



## Curriculum Vitae

VIDAS DAUDARAVIČIUS

(updated on October 2017)

### PERSONAL DATA

*Home address* Tamaros st 20, Virbaliūnai, Kauno raj., Lithuania  
*Work address* UAB VTEX, Mokslininku 2a, LT-08412 Vilnius, Lithuania  
*E-mail* vidas.daudaravicius@vtx.lt  
*Linkedin* <https://lt.linkedin.com/in/vidas-daudaravicius-28371159>

### EDUCATION

2008 - 2012 PhD student, Informatics  
Faculty of Informatics, Vytautas Magnus University.  
Dissertation title: *Collocation Segmentation for Text Chunking*, Defended: 31 Jan 2013

2000–2002 Master degree of Applied Informatics  
Thesis Title: *Syntactical Analysis of Lithuanian Language*. Department of Applied  
Faculty of Informatics, Vytautas Magnus University

1995–2000 Bachelor degree of the Informatics  
Faculty of Informatics, Vytautas Magnus University

1992–1995 Religious studies  
St. Anthony Religious Studies Institute at the Faculty of Catholic Theology  
Vytautas Magnus University

### WORK EXPERIENCE

2015-... VTeX, **Research manager**.

2011-2015 VTeX, **Researcher**.

2013 Vytautas Magnus University, FP7 *Learning and Execution of Action Categories ACAT* (Grant Agr. No. 600578), **researcher**.

2004-2014 Vytautas Magnus University. Faculty of Informatics. **Lecturer**. *Natural Language processing* course.

2000-2008 Vytautas Magnus University. Centre of Computational Linguistics. **Senior engineer-programmer**.

2006-2007 Lithuanian State Commission of Lithuanian Language. **Expert**.

2005-2007 Vytautas Magnus University. **Machine translation expert**.

2000 Vytautas Magnus University. Education Studies Institute. **Administrator**.

1997 “Work&Travel USA” program: “Merriam Park Painting”, St. Paul, USA. **House painter**.

1997 “Metro Produce Dist.”. Minneapolis, USA. **Driver, deliverer**.

1995–1996 St. Anthony Religious Studies Institute at the Faculty of Catholic Theology  
Vytautas Magnus University. **Technician** of a computer classroom.

1990–1991 Personal company “Nuotrauka”. **Owner, photography**.

### RESEARCH INTERESTS

Natural Language Processing  
Machine Learning  
Information Extraction and Information Retrieval  
Text Classification and Clusterization  
Text Mining

Collocation Extraction and Collocation Processing  
Grammar Error correction  
Natural Language Writing Evaluation  
Functional programming

## LANGUAGE SKILLS

Lithuanian – native      English – proficient      Russian – moderate      French – background

## OTHER SKILLS

Programming languages: C++, Haskell, Java, JavaScript, Python, Perl, PHP.  
Mark-up languages: HTML, XML, SGML.  
Distributed computing: MapReduce, Hadoop.  
Parallel programming: MPI.  
Databases: MySQL, MongoDB.  
Other: LaTeX, OpenOffice, MS Word, MS Access, MS Excel, Corel Draw.

## SERTIFICATES

2005      *High Performance Computing: Application Tuning for Clusters*. Intel Software College  
2015      *Machine Learning*. Coursera Verified Certificates License GDUK92FKBLA7  
2016      *Hadoop Platform and Application Framework*. Coursera Course Certificates License  
GNSUXPNFQTGG  
2017      *DeepLearn 2017 International Summer School on Deep Learning*. Univesity of Deusto

## PARTICIPATION IN PROJECTS

2017-...      MC member of COST Action *European Network for Combining Language Learning with Crowdsourcing Techniques*. CA16105.  
2013-2015      *Preparation of scientific publishing data and application of intelligent technologies at VTeX*. EU SF VP2-1.3-ŪM-02-K-04-019.  
2013      FP7 *Learning and Execution of Action Categories ACAT* (Grant Agr. No. 600578)  
2011-2012      *Application of machine learning in the enterprise activities – feasibility study*, EU SF VP2-1.3-ŪM-01-K-02-232.  
2009 – 2012      HEalth teXt Analysis network in the Nordic and Baltic countries (HEXAnord). NordForsk foundation. Researcher networks 2009  
2007 – 2008      *Internet resources: Annotated corpus of the Lithuanian language and tools of annotation (ALKA 2)*. Lithuanian State Science and Studies Foundation  
2006 - 2006      6FP IST, Network of excellence project *Resilience for Survivability in IST (ReSIST)* – (IST-4-026764-NOE)  
2005 – 2007      *Webpage Information Translation Facility*. BPD2004-ERPF-3.3.0-02-04/0005. (<http://vertimas.vdu.lt>)  
2003 – 2006      *Parallel Lithuanian Language Corpus*. Lithuanian State Commission of the Lithuanian language  
2004      *WITFA (Webpage Information Translation Facility)*. PHARE - 2002/000-620.05.01-03.07.  
2005 – 2006      *The survivability of Lithuanian language under the globalisation circumstances: annotated Lithuanian language corpus (ALKA)*. Lithuanian State Science and Studies Foundation  
2001      EU INCO. TELRI-II (Trans European Language Resources Infrastructure – II). Project *Electronic dictionary of Czech-Lithuanian , Lithuanian-Czech*.

## ONLINE TOOLS AND CORPORA

2015-2017      [LaTeX to Text converter](#). It expands macros. A must tool when working with LaTeX files. Implemented in Haskell.  
2016-2017      [Keyphrase extraction](#) from a set of scientific articles.  
2016      [Automated Evaluation of Scientific Writing Data Set](#).  
2005-2008      [Tagger for Lithuanian](#)

## SOCIAL ACTIVITIES

2017

1. **Program Committee Member** of *The 12th Workshop on Innovative Use of NLP for Building Educational Applications*. EMNLP workshop. 2017. Copenhagen. Denmark.
2. **Program Committee Member** of *The 4th Workshop on NLP Techniques for Educational Applications (NLPTEA 2017)*. December 1, 2017, Taipei, Taiwan. IJCNLP 2017 Workshop. (<https://sites.google.com/view/nlptea2017/>)

2016

3. **Organizer** of Automated Evaluation of Scientific Writing (AESW) 2016 Shared Task. Vidas Daudaravicius, Rafael E. Banchs, Elena Volodina, Courtney Napoles. In connection with the 11th Workshop on Innovative Use of NLP for Building Educational Applications, ACL, San Diego, CA, USA, June 16, 2016. (<http://textmining.lt/aesw/index.html>)
4. **Program Committee Member** of *The 3rd Workshop on Natural Language Processing Techniques for Educational Applications (NLP-TEA-3)*. December 12 at Osaka, Japan in conjunction with COLING 2016. (<http://nlptea2016.weebly.com/>)

## PUBLICATIONS

2016

1. **Vidas Daudaravicius**, Rafael E. Banchs, Elena Volodina, Courtney Napoles. *A report on the AESW 2016 Shared Task*. In Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications, ACL, San Diego, CA, USA, June 16, 2016
2. **Vidas Daudaravicius**. *A framework for keyphrase extraction from scientific journals*. Semantics, Analytics, Visualization. Enhancing Scholarly Data: Second International Workshop, SAVE-SD 2016, Montreal, QC, Canada, April 11, 2016, Revised Selected Papers. LNCS 9792. Springer International Publishing

2015

3. **Vidas Daudaravicius**. *Automated Evaluation of Scientific Writing: AESW Shared Task Proposal*. In Proceedings of the Tenth Workshop on Innovative Use of NLP for Building Educational Applications (BEA'10), Denver, Colorado, June, 2015, p. 56-63

2014

4. **Vidas Daudaravicius**. *Language Editing Dataset of Academic Texts*. In Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14). Reykjavik, Iceland. May 2014, p. 1738-1742.
5. A. Henriksson, H. Moen, M. Skeppstedt, **V. Daudaravicius**, M. Duneld. *Synonym extraction and abbreviation expansion with ensembles of semantic spaces*. Journal of biomedical semantics. London : BioMed Central. 2014, 5:6, p. 1-25.

2013

6. **Vidas Daudaravicius**. *VTEX System Description for the NLI 2013 Shared Task*. In Proceedings of the Eighth Workshop on Innovative Use of NLP for Building Educational Applications. Atlanta, Georgia. Jun 2013, p. 89-95.
7. **Vidas Daudaravicius**. *Collocation segmentation for text chunking*. PhD Thesis. 2013.

2012

8. Aron Henriksson, Hans Moen, Maria Skeppstedt, Ann-Marie Eklund, **Vidas Daudaravičius** and Martin Hassel. *Synonym Extraction of Medical Terms from Clinical Text Using Combinations of Word Space Models*. Proceedings of the 5th International Symposium on Semantic Mining in Biomedicine (SMBM 2012). Edited by: Ananiadou, Sophia; Pyysalo, Sampo; Rebholz-Schuhmann, Dietrich; Rinaldi, Fabio; Salakoski, Tapio. Zurich, 2012. ISBN 978-3-033-03823-3.
9. **Vidas Daudaravicius**. *Applying collocation segmentation to the ACL Anthology Reference Corpus*. In Proceedings of the ACL-2012 Special Workshop on Rediscovering 50 Years of Discoveries, pages 66–75, Jeju Island, Korea, July 2012. Association for Computational Linguistics. ISBN 978-1-937284-29-9
10. **Daudaravicius, Vidas**. *VTEX Determiner and Preposition Correction System for the HOO 2012 Shared Task*. In Proceedings of the Seventh Workshop on Building Educational Applications Using NLP, June, 2012, Montreal, Canada, Association for Computational Linguistics, p. 225-232.
11. **Daudaravičius, Vidas**. *Automatic multilingual annotation of EU legislation with Eurovoc descriptors // LREC 2012 : 8th international conference on Language resources and evaluation, 21-27 May, 2012, Istanbul, Turkey : proceedings*. Paris : European Language Resources Association. ISBN 9782951740877. p. 14-20.

2011

12. Allvin, Helen; Carlsson, Elin; Dalianis, Hercules; Danielsson-Ojala, Riitta; **Daudaravičius, Vidas**; Martin, Hassel; Kokkinakis, Dimitrios; Lundgren-Laine, Heljä; Nilsson, Gunnar; Nytrø, Øystein; Salanterä, Sanna; Skeppstedt, Maria; Suominen, Hanna; Velupillai, Sumithra. *Characteristics of Finnish and Swedish intensive care nursing narratives : a comparative analysis to support the development of clinical language technologies* // Journal of biomedical semantics. London : BioMed Central. ISSN 2041-1480. vol. 2, suppl. 3, p. 1-11.

2010

13. Allvin, Helen; Carlsson, Elin; Dalianis, Hercules; Danielsson-Ojala, Riitta; **Daudaravičius, Vidas**; Martin, Hassel; Kokkinakis, Dimitrios; Lundgren-Laine, Heljä; Nilsson, Gunnar; Nytrø, Øystein; Salanterä, Sanna; Skeppstedt, Maria; Suominen, Hanna; Velupillai, Sumithra. *Characteristics and Analysis of Finnish and Swedish Clinical Intensive Care Nursing Narratives* // Louhi'10 : Proceedings of the NAACL HLT 2010, Second Louhi Workshop on Text and Data Mining of Health Documents, June 5, 2010, Los Angeles, California. Stroudsburg, USA : The Association for Computational Linguistics, 2010. p. 53-60.
14. Costa-jussà, Marta R., **Daudaravičius, Vidas**, Banchs, Rafael E., *Integration of statistical collocation segmentations in a phrase-based statistical machine translation system*. EAMT 2010 : Proceedings of the 14th Annual Conference of the European Association for Machine Translation, 27-28 May 2010, Saint-Raphaël, France
15. C. Henríquez, M.R. Costa-jussà, **V. Daudaravičius**, R.E. Banchs, J.B. Mariño (2010) "*UPC-BMIC-VDU system description for the IWSLT 2010: testing several collocation segmentations in a phrase-based SMT system*", in Proceedings of the 7th International Workshop on Spoken Language Translation (IWSLT), pages 189-195, Paris
16. Henríquez, Carlos A. Q.; Costa-jussà, Marta R.; **Daudaravičius, Vidas**; Banchs, Rafael E.; Mariño, B. José. *Using collocation segmentation to augment the phrase table* // WMT'10 : Proceedings of the Joint Fifth Workshop on Statistical Machine Translation and MetricsMATR, Uppsala, Sweden, 15-16 July 2010. Morristown, USA : Association for Computational Linguistics. ISBN 9781932432718. p. 98-102.
17. Costa-jussà, Marta R.; **Daudaravičius, Vidas**; Banchs, Rafael E.. *Using collocation segmentation to extract translation units in a phrase-based statistical machine translation system* // Procesamiento del Lenguaje Natural. Barcelona : Sociedad Espanola para el Procesamiento del Lenguaje Natural. ISSN 1135-5948. 2010, no. 45, p. 215-220.
18. **Daudaravičius, Vidas**. *The influence of collocation segmentation and top 10 items to keyword assignment performance* // Computational Linguistics and Intelligent Text Processing : 11th International Conference, CICLing 2010, Iasi, Romania, March 21-27, 2010 : Proceedings. Berlin : Springer. ISBN 9783642121159. p. 648-660.
19. **Daudaravičius, Vidas**. *Automatic identification of lexical units* // Informatica : An International Journal of Computing and Informatics. Ljubljana : Slovensko društvo Informatika. ISSN 0350-5596. Vol. 34, no. 1 (2010), p. 85-91.

2007

20. Utkā A., Kovalevskaitė J., Rimkutė E., **Daudaravičius V.** *Bilingual Parallel Corpora for English, Czech and Lithuanian* – The Third Baltic Conference on Human Language Technologies proceedings. Kaunas, 2007, 319–326
21. Gindiyeh M., Čulo O., Grigonytė G., Haller J., Avižienis A., Marcinkevičienė R., **Daudaravičius V.** *A document classification tool for the resilience knowledge base of the resist noe project* – The Third Baltic Conference on Human Language Technologies
22. E. Rimkutė, **V. Daudaravičius**, A. Utkā *Morphological Annotation of the Lithuanian Corpus*. – In Proceedings of the Workshop on Balto-Slavonic Natural Language Processing 2007, Special Theme: Information Extraction and Enabling Technologies, P. 94-99. Association of Computational Linguistics.

2006

23. **Daudaravičius V.** *Opening to infinity: Machine Translation and Lithuanian Language*. (In Lithuanian). Deeds and Days, 2006, Nr. 45, P. 7–18.
24. Rimkutė E., Kovalevskaitė J. **Daudaravičius V.** *The usage and application of multilingual corpora*. (In Lithuanian). Deeds and Days, 2006, Nr. 45, P. 41–62.

2005

25. Zinkevičius V., **Daudaravičius V.**, Rimkutė E. *The Morphologically annotated Lithuanian Corpus – The Second Baltic Conference on Human Language Technologies proceedings*. Tallinn, 2005, 365–370

2004

26. **V. Daudaravičius**, R. Marcinkevičienė Gravity Counts for the Boundaries of Collocations. In

*International Journal of Corpus Linguistics*, 2004, Nr 9-2. Amsterdam: John Benjamins Publishing Company

27. Marcinkevičienė R., Bielinskienė A., **Daudaravičius V.**, Rimkutė E. *Corpora for Lithuanian Language Technologies* – In proceedings of *The First Baltic Conference Human Language Technologies. The Baltic Perspective*, 2004, P. 21–24.