Vilniaus universitetas

LIETUVOS TEISMO EKSPERTIZĖS CENTRAS

Ataskaitinė informatikos krypties doktorantų konferencija 2023-03-21

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

Vilniaus universitetas

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.

2022 m. spalio mėn. 1 d. – 2023 m. kovo mėn. 31 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[1])			
I (2020/2021) Antras pusmetis	1	1	1 (L)	1 (T[2])		1 (KD/R[3])	Publikuota
II (2021/2022) Pirmas pusmetis	1	1		1 (L)	1 (KD[4])		
II (2021/2022) Antras pusmetis	1	1	1 (L)	2 (T)	1 (KD / R)	1 (KD)	Publikuota
III (2022/2023) Pirmas pusmetis							
III (2022/2023) Antras pusmetis			1 (T)		1 (CA WoS[5])		
IV (2023/2024) Pirmas pusmetis							
IV (2023/2024) Antras pusmetis			1 (T)		1 (CA WoS)		

[1] Tarpinių rezultatų pristatymas konferencijoje Lietuvoje.

[2] Tyrimo rezultatų pristatymas tarptautinėje mokslinėje konferencijoje.

[3] Tarpinių rezultatų publikavimas (recenzuojamoje konferencijos darbų medžiagoje).

[4] Mokslinių tyrimų disertacijos tema apžvalga (konferencijos darbų medžiagoje).

[5] Rezultatų publikavimas (recenzuojamame periodiniame leidinyje CA WoS su Impact Factor).

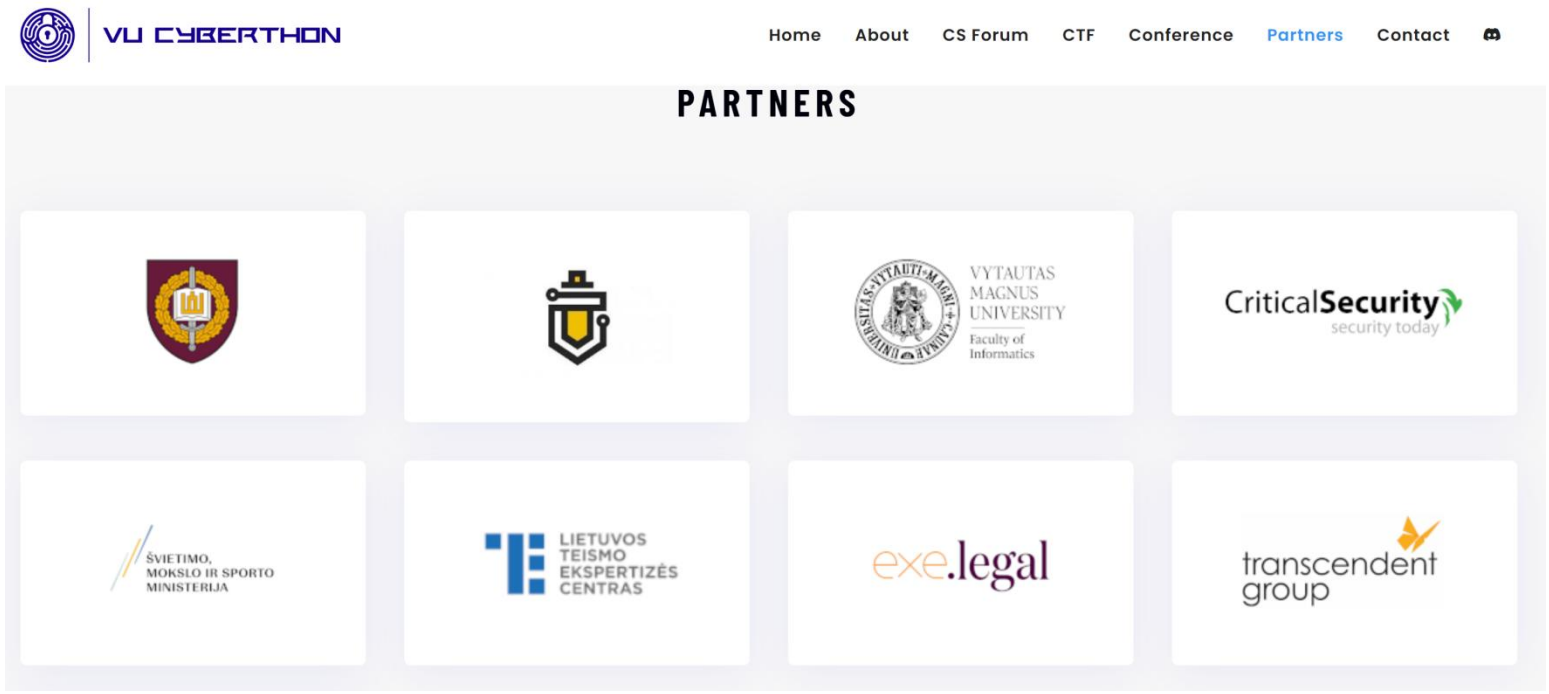
Vilniaus universitetas

Ataskaitinių metų darbo planas ir jo vykdymo suvestinė

Egzaminai			
Visi egzaminai išlaikyti.			
Dalyvavimas konferencijose			
Kol kas visose suplanuotose konferencijose sudalyvauta, kitais metais planuojamas naujų gautų rezultatų pristatymas tarptautinėje konferencijoje.			
Publikacijos			
Planas	**Vykdoma**	**Būklė**	**Publikacijos tipas**
Pagal planą III 2022/2023 antrame pusmetyje, numatytas rezultatų publikavimas recenzuojamame periodiniame leidinyje Clarivate Analytics Web of Science (CA WoS) referuojamame ir turinčiame citavimo rodiklį (Impact Factor) žurnale.	Šiame pusmetyje vykdomas skirtingų algoritmų (metodų) palyginimas. Atsižvelgiant į mokslinę literatūrą sukurtas (testuojamas) RNN LSTM slaptažodžių generatorius.	-	-

Kvalifikacijos kėlimas

- Darbas su VU Kauno fakulteto III k. studentais, dėstomas kursas “Skaitmeninio turinio teisminė analizė” (2023 pavasaris).
- Dalyvavimas Cyberthon 2023 metų užduočių dalyviams rengime.



Vilniaus universitetas

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.	Atlikta, apibendrinti rezultatai mokslinėje ataskaitoje.

Visų mokslinių tyrimų ir disertacijos rengimo etapai

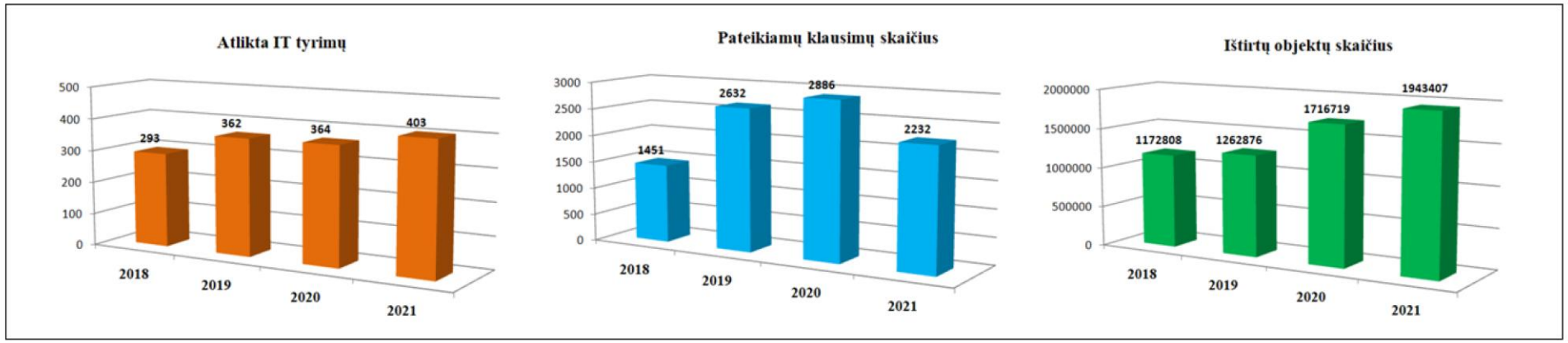
	Darbo pavadinimas	Atlikimo terminai	Pastabos
	2.2. Teorinis tyrimas: 2.2.1. Mašininio mokymosi metodų naudojamų automatizuotame slaptažodžių parinkime tyrimas. 2.2.2. Semantinės slaptažodžių analizės ir šablonų parinkimo metodų tyrimas. 2.2.3.Slaptažodžių parinkimo algoritmų taikant semantinę analizę tyrimas.	2022 m. sausio mėn. – 2022 m. rugsėjo mėn.	Vykdomas apibendrinti rezultatai mokslinėje ataskaitoje.
	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.	Vykdomas - pradėtas skirtingų algoritmų (metodų) palyginimas.
	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.	

Ekspertiniai tyrimai

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



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.

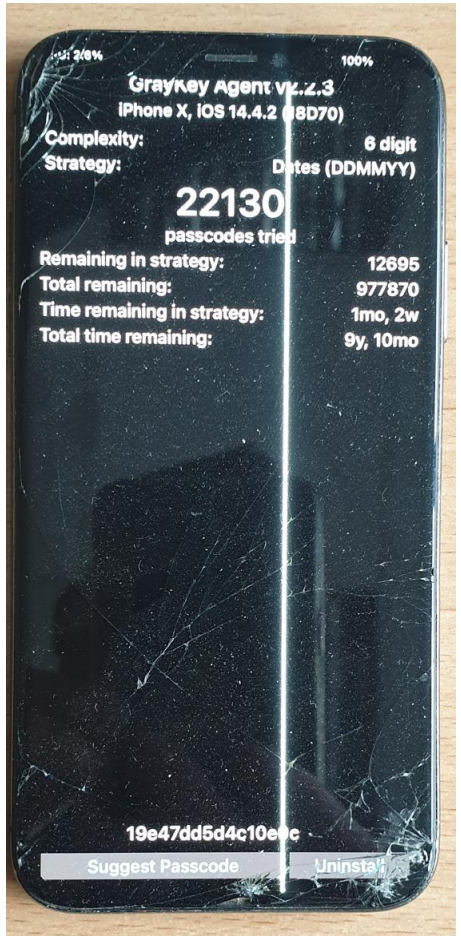
Slaptažodžių parinkimo priemonės:

- Laboratorijos aparatūrinė įranga;
- Laboratorijos programinė įranga.

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

Galimybės

Pilno perrinkimo ataka ≠ teigiamas rezultatas.



Vilniaus universitetas

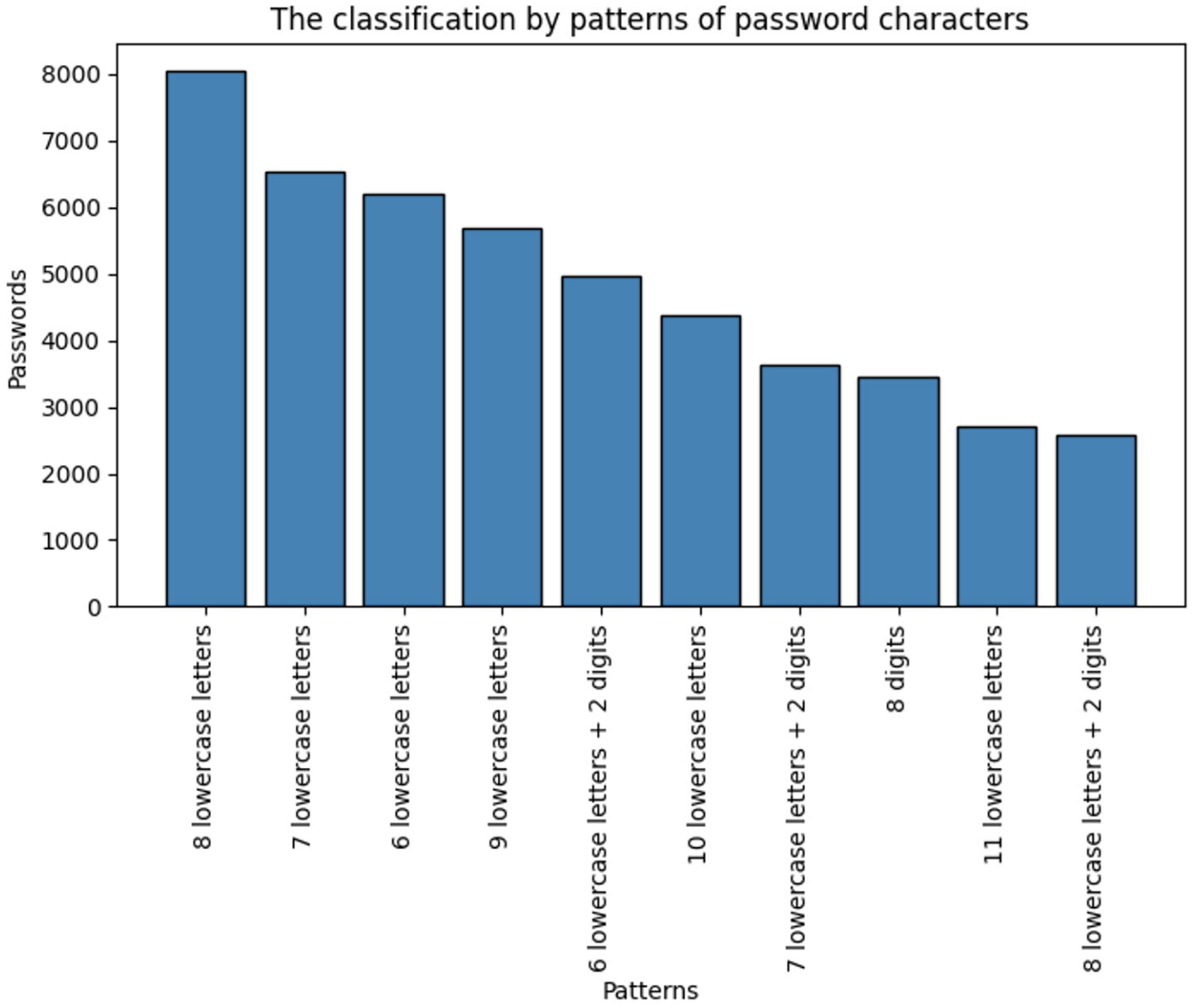
Objektai

Pagrindiniai tyrimo objektai yra:

“lietuviškas” slaptažodis

(ir jį atitinkantis šablonas, būdingas mūsų regiono vartotojams),

bei mašininio mokymosi **algoritmas** skirtas jo generavimui.



Tikslai ir uždaviniai

Disertacijos tikslas:

Sukurti naują arba modifikuoti (patobulinti) jau esamą slaptažodžių parinkimo metodą, adaptuotą mūsų regiono vartotojų slaptažodžių ypatybėms, pritaikytą teisminės IT ekspertizės ir tyrimo uždaviniams.

Disertacijos uždaviniai:

- Atlikti naujausių slaptažodžių parinkimo metodų apžvalgą, siekiant nustatyti tinkamiausius lietuviškų slaptažodžių parinkimui.
- Atlikti eksperimentus su empirinėmis duomenimis taikant atrinktus (pirmajame žingsnyje) slaptažodžių parinkimo metodus, siekiant rasti efektyviausius.
- Pasiūlyti metodą, naudojantį kontekstinę vartotojo informaciją, ištirti jo veikimą ir palyginti gautus rezultatus su alternatyviais metodais.
- Pasiūlyti neuroninių tinklų taikymų grįstą metodą, naudojantį kontekstinę naudotojo informaciją ir slaptažodžių šablonus, ištirti jo veikimą ir palyginti gautus rezultatus su alternatyviais metodais.

Slaptažodžių parinkimo metodai

Metodas	Algoritmas/programa
Taisyklėmis pagrįstas slaptažodžių parinkimo metodas	John the Ripper Hashcat
Markovo grandinės	OMEN
PCFG (angl. k. probabilistic context-free grammars) - tikimybiniai gramatikos taisyklių rinkiniai.	PCFG cracker
Pasikartojantys neuroniniai tinklai - Recurrent Neural Networks (RNNs).	RNN LSTM
Generatyviniai besivaržantys tinklai - Generative adversarial networks (GAN).	PassGan

Empiriniai duomenys

Duomenų bazės pavadinimas	Bendras slaptažodžių skaičius	Atrinkti slaptažodžiai su spausdinamais simboliais	Unikalių slaptažodžių skaičius	Šaltinis
LT1	110,302	-	97,657	https://raidforums.com
LT2	157,617	157,578 (39 išrinkti)	114,861	https://raidforums.com
LT3	645,973	645,463 (510 išrinkti)	469,821	https://raidforums.com
LT4	512,271	511,752 (519 išrinkti)	511,274	https://github.com/lexcor/LT-SecList
LT2+LT3+LT4	1,315,861	1,314,793 (1068 išrinkti)	708,766	-
Jungtinė duomenų bazė	1,314,032 (ilgis nuo 3 iki 20)	-	708,340 (ilgis nuo 3 iki 20)	-
Rockyou	32,603,388	-	14,344,391	http://downloads.skullsecurity.org/passwords

Galimi skirtingi rezultatai

Svarbu pasirinkti tinkamus duomenis.

	Algorithm	Train Set	#guess	Test Set		
				RY-2.6M	*MS*	*FB*
1	Omen-4Gram [8]	RY-30M	1.0E+10	80.40%	77.06%	66.75%
2	PCFG [8]	RY-30M	1.0E+09	32.63%	51.25%	36.4%
3	JtR-Markov [8]	RY-30M	1.0E+10	64%	53.19%	61%
4	JtR-inc [8]	RY-30M	1.0E+10	54%	25.17%	14.8%
5	LSTM	RY-30M	3.4E+09	81.52%	77.98%	59.71%
6	LSTM	RY-30M	9.8E+08	73.77%	69.30%	50.35%

Šaltinis: L. Xu et al., “Password guessing based on LSTM recurrent neural networks,” Proc. - 2017 IEEE Int. Conf. Comput. Sci. Eng. IEEE/IFIP Int. Conf. Embed. Ubiquitous Comput. CSE EUC 2017, vol. 1, pp. 785–788, 2017, doi: 10.1109/CSE-EUC.2017.155.

Generatorius

Slaptažodžių generatoriaus, grįstas natūralios kalbos teksto generavime taikomu rekurentinio neuroninio tinklo (RNN) ilgos trumpos atminties (LSTM) modeliu (char to char), kuris apmokomas parinkti sekantį sekos simbolį, pagal Softmax aktivacijos funkcijos paskaičiuojamas mūsų simbolių žodyno kiekvienos klasės tikimybes.

Pagrindiniai slaptažodžio generatoriaus modelio principai ir idėjos:

- Kiekvienas L ilgio slaptažodis $C_1C_2C_3...C_L$ yra seka x(1) = C_1, x(2)=C_2, ..x(L)=C_L;
- Naudojant seką x(1), x(2), ...,x(t-1), kaip įėjimus, mūsų neuroninis tinklas gali numatyti x(t) tikimybių pasiskirstymą;
- Modelis apmokytas nutekintų slaptažodžių duomenimis;
- Generavimas pradedamas nuo slaptažodžių pradžios reikšmės <bos>;
- Pirmi ir tolesni simboliai parenkami nuosekliai pagal numatytą atrankos kriterijų (slenkstį), pagal paskaičiuotą tikimybių pasiskirstymą.

Pradiniai rezultatai

Slaptažodžių kandidatų skaičius	Parinkta slaptažodžių (vnt.)	Parinkta slaptažodžių (proc.)	Nustatytas slenkstis
20465	77	0.05	10^{-6}
122855	329	0.23	10^{-7}
816127	1127	0.8	10^{-8}
4896160	3257	2.30	10^{-9}

Kito pusmečio darbo planas

1. Eksperimentų pagal sudarytą teorinį ir empirinį tyrimų planą atlikimas.
2. Toliau vykdomas skirtingų algoritmų (metodų) palyginimas. Atsižvelgiant į mokslinę literatūrą kuriamas (tobulinamas) ir testuojamas RNN LSTM slaptažodžių generatorius.
3. Gautų rezultatų analizė, apibendrinimas ir publikacijos (CA WoS) parengimas.

Vilniaus universitetas

Ačiū už dėmesį

Andrius Chaževskas

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