

Tekstynas **GROŽINĖ_INDV** skirtas individualių autorių autorystės nustatymo tyrimams.

Tekstyną sudaro grožinių kūrinių autorių tekstai

Autorių kiekis: 186 autoriai

Minimalus teksto ilgis: 2000 simbolių (įskaitant tarpus)

Eilutės kiekviename iš tekstų atitinka teksto savybes, naudojamas autorystės nustatymo tyrimuose (daugiau apie savybes skaitykite [1]):

- 1) vidutinis sakinio ilgis tekste
- 2) vidutinis žodžio ilgis tekste
- 3) skirtingų žodžių ir visų žodžių tekste santykis
- 4) tekstas
- 5) lemuotas tekstas (lemuota automatinio įrankiu "Lemuoklis"[2]: atpažįstamas žodis verčiamas lema, bendrinis žodis taip pat verčiamas mažosiomis raidėmis; jeigu žodis neatpažįstamas – paliekama neliesta jo forma)
- 6) kalbos dalys (nustatytos "Lemuoklio" pagalba)
- 7) priklausomybių gramatikos sintaksinės žymos kiekvienam iš žodžių (apdorota automatinio įrankiu MaltParser¹, prieš tai jį apmokius su lietuviškuoju medžių banku² bei didžiausią tikslumą duodančiais parametrais [3])
- 8) funkciniai žodžiai (jungtukai, jaustukai, dalelytės, prielinksniai, įvardžiai: nustatyti automatinio lemavimo įrankiu "Lemuoklis")
- 9) simbolių 2-gramos (tekstinio dokumento mastu). Tekstas suskaidomas į elementus po du simbolius (pritaikius slenkančio lango metodą per vieną simbolį), tarpo simboliai prieš tai keičiami "_". Pvz.: "visas_tekstas": "vi", "is", "sa", "as", "s_", "_t", "te" ir t.t.
- 10) simbolių 3-gramos (tekstinio dokumento mastu)
- 11) simbolių 4-gramos (tekstinio dokumento mastu)
- 12) simbolių 5-gramos (tekstinio dokumento mastu)
- 13) simbolių 6-gramos (tekstinio dokumento mastu)
- 14) simbolių 7-gramos (tekstinio dokumento mastu)
- 15) simbolių 2-gramos (žodžio ribose). Tekstas suskaidomas į elementus po du simbolius (pritaikius slenkančio lango metodą per vieną simbolį). Pvz.: "visas_tekstas": "vi", "is", "sa", "as", "te", "ek", "ks", "st", "ta", "as".
- 16) simbolių 3-gramos (žodžio ribose)
- 17) simbolių 4-gramos (žodžio ribose)
- 18) simbolių 5-gramos (žodžio ribose)
- 19) simbolių 6-gramos (žodžio ribose)
- 20) simbolių 7-gramos (žodžio ribose)
- 21) tekstas junginiais: žodžiais ir jų kalbos dalimis
- 22) tekstas junginiais: lemomis ir žodžių kalbos dalimis
- 23) tekstas junginiais: žodžiais, jų kalbos dalimis bei kita morfologine informacija (morfologinė informacija, pvz. linksnis, skaičius, laipsnis, laikas ir kt. nustatyta "Lemuoklio" pagalba)
- 24) tekstas junginiais: lemomis, žodžių kalbos dalimis bei kita morfologine informacija
- 25) tekstas junginiais: kalbos dalimis kartu su morfologine informacija

Nuorodos:

[1] Kapočiūtė-Dzikienė, Jurgita; Utkā, Andrius; Šarkutė, Ligita. 2014. Seimo posėdžių stenogramų tekstynas autorystės nustatymo bei autoriaus profilio sudarymo tyrimams. *Kalbotyra*, 66: 27–45.

[2] Zinkevičius, Vytautas. 2000. Lemuoklis – morfologinei analizei. Gudaitis, L. (red.) *Darbai ir Dienos*, 24: 246–273.

[3] Kapočiūtė-Dzikienė, Jurgita; Nivre, Joakim; Krupavičius, Algis. 2013. Lithuanian Dependency Parsing with Rich Morphological Features. *Empirical Methods in Natural Language Processing – 4th Workshop on Statistical Parsing of Morphologically Rich Languages (SPMRL'2013)*, psl. 12–21.

¹ Įrankis MaltParser parsisiųstas iš <http://www.maltparser.org/>.

² Medžių bankas sukurtas VDU vykdyto Lietuvos valstybinio mokslo ir studijų fondo Lituaništos mokslinių tyrimų prioriteto įgyvendinimo 2007–2008 metų programos projekto "Internetiniai ištekliai: anototas lietuvių kalbos tekstynas ir anotavimo priemonės (ALKA 2)" metu.

Statistika

klasė	pavyzdžių kiekis	žodžių kiekis	skirtingų žodžių kiekis	vid. pvz. ilgis žodžiais
autorius1	161	47592	15825	295.60
autorius2	141	40698	12080	288.64
autorius3	193	59497	12414	308.27
autorius4	155	46098	8505	297.41
autorius5	159	46287	14569	291.11
autorius6	121	34150	12434	282.23
autorius7	184	52201	18961	283.70
autorius8	69	19355	8780	280.51
autorius9	71	19840	8921	279.44
autorius10	53	15075	6558	284.43
autorius11	89	26116	8120	293.44
autorius12	177	50903	16709	287.59
autorius13	126	37832	8822	300.25
autorius14	182	52401	16074	287.92
autorius15	70	21217	8187	303.10
autorius16	45	13260	5416	294.67
autorius17	229	64959	20403	283.66
autorius18	136	39569	12077	290.95
autorius19	101	29480	10618	291.88
autorius20	194	57745	17846	297.65
autorius21	36	10393	4460	288.69
autorius22	459	130161	35160	283.58
autorius23	365	102365	29864	280.45
autorius24	99	29500	8398	297.98
autorius25	164	45222	17620	275.74
autorius26	74	21039	8666	284.31
autorius27	191	53599	16666	280.62
autorius28	132	37949	9841	287.49
autorius29	163	48473	17157	297.38
autorius30	269	77114	20879	286.67
autorius31	212	57366	17518	270.59
autorius32	88	24666	10873	280.30
autorius33	86	24524	10334	285.16
autorius34	232	81004	22124	349.16
autorius35	225	63103	22555	280.46
autorius36	205	58178	14417	283.80
autorius37	43	11508	4485	267.63
autorius38	82	23642	8720	288.32
autorius39	89	24886	9833	279.62
autorius40	108	32013	11811	296.42
autorius41	225	67396	14127	299.54
autorius42	352	98549	31030	279.97
autorius43	147	42557	10436	289.50
autorius44	1077	298088	59131	276.78
autorius45	282	80911	23531	286.92
autorius46	216	62353	20687	288.67
autorius47	86	25292	10525	294.09
autorius48	159	46221	14319	290.70
autorius49	240	71765	14868	299.02
autorius50	222	63839	19109	287.56
autorius51	549	158251	38596	288.25
autorius52	76	21866	8558	287.71
autorius53	145	42366	13967	292.18
autorius54	229	65222	18463	284.81
autorius55	190	54154	14277	285.02
autorius56	205	60591	19154	295.57
autorius57	54	16689	7193	309.06
autorius58	49	14503	6019	295.98
autorius59	196	57330	18646	292.50
autorius60	13	4051	1594	311.62
autorius61	104	33357	11773	320.74
autorius62	394	114524	25474	290.67

autorius63	208	57325	19199	275.60
autorius64	62	16779	7789	270.63
autorius65	21	6357	2453	302.71
autorius66	111	32151	9463	289.65
autorius67	156	46736	15678	299.59
autorius68	43	12341	3839	287.00
autorius69	358	108485	25101	303.03
autorius70	212	60918	17444	287.35
autorius71	168	46095	15621	274.38
autorius72	153	42810	14436	279.80
autorius73	40	11672	4409	291.80
autorius74	342	99752	24683	291.67
autorius75	267	76572	19994	286.79
autorius76	234	64432	20170	275.35
autorius77	67	20781	7853	310.16
autorius78	205	61745	16391	301.20
autorius79	72	20472	7417	284.33
autorius80	113	33106	9100	292.97
autorius81	437	121139	28578	277.21
autorius82	167	48190	14323	288.56
autorius83	48	14400	6268	300.00
autorius84	91	26335	9705	289.40
autorius85	108	30945	10277	286.53
autorius86	484	135597	40004	280.16
autorius87	108	31109	11087	288.05
autorius88	13	11812	4424	908.62
autorius89	92	26663	8596	289.82
autorius90	204	57237	16697	280.57
autorius91	166	50410	12806	303.67
autorius92	103	29686	12302	288.21
autorius93	53	16203	5380	305.72
autorius94	393	115354	22219	293.52
autorius95	74	20065	9490	271.15
autorius96	366	105265	26141	287.61
autorius97	222	63517	15487	286.11
autorius98	224	62853	18667	280.59
autorius99	175	51697	18943	295.41
autorius100	396	112682	20885	284.55
autorius101	254	72867	19565	286.88
autorius102	239	68174	21008	285.25
autorius103	154	45789	14458	297.33
autorius104	191	54464	18517	285.15
autorius105	168	47517	17175	282.84
autorius106	110	29586	12371	268.96
autorius107	203	59100	13719	291.13
autorius108	373	106343	24317	285.10
autorius109	190	54985	15919	289.39
autorius110	292	84395	28985	289.02
autorius111	184	52724	16312	286.54
autorius112	174	49005	13986	281.64
autorius113	174	49019	16907	281.72
autorius114	263	77452	21786	294.49
autorius115	322	94903	22565	294.73
autorius116	66	20670	6699	313.18
autorius117	184	54181	16144	294.46
autorius118	318	90546	22655	284.74
autorius119	69	19663	7566	284.97
autorius120	164	46560	15386	283.90
autorius121	129	36367	12965	281.91
autorius122	97	28717	11633	296.05
autorius123	175	51967	14242	296.95
autorius124	155	45772	17837	295.30
autorius125	168	48673	17341	289.72
autorius126	278	80694	26983	290.27
autorius127	67	19457	7601	290.40

autorius128	277	75386	21790	272.15
autorius129	132	38428	13650	291.12
autorius130	91	28160	8271	309.45
autorius131	189	58190	13204	307.88
autorius132	152	41591	12890	273.63
autorius133	246	71058	23680	288.85
autorius134	137	40110	12585	292.77
autorius135	29	8005	3962	276.03
autorius136	99	29186	9969	294.81
autorius137	149	42094	15705	282.51
autorius138	110	33700	10777	306.36
autorius139	349	106254	24568	304.45
autorius140	85	26010	8567	306.00
autorius141	58	15962	5973	275.21
autorius142	72	22415	8770	311.32
autorius143	86	25234	10240	293.42
autorius144	213	60700	19895	284.98
autorius145	77	22822	8871	296.39
autorius146	168	49880	16657	296.90
autorius147	105	34855	11048	331.95
autorius148	64	17955	7141	280.55
autorius149	164	50503	12416	307.95
autorius150	177	51656	14251	291.84
autorius151	274	79129	20644	288.79
autorius152	167	51366	14792	307.58
autorius153	174	47518	18780	273.09
autorius154	131	38761	13386	295.89
autorius155	165	50547	12773	306.35
autorius156	55	17100	5994	310.91
autorius157	88	26227	10130	298.03
autorius158	130	39419	11927	303.22
autorius159	253	76438	21713	302.13
autorius160	115	33258	11724	289.20
autorius161	128	39573	11962	309.16
autorius162	530	155691	36424	293.76
autorius163	244	71016	19797	291.05
autorius164	283	83460	22629	294.91
autorius165	335	96580	25991	288.30
autorius166	206	56946	22278	276.44
autorius167	217	63502	18963	292.64
autorius168	177	51235	17740	289.46
autorius169	32	9320	4288	291.25
autorius170	343	100113	25939	291.87
autorius171	196	54965	19135	280.43
autorius172	224	65293	20727	291.49
autorius173	170	47539	16758	279.64
autorius174	262	78992	21487	301.50
autorius175	90	26083	9913	289.81
autorius176	171	52513	11116	307.09
autorius177	208	59596	18004	286.52
autorius178	146	40508	17595	277.45
autorius179	108	30807	11811	285.25
autorius180	383	114552	28679	299.09
autorius181	376	112892	30040	300.24
autorius182	65	19312	7004	297.11
autorius183	310	92823	24582	299.43
autorius184	164	46429	13188	283.10
autorius185	201	59406	19446	295.55
autorius186	185	52881	16589	285.84
Visi	33640	9762077	504156	290.19