



M C M X X I I

VYTAUTO DIDŽIOJO
UNIVERSITETAS

**VYTAUTO DIDŽIOJO UNIVERSITETO
SENATAS**

NUTARIMAS

**DĖL PRITARIMO SKAITMENINIŲ IŠTEKLIŲ IR TARPDISCIPLININIŲ TYRIMŲ
INSTITUTO MTEP TEMATIKŲ PLANAMS**

2024 m. balandžio 10 d. Nr. SEN-N-19

Kaunas

Vadovaudamasis Vytauto Didžiojo universiteto Statuto 30.18 ir 30.2 punktais bei atsižvelgdamas į Senato Mokslo komiteto nuomonę Senatas **n u t a r i a**

pritarti Skaitmeninių išteklių ir tarpdisciplininių tyrimų instituto MTEP tematikų planams (pridedama).

Pirmininkas



Ričardas Krikštolaitis

**SKAITMENINIŲ IŠTEKLIŲ IR TARPDISCIPLININIŲ TYRIMŲ INSTITUTO (SITTI)
MOKSLINIŲ TYRIMŲ TEMATIKOS VEIKLOS PLANAS**

2024–2028 m. Skaitmeninių išteklių ir tarpdisciplininių tyrimų instituto (SITTI) biudžeto lėšomis finansuojamos mokslinių tyrimų tematikos „Kalbos vartosenos tyrimai, ištekliai, technologijos ir mašininis mokymasis kalbos technologijoms“ veiklos planas

Mokslinių tyrimų tematika	Kalbos vartosenos tyrimai, ištekliai, technologijos ir mašininis mokymasis kalbos technologijoms
Mokslinių tyrimų tematikos vykdymo laikotarpis	2024–2028
Mokslo sritis, kryptis	Humanitariniai mokslai (Filologija) Socialiniai mokslai (Teisė), Gamtos ir technikos mokslai (Informatika)
Tematikos aktualumas ir ištyrimo lygis	Tekstynų lingvistikos ir kompiuterinės lingvistikos tyrimai VDU pradėti nuo 1992 m. Nuo tų metų iki šiol rengiami tekstynai, išvestiniai jų produktai (pvz., dažniniai žodynai, dažniniai sąrašai, pastoviųjų žodžių junginių žodynai); kuriami arba pritaikomi automatinės kalbos analizės įrankiai (pvz., morfologiniai, sintaksiniai anotatoriai, žodžių sudedamąsias dalis nustatantys įrankiai). VDU SITTI mokslininkus galima laikyti tekstynų lingvistikos pradininkais Lietuvoje, turinčiais patirties ir mašininio, giliojo mokymo srityse. VDU Kompiuterinės lingvistikos centre, kurio pagrindu įsteigtas SITTI, parengtas pirmasis, iki šiol lingvistų plačiai naudojamas, tekstynų etalonu laikomas „Dabartinės lietuvių kalbos tekstynas“, keletas lygiagrečiųjų, palyginamųjų, specializuotų, anotuotų ir neanotuotų tekstynų. Didelė tekstynų dalis yra prieinami vartotojams per vartotojo sąsają, beveik visi ištekliai prieinami ne tik SITTI svetainėje, bet ir tarptautinės infrastruktūros CLARIN svetainėje su išsamiais aprašais, vartotojo teises nurodančiomis licencijomis. Sparčiai tobulėjant mašininio mokymosi technologijoms ir atsiradus giliajam mokymuisi, išaugo didžiulis lietuvių kalbos duomenų poreikis.
Tematikos tyrimų tikslas ir uždaviniai	Tikslas – rengti tekstynus ir jų išvestinius produktus, kalbos analizės įrankius, rengti išteklius dirbtinio intelekto tobulinimui ir jų pagrindu atlikti tyrimus. Uždaviniai: 1. Išsiaiškinti kitų kalbų tekstynų sandaros ir funkcionalumo aspektus. 2. Papildyti, pertvarkyti „Dabartinės lietuvių kalbos tekstyną“, pritaikyti kitų kalbų gerąją patirtį, technologines galimybes patogesnei paieškai.

	3. Remiantis tekstynų ir jų išvestinių produktų duomenimis, parengti mokslinę produkciją, tyrimų rezultatus pristatyti nacionalinėse ir tarptautinėse konferencijose, dalyvauti studijų procese, populiarinti SITTI tyrimus visuomenei. 4. Kaupti lietuvių kalbos išteklius naujųjų giliojo mokymosi technologijų naudojimui. 5. Kurti ir tobulinti lietuvių kalbos standartus, dalyvauti formuojant kalbos politiką, atvirojo mokslo judėjime.
Tematikos pridėtinė vertė (VDU strateginių tikslų įgyvendinimui, VDU prioritetinių mokslo krypčių vystymui, MTEP vertinamųjų vienetų, mokslo krypčių MTEP kokybei ir poveikiui, ir pan.)	Pridėtinė tematikos vertė atitinka Vytauto Didžiojo universiteto 2021–2027 m. Strateginio veiklos plano antrąją dalį. Be to, SITTI priklauso Atvirojo mokslo institucijų tinklui, nes rengia visiems atvira prieinamus kalbos išteklius.
Pagrindinių darbų atlikimo kalendorinis planas	1. 2024-2026 m. parengti 10 mln. žodžių morfologiškai ir sintaksiškai anotuotus tekstynus. 2. 2024-2026 m. sukaupti 3,5 mlrd. žodžių bendrąjį lietuvių k. tekstyną. 3. 2025-2026 m. pertvarkyti „Dabartinės lietuvių kalbos tekstyną“, parengti vartotojams patogesnę, funkcionalesnę vartotojo sąsają. 4. 2024-2028 m. koordinuoti Lietuvos narystę tarptautinėje mokslinių tyrimų infrastruktūros konsorciуме CLARIN-LT, administruoti saugyklos funkcionavimą. 5. Publikavimo gairės pateikiamos individualiai SITTI mokslo darbuotojų veiklos planuose.
Naudojama MTEP infrastruktūra	<u>Techninė infrastruktūra</u>: 8 fiziniai serveriai, kuriuose sukasi 11 virtualių mašinų. Taip pat administruojamos 3 svetainės (sitti.vdu.lt, clarin-lt.lt, clarin.vdu.lt). <u>Ištekliai</u>: SITTI infrastruktūroje prieinami: 13 tekstynų (9 – CLARIN saugykloje); 15 kalbos analizės įrankių (7), 4 duomenų rinkiniai (4), 5 duomenų bazės, 8 elektroniniai žodynai (1), 6 dažniniai sąrašai (5), 3 vektorizuoti žodžių modeliai (1), 2 standartai
Numatomų VDU biudžeto lėšų panaudojimo pagrindimas	Šioje temoje dirba konkursą praėję keturi vyresnieji mokslo darbuotojai (po 0,5 etato, vienas 0,25); du praėję atranką (0,25 ir 0,5 etato). Iš viso 5 etatai, tikėtina, kad prasidėjus projektiniam finansavimui turimų etatų dalys sumažės. „Dabartinės lietuvių kalbos tekstynui“ svarbus tęstinumas, t. y. atnaujinimas, dabarties aktualijas atspindinčių tekstų atranka, sutvarkymas, įkėlimas. Tam reikia kompiuterinių lingvistų, duomenų

	analitikų ir pan. darbuotojų. Taip pat svarbi nuolatinė tekstyno paieškos funkcionalumo plėtra, galimybė rengti ir viešinti išvestinius produktus. Tam reikalingos tęstinės VDU skiriamos biudžeto lėšos. Taip pat ir komandiruotėms, įrangai ir kitoms reikmėms. Mokslo darbuotojų etatai, skiriami bendra tvarka. 2024-2025 numatyta 5,19 įvairių mokslo darbuotojų etatų, iš jų 2 per atranką.
Papildomų šaltinių lėšų panaudojimas uždaviniams įgyvendinti	Morfologiškai ir sintaksiškai anotuotiems tekstynams rengti, Didelės apimties tekstynui sukaupti bus panaudotos Ekonomikos gaivinimo ir atsparumo didinimo priemonės (EGADP) subsidijos lėšos. Lietuvos mokslo taryba parėmė VDU drauge su CLARIN-LT konsorciumo narystę tarptautinėje mokslinių tyrimų infrastruktūroje CLARIN ERIC 2024-2028 metams. Ši MTI skirta kalbos išteklių ir jų analizės priemonių gamybai ir palaikymui.
Tematikos vadovas	Andrius Utka
Tematikos vykdytojai (konkurso būdu pareigas užėmę mokslo darbuotojai)	Erika Rimkutė, Jolanta Kovalevskaitė, Saulė Milčiuvienė, Andrius Utka
Mokslo darbuotojų MTEP tematikoje poreikio pagrindimas (atsižvelgiant į Institutui išskirtas papildomų mokslo darbuotojų etatus/išteklis)	Būtina atnaujinti turimus išteklius, rengti jų išvestinius produktus, sekti naujausias kitų šalių mokslinių tyrimų tendencijas ir jas pritaikyti lietuvių kalbos tyrimams tiek teoriniuose, tiek praktiniuose darbuose. Iki šiol VDU lingvistai yra pristatomi kaip žymiausi Lietuvos tekstynų lingvistikos atstovai. Norint išlaikyti šį statusą, reikalingi nuolatiniai tekstynų lingvistikos srityje dirbantys mokslininkai, tekstynų ir kompiuterinės lingvistikos specialistai, ypač programuotojai ir duomenų analitikai.
Vadovavimas jauniems mokslininkams, jaunųjų tyrėjų įtraukimas	SITTI mokslininkai vadovauja penkiems doktorantams, kurių tyrimai yra susiję su tekstynų lingvistika, lietuvių kalbos automatinės analizės įrankiais. Jie naudojami SITTI parengtais ištekliais, kad išanalizuotų vartosenos dėsningumus. Tokie darbai gali prisidėti prie tekstynų, t. y. dabartinę lietuvių kalbos vartoseną atspindinčiais, duomenimis pagrįstų leksikos ir gramatikos tyrimų. Doktorantai: Laima Jancaitė, Sandra Paškevičiūtė, Edgaras Dambrauskas, Miglė Žemrietė, Aušra Urbanavičiūtė.
Pagrindinės partnerinės institucijos užsienyje ir Lietuvoje	CLARIN žinių centras morfologiškai turtingų kalbų sistemoms ir priemonėms (SAFMORIL), Helsinkio universitetas, Suomija. Gdanskio universitetas (Computational Discourse Lab (CDS Lab)). Kauno technologijos universitetas, Mykolo Romerio universitetas, Vilniaus universitetas,

	Klaipėdos universitetas, Baltijos pažangių tyrimų institutas.	
Planuojama pateikti MTEP produkcija:	Tarpinė (2024–2025 m. laikotarpio) MTEP produkcija	MTEP produkcija baigtiniame etape
Tarptautinės mokslo monografijos ar jų dalys	1	3
Kitos mokslo monografijos ar jų dalys, mokslo studijos, sintezės mokslo darbai	1	2
Mokslo straipsniai Q1-Q2 kvartilių mokslo leidiniuose, referuojamuose CA WoS (priklausančių SCIE ir (ar) SSCI žurnalų rinkiniams) ir (ar) Scopus ¹ duomenų bazėse, turinčiuose cituojamumo rodiklį:	3	4
iš jų su bendraautoriais iš užsienio institucijų;	2	
iš jų tarpkryptiniai straipsniai;	Visi SITTİ straipsniai yra iš esmės tarpkryptiniai	
iš jų su bendraautoriais iš kitų Lietuvos institucijų, verslo įmonių;		
iš jų kaip pagrindinis autorius;		
Mokslo straipsniai prestižiniuose tarptautiniuose mokslo leidiniuose (pavyzdžiui, NATURE, SCIENCE, LANCET, CANCER ir pan.)		
Mokslo straipsniai Q3-Q4 kvartilių mokslo leidiniuose, referuojamuose CA WoS (priklausančių SCIE ir (ar) SSCI žurnalų rinkiniams) ir (ar) Scopus ¹ duomenų bazėse	1	3
Mokslo straipsniai mokslo leidiniuose, referuojamuose CA WoS (priklausančių SCIE ir (ar) SSCI žurnalų rinkiniams) ir (ar) Scopus ¹ duomenų bazėse, neturinčiuose cituojamumo rodiklio		
Mokslo straipsniai kituose referuojamuose mokslo leidiniuose	2	3
Mokslo straipsniai recenzuojamuose leidiniuose		
Mokslo populiarinimo straipsniai		5
Patentinės paraiškos ir patentai, tarptautinę ekspertizę praėjusių augalų veislės ar mokslui naujos organizmų rūšys		
Mokslo srities ar krypties terminų žodynai, moksliniai žinynai, enciklopedijos	1	1
Aukštosios mokyklos vadovėliai		
Moksliškai susisteminti atviros prieigos tyrimų duomenų rinkiniai	19 [CLARIN_LT]	21
Plenariniai pranešimai tarptautinėse mokslinėse konferencijose	3	3
Pranešimai tarptautinėse mokslinėse konferencijose	5	10
Kita MTEP sklaidos ir organizacinė veikla: internetinės svetainės; profesiniai seminarai; viešos paskaitos akademinėi bendruomenei ir plačiajai visuomenei	Tęsiamas trijų interneto svetainių ir Meta paskyros palaikymas ir plėtra: sitti.vdu.lt, clarin-lt.lt, clarin.vdu.lt/xmlui. 2 seminarai	Tęsiama svetainių priežiūra

	2 viešos paskaitos	3 seminarai 3 viešos paskaitos
Tarptautinių MTEP projektų paraiškų rengimas	3	5
Nacionalinių MTEP projektų (konkursinio MTEP finansavimo) paraiškų rengimas	2	3
MTEP sutarčių su ūkio subjektais vykdymas	2	2
Mokslinių konferencijų organizavimas		1 Tarptautinė Baltijos šalių kalbos technologijų konferencija (2027 HLT)
Paraiškų rengimas podoktorantūros stažuotėms, konkursinėms doktorantūros vietoms, studentų mokslinių tyrimų ir mokslinės praktikos konkursams	1	5
Planuojamos paraiškos verslo finansavimui pritraukti		
Kita veikla: dalyvauti ekspertinių mokslo institucijų veikloje; mokslinių tyrimų infrastruktūrose, asociacijose ir darbo grupėse;	10	10

MTEP tematikos veiklos planas apsvarstytas Skaitmeninių išteklių ir tarpdisciplininių tyrimų instituto tarybos susirinkime 2022 m. lapkričio mėn. 14 d., protokolo Nr. 1 (pagrindinės SITTİ mokslinių tyrimų temos ir veiklos kryptys).

Skaitmeninių išteklių ir tarpdisciplininių tyrimų instituto vadovė prof. habil. dr. Rūta Petrauskaitė

Dokumentą elektroniniu parašu
pasirašė RŪTA, PETRAUSKAITĖ
Data: 2024-03-29 13:26:43
Paskirtis: VDU
Vieta: Kaunas
Kontaktinė informacija:
ruta.petrauskaite@vdu.lt

SKAITMENINIŲ IŠTEKLIŲ IR TARPDISCIPLININIŲ TYRIMŲ INSTITUTO (SITTI) MOKSLINIŲ TYRIMŲ TEMATIKOS VEIKLOS PLANAS

2024–2028 m. Skaitmeninių išteklių ir tarpdisciplininių tyrimų instituto (SITTI) biudžeto lėšomis finansuojamos mokslinių tyrimų tematikos „Taikomieji psicho- ir sociolingvistiniai tyrimai“ veiklos planas

Mokslinių tyrimų tematika	Taikomieji psicho- ir sociolingvistiniai tyrimai
Mokslinių tyrimų tematikos vykdymo laikotarpis	2024-2028
Mokslo sritis, kryptis	Humanitariniai mokslai, filologija
Tematikos aktualumas ir ištyrimo lygis	Kalbos kaita daugiakalbėje ir daugiakultūroje aplinkoje, ypač jaunesnio amžiaus bei paauglių kalbinės nuostatos, tapatybės pokyčiai bei kalbos(-ų) įsisavinimo, mokymosi ir vartosenos polinkiai yra aktualūs klausimai, dominantys šių dienų įvairių sričių mokslininkus, politikus ir praktikus. Šios temos svarbios valstybei, institucijoms, pvz., švietimo sistemos sektoriui, bei šeimoms ir jų nariams. Akivaizdu, kad tokios globalios temos reikalauja įvairialypių ir nuoseklių tyrimų, o jų esama labai nedaug, ypač kalbinių klausimų perspektyvos požiūriu. Taigi, ištirtumo lygis būtų vertinamas kaip nepakankamas, todėl būtinas.
Tematikos tyrimų tikslas ir uždaviniai	Tikslai ir uždaviniai: įvertinti globalizacijos, daugiakultūriškumo, imigracijos ir emigracijos poveikį šiandieninei kalbos situacijai, kalbos kitimui ir kalbos raidai, tirti ir analizuoti klausimus, susijusius su kalbos ir švietimo politika; identifikuoti ir analizuoti su daugiakultūriškumu ir daugiakalbyste susijusias kalbos politikos, kalbinių nuostatų, įsisavinimo ir vartosenos problemas Lietuvoje ir tirti jas Europos kontekste, rengti eksperimentus, sociolingvistines apklausas, skirtas įvairių amžiaus ir socialinių grupių tyrimams. Plėtoti taikomuosius tyrimus ir kurti praktines bei metodines priemones, šaltinius (įvairaus pobūdžio elektroninius tekstynus, testus, įvairias kalbos paskatinimo užduotis, rekomendacijas ir kt.) sakytinės kalbos daugiakultūroje visuomenėje analizei.
Tematikos pridėtinė vertė (VDU strateginių tikslų įgyvendinimui, VDU prioritetinių mokslo krypčių vystymui, MTEP vertinamųjų vienetų, mokslo krypčių MTEP kokybei ir poveikiui, ir pan.)	VDU orientuojasi ir visą savo gyvavimo laikotarpį didelį dėmesį skyrė daugiakalbystei, kultūrinei įvairovei, todėl siūloma tematika visais požiūriais dera su strateginių tikslų įgyvendinimu. Kita sritis – mokslo kokybė ir tarptautiškumas. Jau įvykdytais bei vykdomais darbais, projektais ir publikacijomis bei pripažinimu šios tematikos mokslininkų indėlis ryškiai prisidėjo prie mokslo darbų kokybės gerinimo, dalyvaujant tarptautiškose tyrėjų grupėse, publikuojant straipsnius drauge su tarptautinėmis komandomis. Daugelis

	publikacijų yra su poveikio indeksais, priklausantys Q1-4.	
Pagrindinių darbų atlikimo kalendorinis planas	2024-2028	
Naudojama MTEP infrastruktūra	Techninė infrastruktūra: 8 fiziniai serveriai, kuriuose sukasi 11 virtualių mašinų. Ištekliai: SITTI infrastruktūroje prieinami: 13 tekstynų (9 – CLARIN saugykloje); 15 kalbos analizės įrankių (7), 4 duomenų rinkiniai (4), 5 duomenų bazės, 8 elektroniniai žodynai (1), 6 dažniniai sąrašai (5), tekstynai.	
Numatomų VDU biudžeto lėšų panaudojimo pagrindimas	Biudžetas reikalingas žmogiškiesiems ištekliams, komandiruotėms (konferencijos, stažuotės, konsultaciniai vizitai)	
Papildomų šaltinių lėšų panaudojimas uždaviniams įgyvendinti	Tyrimų projektų finansavimas. 2024 planuojama pateikti paraiškas keliems projektams (Nordic program cooperation, Šveicarijos-Lietuvos programa), kitos jau pateiktos. Pvz., Lenkijos vyriausybės programa NAWA, Lietuvos mokslo tarybai du projektai	
Tematikos vadovas	Prof. Ineta Dabašinskienė	
Tematikos vykdytojai (konkurso būdu pareigas užėmę mokslo darbuotojai)	Prof. Ineta Dabašinskienė (0,5); kiti įdarbinti pagal atranką 2 metams ir jų sutartis baigiasi 2024 vasarą.	
Mokslo darbuotojų MTEP tematikoje poreikio pagrindimas (atsižvelgiant į Institutui išskirtas papildomų mokslo darbuotojų etatus/išteklis)	Aukščiau suminėti tikslai, uždaviniai ir jau šios grupės narių įvertintas indėlis moksle, tikėtina, pakankamas pagrindas tikėtis šiai grupei iš viso 3 etatų. 0,5 vyriausiojo jau yra darbuotojas praėjęs konkursą, tikimės dar 2 etatų (vyresniojo ir vyriausiojo) bei 0,5 arba 1,0 pilno etato mokslininko asistento ar pan.	
Vadovavimas jaunesiems mokslininkams, jaunųjų tyrėjų įtraukimas	Grupėje dalyvauja 2 doktorantai, magistrai, Academia cum laude studentės.	
Pagrindinės partnerinės institucijos užsienyje ir Lietuvoje	Vienos universitetas, Leibnico tyrimų centras ZAS institutas Berlyne, Varšuvos universitetas, Tromsės (Arktikos) universitetas, CUNY Niujorke ir daug kitų.	
Planuojama pateikti MTEP produkcija:	Tarpinė (2024–2025 m. laikotarpio) MTEP produkcija	MTEP produkcija baigtiniame etape
Tarptautinės mokslo monografijos ar jų dalys		
Kitos mokslo monografijos ar jų dalys, mokslo studijos, sintezės mokslo darbai	1 mokslo monografija	2-3 mokslo monografijos
Mokslo straipsniai Q1-Q2 kvartilių mokslo leidiniuose, referuojamuose CA WoS (priklausančių SCIE ir (ar) SSCI žurnalų rinkiniams) ir (ar) Scopus ¹ duomenų bazėse, turinčiuose cituojamumo rodiklį:	2	5
iš jų su bendraautorais iš užsienio institucijų;	2	5
iš jų tarpkryptiniai straipsniai;	2	4
iš jų su bendraautorais iš kitų Lietuvos institucijų, verslo įmonių;		
iš jų kaip pagrindinis autorius;	1	3

Mokslo straipsniai prestižiniuose tarptautiniuose mokslo leidiniuose (pavyzdžiui, NATURE, SCIENCE, LANCET, CANCER ir pan.)		1
Mokslo straipsniai Q3-Q4 kvartilių mokslo leidiniuose, referuojamuose CA WoS (priklausančių SCIE ir (ar) SSCI žurnalų rinkiniams) ir (ar) Scopus' duomenų bazėse	4	12
Mokslo straipsniai mokslo leidiniuose, referuojamuose CA WoS (priklausančių SCIE ir (ar) SSCI žurnalų rinkiniams) ir (ar) Scopus' duomenų bazėse, neturinčiuose cituojamumo rodiklio	3	6
Mokslo straipsniai kituose referuojamuose mokslo leidiniuose	2	4
Mokslo straipsniai recenzuojamuose leidiniuose	2	
Mokslo populiarinimo straipsniai	1	6
Patentinės paraiškos ir patentai, tarptautinę ekspertizę praėjusių augalų veislės ar mokslui naujos organizmų rūšys		
Mokslo srities ar krypties terminų žodynai, moksliniai žinyai, enciklopedijos		
Aukštosios mokyklos vadovėliai		
Moksliškai susisteminti atviros prieigos tyrimų duomenų rinkiniai		1
Plenariniai pranešimai tarptautinėse mokslinėse konferencijose	1	1-2
Kitos tarptautinės konferencijos	20	70
Kita MTEP sklaidos ir organizacinė veikla:		
Tarptautinių MTEP projektų paraiškų rengimas	3	7
Nacionalinių MTEP projektų (konkursinio MTEP finansavimo) paraiškų rengimas	2	6
MTEP sutarčių su ūkio subjektais vykdymas		3
Mokslinių konferencijų organizavimas		3
Paraiškų rengimas podoktorantūros stažuotėms, konkursinėms doktorantūros vietoms, studentų mokslinių tyrimų ir mokslinės praktikos konkursams	2	8
Planuojamos paraiškos verslo finansavimui pritraukti		
Kita veikla ...		

MTEP tematikos veiklos planas apsvaistytas Skaitmeninių išteklių ir tarpdisciplininių tyrimų instituto tarybos susirinkime 2022 m. lapkričio mėn. 14 d., protokolo Nr. 1 (pagrindinės SITT mokslinių tyrimų temos ir veiklos kryptys).

Skaitmeninių išteklių ir tarpdisciplininių tyrimų instituto vadovė prof. habil. dr. Rūta Petrauskaitė

Dokumentą elektroniniu parašu
pasirašė RŪTA, PETRAUSKAITĖ
Data: 2024-03-29 13:30:12
Paskirtis: VDU
Vieta: Kaunas
Kontaktinė informacija:
ruta.petrauskaite@vdu.lt