

Ulrich Schäfer

1 Coordinates and Short CV

Dr. Ulrich Schäfer
Senior Engineer
Language Technology Lab
German Research Center for Artificial Intelligence (DFKI)
✉ Campus D 3 1, D-66123 Saarbrücken, Germany
☎ +49-681-85775-5154
email ulrich.schaefer@dfki.de
Homepage <http://www.dfki.de/%7Euschaefer>

Ulrich Schäfer studied Computer Science and Computational Linguistics (minor subject) at Saarland University, Saarbrücken. He received the Computer Science diploma (equivalent to MSc) in 1995, with a thesis on *Parameterized Type Expansion in the Typed Feature Structure formalism TDL* that he had developed under the supervision of Dr. Hans-Ulrich Krieger.

From 1995 to 2000, he worked with Dokumenta S.A., a Luxembourg-based IT supplier, as application developer and consultant. He developed multilingual document editing software and did consulting and training in EDI, messaging and directory services for EU institutions and companies in Belgium, France, Germany and Luxembourg.

Since 2000, he is working as Senior Engineer in research in the Language Technology Lab of the German Research Center for Artificial Intelligence (DFKI), based in Saarbrücken. He co-developed a platform for multilingual information extraction and structured named entity recognition (*SProUT*) and devised the hybrid NLP middleware *Heart of Gold*, combining multilingual shallow and deep NLP components. Both were employed in many applications (information extraction, question answering, document authoring support) in EU and BMBF-funded research projects such as Whiteboard, DeepThought, HyLaP and Quetal (BMBF is the German Ministry of Education and Research).

In 2007, he defended his dissertation *Integrating Deep and Shallow Natural Language Processing Components – Representations and Hybrid Architectures*, supervised by Prof. Wolfgang Wahlster and Prof. Hans Uszkoreit, in the Faculty of Mathematics and Computer Science of Saarland University, with grade *very good*.

Since 2009, he is project leader of the BMBF-funded project TAKE, working on parsing and search in digital libraries. In the context of TAKE, the international DELPH-IN collaboration on deep linguistic processing and the Saarbrücken-based Cluster of Excellence *Multimodal Computing and Interaction* (M2CI) funded by the DFG, he also devised the *ACL Anthology Searchbench*, publicly available at <http://aclasb.dfki.de>, a novel search engine for semantic, full-text and bibliographic search in the open digital library of research papers in Computational Linguistics and Language Technology, the *ACL Anthology*.

Since 2004, he regularly lectures Java and Python Programming courses and seminars in the Computational Linguistics Department of Saarland University and also taught two summer school courses at ESSLLI-2009 (Bordeaux) and ELSNET-2006 (Hamburg). He supervised BSc and MSc theses in Computer Science and Computational Linguistics, and served as journal and book reviewer, and as conference and workshop program committee member.

2 Publications

Journal Articles

- [1] Anette Frank, Hans-Ulrich Krieger, Feiyu Xu, Hans Uszkoreit, Berthold Crismann, and Ulrich Schäfer. Question answering from structured knowledge sources. *Journal of Applied Logics, Special Issue on Questions and Answers: Theoretical and Applied Perspectives*, 5(1):20–48, March 2007.
- [2] Witold Drozdowski, Hans-Ulrich Krieger, Jakub Piskorski, Ulrich Schäfer, and Feiyu Xu. Shallow processing with unification and typed feature structures — Foundations and applications. *Künstliche Intelligenz*, 1:17–23, 2004.

Book Chapters

- [3] Ulrich Schäfer and Bernd Kiefer. Advances in deep parsing of scholarly paper content. In Raffaella Bernardi, Sally Chambers, Björn Gottfried, Frédérique Segond, and Ilya Zaihrayeu, editors, *Advanced Language Technologies for Digital Libraries*, number LNCS 6699 in LNCS Theoretical Computer Science Series, pages 135–153. Springer, 2011.
- [4] Arif Bramantoro, Ulrich Schäfer, and Toru Ishida. Pipelining software and services for language processing. In Toru Ishida and Donghui Lin, editors, *The Language Grid: Service-Oriented Collective Intelligence for Language Resource Interoperability*, LNCS Cognitive Technologies Series, chapter 16. pages 247–262. Springer, 2011.

Refereed Conferences

- [5] Ulrich Schäfer, Bernd Kiefer, Christian Spurk, Jörg Steffen, and Rui Wang. The ACL Anthology Searchbench. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL HLT 2011), System Demonstrations*, pages 7–13. Association for Computational Linguistics (ACL), June 2011.
- [6] Peter Adolphs, Martin Theobald, Ulrich Schäfer, Hans Uszkoreit, and Gerhard Weikum. YAGO-QA: Answering questions by structured knowledge queries. In *Fifth IEEE International Conference on Semantic Computing*, pages 158–161. Los Alamitos, CA, USA, September 2011. IEEE Computer Society.
- [7] Cailing Dong and Ulrich Schäfer. Ensemble style self-training on citation classification. In *Proceedings of the 5th International Joint Conference on Natural Language Processing (IJCNLP)*, pages 623–631, Chiang Mai, Thailand, November 2011.
- [8] Magdalena Wolska, Ulrich Schäfer, and The Nghia Pham. Bootstrapping domain-specific terminological taxonomies from scientific text. In *Proceedings of the 9th International Conference on Terminology and Artificial Intelligence*, pages 17–23, Paris, France, November 2011.
- [9] Ulrich Schäfer and Christian Spurk. TAKE scientist’s workbench: Semantic search and citation-based visual navigation in scholar papers. In *Fourth IEEE International Conference on Semantic Computing*, pages 317–324, Los Alamitos, CA, USA, September 2010. IEEE Computer Society.
- [10] Hans-Ulrich Krieger and Ulrich Schäfer. DL meet FL: A bidirectional mapping between ontologies and linguistic knowledge. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING-2010)*, pages 588–596, Beijing, China, August 2010.

- [11] Ulrich Schäfer Arif Bramantoro and Toru Ishida. Towards an integrated architecture for composite language services and multiple linguistic processing components. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC-2010)*, pages 3506–3511, Valletta, Malta, May 2010. European Language Resources Association (ELRA).
- [12] Arif Bramantoro, Masahiro Tanaka, Yohei Murakami, Ulrich Schäfer, and Toru Ishida. A hybrid integrated architecture for language service composition. In *Proceedings of the IEEE 2008 International Conference on Web Services (ICWS-2008 Research Track)*, pages 345–352, Beijing, China, October 2008. IEEE Computer Society.
- [13] Ulrich Schäfer, Hans Uszkoreit, Christian Federmann, Torsten Marek, and Yajing Zhang. Extracting and querying relations in scientific papers. In *Proceedings of the 31st Annual German Conference on Artificial Intelligence, KI-2008*, pages 127–134, Kaiserslautern, Germany, September 2008. Springer LNAI 5243.
- [14] Ulrich Schäfer, Hans Uszkoreit, Christian Federmann, Torsten Marek, and Yajing Zhang. Extracting and querying relations in scientific papers on language technology. In *Proceedings of LREC-2008*, Marrakesh, Morocco, May 2008. ELRA.
- [15] Ulrich Schäfer and Daniel Beck. Automatic testing and evaluation of multilingual language technology resources and components. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC-2006)*, pages 173–178, Genoa, Italy, May 2006. ELRA.
- [16] Ulrich Schäfer. OntoNERdIE – Mapping and linking ontologies to named entity recognition and information extraction resources. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC-2006)*, pages 1756–1761, Genoa, Italy, May 2006. ELRA.
- [17] Ben Waldron, Ann Copestake, Ulrich Schäfer, and Bernd Kiefer. Preprocessing and tokenisation standards in DELPH-IN tools. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC-2006)*, pages 2263–2268, Genoa, Italy, May 2006.
- [18] Hans-Ulrich Krieger, Witold Drozdzyński, Jakub Piskorski, Ulrich Schäfer, and Feiyu Xu. A bag of useful techniques for unification-based finite-state transducers. In Ernst Buchberger, editor, *Proceedings of 7th KONVENS*, pages 105–112, 2004.
- [19] Hans Uszkoreit, Ulrich Callmeier, Andreas Eisele, Ulrich Schäfer, Melanie Siegel, and Jakob Uszkoreit. Hybrid robust deep and shallow semantic processing for creativity support in document production. In *Proceedings of KONVENS-2004*, pages 209–216, Vienna, Austria, September 2004. ÖGAI.
- [20] Anette Frank, Kathrin Spreyer, Witold Drozdzyński, Hans-Ulrich Krieger, and Ulrich Schäfer. Constraint-based RMRS construction from shallow grammars. In Stefan Müller, editor, *Proceedings of the HPSG-04 Conference Workshop on Semantics in Grammar Engineering*, pages 393–413, Leuven, Belgium, August 2004. Center for Computational Linguistics, Katholieke Universiteit Leuven, CSLI Publications, Stanford, CA.
- [21] Witold Drozdzyński, Hans-Ulrich Krieger, Jakub Piskorski, and Ulrich Schäfer. A multilingual content production tool for the semantic web. In *European Conference on Knowledge Engineering and Knowledge Management (EKAW-04)*, Northamptonshire, UK, October 2004.
- [22] Ulrich Callmeier, Andreas Eisele, Ulrich Schäfer, and Melanie Siegel. The DeepThought core architecture framework. In *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC) 2004*, pages 1205–1208, Lisbon, Portugal, May 2004. ELRA.
- [23] Anette Frank, Markus Becker, Berthold Crysmann, Bernd Kiefer, and Ulrich Schäfer. Integrated shallow and deep parsing: TopP meets HPSG. In *Proceedings of ACL-2003*, pages 104–111, Sapporo, Japan, 2003. ACL.

- [24] Stephan Busemann, Witold Drozdzyński, Hans-Ulrich Krieger, Jakub Piskorski, Ulrich Schäfer, Hans Uszkoreit, and Feiyu Xu. Integrating information extraction and automatic hyperlinking. In *Proceedings of the Interactive Posters/Demonstration at ACL-03*, pages 117–120, Sapporo, Japan, 2003. ACL.
- [25] Markus Becker, Witold Drozdzyński, Hans-Ulrich Krieger, Jakub Piskorski, Ulrich Schäfer, and Feiyu Xu. SProUT – Shallow processing with typed feature structures and unification. In *Proceedings of the International Conference on NLP (ICON-2002)*, Mumbai, India, December 2002.
- [26] Berthold Crysmann, Anette Frank, Bernd Kiefer, Hans-Ulrich Krieger, Stefan Müller, Günter Neumann, Jakub Piskorski, Ulrich Schäfer, Melanie Siegel, Hans Uszkoreit, and Feiyu Xu. An integrated architecture for shallow and deep processing. In *Proceedings of ACL-2002*, Philadelphia, PA, July 2002.
- [27] Hans-Ulrich Krieger and Ulrich Schäfer. TDL – A type description language for constraint-based grammars. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING -94)*, volume 2, pages 893–899, Kyoto, Japan, 1994.

Theses

- [28] Ulrich Schäfer. *Integrating Deep and Shallow Natural Language Processing Components – Representations and Hybrid Architectures*. PhD thesis, Saarland University, Faculty of Mathematics and Computer Science, Saarbrücken, Germany, June 2007.
- [29] Ulrich Schäfer. Parameterized type expansion in the feature structure formalism TDL. Master’s thesis, Saarland University, Computer Science Department, Saarbrücken, Germany, 1995.

Workshops

- [30] Ulrich Schäfer and Uwe Kasterka. Scientific authoring support: A tool to navigate in typed citation graphs. In *Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing: Writing Processes and Authoring Aids (CL&W-2010)*, pages 7–14, Los Angeles, CA, June 2010. ACL.
- [31] Ulrich Schäfer. Shallow, deep and hybrid processing with UIMA and Heart of Gold. In *Proceedings of the LREC-2008 Workshop Towards Enhanced Interoperability for Large HLT Systems: UIMA for NLP*, pages 43–50, Marrakesh, Morocco, May 2008. ELRA.
- [32] Ulrich Schäfer. Middleware for creating and combining multi-dimensional NLP markup. In *Proceedings of the 5th Workshop on NLP and XML (NLPXML-2006): Multi-dimensional Markup in Natural Language Processing, 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL-2006)*, pages 81–84, Trento, Italy, April 2006.
- [33] Christian Bering and Ulrich Schäfer. JTaCo & SProUTomat: Automatic evaluation and testing of multilingual language technology resources and components. In *Proceedings of the LREC-2006 Workshop on Quality Assurance and Quality Measurement for Language and Speech Resources*, pages 42–47, Genoa, Italy, May 2006.
- [34] Anette Frank, Hans-Ulrich Krieger, Feiyu Xu, Hans Uszkoreit, Berthold Crysmann, Brigitte Jörg, and Ulrich Schäfer. Querying structured knowledge sources. In *Proceedings of the AAIL-05 Workshop on Question Answering in Restricted Domains*, pages 10–19, Pittsburgh, PA, July 2005.
- [35] Witold Drozdzyński, Hans-Ulrich Krieger, Jakub Piskorski, and Ulrich Schäfer. SProUT – a general-purpose NLP framework integrating finite-state and unification-based grammar formalisms. In *Proceedings of the 5th International Workshop on Finite-State Methods and Natural Language Processing*, Helsinki, Finland, September 2005. Springer Lecture Notes in Computer Science.

- [36] Ulrich Schäfer. Using XSLT for the integration of deep and shallow natural language processing components. In Kiril Simov and Erhard Hinrichs, editors, *Proceedings of the ESLLI-2004 Workshop on Combining Shallow and Deep Processing for NLP*, pages 31–40, Nancy, France, August 2004. ESLLI 2004.
- [37] Kiyong Lee, Lou Burnard, Laurent Romary, Eric de la Clergerie, Ulrich Schäfer, Thierry Declerck, Syd Bauman, Harry Bunt, Lionel Clément, Tomaz Erjavec, Azim Roussanaly, and Claude Roux. Towards an international standard on feature structure representation (2). In *Proceedings of the LREC-2004 workshop on A Registry of Linguistic Data Categories within an Integrated Language Resources Repository Area*, pages 63–70, Lisbon, Portugal, May 2004. ELRA.
- [38] Ulrich Schäfer. WHAT: An XSLT-based infrastructure for the integration of natural language processing components. In *Proceedings of the Workshop on the Software Engineering and Architecture of Language Technology Systems (SEALTS), HLT-NAACL03, May 31*, pages 9–13, Edmonton, Canada, 2003.
- [39] Christian Bering, Witold Drozdowski, Gregor Erbach, Clara Guasch, Petr Homola, Sabine Lehmann, Li Hong, Hans-Ulrich Krieger, Jakub Piskorski, Ulrich Schäfer, Atsuko Shimada, Melanie Siegel, Feiyu Xu, and Dorothee Ziegler-Eisele. Corpora and evaluation tools for multilingual named entity grammar development. In *Proceedings of Multilingual Corpora Workshop at Corpus Linguistics*, pages 42–52, Lancaster, UK, March 2003.
- [40] Hans-Ulrich Krieger and Ulrich Schäfer. TDL – A type description language for unification-based grammars. In Harold Boley, François Bry, and Ulrich Geske, editors, *Proceedings of the Workshop on “Neuere Entwicklungen der deklarativen KI-Programmierung”*, number RR-93-35 in Research Report, pages 67–82, Saarbrücken, Germany, 1993. DFKI.

Techreports, Misc.

- [41] Ulrich Schäfer. *Integrating Natural Language Processing Components with XML and XSLT: Representations and Hybrid Architectures*. VDM Verlag Dr. Müller, Saarbrücken, Germany, April 2008. Book.
- [42] Günter Neumann and Ulrich Schäfer. Whiteboard – Eine XML-basierte Architektur für die Analyse natürlichsprachlicher Texte. In Stefan Jänichen, editor, *Proceedings of Online 2002, 25th European Congress Fair for Technical Communication*, volume C, pages 635.01–635.12, Düsseldorf, Germany, 2002. Online GmbH Kongresse und Messen für technische Kommunikation.
- [43] Hans-Ulrich Krieger and Ulrich Schäfer. Efficient parameterizable type expansion for typed feature formalisms. Research Report RR-95-18, DFKI, Saarbrücken, Germany, 1995.
- [44] Hans-Ulrich Krieger and Ulrich Schäfer. TDL – A type description language for HPSG. Part 2: User guide. Document D-94-14, DFKI, Saarbrücken, 1994.
- [45] Hans-Ulrich Krieger and Ulrich Schäfer. TDL – A type description language for HPSG. Part 1: Overview. Research Report RR-94-37, DFKI, Saarbrücken, 1994.
- [46] Hans-Ulrich Krieger and Ulrich Schäfer. TDL ExtraLight user’s guide. Document D-93-09, DFKI, Saarbrücken, Germany, 1993.

3 Research Projects

- 2009 – 2011: TAKE (BMBF) – *Technologies for Advanced Knowledge Extraction for Aggregating Semantic Models of Human Life and Thought*
- 2008 – : M2CI (DFG) – *Cluster of Excellence Multimodal Computing and Interaction: Robust, Efficient and Intelligent Processing of Text, Speech, Visual Data and High Dimensional Representations*

- 2006 – 2008: HyLaP (BMBF) – *Hybrid language processing technologies for a personal associative information access and management application*
- 2007 – 2008: Language Grid (NICT) – *Integrating Language Grid and Heart of Gold as Hybrid Language Processing Service*
- 2003 – 2005: QUETAL (BMBF) – *Multilingual Hybrid Question Answering*
- 2002 – 2004: DeepThought (EU) – *Hybrid Deep and Shallow Methods for Knowledge-intensive Information Extraction*
- 2000 – 2002: Whiteboard (BMBF) – *Multilevel Annotation for Dynamic Free Text Processing*

4 Systems

- ACL Anthology Searchbench – *Semantic, Fulltext and Bibliographic Search in the ACL Anthology:*
<http://take.dfki.de>
- Scientist’s Workbench – *Semantic Search and Citation-based Visual Navigation in Scientific Papers:*
<http://hylap.dfki.de>
- SProUT – *Shallow Processing with Unification and Typed Feature Structures:*
<http://sprout.dfki.de>
- Heart of Gold – *Middleware for Combining Deep and Shallow NLP Components:*
<http://heartofgold.dfki.de>

5 Courses

- Course taught at ESSLLI-2009, the 21st *European Summer School in Logic, Language and Information on Standard XML Query Languages for Natural Language Processing*, July 20-31, 2009, Bordeaux, France.
- Invited course taught at the 12th *ELSNET European Summer School on Language and Speech Communication – Information Fusion in Natural Language Systems on XML-based Integration of Natural Language Processing Components*, July 3-14, 2006, Computer Science Department, University of Hamburg, Germany.

6 Lectures and Seminars

- Wintersemester 2011/2012: Projektseminar *Natural Language Processing for Scientific Discourse*
- Wintersemester 2011/2012: *Java II*
- Wintersemester 2010/2011: Projektseminar *Parsing Scientific Texts in the Mathematics and Computational Linguistics Domain*
- Wintersemester 2010/2011: *Java II*
- Wintersemester 2009/2010: Projektseminar *NLP + ML for Science Information Systems*
- Wintersemester 2009/2010: *Java II*
- Wintersemester 2009/2010: *Introduction to Python* (part of the Preparatory Course for new Master Students; second week)

- Wintersemester 2008/2009: *Java II*
- Wintersemester 2008/2009: *Introduction to Python* (part of the Preparatory Course for new Master Students; second week)
- Wintersemester 2007/2008: *Java II*
- Wintersemester 2006/2007: *Java II*
- Wintersemester 2005/2006: *Java II*
- Wintersemester 2004/2005: *Java II*

7 Program Committees

- ACL 2012 – Main Conference Workshop “*Rediscovering 50 Years of Discoveries*”, July 8–14, 2012, Jeju, Korea.
- PACLING 2007 – *Conference of the Pacific Association for Computational Linguistics*, September 19-21, 2007, Melbourne, Australia.
- LREC 2008 Workshop on *Partial Parsing: Between Chunking and Deep Parsing*, June 1, 2008, Marrakech, Morocco.

8 Reviews

- CSLI Publications: *Collaborative Language Engineering – A Case Study in Efficient Grammar-Based Processing* (book)
- Elsevier journals: *Artificial Intelligence, Knowledge-Based Systems*

9 Supervised Theses

- Daniel Beck: *OnTarget – A Framework for Editing Text Assisted by Visualization and Navigation of Ontology Data*. Bachelor Thesis, Faculty of Mathematics and Computer Science, Saarland University, 2006.
- Daniel Beck: *Hierarchical Classification Using NLP Techniques*. Master’s Thesis, Faculty of Mathematics and Computer Science, Saarland University, 2007.
- Uwe Kasterka: *Ein generisches Framework für die Visualisierung von und Navigation in getypten Zitationsgraphen*. Bachelor Thesis, Faculty of Mathematics and Computer Science, Saarland University, 2010.
- The Nghia Pham: *NLP-based Extraction of Ontology Information from Scientific Papers on Language Technology*. Master’s Thesis, Computational Linguistics Department, Saarland University, 2011.
- Melanie Reiplinger: *Automatic Glossary Extraction from Scientific Papers*. Bachelor Thesis, Computational Linguistics Department, Saarland University, 2011.
- Benjamin Weitz: *Ein grafischer Zitationsbrowser für die ACL Anthology*. Bachelor Thesis, Computational Linguistics Department, Saarland University, 2011.
- Øyvind Raddum Berg: *High precision text extraction from PDF documents*. 1-year Master’s Thesis, University of Oslo, Department of Informatics, 2011. (my rôle: external reviewer)