

Biomedical NLP Workshop associated with RANLP 2017

8 September 2017, Varna, Bulgaria



Program

9 ⁰⁰ -10 ⁰⁰	Invited Talk Understanding Medical Terminology Using the Web as a Corpus <i>Preslav Nakov</i>
10 ⁰⁰ -10 ³⁰	Understanding of unknown medical words <i>Natalia Grabar and Thierry Hamon</i>
10 ³⁰ -11 ⁰⁰	Document retrieval and question answering in medical documents. A large-scale corpus challenge. <i>Curea Eric</i>
11 ⁰⁰ -11 ³⁰	Coffee break
11 ³⁰ -12 ⁰⁰	Discourse-Wide Extraction of Assay Frames from the Biological Literature <i>Dayne Freitag, Paul Kalmar and Eric Yeh</i>
12 ⁰⁰ -12 ³⁰	Towards Confidence Estimation for Typed Protein-Protein Relation Extraction <i>Camilo Thorne and Roman Klinger</i>
12 ³⁰ -13 ⁰⁰	Adapting the TTL Romanian POS Tagger to the Biomedical Domain <i>Maria Mitrofan and Radu Ion</i>
13 ⁰⁰ -14 ⁰⁰	Lunch break
14 ⁰⁰ -14 ³⁰	POMELO: Medline corpus with manually annotated food-drug interactions <i>Thierry Hamon, Vincent Tabanou, Fleur Mouglin, Natalia Grabar and Frantz Thiessard</i>
14 ³⁰ -15 ⁰⁰	Classification based extraction of numeric values from clinical narratives <i>Maximilian Zubke</i>
15 ⁰⁰ -15 ³⁰	Identification of Risk Factors in Clinical Texts through Association Rules <i>Svetla Boytcheva, Ivelina Nikolova, Galia Angelova, Zhivko Angelov and Dimitar Tcharaktchiev</i>
15 ³⁰ -16 ⁰⁰	Coffee break
16 ⁰⁰ -16 ³⁰	Entity-Centric Information Access with Human in the Loop for the Biomedical Domain <i>Seid Muhie Yimam, Steffen Remus, Alexander Panchenko, Andreas Holzinger and Chris Biemann</i>
16 ³⁰ -16 ⁵⁵	One model per entity: using hundreds of machine learning models to recognize and normalize biomedical names in text <i>Victor Bellon and Raul Rodriguez-Esteban</i>
16 ⁵⁵ -17 ²⁰	Annotation of Clinical Narratives in Bulgarian language <i>Ivaylo Radev, Kiril Simov, Galia Angelova, and Svetla Boytcheva</i>

Understanding Medical Terminology Using the Web as a Corpus

Abstract: Today, the Web is the biggest available corpus, and, in corpus-based natural language processing, size does matter. In this talk, I will reveal some of the hidden potential of the Web that lies beyond the n-gram, with focus on interpreting noun compounds, which are abundant in medical texts. First, I will present a highly accurate lightly supervised approach based on surface markers and linguistically-motivated paraphrases that yields state-of-the-art results for noun compound bracketing: e.g., "[[liver cell] antibody]" is left-bracketed, while "[liver [cell line]]" is right-bracketed. Second, I will present a simple unsupervised method for mining implicit predicates that can characterize the semantic relations holding between the nouns in noun compounds, e.g., "malaria mosquito" is a "mosquito that carries / spreads / causes / transmits / brings / infects with /... malaria". I will also show how these ideas can be applied to other NLP problems.

Bio: Dr. Preslav Nakov is a Senior Scientist at the Qatar Computing Research Institute, HBKU. His research interests include computational linguistics and natural language processing (for English, Arabic and other languages), machine translation, question answering, fact-checking, sentiment analysis, lexical semantics, Web as a corpus, and biomedical text processing.

Preslav Nakov co-authored a Morgan & Claypool book on Semantic Relations between Nominals, two books on computer algorithms, and many research papers in top-tier conferences and journals. He received the Young Researcher Award at RANLP'2011. He was also the first to receive the Bulgarian President's John Atanasoff award, named after the inventor of the first automatic electronic digital computer.

Preslav Nakov is Secretary of ACL SIGLEX, the Special Interest Group on the Lexicon of the Association for Computational Linguistics. He is also a Member of the Editorial Board of the Journal of Natural Language Engineering, an Associate Editor of the AI Communications journal, and an Editorial Board member of the Language Science Press Book Series on Phraseology and Multiword Expressions. He served on the program committees of the major conferences and workshops in computational linguistics, including as a co-organizer and as an area/publication/tutorial/shared task chair, Senior PC member, student faculty advisor, etc.; he co-chaired SemEval 2014-2016 and was an area co-chair of ACL, EMNLP, NAACL-HLT, and *SEM, a Senior PC member of IJCAI, and a shared task co-chair of IJCNLP'2017.

Preslav Nakov received a PhD degree in Computer Science from the University of California at Berkeley (supported by a Fulbright grant and a UC Berkeley fellowship), and a MSc degree from the Sofia University. He was a Research Fellow at the National University of Singapore, a honorary lecturer in the Sofia University, and a research staff in the Bulgarian Academy of Sciences.