

Marek Rei

Homepage: www.marekrei.com

Email: marek@marekrei.com

Experience

Senior Research Associate, University of Cambridge

July 2016 - current

My research is on machine learning models and advanced neural network architectures, with a focus on sequence labelling, language modelling and multimodal frameworks. I am part of the ALTA project – an industry collaboration with Cambridge Assessment, with the goal of creating innovative new technologies for language learning and teaching. I supervise MPhil and undergraduate students for their course projects, and teach as part of the MPhil course on Advanced Computer Science.

Research Associate, University of Cambridge

November 2014 - July 2016

Performing research on machine learning solutions for automatically evaluating, analysing and correcting the writing skills of language learners. Technologies I developed during this time include a neural sequence labelling model for detecting word-level writing errors, and an online representation learning framework for improved language modelling. I collaborated closely with my team members, and our work has led to a number of publications and a practical free tool called Write & Improve. I also supervised two MPhil students for their 5-month long course projects, involving novel research and culminating in a thesis.

Visiting Lecturer, University of Tartu

March 2015 - June 2015

As a visiting lecturer, I created and delivered a new course on machine learning and language modelling. The lectures cover topics such as n-gram language modelling, smoothing techniques, language model evaluation, neural network language models, recurrent models, backpropagation, gradient descent, representation learning, and an overview of current state-of-the-art methods for language modelling. The course was attended by Master's and PhD students, and also attracted interest from academics.

Language Research Engineer, SwiftKey

November 2012 - November 2014

As a member of the Research team in SwiftKey, our goal was to focus on researching potential future technologies. I was able to apply a variety of natural language processing techniques and worked on different machine learning tasks, such as language modelling, clustering algorithms, distributional similarity measures, artificial neural networks, classifiers and collaborative filtering. My work produced a patent application on efficient neural network models for text prediction on mobile devices.

One of the main projects I worked on was the neural network version of SwiftKey, which was recently released to very positive reviews. Since then, SwiftKey has been acquired by Microsoft, largely based on the strength of its technology.

Supervision Tutor, University of Cambridge

January 2011 - March 2012

I organised tutoring sessions for undergraduate students in the subject of Information Retrieval. This involved regular tutorials in small groups, and providing them with theoretical and practical homework

assignments.

Public Relations Manager, Board of European Students of Technology

September 2007 - July 2008

Organising BEST General Assembly 2008 - an international student congress with representatives from 76 universities. Handled relations with the media, companies and universities, and designed various promotional materials. Worked in a core team of 5 people together with over 40 volunteers.

Software Analyst, Webmedia

June 2007 - August 2007

Internship as a software analyst. Worked on enhancing a self-service system for a mobile service provider, designed and performed usability tests and quality control, wrote specifications and created a prototype for a new user interface.

Project manager, Board of European Students of Technology

October 2006 - May 2007

Main organiser of Key to the Future (Võti Tulevikku) 2007, a national career project for graduates in Estonia with 74 participating companies. As the project manager I led a core team of 5 people plus many volunteers, and was involved in every aspect of the project.

Software Developer, Movies.ee

June 2006 - April 2007

Developed a custom-built content management system for managing movies, actors, news, galleries, contests and live cinema timetables.

Software Developer at Geomedia

February 2006 - June 2006

Developed software for conducting large-scale population surveys.

Education

PhD, University of Cambridge

2009 - 2013

My thesis is about minimally supervised dependency-based methods for natural language processing, under the supervision of Prof Ted Briscoe. I worked on topics such as distributional and semantic similarity, speculation detection, entailment detection, and unsupervised parse reranking. I applied various machine learning techniques and developed systems that achieved state-of-the-art performance.

MPhil, University of Cambridge

2008 - 2009

MPhil in Computer Speech, Text and Internet Technologies. Covered topics such as speech recognition, speech synthesis, parsing, discourse, dialogue systems, machine translation and question answering.

BSc, Tallinn University of Technology

2005 - 2008

Bachelor's Degree in Computer Science from the Tallinn University of Technology. Specialised in Network

Applications and Logic, graduated *cum laude*.

Program Committee

I am part of the program committee for the following conferences and workshops:

- The Sixth Joint Conference on Lexical and Computational Semantics (*SEM 2017)
- The 12th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2017)
- Workshop on Evaluating Vector Space Representations for NLP 2017 (RepEval 2017)
- Workshop on Subword and Character Level Models in NLP 2017 (SCLeM 2017)
- Workshop on Evaluating Vector Space Representations for NLP 2016 (RepEval 2016)
- The 11th Workshop on Innovative Use of NLP for Building Educational Applications (BEA 2016)

Reviewing

I review for the following journals, conferences and workshops:

- Journal of Natural Language Engineering (JNLE)
- The Association for Computational Linguistics (ACL)
- The European Chapter of the Association for Computational Linguistics Conference (EACL)
- Conference on Empirical Methods in Natural Language Processing (EMNLP)
- The Joint Conference on Lexical and Computational Semantics (*SEM)
- The SIGNLL Conference on Computational Natural Language Learning (CoNLL)
- Workshop on Innovative Use of NLP for Building Educational Applications (BEA)
- Workshop on Metaphor in NLP
- Workshop on Evaluating Vector Space Representations for NLP (RepEval)

Supervision

Students whom I supervise for their thesis or course project:

Domain Adaptation for Neural Named Entity Recognition

Stefan Hosein

MPhil in Advanced Computer Science, University of Cambridge, to finish in 2017.

Hybrid Language Modeling for Text Prediction

Devan Kuleindiren

Computer Science Tripos, University of Cambridge, to finish in 2017.

Convolutional Neural Networks for Automated Essay Assessment

Younna Farag

MPhil in Advanced Computer Science, University of Cambridge, 2016.

Predicting Word Difficulty for Lexical Acquisition

Shushan Arakelyan

*MPhil in Advanced Computer Science, University of Cambridge, 2016.***Awards, Grants and Scholarships**

Language Sciences Research Incubator Fund, 2017

EPSRC Doctoral Training award, 2009-2012

Cambridge European Trust award, 2009-2012

Computer Laboratory Local Industry Fund, 2009-2012

EPSRC Collaborative Training award, 2008

Cambridge European Trust award, 2008

AS Elion Scholarship, 2007

Book Chapters**Intelligent Information Access from Scientific Papers**

Ted Briscoe, Karl Harrison, Andrew Naish-Guzman, Andy Parker, Marek Rei, Advait Siddharthan, David Sinclair, Mark Slater and Rebecca Watson

*Current Challenges in Patent Information Retrieval, edited by Mihai Lupu, Katja Mayer, John Tait and Anthony J. Trippe, 2011.***Peer-reviewed Publications****Semi-supervised Multitask Learning for Sequence Labeling**

Marek Rei

*In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL-2017)***Attending to Characters in Neural Sequence Labeling Models**

Marek Rei, Sampo Pyysalo and Gamal K.O. Crichton

*In Proceedings of the 26th International Conference on Computational Linguistics (COLING-2016)***A Joint Model for Word Embedding and Word Morphology**

Kris Cao and Marek Rei

*In Proceedings of the 1st Workshop on Representation Learning for NLP (RepL4NLP-2016)***Compositional Sequence Labeling Models for Error Detection in Learner Writing**

Marek Rei and Helen Yannakoudakis

*In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL-2016)***Automatic Text Scoring Using Neural Networks**

Dimitrios Alikaniotis, Helen Yannakoudakis and Marek Rei

*In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (ACL-2016)***Sentence Similarity Measures for Fine-Grained Estimation of Topical Relevance in Learner Essays**

Marek Rei and Ronan Cummins

In Proceedings of the 11th Workshop on Innovative Use of NLP for Building Educational Applications (BEA), 2016

Online Representation Learning in Recurrent Neural Language Models

Marek Rei

*In Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing (EMNLP), 2015.***Looking for hyponyms in vector space**

Marek Rei and Ted Briscoe

*In Proceedings of the Eighteenth Conference on Computational Natural Language Learning (CoNLL-14), 2014.***Parser lexicalisation through self-learning**

Marek Rei and Ted Briscoe

*In Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), 2013.***Unsupervised Entailment Detection between Dependency Graph Fragments**

Marek Rei and Ted Briscoe

*In Proceedings of the 2011 Workshop on Biomedical Natural Language Processing (BioNLP-11).***Combining Manual Rules and Supervised Learning for Hedge Cue and Scope Detection**

Marek Rei and Ted Briscoe

The 14th Conference on Natural Language Learning, 2010.

Theses

Minimally Supervised Dependency-based Methods for Natural Language Processing

PhD thesis, University of Cambridge

*Cambridge, United Kingdom, 2013***Adaptive Interactive Information Extraction**

MPhil thesis, University of Cambridge

*Cambridge, United Kingdom, 2009***Audio-visual Speech Synthesis**

Bachelor's thesis, Tallinn University of Technology

Tallinn, Estonia, 2008

Invited Talks

Detecting Writing Errors with Neural Networks*Machine Learning for Further Language Learning, summer school in Chania, Greece. July 2017***Neural Architectures for Sequence Labelling***Apple Cambridge. June 2017***Neural Architectures for Sequence Labelling***University College London. May 2017***Neural Architectures for Sequence Labelling***University of Cambridge. May 2017***Character-based and Multi-task Learning for Neural Sequence Labeling***University of Sheffield. March 2017*

Programming Skills

Main experience: Python, Java, PHP, SQL, HTML, CSS, JavaScript

Limited experience: C++, C, Android, Clojure, Perl, JSP, Lisp, Prolog, Haskell

Selected Projects

Encode Explorer: An open-source PHP script for displaying and managing files in the web server. It has been downloaded over 55,000 times.

Website: encode-explorer.siineiolekala.net

Sequence Labeler

An open-source framework for training sequence labeling models using neural network architectures. It can be used for tasks such as named entity recognition, POS-tagging, chunking, and error detection.

Website: github.com/marekrei/sequence-labeler

Tonu: A system for displaying a 3-dimensional talking head with realistic mouth movements, designed to integrate with a speech synthesiser for people with disabled hearing or noisy work environments.

Website: tonu.siineiolekala.net

Nifty Events: An event recommendation system. The application retrieves various upcoming events, using machine learning and neural networks to rank them according to the interests of each specific user. It can also find events that multiple users might like to attend together, based on Facebook integration.