

PRIVATUMO PRANEŠIMAS

Projekto tikslas - surinkti reikiamus lingvistinius resursus, juos tinkamai sutvarkyti ir parengti išsamų Bendrąjį lietuvių kalbos tekstyną, tenkinantį dirbtinio intelekto technologijų vystymo ir skaitmeninių bei statistinių kalbos tyrimų poreikius, ir jų pagrindu sukurti iš anksto apmokytus (angl. pre-trained) neuroninius lietuvių kalbos modelius, turinčius išsamias faktines žinias apie lietuvių kalbą bei tenkinančius dirbtinio intelekto technologijų vystymo ir skaitmeninių bei statistinių kalbos tyrimų poreikius.

Kadangi asmens duomenys šiame Projekte gaunami ne tiesiogiai iš duomenų subjektų, taikomas Bendrojo duomenų apsaugos reglamento 14 straipsnis. Atsižvelgiant į labai didelę tvarkomų tekstų apimtį, šaltinių kiekį, duomenų subjektų skaičių ir tai, kad daugeliu atvejų duomenys yra tik neatskiriama viešai paskelbtų tekstų dalis, informacijos pateikimas kiekvienam duomenų subjektui atskirai būtų neįmanomas arba pareikalautų neproporcingų pastangų. Viso, tekstai paskelbti visuomenės informavimo priemonių sudarys apie 3 941 518 292 žodžių, t.y. 97,83 proc. viso tekstyno. Todėl informacija apie duomenų tvarkymą teikiama viešai šiame privatumo pranešime, vadovaujantis BDAR 14 straipsnio 5 dalies b punktu.

Projekto veiklų įgyvendinimo pradžia: 2024 m. rugsėjo 19 d.

Projekto veiklų įgyvendinimo pabaiga: 2026 m. balandžio 15 d.

1. Duomenų valdytojai

Duomenis tvarko Projekto partneriai, kartu veikiantys kaip bendri duomenų valdytojai:

- Vytauto Didžiojo universitetas;
- UAB „Tilde Informacinės Technologijos“;
- UAB „Neurotechnology“;
- MB „Krilas“.

Bendri duomenų valdytojai yra tarpusavyje susitarę dėl atsakomybės pasidalijimo, įskaitant informacijos teikimą duomenų subjektams ir duomenų subjektų teisių nagrinėjimą. Duomenų subjektai gali kreiptis į bet kurį iš bendrų duomenų valdytojų, tačiau dėl praktinių priežasčių kontaktinis taškas klausimams dėl duomenų apsaugos yra:

Vytauto Didžiojo universitetas

El. paštas: andrius.utka@vdu.lt

2. Duomenų tvarkymo tikslai

Duomenys tvarkomi šiais tikslais:

- sukurti didelės apimties Bendrąjį lietuvių kalbos tekstyną;
- atlikti lingvistinę, statistinę ir kompiuterinę tekstų analizę;
- mokyti, testuoti ir tobulinti lietuvių kalbos natūralios kalbos apdorojimo ir dirbtinio intelekto modelius;
- sudaryti sąlygas naudoti sukurtus kalbinius išteklius kalbos technologijų ir kitų skaitmeninių sprendimų kūrimo tikslais.

3. Duomenų tvarkymo teisinis pagrindas

Duomenys tvarkomi vadovaujantis BDAR 6 straipsnio 1 dalies e punktu, t. y. kai duomenų tvarkymas yra būtinas vykdant užduotį, atliekamą viešojo intereso labui. Duomenų tvarkymas atliekamas vykdant Projektą, kuris prisideda prie 2021–2030 metų valstybės skaitmeninimo plėtros programos pažangos priemonės „Kalbinių išteklių dirbtinio intelekto technologijų sprendimų poreikiams plėtra“ įgyvendinimo.

Kadangi tekstynas sudaromas iš įvairių viešai prieinamų tekstinių šaltinių, tam tikrais atvejais tvarkomuose tekstuose gali pasitaikyti ir specialių kategorijų asmens duomenų, kurie yra paskelbti viešuose šaltiniuose. Tokie duomenys nėra renkami tikslingai ir nėra savarankiškas duomenų rinkimo objektas; jie tvarkomi tik tiek, kiek neišvengiamai pasitaiko tvarkomuose tekstuose. Tekstai į tekstyną įtraukiami iš teisėtai prieinamų šaltinių, įskaitant šaltinius, su kuriais projekto vykdytojai yra sudarę duomenų pateikimo ar naudojimo susitarimus. Tvarkant tekstinius duomenis siekiama laikytis duomenų minimizavimo principo ir užtikrinti, kad asmens duomenys nebūtų tvarkomi plačiau nei būtina projekto tikslams pasiekti.

4. Šaltiniai, iš kurių gaunami duomenys

Asmens duomenys, jeigu jie pasitaiko tvarkomuose tekstuose, gaunami ne tiesiogiai iš duomenų subjektų, o iš viešai prieinamų šaltinių, naudojamų Bendrojo lietuvių kalbos tekstyno kūrimui.

Nr.	Šaltinio pavadinimas	Tekstai	Žodžių sk.	%, tekстыne
1	15min.lt	559107	230933602	5,86%
2	autogidas.lt	1846	735416	0,02%
3	bernardinai.lt	114040	65085920	1,65%
4	BNS	352946	83877818	2,13%
5	CRIS-DB	6483	34630986	0,88%
6	data_from_stat.gov.lt	21449	6207750	0,16%
7	delfi.lt	1714345	606662831	15,39%
8	elta.lt	1235494	254356849	6,45%
9	epaveldas.lt	265	11339315	0,29%
10	etalpykla.vilniustech.lt.lt	258	210746	0,01%
11	e-tar.lt	307240	567490596	14,40%
12	eur-lex.europa.eu	141887	319901553	8,12%
13	Gargždapis	99	98983	0,00%
14	gagas.lt	5454	2022545	0,05%
15	gov.lt	493727	156570014	3,97%
16	Hubris	16	1118951	0,03%
17	Humanitarų meka	64	68586	0,00%
18	Internet archive	112	5592994	0,14%
19	lv_auoriai	182	1226753	0,03%
20	kaunas.kasvyksta.lt	49662	15141979	0,38%
21	Lituanistika_DB	286	1036134	0,03%
22	LLTI	2	867891	0,02%
23	lrytas.lt	1140602	380082600	9,64%
24	lrkt.lt	1029	4162009	0,11%

25	lrt.lt	896722	363221463	9,22%
26	MIDAS	42	39747	0,00%
27	pirkimai.eviesieji.pirkimai.lt	46288	117844194	2,99%
28	Raudonoji knyga	396	103296	0,00%
29	sena.lt	88	31883	0,00%
30	Švenčionių kraštas	11046	5253903	0,13%
31	tautinispaveldas.lt	392	59793	0,00%
32	teismai.lt	715723	499989396	12,69%
33	vdu.lt	5937	709234	0,02%
34	Vakarų ekspresas	427947	160571897	4,07%
35	Vikipedija	170718	34741743	0,88%
36	zmones.15min.lt	16310	9528922	0,24%
		8438204	3941518292	100,00%

5. Kokie duomenys gali būti tvarkomi

Projektas nėra skirtas rinkti duomenis apie konkrečius asmenis. Tačiau viešai paskelbtuose tekstuose gali būti:

- vardai ir pavardės;
- autoriaus duomenys;
- pareigos, profesinė veikla ar kiti viešai paskelbti identifikaciniai duomenys;
- viešų publikacijų turinyje pateikta kita informacija apie fizinius asmenis;
- metaduomenys, tokie kaip šaltinis, publikavimo data, dokumento pavadinimas, nuoroda į šaltinį.

Tam tikrais atvejais viešuose tekstuose gali pasitaikyti ir specialių kategorijų asmens duomenų arba informacijos apie nusikalstamas veikas (kaip tai apibrėžia BDAR 9 ir 10 straipsniai), tačiau tokie duomenys nėra renkami tikslingai ir nėra atskirai klasifikuojami kaip savarankiškas projekto objektas.

6. Duomenų gavėjai

Prieiga prie nepublikuotų duomenų suteikiama tik tiems projekto dalyviams, darbuotojams, tyrėjams ir paslaugų teikėjams, kuriems tokia prieiga būtina pagal jų funkcijas.

Įgyvendinus projektą suformuotas tekstynas ir modelis bus perduotas Valstybės skaitmeninių išteklių agentūrai bei paskelbtas mažiausiai dviejose atviros prieigos platformose (pvz., „Hugging Face“, CLARIN ar kt.) ir Lietuvos atvirų duomenų portale (data.gov.lt) visiškai atveriant jame esančius duomenis. Kadangi Projekto rezultatai bus skelbiami atvirose duomenų platformose, paskelbti duomenys gali būti prieinami ir naudojami už Europos ekonominės erdvės ribų.

7. Duomenų saugojimo terminas

Duomenys pas Projekto vykdytojus pasibaigus projektui saugomi tiek, kiek būtina Projekto tikslams pasiekti, Projekto rezultatams validuoti, perduoti užsakovui, taip pat tiek, kiek to reikia teisinių reikalavimų vykdymui ar teisiniams reikalavimams pareikšti, vykdyti ar ginti.

Projekto rezultatus perdavus Valstybės skaitmeninių išteklių agentūrai ir juos paskelbus atviros prieigos platformose, jų prieinamumo trukmė yra neterminuota.

8. Duomenų subjektų teisės

Duomenų subjektai turi teisę, kiek tai neriboja techninės galimybės ir nesukuria neproporcingos naštos duomenų valdytojui:

- gauti informaciją apie duomenų tvarkymą;
- prašyti susipažinti su savo asmens duomenimis;
- prašyti ištaisyti netikslūs duomenis;
- prašyti ištrinti duomenis arba apriboti jų tvarkymą, kai tam yra BDAR nustatytos sąlygos;
- nesutikti su duomenų tvarkymu, kai tai taikoma pagal BDAR.

Atkreipiame dėmesį, kad kai kurios teisės gali būti ribojamos tiek, kiek tai leidžiama pagal BDAR tais atvejais, kai duomenys tvarkomi mokslinių tyrimų ar statistikos tikslais ir kai tokių teisių įgyvendinimas padarytų neįmanomą arba labai apsunkintų Projekto tikslų pasiekimą, tačiau tik taikant tinkamas apsaugos priemonės.

Jei nerimaujate dėl jūsų asmens duomenų tvarkymo, susisiekite su Andrius Utką andrius.utka@vdu.lt. Sprendimas dėl Jūsų kreipimosi bus priimtas ir apie tai būsite informuotas per 20 darbo dienų.

Jeigu manote, kad Jūsų asmens duomenis Vytauto Didžiojo universitete tvarko pažeisdamas duomenų apsaugos teisės aktų reikalavimus, rekomenduotina visuomet pirmiausia kreiptis tiesiogiai į mus šiais nurodytais kontaktais: telefonu: +370 37 788196 , el. paštu: dap@vdu.lt. Jeigu Jūsų netenkina Universiteto siūlomas problemos išsprendimo būdas arba manote, kad Universitetas nesiima pagal pateiktą prašymą būtinų veiksmų, turite teisę pateikti skundą priežiūros institucijai, kuri Lietuvos Respublikoje yra Valstybinė duomenų apsaugos inspekcija (L. Sapiegos g. 17, LT-10312 Vilnius; tel. (0 5) 212 7532; el. paštas: ada@ada.lt , internetinė svetainė: <https://vdai.lrv.lt/lt/>).