

LIETUVOS VALSTYBINIS MOKSLO IR STUDIJŲ FONDAS
Prašymas skirti paramą moksliniams tyrimams

Registracijos Nr.
(pildo Fondas)

Registracijos data
(pildo Fondas)

Mokslų sritis:

Darbo vadovas (mokslininkų grupės vadovas) (pedagoginis vardas, mokslo laipsnis, vardas, pavardė):

Darbo vadovo registracijos anketos numeris:

PRAŠYMO TURINYS

Prašu skirti fondo paramą mokslinio tyrimo darbui:

1. Mokslinio tyrimo darbo pavadinimas:

2. Mokslų šakų kodai pagal Europos Sąjungos "Mokslų klasifikaciją":

3. Darbą baigsime ir ataskaitą Fondui įteiksime (metai ir mėnuo - skaitmenimis)

4. Darbą vykdys šie mokslininkų grupės nariai:

Eil. Nr.	Pedagoginis vardas ir mokslo laipsnis	Vardas ir pavardė	Darbovietė	Pastabos
1.	dr.	Gailius Raškinis	VDU IF	
2.	dr.	Asta Kazlauskienė	VDU HMF	
4.	dokt.	Giedrius Norkevičius	VDU IF	

5. Darbo aprašymas:

5.1. Tikslai, kurių siekia mokslininkų grupė šiuo darbu. Į kokius svarbius mokslui ar mokslo taikomajai veiklai klausimus norite atsakyti įvykdę planuojamus tyrimus.

Tikslas: Paramos prašoma mokslo tyrimams, kurių tikslas – sukurti metodus, automatiškai kirčiuojančius bendrinės lietuvių kalbos tekstą.

Aktualumas: Sukurti metodai bus įdiegti Internete, o automatinio kirčiavimo galimybės taps viešai prieinamos Lietuvos ir pasaulio vartotojams. Jomis galės naudotis radijo ir televizijos laidų vedėjai, diktoriai, aktoriai, mokiniai, studentai, užsieniečiai ir visi kiti, besimokantys taisyklingo lietuvių kalbos kirčiavimo.

Kuriami automatinio kirčiavimo metodai yra svarbūs ir kalbos technologijų tyrimams, tokiems kaip šnekos atpažinimas ir sintezė.

Šis projektas tiesiogiai susijęs su lituanistikos tyrinėjimais, kurie yra nuolatinių Fondo prioritetų sąrašė.

5.2. Kuo šis darbas ir jame atliekami tyrimai skiriasi nuo kitų, jau padarytų arba šiuo metu kitų mokslininkų vykdomų analogiškų tyrimų. Apibūdinkite, kas šia tema padaryta Lietuvoje ir pasaulyje.

Fiksuoto kirčio kalbose (latvių, čekų, slovakų, islandų, estų, suomių, vengrų, prancūzų, armėnų, lenkų ir kt.) tekstų kirčiavimo problema nėra aktuali. Ne toks problemiškas kirčiavimas ir kai kuriose laisvojo kirčio kalbose, kur kirtis žodžio paradigmoje nekeičia vietos (anglų, vokiečių) arba vienodas galūnes turintys žodžiai paprastai kirčiuojami vienodai (italų).

Lietuvių kalbos kirtis yra laisvas, t.y. nesusijęs su konkrečiu, visada tuo pačiu skiemeniu. Lietuvių kalbos kirčiavimą sunkina ir tai, kad, skirtingai nei kitos laisvojo kirčio kalbos, lietuvių kalba turi priegaides: skirtingą ilgųjų kirčiuotų skiemenų tarimą, nuo kurio dažnai ir priklauso kirčio vieta paradigmoje.

Lietuvių kalbos kirčio ir priegaidės kitimo dėsningumai buvo išsamiai nagrinėti lietuvių kalbos mokslo tyrimo darbuose. Lietuvių kalbos gramatika pateikia daug detalių kirčiavimo taisyklių ir jų taikymo išimčių. Tačiau, gramatikoje pateikiamos kirčiavimo taisyklės negali būti tiesiogiai panaudotos automatinio kirčiavimo metodams kurti dėl dviejų pagrindinių priežasčių:

a) Gramatikose aprašytos kirčiavimo taisyklės skirtos žmonėms turintiems “kalbos jausmą” (pvz., daiktavardžio kirčiuotė nustatoma pagal tai, kurioje žodžio dalyje – kamiene ar galūnėje – kirčiuojamos daugiskaitos naudininko ir galininko formos. Daroma prielaida, kad kirčiuojantysis (žmogus) pajėgia tai nustatyti). Kad galėtų panaudoti kirčiavimo taisykles, automatinio kirčiavimo metodams **reikalingos papildomos žinios**.

b) Gramatikose pateikiamų taisyklių taikymo sritys persidengia, o jų tarpusavio taikymo tvarka visais galimais atvejais nėra aiški (pvz. priešdėliniams, sangražiniams ar priesaginiams veiksmažodžiams kirčiuoti nurodomos atskiros taisyklės, tačiau kuriai taisyklei teikti prioritetą, norint sukirčiuoti veiksmažodį “persigalvoti”, kuris yra ir priešdėlinis, ir sangražinis, ir priesaginis vienu metu, nuorodų nerandame). Automatinio kirčiavimo metodams **reikalingas vienareikšmiškumas**, t.y. žinios apie kirčiavimo taisyklių taikymo tvarką.

Automatinio kirčiavimo metodams iš dalies skirta dr. P. Kasparaičio disertacija, kuri šalia projekto teikėjų darbų yra vienintelis tokio pobūdžio tyrimas, skirtas lietuvių kalbai. Ten pasiūlytas automatinio daiktavardžių ir būdvardžių kirčiavimo taisyklių sudarymo algoritmas, taikomas siauresnėje srityje – šnekos sintezėje.

Metodų, kurie automatiškai kirčiuotų ištisus bendrinės lietuvių kalbos tekstus ir kuriais galėtų naudotis visi, gebantys surinkti tekstą kompiuteriu ir turintys galimybę naudotis internetu, nėra.

5.3. Kokių mokslinių rezultatų tyrimų tema jau turi mokslininkų grupė? Ką naujo norite sužinoti ar išaiškinti vykdydami siūlomus tyrimus?

Projekto teikėjų grupė yra:

1. Surinkusi daug kompiuterinių žinių, kurios bus panaudotos, kuriant automatinio teksto kirčiavimo metodus (išanalizuoti pirminiai ir mišrieji veiksmažodžiai, dviskiemeniai vardažodžiai, akcentinės priešdėlių, priesagų ypatybės, priegaidžių kaitos svarbesnieji atvejai, kirčiavimo išimtis ir gretimybės, kirčio sąsaja su morfologinėmis žodžio savybėmis, kirčio semantinės ypatybės);
2. Sukūrusi teksto skiemenavimo metodus ir atskirų žodžių kirčiavimo algoritmo prototipą;

5.4. Išdėstykite, kaip planuojate vykdyti tyrimus? Kokius metodus, aparatūrą naudosite, kokia numatoma darbų seka, kokių rezultatų laukiate baigę darbą? Jei bus bendradarbiaujama su užsienio mokslininkais, kokius tyrimus jie atliks ir kokius duomenis pateiks?

Tyrimų vykdymo planas:

- I. Papildomų kompiuterinių žinių ir kalbos resursų, reikalingų automatinio kirčiavimo metodams, surinkimas (atsakinga A. Kazlauskienė):
- 1.1. Papildomų žinių apie bendrinės lietuvių kalbos žodžių kirčiavimą surinkimas (antraštinė žodžio forma, kalbos dalis, kirčiuotė, kirčio vieta ir priegaidė).
 - 1.2. Žinių apie vietovardžių ir vardų kirčiavimą surinkimas.
 - 1.3. Papildomų žinių apie priesaginių daiktavardžių ir būdvardžių kirčiavimą kompiuterinis pateikimas.
 - 1.4. Papildomų žinių apie pirminių ir mišriųjų veiksmažodžių tris pagrindines formas (kirčiuotas) surinkimas ir kompiuterinis pateikimas.
 - 1.5. Žinių apie sudėtingesnius skiemenavimo atvejus (dviejų balsių sandūra) surinkimas.
- II. Gramatikose pateikiamų kirčiavimo taisyklių struktūrinė analizė ir pritaikymas automatinio kirčiavimo metodams (atsakingi G. Raškinis, G. Norkevičius):
- 2.1 Gramatikose pateikiamų veiksmažodžių ir dalyvių kirčiavimo taisyklių prioritetų analizė (atsakinga A. Kazlauskienė).
 - 2.2 Daiktavardžių, būdvardžių, veiksmažodžių, dalyvių ir kitų kalbos dalių (skaitvardžio, įvardžio,rieveiksmio) automatinio kirčiavimo metodų konstravimas ir korekcija.
- III. Kirčiavimo algoritmų kompiuterinė realizacija ir pateikimas.
- 3.1. Kirčiavimo metodų kompiuterinė realizacija (atsakingi G. Raškinis, G. Norkevičius).
 - 3.2 Kirčiavimo metodų veikimo tikslumo įvertinimas (atsakinga A. Kazlauskienė).
 - 3.3 Kirčiavimo technologijos pateikimas internete (atsakingas G. Norkevičius).
- IV. Mokslinės publikacijos anglų k. tarptautiniame recenzuojamame leidinyje rengimas, baigiamosios ataskaitos Fondui rengimas (atsakingi G. Raškinis, A. Kazlauskienė).

Laukiami rezultatai

Moksliniai:

1. Bus ištirta dabartinės lietuvių bendrinės kalbos kirčiavimo taisyklių visuma, išryškintos persidengiančios šių taisyklių taikymo sritys ir nustatyti taisyklių taikymo prioritetai.
2. Bus sukurta duomenų bazė, kurioje bus surinktos ir kompiuterine kalba užrašytos automatiniam kirčiavimui reikalingos papildomos žinios.
3. Bus sudarytas ir programiškai realizuotas visas lietuvių kalbos dalis apimantis automatinis kirčiavimo algoritmas, taisyklingai kirčiuojantis ne mažiau kaip 95% lietuvių kalbos teksto žodžių.
4. Bus ištirtas kirčiavimo algoritmo veikimo tikslumas ir tikslumo priklausomybė nuo svarbiausių algoritmo parametrų.

Praktiniai: Darbo metu sukurti automatinio kirčiavimo metodai bus pateikti interneto serveryje. Tai užtikrins, kad automatinio kirčiavimo galimybės bus viešai prieinamos visiems interneto vartotojams. Automatinio kirčiavimo priemonės kirčiuos tik bendrinės kalbos žodžius. Labai retai vartojami tarptautiniai žodžiai, asmenvardžiai, tarminiai ar šnekamosios buitinės kalbos žodžiai bus kirčiuojami iš dalies. Kol kas nebus kirčiuojami homografi (vienodos rašybos, bet skirtingą kirčio vietą ar priegaidę turintys žodžiai). Nustatant žodžių kirčiavimo tipą bus remiamasi „Dabartinės lietuvių kalbos žodyno“ IV leidimu (2000). Jeigu žodis gali būti kirčiuojamas dvejopai, programa sukirčiuos pagal pirmąjį minėtame žodyne pateiktą kirčiavimo modelį.

5.5. Ką dar, be privalomos mokslinės ataskaitos, pateiksite Fondui baigę darbą (kiek ir kokių mokslinių publikacijų numatote parengti, maketus, technologijas, patentų paraiškas ir panašiai)?

1. Darbo pagrindu numatome parengti mokslinę publikaciją anglų kalba leidinyje "Informatica".
2. Darbo metu sukurtus automatinio kirčiavimo metodus numatome pateikti interneto serveryje, užtikrinant viešą priėjimą prie automatinių lietuvių kalbos teksto kirčiavimo galimybių.

6. Nurodykite mokslinio tyrimo darbus, kuriuose mokslininkų grupės nariai vykdys kartu su šiuo darbu ir jų finansavimo šaltinius. Kuo šis darbas skiriasi nuo kitų tuo pat metu vykdomų darbų ir kokią šio darbo dalį numatote vykdyti kitų finansavimo šaltinių lėšomis?

Vytauto Didžiojo universitetas žada paremti teikiamą projektą 2500 Lt techninei įrangai įsigyti. 2005 m. A. Kazlauskienė ir G. Raškinis gali dalyvauti projekte „Automatinis lietuvių šnekamosios kalbos segmentavimas ir lietuvių šnekos automatinė transkripcija“, jei šį projektą finansuos Valstybinė lietuvių kalbos komisija. Pastarasis projektas susijęs su lietuvių šnekos garso įrašų analize ir nesusijęs su automatiniu kirčiavimo metodų kūrimu.

7. Pateikite svarbiausiųjų mokslinių publikacijų, kuriuos per praėjusius penkerius metus paskelbė visi autorių grupės nariai (straipsnio autoriai, pavadinimas, kur paskelbta, leidinio duomenys, išleidimo metai, puslapiai leidinyje)

Savickienė, I., A. Kazlauskienė ir L. Kamandulytė. Naujas požiūris į lietuvių kalbos daiktavardžio linksniavimo tipus pagal natūraliosios morfologijos teoriją // Acta Linguistica Lituanica. ISSN 1648-4444. Vilnius: Lietuvių kalbos institutas, 2004, T.50, p. 79–98.

Šilingas, D., G. Raškinis ir L. Telksnys. Review of Lithuanian Speech and Language Processing // Proceedings of the 1st Baltic Conference “Human Language Technologies”, the Baltic Perspective, Riga University, Riga, Latvia, p. 144-149.

Vaičiūnas, A., G. Raškinis ir V. Kaminskas. Statistical Language Models of Lithuanian based on word clustering and morphological decomposition // priimtas spausdinti Informatica leidinyje.

Pajarskaitė, G, V. Gričiūtė, G. Raškinis ir J. Kuper. Designing HMM-based Part-of-Speech Tagger for Lithuanian Language // Informatica. Vilnius: Matematikos ir Informatikos institutas, 2004, 15(2), p. 231-242.

Kazlauskienė, A., K. Veličkaitė. Pastabos dėl lietuvių kalbėjimo tempo // Acta Linguistica Lituanica. ISSN 1648-4444. Vilnius: Lietuvių kalbos instituto leidykla, 2003, T.48, p. 49–58.

Kazlauskienė, A., Dūrinių kirčiavimo gretimybės studentų kalboje // Kalbos kultūra. ISSN 0130-2795. Vilnius: Lietuvių kalbos instituto leidykla, 2003, nr. 76, p. 108–113.

Raškinis A., G. Raškinis ir A. Kazlauskienė. SAMPA (Speech Assessment Methods Phonetic Alphabet) for Encoding Transcriptions of Lithuanian Speech Corpora // Information technology and control. ISSN 1392-124X. Kaunas: Technologija, 2003, nr. 4(29), p. 52–56.

Raškinis, G., D. Raškinienė. Development of Medium-Vocabulary Isolated-Word Lithuanian HMM Speech Recognition System // Informatica. Vilnius: Matematikos ir Informatikos institutas, 2003, 14(1), P. 75-84.

Šilingas, D., G. Raškinis ir L. Telksnys. Speech and Language Processing for Lithuanian: a Review // Speech Processing Workshop, German Association for Pattern recognition (DAGM). Otto-von-Guericke University, Magdeburg, Germany, 2003, p. 57-64.

Kazlauskienė, A. Priesaginių veiksmažodžių kirčiavimo variantai studentų kalboje // Kalbos kultūra. ISSN 0130-2795. Vilnius: Lietuvių kalbos instituto leidykla, 2002, nr. 75, p. 131–137.

Kazlauskienė, A. Pietinių vakarų aukštaičių tarmės balsių spektras ir prozodija // Lituania. ISSN 0235-716X. Vilnius: Lietuvos mokslų akademijos leidykla, 2001, nr.3 (47), p. 39–45.

Kazlauskienė, A. Šalutiniai kirčiai vakarų aukštaičių kauniškių tarmėje // Kalbotyra. ISSN 1392-1517. Vilnius: Vilniaus universiteto Bendrosios kalbotyros katedra, 2001, T. 50 (1), p. 53 – 66.

Kazlauskienė, A. Gretimų garsų įtaka balsių kiekybei pietinių vakarų aukštaičių tarmėje // Kalbotyra. – Vilnius, 2000.– T. 48 (1)–49 (1).– P. 63–70.

Kazlauskienė, A. Daiktavardžio kirčiuočių dažnumas // Darbai ir dienos. – Kaunas, 2000. – T. 24.– 83–88.

8. Nurodykite, kokią Fondo paramą yra gavę autorių grupės nariai (darbų vadovai, darbų pavadinimai, sutarčių su Fondu arba registracijos Nr., paramos suma). Nurodykite, kokias mokslines publikacijas paskelbėte ar kitą mokslinę produkciją sukūrėte, vykdydami su Fondo parama anksčiau atliktus darbus.

G.Raškinis buvo dalyvis programos, kuri remiama Fondu, kaip mokslo tyrimas, vykdomas pagal tarpvalstybinius susitarimus (2001 03 01-2001 12 30, sutarties nr. V-025) Europos bendrijos INCO programa TELRI-II "Bendraeuropinė kalbų resursų infrastruktūra" "Elektroninis čekų-lietuvių, lietuvių-čekų kalbų žodynas", pareiškėjas – dr. R.Marcinkevičienė, kontraktas nr. PL977085. Projektui suteikta parama iš viso sudaro 28000 Lt.

Fondas suteikė paramą G.Raškiniui tyrimo darbui „Liaudies dainų ritmo atpažinimo metodai ir jų tyrimas“ (2002 m, autorinė sutartis nr. 536). Projektui suteikta parama iš viso sudaro 8000 Lt.

Mokslinė produkcija (ritmo atpažinimo metodai), gauti, remiant Fondui, įdiegta liaudies dainų užrašymo natomis sistemoje, kuri patalpinta interneto svetainėje (www.recognisoft.com). Ši sistema prieinama visiems Interneto vartotojams ir nemokama Lietuvos vartotojams tyrimo ir mokymo tikslams.

Sukurti metodai taip pat pritaikyti gretutinėms tyrimo sritims - Informatica leidinyje paskelbta publikacija:

Pajarskaitė, G, V. Griciūtė, G. Raškinis and J. Kuper. Designing HMM-based Part-of-Speech Tagger for Lithuanian Language.

9. Šiam mokslinio tyrimo darbui prašome skirti **Lt**
(nurodyti apvalinant iki šimtų Lt)

Ši suma bus panaudota tokioms išlaidoms (žr. "Prašymų teikimo, jų nagrinėjimo, paramos skyrimo ir ataskaitų pateikimo reglamento" 43 punktą):

9.1. Darbui ir socialiniam draudimui apmokėti (43.1 punktas) **Lt**
(paaiškinimai ir skaičiavimas - 1 priedas)

9.2. Medžiagoms ir kitoms prekėms įsigyti (43.2. punktas) **Lt**
(sąrašas, pagrindimas ir kainos - 1 priedas)

9.3. Būtinoms ekspedicijoms ir išvykoms (43.3 punktas) **Lt**

9.4. Sutarties aptarnavimo ir institucijos kitos išlaidos (43.4) **Lt**
(ne daugiau kaip 10% išlaidų, nurodytų 9.1, 9.2 ir 9.3 punktuose)

Priedai:

- Išlaidų darbui ir socialiniam draudimui apmokėti paaiškinimai ir skaičiavimas
- Trumpa darbo anotacija lietuvių ir anglų kalbomis (pateikiama, jei autoriai pageidauja, kad apie vykdomus tyrimus būtų paskelbta Fondo INTERNET svetainėje)

Darbo vadovas

..... / dr. Gailius Raškinis
(parašas) (pavardė)

Mokslininkų grupės nariai:

..... / dr. Asta Kazlauskienė
(parašas) (pavardė)

..... / Giedrius Norkevičius
(parašas) (pavardė)

Užpildymo data