and Digital Technologies.

dr. Virginijus Marcinkevičius
senior researcher at Vilnius University, Institute of Data Science and Digital Technologies.

Vilniaus universitetas

Pagrindiniai uždaviniai

- Atlikti nutekėjusių lietuviškų slaptažodžių analizę;
- Atskleisti pagrindinius slaptažodžių modelius ir sudėtingumą;
- Pabandyti nustatyti skirtingų vartotojų grupių slaptažodžių kūrimo tendencijas ir palyginkite jas su pasaulinėmis tendencijomis.
- Gautus rezultatus pritaikyti užšifruotos informacijos strategijos ir slaptažodžių atspėjimo procedūrose, taikomose teismo kriminalistiniuose tyrimuose.

Tyrimas

Buvo atlikta vieno iš Lietuvos socialinių paslaugų teikėjo nutekintų duomenų analizė. Kadangi tai buvo svetainė, kurioje teikiamos socialinio žmonių bendravimo paslaugos, nutekėjusiose duomenų bazėse buvo informacija susijusi su:

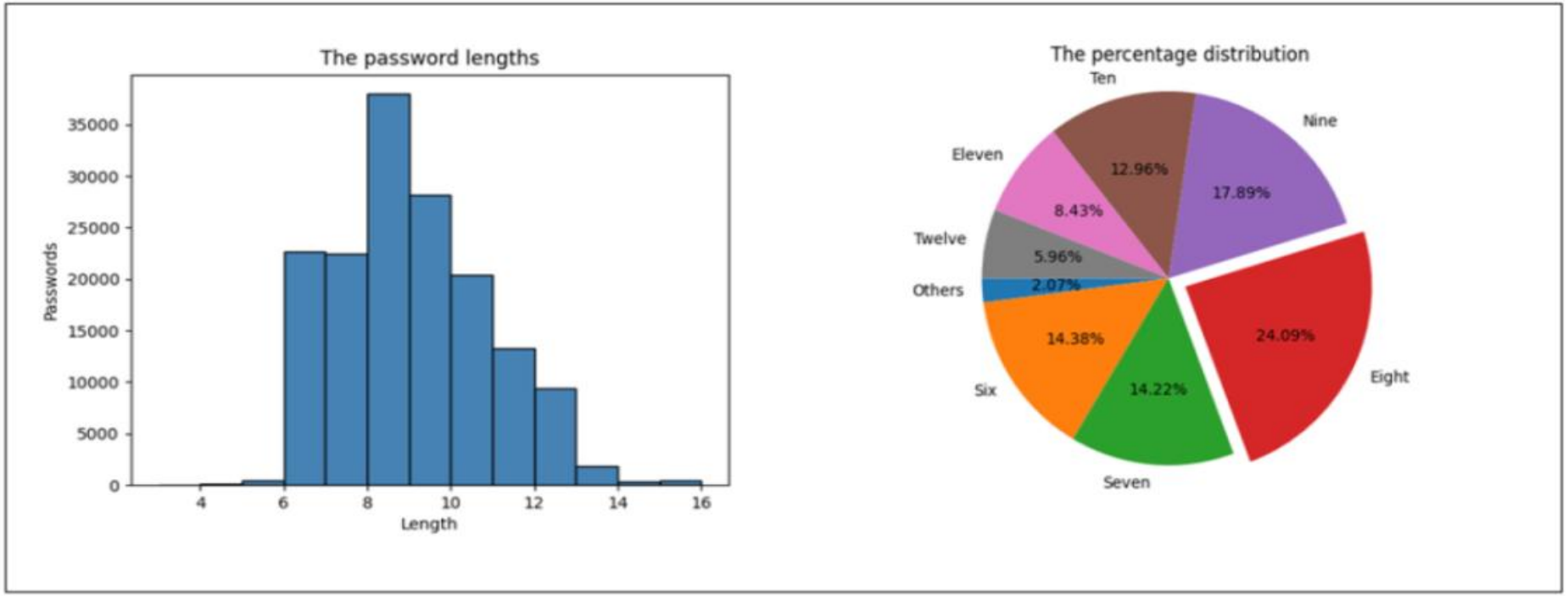
- Vartotojų vardais;
- E-pašto adresais;
- Slaptažodžiais.

Papildomai mes turėjome daug informacijos apie pačius vartotojus:

- Vardai ir pavardės;
- Tautybė ir religija;
- Amžius ir lytis;
- Išsilavinimo lygis;
- Šeimyninė padėtis;

Slaptažodžių ilgiai

Duomenų rinkinys, kuriame iš viso buvo 157617 slaptažodžių, buvo klasifikuojamas (sugrupuojamas) pagal slaptažodžių ilgių dažnį.

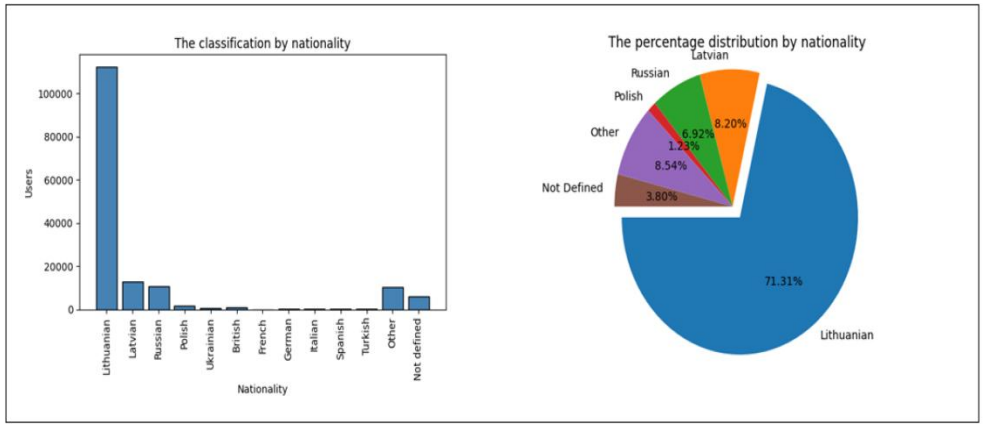


Vilniaus universitetas

Vartotojų grupės

Nustatyti dažniausiai naudojami slaptažodžiai skirtingoms vartotojų grupėms.

Vartotojai buvo sugrupuoti pagal:



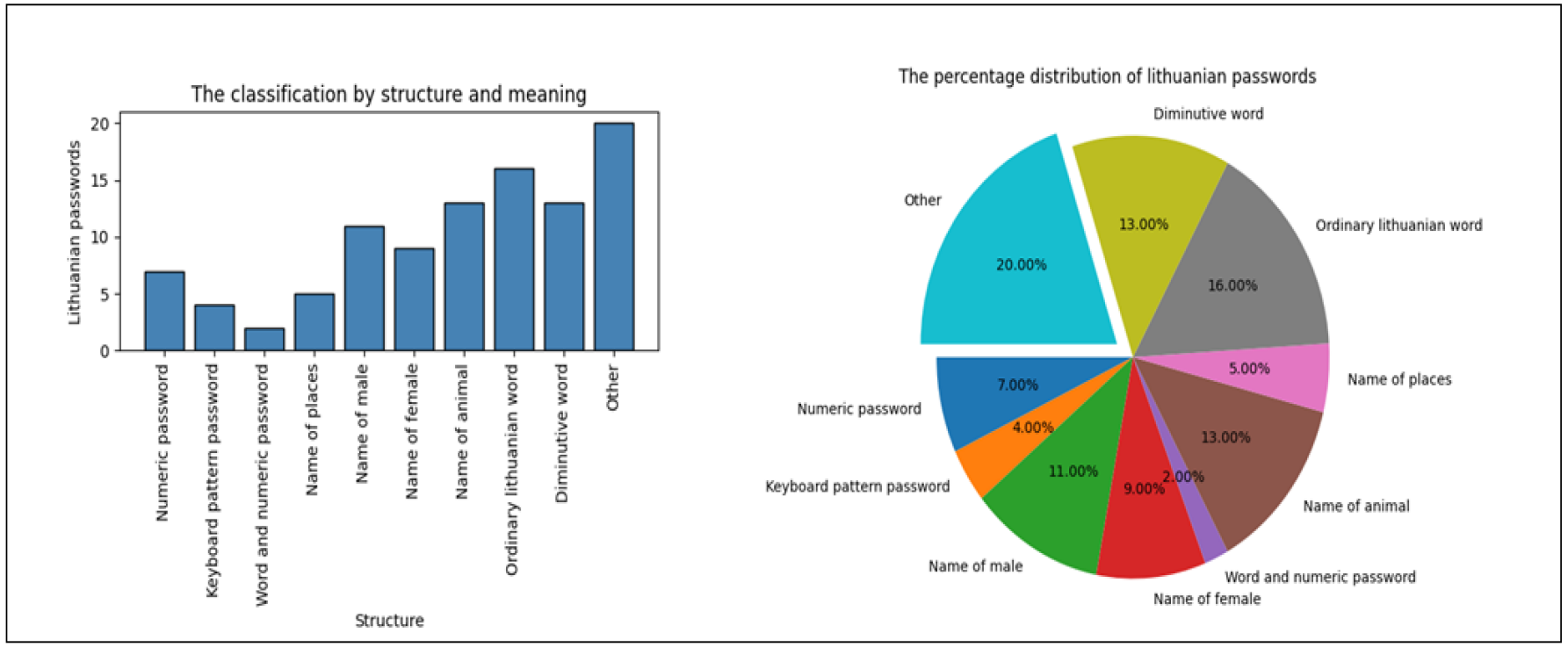
- **Lytį (vyrai, moterys, neapibrėžti);**
- **Vartotojai sugrupuoti pagal amžių;**
- **Vartotojai sugrupuoti pagal tautybę (lietuvių, latvių, slavų, kiti);**

Pastaba: mūsų tyrime terminas - slavų tautybė apibrėžiamas kaip tautų grupė, kurios pilietybė yra rusų, lenkų ar ukrainiečių.

Vilniaus universitetas

Lietuviški slaptažodžiai

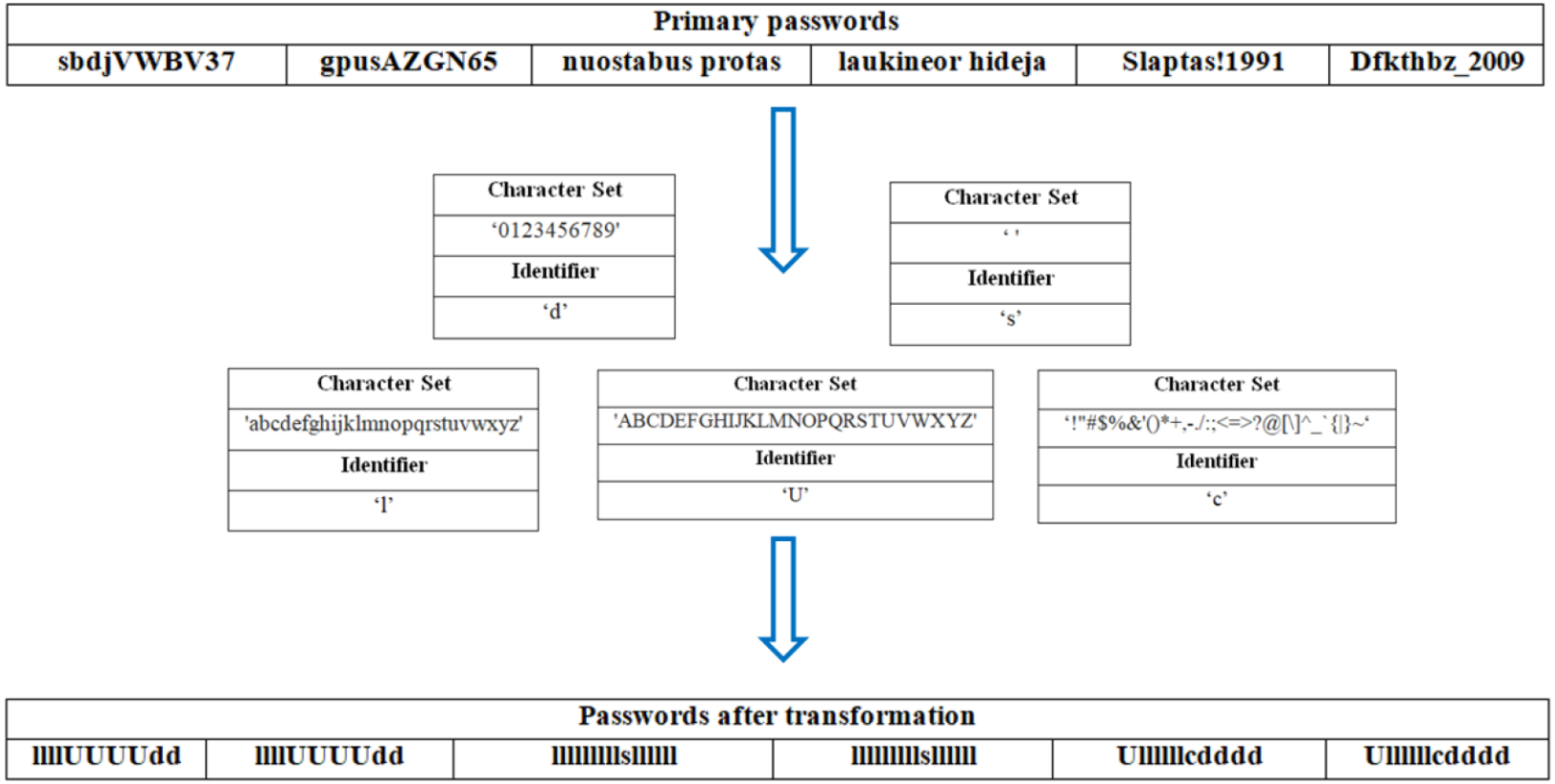
Žemiau pateikiamas šimtas dažniausiai pasitaikančių lietuviškų slaptažodžių, klasifikuojamų pagal įvairias kompozicijas, atsižvelgiant į jų struktūrą ir reikšmę.



Vilniaus universitetas

Slaptažodžių struktūros

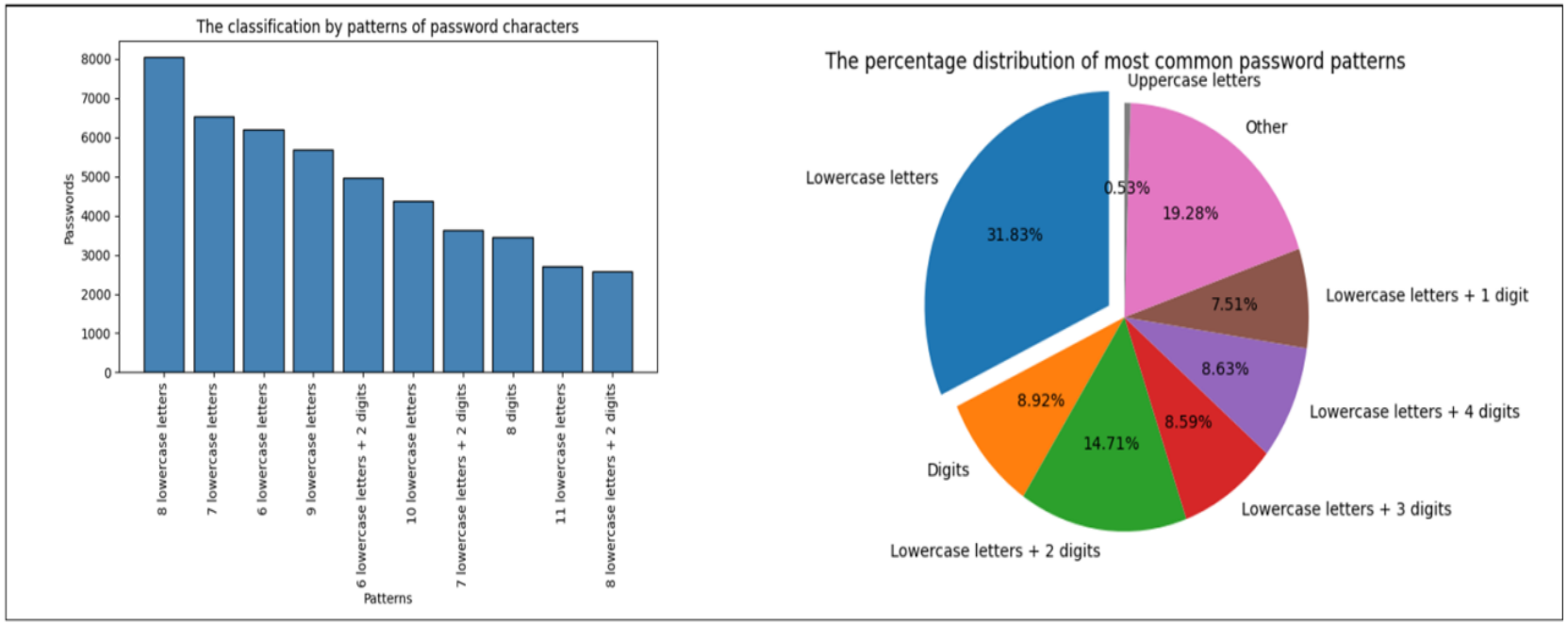
Norėdami atskirti tipiškiausius modelius, naudojant „Python 3.7“ „String“ biblioteką transformavome juos:



Vilniaus universitetas

Slaptažodžių medeliai

Toliau pateikiami dešimt dažniausiai pasitaikančių modelių (kairėje) ir visų modelių, turinčių skirtingo ilgio simbolių rinkinius (pvz., Visų ilgių mažosios raidės, visų ilgių skaitmenys, visų ilgių didžiosios raidės ir t.t), procentinis pasiskirstymas (dešinėje).



Vilniaus universitetas

Rezultatai ir išvados

1. Rezultatai patvirtina, kad daugelis vartotojų linkę rinktis paprastus slaptažodžius, kuriuos lengva įsiminti. Apie 27% nutekintos duomenų bazės slaptažodžių buvo dubliuojami vieną ar daugiau kartų. 29% visų skirtingų slaptažodžių buvo galima lengvai atspėti naudojant brutalios jėgos išpuolius per mėnesį ar greičiau.
2. Slaptažodžių klasifikacija pagal skirtingas vartotojų kategorijas rodo, kad galima atskirti kai kurias tipiškų charakteristikų slaptažodžių grupes. Pavyzdžiui, 12% 50 dažniausiai naudojamų vyrų slaptažodžių sudaro vyrų vardai, o 18% 50 dažniausiai naudojamų moterų slaptažodžių yra moterų vardai.
3. Vartotojų klasifikacija pagal tautybę (žr. Dažniausiai pasitaikančius lietuvių ir latvių slaptažodžius) rodo, kad vartotojai linkę naudoti slaptažodžius pagal savo nacionalinės kalbos žodynus.
4. Dažniausiai pasitaikantys slaptažodžių modeliai gali būti naudojami realiuose šifruotuos infprmacijos tyrimuose, taikant žodyną ir “protingas” pilno perrinkimo atakas.
5. Visi, šiame tyrime, išskirti slaptažodžių modeliai yra geras tolesnių tyrimų šaltinis, siekiant atlikti gilesnę semantinę analizę ir nacionalinės kalbos vartojimo įtaką slaptažodžių struktūrose.

Kito pusmečio darbo planas.

1. Tyrimo metodikos sudarymas. Teorinio ir empirinio tyrimų suplanavimas pagal pasirinktą metodiką (spalis-gruodis).

2. Eksperimentų pagal sudarytą planą atlikimas (gruodis-birželis).

3. Gautų rezultatų analizė. Publikacijos (CA WoS) ruošimas (vasaris-birželis).

4. Tarpinių rezultatų pristatymas 12-oje tarptautinėje konferencijoje „Duomenų analizės metodai programų sistemoms“ (gruodis).

5. Išlaikyti privalomojo dalyko „Fundamentalieji informatikos ir informatikos inžinerijos mokslų metodai“ egzaminą (sausis).

6. Esant palankiai epidemiologiniai situacijai, pateikti paraišką stažuotei į Bolonijoje esantį didžiausią Italijos skaičiavimo centrą CINECA.

Vilniaus universitetas

Vilniaus universitetas

Ačiū už dėmesį

Andrius Chaževskas

VU DMSTI doktorantas
Andrius.Chazevskas@mif.stud.vu.lt