

## **Jana Straková (Kravalová)**

**Work Position:** research assistant at Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University

**Work Address:** ÚFAL MFF UK, Malostranské náměstí 25, Prague 1, 11800, Czech Republic

**Phone:** +420 720 453 696

**E-mail:** [strakova@ufal.mff.cuni.cz](mailto:strakova@ufal.mff.cuni.cz)

**WWW:** <http://ufal.mff.cuni.cz/jana-strakova>

### **Education:**

- 2009 – 2017: graduated (Ph.D.) in Computer Science from Institute of Formal and Applied Linguistics, Faculty of Mathematics and Physics, Charles University, with thesis "Neural Network Based Named Entity Recognition",
- 2007 – 2009: graduated (master's degree) in Computer Science from Faculty of Mathematics and Physics, Charles University, with thesis "Syntax Based Information Retrieval",
- 2004-2007: graduated (bachelor's degree) in Computer Science from Faculty of Mathematics and Physics, Charles University, with thesis "Word Alignment using Minimum Weight Edge Cover".

### **Work Experience and Internships:**

- since 2009: research assistant at Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University in Prague,
- October – November 2010: internship at Institute of Cognitive Science, University of Colorado at Boulder, Boulder, USA,
- January – March 2010: internship at Max-Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany,
- July – September 2009: internship at Google London as software engineer intern, London, UK,
- June – September 2008: internship at Google Zurich as software engineer intern, Zurich, Switzerland.

### **Main Research Interests and Skills:**

- advanced knowledge of newest methods of NLP deep learning,
- research and development of several NLP models with focus on tagging and named entity recognition using artificial neural networks,
- co-author of open-source tools MorphoDiTa (tokenization, lemmatization and POS tagging tool), NameTag (named entity recognition) and UDPipe (NLP pipeline),
- programming skills in C++, Python, Perl, Java,
- everyday experience in Linux programming.

### **Selected Bibliography:**

- Agirre Eneko, Alfonseca Enrique, Hall Keith Brendan, Kravalová Jana, Pasca Marius, Soroa Aitor: A Study on Similarity and Relatedness Using Distributional and WordNet-based Approaches. In: Proceedings of NAACL-HLT 09, Copyright © Association for Computational Linguistics, Boulder, CO, USA, ISBN 978-1-932432-41-1, pp. 19-27, 2009,
- Straková Jana, Straka Milan, Hajič Jan: Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, Copyright © Association for Computational Linguistics, Stroudsburg, PA, USA, ISBN 978-1-941643-00-6, pp. 13-18,

2014,

- Straka Milan, Hajič Jan, Straková Jana: UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing. In: Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016), Copyright © European Language Resources Association, Paris, France, ISBN 978-2-9517408-9-1, pp. 4290-4297, 2016.