

Torsten Marek

Employments

- 07/2010 – now **Software Engineer**, *Google Switzerland GmbH*, Zürich.
Operative and engineering work in the team that maintains Google's world-wide computing fleet
- 04/2009 – 06/2010 **PhD Candidate**, *Institute of Computational Linguistics, University of Zürich*, Zürich.
Tools for parallel and monolingual treebanks, corpus query formalisms, teaching
- 12/2008 – 03/2009 **Student Assistant**, *Department of Computational Linguistics and Phonetics, Saarland University*, Saarbrücken.
Query formalisms for multi-layer corpora
- 11/2007 – 09/2008 **Student Assistant**, *German Research Center for Artificial Intelligence (DFKI GmbH)*, Saarbrücken.
Information extraction for intelligent authoring systems
- 09/2006 – 09/2007 **Software Engineer**, *Fast Search & Transfer Deutschland GmbH*, Munich.
Eclipse RCP, linguistic processing pipelines, OSGi, test frameworks
- 02/2006 – 03/2006 **External Software Consultant**, *Fast Search & Transfer Deutschland GmbH*, Munich.
Linguistic information extraction in FAST ESP
- 11/2003 – 06/2005 **System administrator**, *Institut für Theoretische Computerlinguistik*, Tübingen.
Linux system administration
- 07/2001 – 09/2003 **Programmer and System administrator**, *van Bürk Datentechnik GmbH*, Gladbeck.
Desktop client programming with VB6, Linux server administration

Education

- 10/2007 – 03/2009 **M.Sc. in Language Science & Technology**, *Saarland University*, Saarbrücken.
- 10/2003 – 08/2006 **B.A. in Computational Linguistics**, *Eberhard-Karls-University*, Tübingen.

Master Thesis

Title *Integration of Light-weight Semantics into a Syntax Query Formalism*
Supervisors Prof. Dr. Manfred Pinkal and Dr. Caroline Sporleder
Grade 1.0

Bachelor Thesis

Title *Analysis of German Compounds Using Weighted Finite State Transducers*
Supervisor Dale Gerdemann, PhD
Grade 1.0

Experience

Teaching

- Spring 2010 **Machine Learning for Natural Language Processing**, *Institute of Computational Linguistics*, University of Zürich.
Semester-long bachelor-level course about machine learning algorithms for natural language processing.
- Fall 2009 **Programming Techniques for Computational Linguistics III**, *Institute of Computational Linguistics*, University of Zürich.
Semester-long bachelor-level course about algorithms and technologies for natural language processing, including exam.
- 04/2007 **Introduction to Programming in Python**, *Department of Linguistics*, Stockholm University.
One-week intensive course about programming in Python, including exam.

International Studies

- 08/2005 – 01/2006 Erasmus Student at the Department of Computational Linguistics at Stockholm University, Stockholm

Internships

- 02/2005 **Programmer**, *Fast Search & Transfer Deutschland GmbH*, Munich.
Linguistic information extraction in FAST ESP

Languages

German	Native	<i>Mother tongue</i>
English	Fluent	<i>Writing & speaking</i>
Swedish	Basic	<i>Six-month Erasmus exchange in Stockholm</i>

Skills

Programming Languages	Python, C++, Java, Haskell, Go, JavaScript, C, Perl, SQL
Databases	SQLite, MySQL
Misc.	Debian packaging, Linux system/server administration, XML

Professional Interests

- Machine Learning
- Parallel computing
- String and graph algorithms
- Metaprogramming
- Data mining

Previous Research Projects

- TreeAligner **A tool for creating, browsing and querying parallel treebanks**, Maintainer.
<http://kitt.ifi.uzh.ch/kitt/treealigner>
- ladon **A Python library for working with syntax graphs**, Principal Author.
<http://kitt.ifi.uzh.ch/kitt/hg/ladon/head>
- SMULTRON **Parallel treebank with syntactic alignments**, Technical Coordinator.
http://www.cl.uzh.ch/research/paralleltreebanks_en.html
- Text + Berg **A project to build and annotate a comparable and parallel corpus of texts on alpinism**, Programmer and Technical Coordinator.
<http://www.textberg.ch>
- ANNIS2 **System for searching and visualizing multilevel corpora**, Contributor.
<http://www.sfb632.uni-potsdam.de/dl/annis/>

Selected Publications

Scientific Articles

- Martin Volk, **Torsten Marek** and Rico Sennrich (2010). Reducing OCR errors by combining two OCR systems. In *ECAI 2010 Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH 2010)*, pp. 61–65. Lisbon, Portugal.
- 2009 **Torsten Marek** (2009). Integration of light-weight Semantics into a Syntax Query Formalism. In *Proceedings of the Biennial GSCL Conference*, pp. 109–114. Potsdam, Germany.
- Torsten Marek**, Gerold Schneider and Martin Volk (2009). A framework for constituent-dependency conversion. In Marco Passarotti, Adam Przepiórkowski, Savina Raynaud and Frank Van Eynde, eds., *Proceedings of the Eighth International Workshop on Treebanks and Linguistic Theories (TLT8)*. Milano, Italy.
- 2008 **Torsten Marek**, Joakim Lundborg and Martin Volk (2008). Extending the TIGER Query Language with Universal Quantification. In Angelika Storrer, Alexander Geyken, Alexander Siebert and Kay-Michael Würzner, eds., *Text Resources and Lexical Knowledge*. Berlin, Germany: Mouton de Gruyter.
- 2007 Joakim Lundborg, **Torsten Marek**, Maël Mettler and Martin Volk (2007). Using the Stockholm TreeAligner. In *Proc. of The 6th Workshop on Treebanks and Linguistic Theories*. Bergen.
- 2006 Martin Volk, Sofia Gustafson-Capková, Joakim Lundborg, **Torsten Marek**, Yvonne Samuelsson and Frida Tidström (2006). XML-based Phrase Alignment in Parallel Treebanks. In *Proceedings of EACL Workshop on Multi-dimensional Markup in Natural Language Processing*. Trento.

Books

- 2005 Nix, Markus, ed. (2005). *Exploring Python*, chapter 1: Generatoren. Software & Support Verlag. ISBN 978-3-935042-69-7.