

LITUND: lietuvių kalbos tekstynas nepatikimų naujienų atpažinimui

Edgaras Dambrauskas



VYTAUTO
DIDŽIOJO
UNIVERSITETAS
VILNIUS

Hmf

Humanitarinių
mokslų fakultetas

LITUND: Lithuanian Unreliable News Detection Corpus

Pilnos apimties, naujienas
imituojantys tekstai.

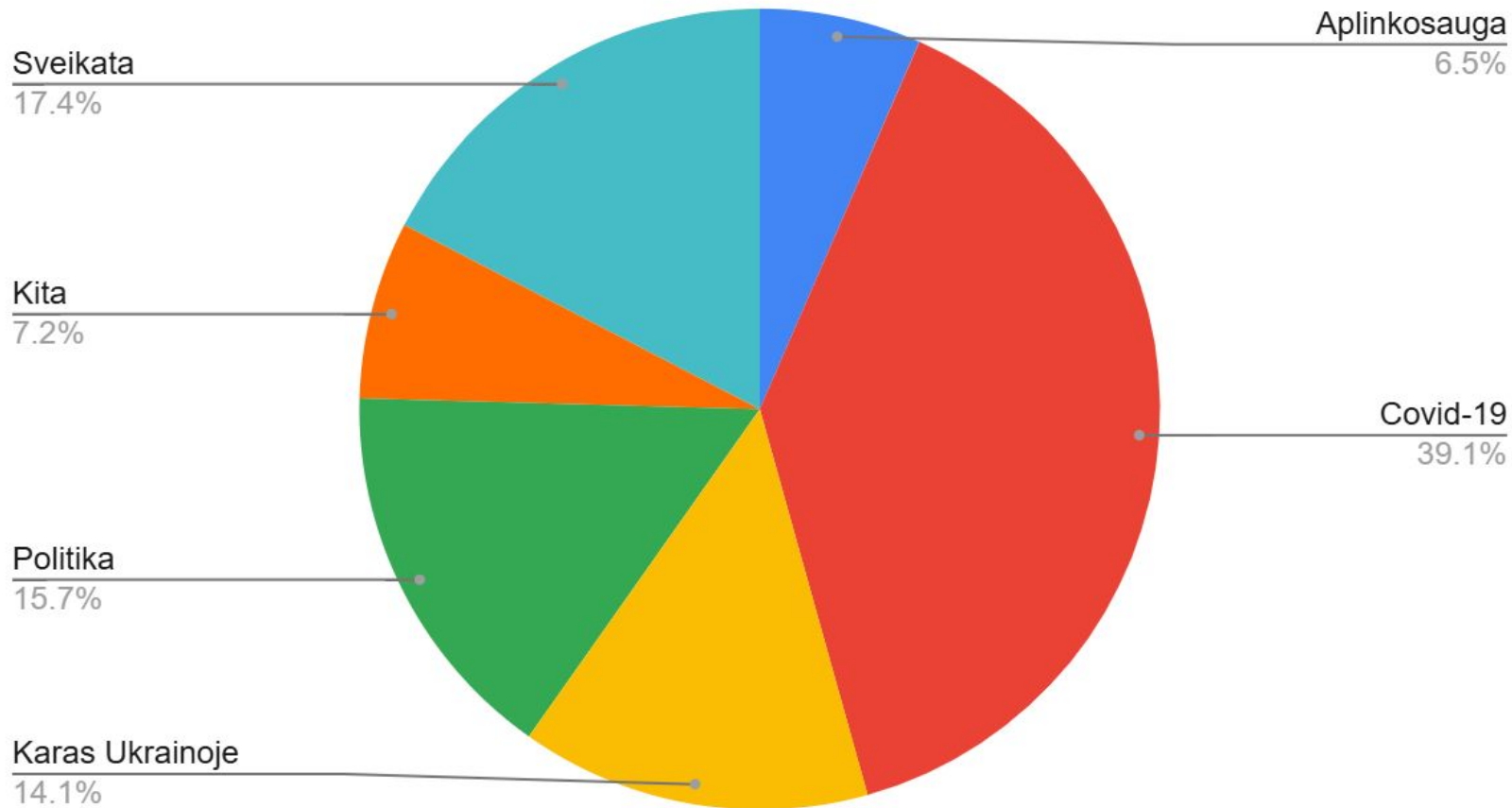
Patikimumą ir kokybę
užtikrina profesionalių
žurnalistų tyrimai.



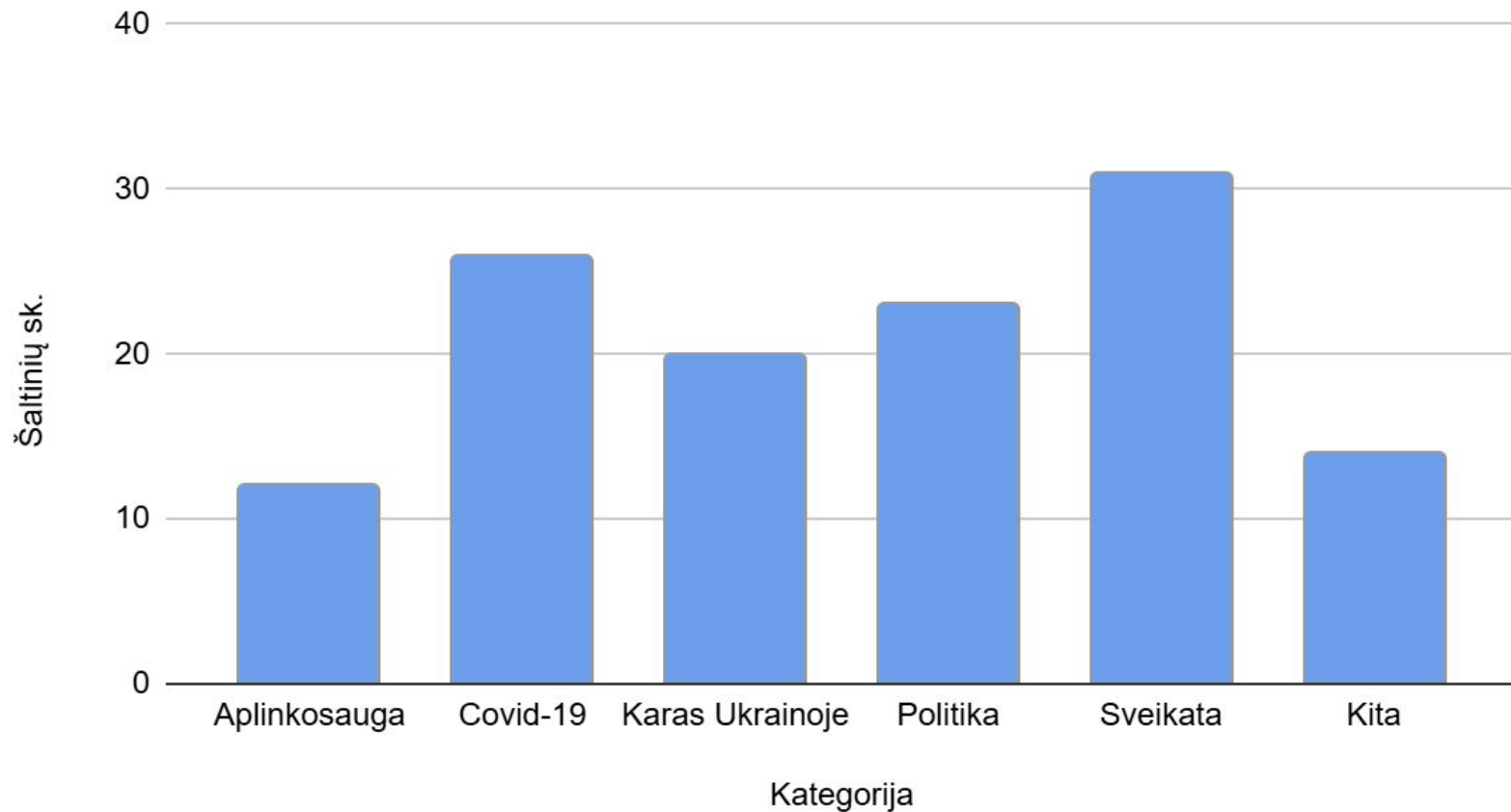
LITUND

- ✓ Autorių ir šaltinių įvairovė;
- ✓ Tekstai turi savo tematinius atitikmenis;
- ✓ Morfologiškai anotuoti;
- ✓ Suskirstyti į 6 kategorijas.

Temų pasiskirstymas



Šaltinių įvairovė



NELABAI PATIKIMŲ NAUJIENŲ NELABAI PATIKIMA PIRAMIDĖ



Temų įvairovė

Bideno šeima pripažįsta, kad jis mirė ir
2019 m. jį pakeitė aktorius

Norite nuplukti? Pasiskiepykite kovidų "vakcina"!

"5G" yra ginklų sistema, sukurta žmonėms žudyti, sako
ginklų ekspertas Markas Steele'as

Laurynas Kasčiūnas nusprendė – kariauti eis ir
moterys, ir pensininkai

Tematinės sąsajos

Lietuva pagal mirštamumą nuo AIDS yra pirmoje vietoje Europos Sąjungoje

Daugėja AIDS atvejų Lietuvoje: viršijamas net Europos Sąjungos vidurkis

Slovakijos premjeras patvirtino, kad Lietuva ir kitos šalys siųs kariuomenę į Ukrainą

Slovakija prie savo sienos su Ukraina dislokuoja papildomus karius

COVID-19 pavyzdys

Žodis	NN / 1000	LRT / 1000	LOG	Žodis	NN / 1000	LRT / 1000	LOG
injekcija	1.972	0.029	2.225501	pandemija	0.909	4.219	-1.74425
vėžys	1.972	0.157	1.911074	atmaina	0.135	1.302	-1.50528
kraujas	1.595	0.107	1.786689	koronavirusas	1.151	3.898	-1.41336
ląstelė	1.912	0.257	1.67075	para	0.027	0.865	-1.37357
vakcina	10.197	3.204	1.530158	protrūkis	0.101	1.037	-1.35472
straipsnis	1.259	0.114	1.517186	situacija	0.249	1.409	-1.34956
kovidas	1.185	0.122	1.438439	įstaiga	0.162	1.166	-1.33221
skiepai	2.288	0.615	1.32242	valstybė	0.175	1.180	-1.31548
DNR	0.875	0.064	1.284753	karantinas	0.108	0.958	-1.26286
širdis	1.030	0.186	1.15722	šalis	0.855	2.725	-1.25105
baltymas	1.063	0.207	1.144182	gyventojas	0.518	1.888	-1.22965
daktaras	0.707	0.093	1.025091	simptomas	0.357	1.323	-1.0894
mirtis	2.652	1.108	0.970568	priemonė	0.444	1.495	-1.07891
aliuminis	0.471	0.000	0.957783	sergamumas	0.114	0.794	-1.07431
poveikis	1.346	0.472	0.925524	metas	1.144	2.954	-1.07064

Karo Ukrainoje pavyzdys

Žodis	NN / 1000	LRT / 1000	LOG
užsienis	2.385	0.539	1.472786
režimas	1.206	0.160	1.370386
organas	0.616	0.000	1.158821
plėtra	0.670	0.040	1.115549
valstija	1.420	0.440	1.031304
interesas	0.670	0.080	1.012548
straipsnis	0.777	0.140	0.997119
legionas	0.563	0.040	0.976913
biolaboratorija	0.482	0.000	0.974353
apsauga	1.018	0.280	0.961483
genocidas	0.429	0.000	0.893412
tarnautojas	0.456	0.020	0.877922
valstybė	2.841	1.319	0.877249
sąjunga	0.965	0.300	0.873132
karininkas	0.831	0.240	0.847126

Žodis	NN / 1000	LRT / 1000	LOG
orlaivis	0.027	1.179	-1.67212
oras	0.509	2.378	-1.5117
dronas	0.322	1.738	-1.44589
mėnuo	0.429	1.698	-1.24298
moteris	0.080	0.839	-1.20622
lėktuvas	0.241	1.119	-1.12708
bandymas	0.402	1.459	-1.11861
įmonė	0.027	0.639	-1.1129
brigada	0.241	1.059	-1.07265
naikintuvas	0.080	0.719	-1.07092
miestas	0.456	1.479	-1.04997
puolimas	0.348	1.139	-0.9499
ponia	0.054	0.539	-0.90893
energetika	0.054	0.539	-0.90893
regionas	0.348	1.059	-0.87777



Ačiū

MCMXXII

VYTAUTO DIDŽIOJO
UNIVERSITETAS

