

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

Tekstyną sudaro registruotų vartotojų pasisakymai forume

Laikotarpis: 2014 m. kovas

Aurių kiekis: 174 autoriai

Minimalus žodžių kiekis tekste: nėra

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) tekstas junginiais: žodžiais ir jų kalbos dalimis
- 16) tekstas junginiais: lemomis ir žodžių kalbos dalimis
- 17) tekstas junginiais: žodžiais, jų kalbos dalimis bei kita morfologine informacija (morfologinė informacija, pvz. linksnis, skaičius, laipsnis, laikas ir kt. nustatyta "Lemuoklio" pagalba)
- 18) tekstas junginiais: lemomis, žodžių kalbos dalimis bei kita morfologine informacija

Nuorodos:

[1] Kapočiūtė-Dzikiėnė, 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-Dzikiėnė, 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 Lituanistikos mokslinių tyrimų prioriteto įgyvendinimo 2007–2008 metų programos projekto "Internetiniai ištekļiai: anototas lietuvių kalbos tekstynas ir anotavimo priemonės (ALKA 2)" metu.

Statistika

klasė	pavyzdžių kiekis	žodžių kiekis	skirtingų žodžių kiekis	vid. pvz. ilgis žodžiais
autorius1	71	3149	1634	44.35
autorius2	15	967	576	64.47
autorius3	11	354	263	32.18
autorius4	10	421	301	42.10
autorius5	216	13100	5519	60.65
autorius6	14	516	366	36.86
autorius7	131	8282	2840	63.22
autorius8	11	500	322	45.45
autorius9	14	573	368	40.93
autorius10	301	6273	2855	20.84
autorius11	26	934	510	35.92
autorius12	38	2829	1563	74.45
autorius13	22	506	404	23.00
autorius14	11	204	168	18.55
autorius15	12	497	348	41.42
autorius16	254	12613	5608	49.66
autorius17	50	1736	924	34.72
autorius18	30	1059	625	35.30
autorius19	12	311	252	25.92
autorius20	59	7518	3323	127.42
autorius21	24	963	557	40.13
autorius22	76	3185	1686	41.91
autorius23	279	15005	5573	53.78
autorius24	24	909	555	37.88
autorius25	261	11364	3861	43.54
autorius26	433	10757	4871	24.84
autorius27	11	189	146	17.18
autorius28	63	2409	1275	38.24
autorius29	20	447	290	22.35
autorius30	19	1043	604	54.89
autorius31	33	1701	941	51.55
autorius32	13	736	485	56.62
autorius33	11	639	419	58.09
autorius34	29	595	418	20.52
autorius35	28	896	467	32.00
autorius36	145	5316	2594	36.66
autorius37	15	288	196	19.20
autorius38	20	815	518	40.75
autorius39	36	2121	1130	58.92
autorius40	143	7579	3229	53.00
autorius41	53	1977	945	37.30
autorius42	41	1907	1006	46.51
autorius43	116	5864	2617	50.55
autorius44	21	697	453	33.19
autorius45	67	2533	1328	37.81
autorius46	54	1344	817	24.89
autorius47	70	4947	2012	70.67
autorius48	15	145	129	9.67
autorius49	60	2565	1343	42.75
autorius50	93	3050	1361	32.80
autorius51	25	1846	990	73.84
autorius52	13	1026	601	78.92
autorius53	24	1802	1047	75.08
autorius54	36	1218	738	33.83
autorius55	298	9869	3561	33.12
autorius56	11	309	221	28.09
autorius57	43	1575	1031	36.63
autorius58	124	4862	2106	39.21
autorius59	15	728	463	48.53
autorius60	24	1010	507	42.08
autorius61	20	1896	834	94.80
autorius62	13	413	271	31.77
autorius63	77	2819	1429	36.61

autorius64	37	1597	986	43.16
autorius65	15	357	238	23.80
autorius66	27	516	357	19.11
autorius67	19	1317	735	69.32
autorius68	56	1644	1059	29.36
autorius69	27	1992	1055	73.78
autorius70	14	1112	481	79.43
autorius71	20	2340	1075	117.00
autorius72	43	2607	1247	60.63
autorius73	1172	47943	11947	40.91
autorius74	14	469	289	33.50
autorius75	12	429	312	35.75
autorius76	21	1418	703	67.52
autorius77	14	1193	670	85.21
autorius78	116	3037	1738	26.18
autorius79	14	894	589	63.86
autorius80	29	1256	824	43.31
autorius81	16	366	229	22.88
autorius82	16	738	287	46.13
autorius83	13	2677	599	205.92
autorius84	11	247	175	22.45
autorius85	87	3866	1814	44.44
autorius86	13	1144	624	88.00
autorius87	10	191	161	19.10
autorius88	10	312	233	31.20
autorius89	11	710	440	64.55
autorius90	18	957	579	53.17
autorius91	39	1837	1005	47.10
autorius92	94	6709	2922	71.37
autorius93	25	2048	1137	81.92
autorius94	49	1722	1085	35.14
autorius95	17	643	411	37.82
autorius96	13	595	425	45.77
autorius97	166	3899	2095	23.49
autorius98	74	2447	1389	33.07
autorius99	43	1729	1016	40.21
autorius100	16	521	380	32.56
autorius101	56	1892	986	33.79
autorius102	30	1087	671	36.23
autorius103	174	8003	3203	45.99
autorius104	20	1346	817	67.30
autorius105	1323	40153	10877	30.35
autorius106	20	917	564	45.85
autorius107	40	1557	930	38.93
autorius108	357	11527	4096	32.29
autorius109	11	376	276	34.18
autorius110	16	881	484	55.06
autorius111	31	1000	638	32.26
autorius112	48	1611	738	33.56
autorius113	38	2306	1278	60.68
autorius114	17	1105	390	65.00
autorius115	49	2310	1093	47.14
autorius116	12	276	227	23.00
autorius117	31	790	559	25.48
autorius118	11	1048	484	95.27
autorius119	271	8207	3131	30.28
autorius120	119	3301	1847	27.74
autorius121	29	1183	765	40.79
autorius122	31	960	581	30.97
autorius123	140	6602	3068	47.16
autorius124	11	899	616	81.73
autorius125	79	2545	1296	32.22
autorius126	34	1967	1034	57.85
autorius127	24	573	385	23.88
autorius128	35	2362	1261	67.49

autorius129	55	2940	1416	53.45
autorius130	18	1143	746	63.50
autorius131	23	830	436	36.09
autorius132	18	860	602	47.78
autorius133	11	312	257	28.36
autorius134	13	205	163	15.77
autorius135	59	1835	1021	31.10
autorius136	352	29427	9844	83.60
autorius137	11	880	564	80.00
autorius138	19	801	496	42.16
autorius139	37	1682	1026	45.46
autorius140	12	1053	692	87.75
autorius141	10	379	267	37.90
autorius142	59	4787	2220	81.14
autorius143	322	21339	7961	66.27
autorius144	32	1825	925	57.03
autorius145	221	7816	3165	35.37
autorius146	11	362	263	32.91
autorius147	118	4045	1955	34.28
autorius148	187	8965	4159	47.94
autorius149	21	676	492	32.19
autorius150	43	2193	1152	51.00
autorius151	103	3439	1819	33.39
autorius152	15	439	317	29.27
autorius153	10	286	225	28.60
autorius154	78	1967	1146	25.22
autorius155	15	744	455	49.60
autorius156	10	286	214	28.60
autorius157	17	364	280	21.41
autorius158	232	17366	4954	74.85
autorius159	17	1159	715	68.18
autorius160	19	809	491	42.58
autorius161	41	1843	1042	44.95
autorius162	14	341	266	24.36
autorius163	39	869	518	22.28
autorius164	84	2129	1176	25.35
autorius165	18	345	283	19.17
autorius166	13	534	324	41.08
autorius167	25	1222	772	48.88
autorius168	10	131	123	13.10
autorius169	17	994	632	58.47
autorius170	68	2858	1423	42.03
autorius171	1072	49672	14130	46.34
autorius172	14	573	381	40.93
autorius173	579	17987	6463	31.07
autorius174	38	2179	1041	57.34
Visi	13952	596855	80099	42.78