

# Milan Straka – Curriculum Vitae

## Osobní údaje

Jméno: RNDr. Milan Straka, Ph.D.  
Datum narození: 11. dubna 1984, Plzeň  
Národnost: česká  
Jazykové znalosti: angličtina (FCE), němčina (pasivně)  
Adresa: Liliová 1514  
Strakonice, 38601  
Česká republika  
E-mail: milan@strakovi.com  
Telefonní číslo: +420 605 435 586



## Vzdělání

- doktorské studium (Ph.D.) v oboru Diskrétní modely a algoritmy dokončeno v září 2013 na Matematicko-fyzikální fakultě Univerzity Karlovy; disertační práce na téma Funkcionální datové struktury a algoritmy
- magisterské studium (Mgr) v oboru Matematické metody počítačové bezpečnosti dokončeno v červnu 2008 na Matematicko-fyzikální fakultě Univerzity Karlovy; všechny zkoušky absolvovány za jedna
- bakalářské studium (Bc) v oboru Obecná informatika dokončeno v červnu 2006 na Matematicko-fyzikální fakultě Univerzity Karlovy; všechny zkoušky absolvovány za jedna
- čtyřikrát v řadě přiděleno prospěchové stipendium

## Pracovní a výzkumné zkušenosti

- odborný asistent (dříve výzkumný pracovník) na Ústavu formální a aplikované informatiky, Matematicko-fyzikální fakulty University Karlovy, od roku 2013
  - výzkum v oblasti hlubokého učení v NLP
  - implementace NLP nástrojů (tagger, lemmatizátor, parser, rozpoznávač pojmenovaných entit, morfologický derivátor)
  - výuka kurzů hlubokého učení a distribuovaného programování
  - vedení diplomových prací na téma strojové učení
  - spoluorganizace CoNLL 2017 shared task in UD parsing
  - smluvní výzkum pro MVČR v roce 2017 (vyhledávání pojmenovaných entit)
- instalace výpočetního clusteru Hadoop a školení o něm, pro ÚFAL MFF UK, leden 2017
  - Hadoop nahrazen clusterem Apache Spark v roce 2014

## Mezinárodní výzkumné pobyty a stáže

- výzkumný pobyt v Google, Zurich, září 2017–březen 2018
  - hluboké učení při zpracování přirozeného jazyka – konverznacionalizace odpovědí asistenta Google Home
- stážista v Microsoft Research Labs, Cambridge, březen–květen 2010
  - práce na funkcionálních datových strukturách v Haskellu a překladači GHC
- výzkumný pobyt na Simon Fraser University, Kanada, únor 2009
  - práce na plně perzistentních datových strukturách
- dvakrát stážista v Google, Zurich, červenec–září 2009 a červen–září 2008
  - distribuované výpočty pomocí MapReduce
  - česká morfologická analýza a české jazykové modely

## Výuka

- přednáška Hluboké učení (přednášky a cvičení v TensorFlow)
  - navazující Seminář z hlubokého učení
- seminář Distribuované zpracování rozsáhlých dat (pomocí Spark a SGE)

- seminář Pokročilé techniky funkcionálního programování (jazyk Haskell)
- seminář Programovací jazyky OCaml a F#
- cvičení k přednáškám
  - programování pro pokročilé
  - algoritmy a datové struktury I a II

## Vědecké články

- Milan Straka, Jana Straková and Jan Hajič: **Prague at EPE 2017: The UDPipe System**. In Proceedings of the 2017 Shared Task on Extrinsic Parser Evaluation at the Fourth International Conference on Dependency Linguistics and the 15th International Conference on Parsing Technologies, Pisa, Italy, September 2017.
- Milan Straka and Jana Straková: **Tokenizing, POS Tagging, Lemmatizing and Parsing UD 2.0 with UDPipe**. In Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, Vancouver, Canada, August 2017.
- Daniel Zeman, Martin Popel, Milan Straka, Jan Hajič, Joakim Nivre et al.: **CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies**. In Proceedings of the CoNLL 2017 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, Vancouver, Canada, August 2017.
- Natalia Klyueva and Antoine Doucet and Milan Straka: **Neural Networks for Multi-Word Expression Detection**. In Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2017, Workshop on Multiword Expressions), Valencia, Spain, April 2017.
- Jana Straková, Milan Straka, Magda Ševčíková, and Zdeněk Žabokrtský: **Czech Named Entity Corpus**. In The Handbook of Linguistic Annotation, editors Nancy Ide and James Pustejovsky, ISBN 9402408797, Springer, Netherlands, 2017.
- Ševčíková Magda, Žabokrtský Zdeněk, Vidra Jonáš, Straka Milan. **Lexikální síť DeriNet: elektronický zdroj pro výzkum derivace v češtině**. In Časopis pro moderní filologii, Vol. 98, No. 1, Charles University, Prague, Czech Republic, ISSN 0008-7386, pp. 62-76, Sep 2016.
- Jana Straková, Milan Straka and Jan Hajič: **Neural Networks for Featureless Named Entity Recognition in Czech**. In Proceedings of the 19th International Conference on Text, Speech and Dialogue (TSD 2016), Brno, Czech Republic, September 2016.
- Milan Straka, Jan Hajič and Jana Straková: **UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing**. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia, May 2016.
- Zdeněk Žabokrtský, Magda Ševčíková, Milan Straka, Jonáš Vidra and Adéla Limburská: **Merging Data Resources for Inflectional and Derivational Morphology in Czech**. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia, May 2016.
- Milan Straka, Jan Hajič, Jana Straková and Jan Hajič jr: **Parsing Universal Dependency Treebanks using Neural Networks and Search-Based Oracle**. In Proceedings of the Fourteenth International Workshop on Treebanks and Linguistic Theories (TLT 14), December 2015.
- Jana Straková, Milan Straka and Jan Hajič: **Open-Source Tools for Morphology, Lemmatization, POS Tagging and Named Entity Recognition**. In Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics (ACL 2014): System Demonstrations, pages 13-18, Baltimore, Maryland, June 2014.
- Jana Straková and Milan Straka: **A New State-of-the-Art Czech Named Entity Recognizer**. In Text, Speech and Dialogue (TSD 2013), Pilsen, Czech Republic, Sep 2013.
- David Mareček and Milan Straka: **Stop-probability estimates computed on a large corpus improve Unsupervised Dependency Parsing**. In Annual Meeting of the Association for Computational Linguistics (ACL 2013), Sofia, Bulgaria, August 2013.

- Milan Straka: *Adams' Trees Revisited – Correct and Efficient Implementation*. In Proceedings of Symposium on Trends in Functional Programming (TFP 2011), Madrid, Spain, May 2011.
- Milan Straka: *The performance of the Haskell containers package*. In Proceedings of Haskell 2010, 3rd ACM Haskell symposium on Haskell, Baltimore, Maryland, Sep 2010.
- Milan Straka: *Optimal worst-case fully persistent arrays*. In Symposium on Trends in Functional Programming (TFP 2009), Komarno, Slovakia, June 2009.
- Martin Mareš, Milan Straka: *Linear-Time Ranking of Permutations*. In Proceedings of ESA 2007, 15th Annual European Symposium, Eilat, Israel, October 2007.

## Závěrečné práce

- doktorská práce: *Funkcionální datové struktury a algoritmy*
- diplomová práce: *Kryptografie založená na kvadratických tělesech*
- bakalářská práce: *Faktorizace polynomů nad konečnými tělesy*

## Informatické znalosti

- programovací schopnosti
  - výborné programovací schopnosti v C++, Python, Perl, Java, C#
  - pokročilé znalosti funkcionálních jazyků, především Haskell, F#, OCaml
  - každodenní zkušenosti s programováním pro Linux a Windows
  - pokročilé znalosti MapReduce paradigmatu a frameworku Spark
  - rozsáhlé zkušenosti s rodinou protokolů TCP/IP
- strojové učení
  - výborná znalost nejnovějších metod hlubokého učení v oblasti zpracování přirozeného jazyka
  - několik open-source implementací NLP nástrojů (tokenizér, tagger, rozpoznávač pojmenovaných entit, parser) založených na hlubokých neuronových sítích, naimplementovaných od základu v C++
  - výborná znalost TensorFlow
  - bayesovské neparametrické modelování
- teoretická informatika
  - algoritmizace a pokročilé grafové algoritmy (vážené perfektní párování, micro/macrotree dekompozice, LCA, testování rovinnosti)
  - pokročilé datové struktury (intervalové, suffixové, Sleator-Tarjanovy stromy; Q-haldy; perzistentní, samovyvažující a amortizované datové struktury)
  - pokročilé znalosti diskrétní matematiky a teorie grafů
  - pravděpodobnostní algoritmy (perfektní hešování, náhodné procházky, derandomizace)
- počítačová grafika
  - zkušenosti s hardwarově akcelerovanou 3D grafikou, především OpenGL and GLSL
- informační bezpečnost
  - symetrické šifry, hešovací funkce, kryptografie s veřejným klíčem (RSA, DSA)
  - algebraická teorie čísel (testování prvočíselnosti, faktorizace, kryptoanalýza)

## Umístění na programovacích soutěžích

- zlatá medaile (14. z 269 soutěžících) na mezinárodní olympiádě International Olympiad in Informatics v roce 2003
- čtyřikrát první místo v národním kole ACM Intern International Collegiate Programming Contest v letech 2004–2007, druhé místo v roce 2003
- čtvrté, páté, páté a šesté místo z 50, 53, 62 a 67 účastníků na střeoevropské soutěži Central European Programming Contest of the ACM International Collegiate Programming Contest v letech 2004–2007
- první nebo druhé místo v šesti národních kolech programovacích soutěží v letech 1999–2003

- první místo z 59 účastníků v Korespondenčním semináři z programování pořádaným Matematicko-fyzikální fakultou University Karlovy v roce 2003
- účast na mezinárodní olympiádě International Olympiad in Informatics v roce 2002
- stříbrná medaile (9. místo) na středoevropské olympiádě Central European Olympiad in Informatics v roce 2002