Joachim Daiber, PhD

Natural Language Processing, Machine Learning

159A Bartlett Street
San Francisco, CA 94110
650-334-7478
daiber.joachim@gmail.com

Personal Details

Born: April 23, 1987 in Bad Waldsee, Germany. Nationality: German.

Employment

since 06/2019 **Tech Lead, Query Understanding at Siri Knowledge**
Machine Learning and Natural Language Processing
Apple, Inc.
1 Apple Park Way, Cupertino.

since 01/2018 **Machine Learning Engineer at Siri Knowledge**
Machine Learning and Natural Language Processing
Apple, Inc.
1 Apple Park Way, Cupertino.

Education

10/2013 – 10/2017 **PhD in Machine Translation and Natural Language Processing**
Statistical Language Processing and Learning Lab.
ILLC, University of Amsterdam, The Netherlands.
Graduation (PhD defense): March 20, 2018.

10/2011 – 09/2013 **MSc and MA in Natural Language Processing**
Grade: *Summa cum laude*, Erasmus Mundus double degree program
Charles University in Prague, Czech Republic (Master of Science)
University of Groningen, The Netherlands (Master of Arts)

10/2007 – 09/2011 **BSc in Computer Science, English Philology**
Free University of Berlin, Germany

Scholarships and Awards

2017 **O-1 Extraordinary Ability Visa**
United States of America

2013 – 2015 **Marie Skłodowska-Curie Fellowship**
EXPERT ITN, European Commission

2011 – 2013 **Erasmus Mundus Scholarship**
European Commission

2011 – 2012 **Charles University Merit Scholarship**
Charles University in Prague (prospěchové stipendium)

Languages and Programming

Languages — German (native), English (fluent), Spanish (conversational), Dutch (basic).

Advanced — Scala, Java, Python/NumPy, Bash, LaTeX, JavaScript, CSS.

Intermediate to Basic — Perl, Haskell, PHP, TensorFlow, C, C++, Objective-C, Go.

Professional Experience (R&D)

03 – 05/2017 **PhD Intern, Lattice Data/Apple**
Menlo Park, USA. 3 months, full time.

08/2016 **PhD Intern, Unbabel**
Lisbon, Portugal. 1 month, full time.

2010 – 2011 **Research Assistant in the Language Technology Lab**
German Research Center for AI. Berlin, Germany. 16 months, part time.

10 – 12/2010 **Intern, Natural Language Processing**
vionto GmbH, Berlin. 3 months, full time.

2012 **Participant, Google Summer of Code 2012**
Organization: DBpedia Spotlight.
Project: *Implementation of an efficient probabilistic model for entity linking.*

Teaching and Mentorship

2014 – 2015 **Teaching Assistant, Natural Language Processing 1**
Prof. Ivan Titov, two iterations of the course.

2014 – 2016 **Teaching Assistant, Natural Language Processing 2**
Prof. Khalil Sima'an and Dr. Wilker Aziz, three iterations of the course.

2016 **Co-supervision of Master thesis (Master of AI)**
Title: *Syntax-Based Markov Models for Word Alignment*; with Prof. Khalil Sima'an.

2015 **Supervision of student project (Master of AI)**
Title: *Splitting German Compounds with Word Embeddings*; with Dr. Stella Frank.

2014 – 2015 **Mentor, Google Summer of Code**
Various information extraction and NLP projects for DBpedia Spotlight.

Summer Schools and Research Visits

09/2015 **Machine Translation Marathon 2015**
Proposed and lead project: *Better unsupervised processing of compound words.*
Prague, Czech Republic.

08/2014 **European Summer School in Logic, Language and Information**
Tübingen, Germany.

07/2012 **Lisbon Machine Learning Summer School (LxMLS)**
Lisbon, Portugal.

05 – 07/2016 **Visiting Researcher, Dublin City University**
Dublin, Ireland. 3 month, full time.

08/2015 – 09/2015 **Visiting Researcher, University of Saarbrücken**
Saarbrücken, Germany. 2 month, full time.

2012 **Visiting Researcher, DBpedia**
Web-based Systems Group (DBpedia project), Free University of Berlin.

Selected Publications (see Google Scholar)

Books

Joachim Daiber. *Typologically robust statistical machine translation: Understanding and exploiting differences and similarities between languages in machine translation*. ILLC dissertation series DS-2018-05. University of Amsterdam, 2018. ISBN 978-94-028-0947-3

Peer-Reviewed Articles: Machine Translation

Joachim Daiber, Miloš Stanojević, and Khalil Sima'an. Universal reordering via linguistic typology. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics*, Osaka, Japan, Dec 2016b

Joachim Daiber, Miloš Stanojević, Wilker Aziz, and Khalil Sima'an. Examining the relationship between preordering and word order freedom in machine translation. In *Proceedings of the First Conference on Machine Translation, Volume 1: Research Papers*, pages 118–130, Berlin, Germany, Aug 2016a

Joachim Daiber and Khalil Sima'an. Machine translation with source-predicted target morphology. In *Proceedings of the 15th Machine Translation Summit (MT Summit 2015)*, pages 283–296, Miami, USA, 2015

Joachim Daiber, Lautaro Quiroz, Roger Wechsler, and Stella Frank. *Proceedings of the 1st Deep Machine Translation Workshop*, chapter Splitting Compounds by Semantic Analogy, pages 20–28. ÚFAL MFF UK, 2015

Joachim Daiber and Khalil Sima'an. *Proceedings of the 1st Deep Machine Translation Workshop*, chapter Delimiting Morphosyntactic Search Space with Source-Side Reordering Models, pages 29–38. ÚFAL MFF UK, 2015

Peer-Reviewed Articles: Entity Extraction

Joachim Daiber, Max Jakob, Chris Hokamp, and Pablo N. Mendes. Improving efficiency and accuracy in multilingual entity extraction. In *Proceedings of the 9th International Conference on Semantic Systems*, I-SEMANTICS '13, pages 121–124, New York, NY, USA, 2013. ACM

Pablo N. Mendes, Joachim Daiber, Rohana Rajapakse, Felix Sasaki, and Christian Bizer. Evaluating the impact of phrase recognition on concept tagging. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC'12)*, Istanbul, Turkey, may 2012

Pablo N. Mendes, Joachim Daiber, Max Jakob, and Christian Bizer. Evaluating dbpedia spotlight for the tac-kbp entity linking task. In *Proceedings of the TACKBP 2011 Workshop*, volume 116, pages 118–120, 2011

Peer-Reviewed Articles: Dependency Parsing

Joachim Daiber and Rob van der Goot. The Denoised Web Treebank: Evaluating dependency parsing under noisy input conditions. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, May 2016