

Curriculum Vitae  
VIDAS DAUDARAVIČIUS  
25 JUL 2014

**PERSONAL DATA**

*Home address* Virbaliūnų k., Batniavos sen., Kauno raj., Lithuania  
*Work address* VTEX,  
Akademijos st. 4, LT-08412 Vilnius, Lithuania  
*E-mail* vidas.daudaravicius@vtex.lt

**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 Informatics  
Faculty of Informatics  
Vytautas Magnus University

1995–2000 Bachelor degree of the Informatics  
Department of Applied 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**

2011-... 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

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
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

## REASEARCH INTERESTS

Natural Language processing  
 Natural Language Parsing  
 Information Extraction  
 Information Retrieval  
 Text Classification and Clusterization  
 Text Mining  
 Collocation Extraction and Collocation Processing  
 Grammar Error correction  
 Functional programming  
 Native Language Identification  
 Language Quality Evaluation

## LANGUAGE SKILLS

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

## PARTICIPATION IN PROJECTS

2013-...	<i>Preparation of scientific publishing data and application of intelligent technologies at VTeX.</i> EU SF VP2-1.3-ŪM-02-K-04-019.
2013	FP7 <i>Learning and Execution of Action Categories ACAT</i> (Grant Agr. No. 600578)
2011-2012	<i>Application of machine learning in the enterprise activities – feasibility study,</i> 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	<i>Internet resources: Annotated corpus of the Lithuanian language and tools of annotation (ALKA 2).</i> Lithuanian State Science and Studies Foundation
2006 -	6FP IST, Network of excellence project <i>Resilience for Survivability in IST (ReSIST)</i> – (IST-4-026764-NOE)
2005 – 2007	<i>Webpage Information Translation Facility.</i> BPD2004-ERPF-3.3.0-02-04/0005. ( <a href="http://vertimas.vdu.lt">http://vertimas.vdu.lt</a> )
2003 – 2006	<i>Parallel Lithuanian Language Corpus.</i> 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*.

## SCHOLARSHIP

- 2007.09 NATO ASI Mining Massive Data Sets for Security
- 2002 02 – 2002 05 Socrates/Erasmus, Faculty of Computer Science, University of Twente, The Netherlands (Olandija)
- 2002 12 Socrates/Erasmus, Faculty of Computer Science, University of Twente, The Netherlands (Olandija)
- 2004 06 Socrates/Erasmus, Faculty of Computer Science, University of Twente, The Netherlands (Olandija)

## OTHER SKILLS

Programming languages: C++, Haskell, Java, PHP, JavaScript.

Mark-up languages: HTML, XM, SGML.

Parallel programming: MPI.

Databases: MySQL. MongoDB.

Other: OpenOffice, MS Word, MS Access, MS Excel, Corel Draw, LaTeX.

## SERTIFICATES

*High Performance Computing: Application Tuning for Clusters*. Intel Software College. September 2005.

## PUBLICATIONS

### 2014

1. 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.
2. 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

3. 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.

4. Vidas Daudaravicius. *Collocation segmentation for text chunking*. PhD Thesis. 2013.

## 2012

5. 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; Rebolz-Schuhmann, Dietrich; Rinaldi, Fabio; Salakoski, Tapio. Zurich, 2012. ISBN 978-3-033-03823-3.
6. 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
7. 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.
8. 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

9. 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

10. 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.
11. 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
12. C. Henríquez, M.R. Costa-jussà, V. Daudaravicius, 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
13. 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.

14. 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 Española para el Procesamiento del Lenguaje Natural. ISSN 1135-5948. 2010, no. 45, p. 215-220.
15. 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.
16. 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

17. E.Rimkutė E., Daudaravičius V. *Morfologinis „Dabartinės lietuvių kalbos tekstyno“ anotavimas.* – Kalbų studijos, 2007, Nr. 11, P. 30–35.
18. Rimkutė E., Daudaravičius V. *Morfologinis „Dabartinės lietuvių kalbos tekstyno“ anotavimas* Kalbų studijos, 2007, Nr. 9, P. 30–35.
19. 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
20. 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
21. 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

22. *Daudaravičius V. Pradžia į begalybę (Mašininis vertimas ir lietuvių kalba).* – *Darbai ir Dienos*, 2006, Nr. 45, P. 7–18.
23. *Rimkutė E., Kovalevskaitė J. Daudaravičius V. Daugiakalbių tekstynų naudojimas ir taikymas.* – *Darbai ir Dienos*, 2006, Nr. 45, P. 41–62.

## 2005

24. 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

25. 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
26. Marcinkevičienė R., Bielinskienė A., Daudaravičius V., Rimkutė E. *Corpora for Lithuanian Language Technologies* – Tarptautinės konferencijos *The First Baltic Conference Human Language Technologies. The Baltic Perspective* pranešimų medžiaga, 2004, P. 21–24.

## 2001

27. V. Daudaravičius *Elektroniniai tekstynai lietuvių kalbai tirti ir kompiuterizuoti*, Informacinės Technologijos 2002, konferencijos pranešimų medžiaga, Kauno Technologijos Universitetas, 2002 01 29-30

## 2000

28. V. Daudaravičius *Interneto teksto paruošimas automatinei analizei*, 6-oji magistrantų ir doktorantų konferencija, konferencijos pranešimų medžiaga, Vytauto Didžiojo Universitetas, 2001 04 26

## OTHER PRESENTATIONS

1. V. Daudaravičius (2007). *The exploration on statistical learning and parsing dependencies with abstract syntactic rules*. NATO ASI Mining Massive Data Sets for Security (Italy)
2. V. Daudaravičius, A. Utko (2005) *Language-independent collocation extraction*. Exploiting parallel corpora in up to 20 languages (26-27 September 2005), JRC, Arona (NO), Italy
3. R. Marcinkevičienė, V. Daudaravičius (2002) *Detection of the Boundaries of Collocations*. 7<sup>th</sup> TELRI seminar Information in Corpora. September 26-28, 2002, Dubrovnik, Croatia.