

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
ORCID: 0000-0003-3295-5576



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 a 2018 shared tasku v UD parsingu
 - spoluorganizace CoNLL 2019 shared tasku MRP
 - 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 – konverzacionalizace 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

- přednáška Hluboké zpětnovazební učení (přednášky a cvičení v TensorFlow)
- přednáška Strojové učení pro zelenáče
- 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

Vybrané vědecké články za posledních 5 let

- David Samuel, Milan Straka: **ÚFAL at MRP 2020: Permutation-invariant Semantic Parsing in PERIN**. In Proceedings of the CoNLL 2020 Shared Task: Cross-Framework Meaning Representation Parsing, CoNLL 2020.
- Kondratyuk Daniel, Straka Milan: **75 Languages, 1 Model: Parsing Universal Dependencies Universally**. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing, EMNLP 2019, Hongkong, China, 2019.
- Straka Milan, Straková Jana: **ÚFAL MRPipe at MRP 2019: UDPipe Goes Semantic in the Meaning Representation Parsing Shared Task**. In Proceedings of the CoNLL 2019 Shared Task: Cross-Framework Meaning Representation Parsing, CoNLL 2019, Hongkong, China, 2019.
- Jana Straková, Milan Straka, Jan Hajič: **Neural Architectures for Nested NER through Linearization**. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics, ACL, Florence, Italy, 2019.
- Straka Milan, Straková Jana, Hajič Jan: **UDPipe at SIGMORPHON 2019: Contextualized Embeddings, Regularization with Morphological Categories, Corpora Merging**. In Proceedings of the 16th SIGMORPHON Workshop on Computational Research in Phonetics, Phonology, and Morphology, SIGMORPHON 2019, Florence, Italy, 2019.
- Milan Straka: **UDPipe 2.0 Prototype at CoNLL 2018 UD Shared Task**. In Proceedings of CoNLL 2018: The SIGNLL Conference on Computational Natural Language Learning, ACL, Stroudsburg, PA, USA, 2018.
- Daniel Zeman, Jan Hajič, Martin Popel, Martin Potthast, Milan Straka, Filip Ginter, Joakim Nivre, Slav Petrov: **CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies**. In Proceedings of the CoNLL 2018 Shared Task: Multilingual Parsing from Raw Text to Universal Dependencies, ACL, Stroudsburg, PA, USA, 2018.
- Daniel Zeman, Martin Popel, Milan Straka, Jan Hajic, 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.
- 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.

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 a zpětnovazebního učení
 - 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 střeoevropské olympiádě Central European Olympiad in Informatics v roce 2002

Komunitní aktivity

- organizace programovacích soutěží
 - Mistrovství České republiky v programování, od roku 2004
 - Matematická olympiáda kategorie P, 2004–2011
 - 4 roky hlavní organizátor Korespondenčního semináře z programování, pořádaného Matematicko-fyzikální fakultou Univerzity Karlovy
 - Střeoevropská olympiáda v programování, Brno 2007
- pořádání táborů s výukou programování
 - soustředění Korespondenčního semináře v programování, 2003–2007
 - Letní soustředění programátorů, 2002–2007