

DR. ANDRIUS UTKA

CURRICULUM VITAE

ASMENS DUOMENYS

Gimimo data

Namų adresas

Darbo adresas Vytauto Didžiojo universiteto HMF
Kompiuterinės lingvistikos centras

El. paštas

DUOMENYS APIE IŠSILAVINIMĄ

- 2004 m. Humanitarinių mokslų srities Filologijos krypties daktaro laipsnis (Vytauto Didžiojo universitetas). Disertacijos tema: „Statistinis tekstų funkcijų nustatymas“.
- 2003 m. Birmingemo universiteto (Jungtinė Karalystė) Tekstynų lingvistikos magistro laipsnis (MPhil).
- 1999 m. Vytauto Didžiojo universiteto Anglų kalbos ir literatūros magistro laipsnis.
- 1997 m. Vytauto Didžiojo universiteto Anglų kalbos ir literatūros bakalauro laipsnis.
- 1990 m. Kauno m., T. Vaižganto vidurinė mokykla.

DARBO PATIRTIS

- Nuo 2009 iki dabar Vytauto Didžiojo universiteto Lietuvių kalbos katedros docentas.
- Nuo 2005 – iki dabar Vytauto Didžiojo universiteto Kompiuterinės lingvistikos centro mokslo darbuotojas (nuo 2010 centro vadovas)
- Nuo 2005 – 2009 Vytauto Didžiojo universiteto Lietuvių kalbos katedros lektorius.
- 2000-2003 (su pertraukomis) Birmingemo universitete (JK) mokslinis bendradarbis.
- 1998-2004 (su pertraukomis) Vytauto Didžiojo universiteto Kompiuterinės lingvistikos centro, kompiuterinis lingvistas

KALBOS

Lietuvių – gimtoji; rusų, anglų – gerai; vokiečių – vidutiniškai.

PEDAGOGINĖ VEIKLA

VDU Lietuvių kalbos katedroje dėstomi dalykai:

- magistro studijų studentams - Automatizuotas vertimas ir redagavimas, Tekstynų lingvistikos teorija ir praktika (kartu su R. Petrauskaite), Natūraliosios kalbos apdorojimo metodai, Kalbos technologijos;
- bakalauro studijų studentams – Kompiuterizuotas vertimas.

EKSPERTINĖ PATIRTIS

- 1) Nuo 2013 m. atstovauja VDU Lietuvos standartizacijos departamento (LST) Technikos LST TK 4 komitete.
- 2) Nuo 2012 m. MITA ekspertas.
- 3) Nuo 2011 m. VLKK Kalbos technologijų pakomisės narys
- 4) Nuo 2011 m. LMT ekspertas
- 5) „Eureka“ programos ekspertas
- 6) COST programos ekspertų duomenų bazėje

PROJEKTINĖ VEIKLA

(pagrindiniai projektai per pastaruosius 5 metus)

1) 2015-2016 „Lietuvos narystė tarptautinėje mokslinių tyrimų infrastruktūroje - Bendrosios kalbos išteklių ir technologijų infrastruktūra Europos mokslinių tyrimų infrastruktūros konsorciumas (CLARIN ERIC)“, vadovas.

2) 2014-2015 Nacionalinės lituanistikos plėtros 2009–2015 metų programos projektas: „Automatiniai autorių ir autorių grupių individualaus stiliaus nustatymo tyrimai (ASTRA)“, vad. J. Kapočiūtė-Dzikienė.

3) 2012-2014 „Lietuvių kalbos sintaksinės-semantinės analizės sistema tekstynui, lietuviškam internetui ir viešojo sektoriaus taikymams“ (VP2-3.1-IVPK-12-K-01-007), kuris yra finansuojamas pagal ES Ekonomikos augimo veiksmų programos 3 prioriteto *Informacinės visuomenės visiems* įgyvendinimo priemonę *Lietuvių kalba informacinėje visuomenėje*, vadovas (Nr. VP2-3.1-IVPK-12-K).

4) 2011-2012 ES tarpvalstybinės Lietuvos-Latvijos bendradarbiavimo programos (en. *Cross-border Cooperation programme*) projekto "Humanitarinių mokslų švietimo infrastruktūra Rytų Latvijoje ir Lietuvoje (Kaunas)" partnerinės dalies koordinatorius.

5) 2010 Nacionalinės programos „Valstybė ir tauta: paveldas ir tapatumas“ projektas „Lituanistinių skaitmeninių išteklių metaduomenų sukūrimas ir suderinimas su CLARIN“ (LIMS), vad. V. Fomin. *Projektas skirtas prisijungimui prie tarptautinės kalbos išteklių infrastruktūros CLARIN*.

6) 2010-2012 Nacionalinės lituanistikos plėtros 2009–2015 metų programos projektas: „Švietimo ir mokslo terminų automatinis identifikavimas (ŠIMTAI2)“, vad. E. Rimkutė. *Projekto tikslas sukurti metodiką automatiniam terminų identifikavimui ir ją panaudoti švietimo ir mokslo terminams identifikuoti*.

7) 2010-2011 Nacionalinės lituanistikos plėtros 2009–2015 metų programos projektas: „Dabartinės lietuvių kalbos priesaginių veiksmažodžių daryba (I dalis)“, vad. J. Pakerys. *Projektas skirtas priesaginių veiksmažodžių darybos tyrimui*.

8) 2009-2012 ES struktūrinių fondų MTTP projektas „Semantinis informacijos valdymo variklis“ (SIVV) (priemonė INTELEKTAS LT). *Projektas skirtas semantinės infrastruktūros kūrimui ir semantinei kalbos duomenų analizei*.

9) 2009-2012 ES struktūrinių fondų projektas „Kalbos, semantinių, verslo modeliavimo technologijų inovacijų žinių teminis tinklas verslo aplinkos gerinimui“ (priemonė INOGEB LT 1).

2010–2016 M. MOKSLINĖS PUBLIKACIJOS

Moksliniai straipsniai

1. Rimkutė, Erika, Kazlauskienė, Asta, Utkā, Andrius. 2016. Morphemic structure of Lithuanian words. *Open linguistics* 2-1, 160-179.
2. Utkā, Andrius, Amilevičius, Darius. 2016. Normalisation of Lithuanian Social Media Texts: Towards Morphological Analysis of User-Generated Comments. *LREC 2016: Proceedings of the Workshop on Normalisation and Analysis of Social Media Texts (NormSoMe)*, 28th May 2016, Portorož, Slovenia / editors Andrius Utkā, Jurgita Vaičenonienė, Rita Butkienė p. 39-44.

3. Utkā, Andrius, Amilevičius, Darius, Krilavičius, Tomas, Vitkutė-Adžgauskienė, Daiva. 2016. Overview of the Development of Language Resources and Technologies in Lithuania (2012–2015). Human language technologies - the Baltic Perspective: Proceedings of the 7th International Conference, Baltic HLT 2016, Riga / editors Inguna Skadiņa, Roberts Rozis, 12-19.
4. Vitkutė-Adžgauskienė, Daiva, Utkā, Andrius, Amilevičius, Darius, Krilavičius, Tomas. 2016. NLP Infrastructure for the Lithuanian Language. LREC 2016: 10th International Conference on Language Resources and Evaluation, 23-28 May, 2016, Portorož, Slovenia: proceedings, 2539-2542.
5. Bielinskienė, Agnė, Boizou, Loic, Grigonytė, Gintarė, Kovalevskaitė, Jolanta, Rimkutė, Erika, Utkā, Andrius. 2015 Lietuvių kalbos terminų automatinis atpažinimas ir apibrėžimas. Kaunas: Vytauto Didžiojo universitetas.
6. Kapočiūtė-Dzikienė, Jurgita, Utkā, Andrius, Šarkutė, Ligita. 2015. Authorship attribution and author profiling of Lithuanian literary texts. // RANLP 2015: 10th international conference on recent advances in natural language processing, BSNLP 2015: 5th workshop on Balto-Slavic natural language processing, 10–11 September 2015, Hissar, Bulgaria: proceedings. Shoumen, Bulgaria : INCOMA Ltd., 96-105.
7. Kapočiūtė-Dzikienė, Jurgita, Utkā, Andrius, Šarkutė, Ligita. 2015. Authorship attribution of internet comments with thousand candidate authors. ICIST 2015: Information and software technologies: 21st international conference, Druskininkai, Lithuania, October 15-16 2015: Proceedings / editors Dregvaitė, Giedrė, Damaševičius, Robertas, 433-448.
8. Kapočiūtė-Dzikienė, Jurgita, Šarkutė, Ligita, Utkā, Andrius. 2015. The Effect of author set size in authorship attribution for Lithuanian. NODALIDA 2015: Proceedings of the 20th Nordic Conference of Computational Linguistics, May 11–13, 2015, Institute of the Lithuanian language, Vilnius / editor Beata Megyesi. 87-96.
9. Kapočiūtė-Dzikienė, Jurgita, Utkā, Andrius, Šarkutė, Ligita. 2014. Feature exploration for authorship attribution of Lithuanian parliamentary speeches. Text, speech and dialogue : 17th international conference, TSD 2014, Brno, Czech Republic, September 8-12, 2014 : proceedings, 93-100.
10. Kapočiūtė-Dzikienė, Jurgita, Šarkutė, Ligita, Utkā, Andrius. 2014. Automatic author profiling of Lithuanian parliamentary speeches : *Proceedings of the Sixth International Conference „Human Language Technologies – The Baltic Perspective“*. Amsterdam: IOS Press, 99-106.
11. Rimkutė, Erika, Utkā, Andrius, Levane-Petrova, Kristine. 2013. Lietuvių–latvių ir latvių–lietuvių kalbų lygiagretusis tekstynas LILA. Kalbų studijos = Studies about languages. 2013, nr. 23, 70-77.
12. Bielinskienė, Agnė, Boizou, Loic, Kovalevskaitė, Jolanta, Utkā, Andrius. 2012: Towards the automatic extraction of term-defining contexts in Lithuanian. Proceedings of the Fifth International Conference „Human Language Technologies – The Baltic Perspective“. Amsterdam: IOS Press, 18–26.
13. Boizou, Loic, Grigonytė, Gintarė, Rimkutė, Erika, Utkā, Andrius. 2012. Automatic inference of base forms for multiword terms in Lithuanian. Human language technologies - the Baltic perspective: the 5th international conference Baltic HLT, Tartu, Estonia, October 4–5, 2012: proceedings, 27-35.
14. Marcinkevičienė, Rūta, Kovalevskaitė, Jolanta, Utkā, Andrius, Vaičenonienė, Jurgita. 2012. An overview of multilingual processing for Lithuanian. Multilingual processing in Eastern and Southern EU languages : low-resourced technologies and translation / ed. Cristina Vertan, Walther v.Hahn. 119-143.
15. Pakerys, Jurgis, Rimkutė, Erika, Utkā, Andrius. 2012. Lietuvių kalbos denominatyviniai veiksmazodžiai ir greta jų vartojami pamatiniai žodžiai (tekstyno tyrimas). Darbai ir dienos. 58 (2012), 81-92.
16. Utkā, Andrius, Levane-Petrova, Kristine, Bielinskienė, Agnė, Kovalevskaitė, Jolanta, Rimkutė, Erika, Vevere, Daira. 2012. Lithuanian-Latvian-Lithuanian parallel corpus. Human language technologies - the Baltic perspective: the 5th international conference Baltic HLT, Tartu, Estonia, October 4–5, 2012: proceedings. 260-264.

17. Fomin, Vladislav, Vitkutė-Adžgauskienė, Daiva, Marcinkevičienė, Rūta, Utkā, Andrius. 2011. Standartization in designing research and application infrastructure for Lithuanian language technologies. Information technologies' 2011 : proceedings of the 17th international conference on Information and Software Technologies, IT 2011, Kaunas, Lithuania, April 27-29, 2011 / edited by R. Butleris, R, 174-181.
18. Grigonytė, Gintarė, Rimkutė, Erika, Utkā, Andrius, Boizou, Loic. 2011. Experiments on Lithuanian term extraction. NODALIDA 2011 : The 18th Nordic conference of computational linguistics, Riga, Latvia, May 11–13, 2011 : proceedings. 2011, Vol. 11 : NEALT Proceedings Series, 82-89.
19. Rimkutė, Erika, Kovalevskaitė, Jolanta, Melninkaitė, Vida, Utkā, Andrius, Vitkutė-Adžgauskienė, Daiva. 2010: Corpus of Contemporary Lithuanian Language – the Standardised Way. *Proceedings of the Fourth International Conference „Human Language Technologies – The Baltic Perspective“*. Amsterdam: IOS Press, 154–160.

Žodynai

1. Bielinskienė, Agnė, Boizou, Loic, Grigonytė, Gintarė, Kovalevskaitė, Jolanta, Markievicz, Irena, Rimkutė, Erika, Utkā, Andrius, Viliūnas, Giedrius, 2013. Švietimo ir mokslo terminų žodynas [elektroninis išteklis]. Kaunas: Vytauto Didžiojo universitetas. (<http://daukantas.vdu.lt/moksliniai-terminai/>)
2. Utkā, Andrius. 2009: Dažninis rašytinės lietuvių kalbos žodynas: 1 milijono žodžių morfologiškai anotuoto tekstyno pagrindu [elektroninis išteklis]. Kaunas: Vytauto Didžiojo universiteto leidykla.