Vuk Batanović

Date of birth: 02.06.1987.
Address: Svetozara Markovića 18, 11000 Belgrade,
Republic of Serbia
Phone number: +381 63 710 2001, +381 11 3231 468
E-mail: vukbatanovic@gmail.com
Website: vukbatanovic.github.io
LinkedIn: linkedin.com/in/vukbatanovic
ResearchGate: researchgate.net/profile/Vuk_Batanovic

Curriculum vitae

Goals and aims

- *Develop NLP solutions for Serbian and other less prominent languages*
- *Apply modern machine learning techniques in NLP research*

Education

- *School of Electrical Engineering, University of Belgrade*

  2012 – present

  PhD studies in natural language processing, **10/10** GPA, Dept. of Software Engineering

  Planned thesis: *A methodology for solving semantic tasks in the processing of short texts written in natural languages with limited resources*

  2010 – 2011

  Master's degree with a **10/10** GPA, Dept. of Computer Science and Engineering

  Thesis: *An expert system for determining the semantic similarity of short texts in Serbian*

  2006 – 2010

  Bachelor's degree with a **9.56/10** GPA, Dept. of Computer Science and Engineering

  Thesis: *A visual simulator of search algorithms*

- *Summer Schools and Seminars*
  - ESSLLI 2016 – *28th European Summer School in Logic, Language and Information*, Free University of Bozen-Bolzano, Italy
  - LxMLS 2016 – *6th Lisbon Machine Learning Summer School*, Instituto Superior Técnico, Portugal
  - ReLDI (*Regional Linguistic Data Initiative*) seminars in Belgrade, Serbia, and Zagreb, Croatia, 2016/2017

- *Online Courses*
  - Natural Language Processing, Stanford University, Coursera
  - Natural Language Processing, Columbia University, Coursera
  - Machine Learning, Stanford University, Coursera

- o Introduction to Natural Language Processing, University of Michigan, Coursera
- o Miracles of Human Language: An Introduction to Linguistics, Universiteit Leiden, Coursera
- o Text Retrieval and Search Engines, University of Illinois at Urbana-Champaign, Coursera
- o Data Mining with Weka, University of Waikato
- o More Data Mining with Weka, University of Waikato

Research

- **Research Projects**
  - o Research collaborator on the ReLDI (*Regional Linguistic Data Initiative*) SNSF SCOPES project 160501
- **Selected Papers (full bibliography available on website)**
  - o Vuk Batanović, Boško Nikolić, Milan Milosavljević: *Reliable Baselines for Sentiment Analysis in Resource-Limited Languages: The Serbian Movie Review Dataset*, in Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC 2016), Portorož, Slovenia, pp. 2688–2696, 2016.
  - o Vuk Batanović, Dragan Bojić: *Using Part-of-Speech Tags as Deep-Syntax Indicators in Determining Short-Text Semantic Similarity*, Computer Science and Information Systems, vol. 12, no. 1, pp. 1–31, 2015.
  - o Bojan Furlan, Vuk Batanović, Boško Nikolić: *Semantic similarity of short texts in languages with a deficient natural language processing support*, Decision Support Systems, vol. 55, no. 3, pp. 710–719, 2013.
- **Created Datasets and Tools**
  - o Serbian Movie Review Dataset (*SerbMR*) – sentiment analysis document collection
  - o SRSentiComments Dataset – sentiment analysis short-text collection
  - o NBSVM-Weka – a multiclass NBSVM implementation for Weka
  - o SCStemmers – a collection of stemmers for Serbian and Croatian
  - o Serbian Paraphrase Corpus (*SRParaphrase*) – semantic similarity short-text collection
  - o LinSTSS and POST STSS - algorithms for measuring the semantic similarity of short texts, particularly suitable for languages with limited access to advanced NLP tools
- **Grants and Scholarships**
  - o ReLDI project grant for the creation of language resources in Serbian
  - o Elsevier grant for ESSLLI 2016 participation
  - o Scholarship of the Fund for Young Talents of the Republic of Serbia

Skills and interests

- **Professional interests**

  Sentiment analysis, semantic similarity, distributional semantics, short-text processing, morphological normalization, NLP for minor languages, multilingual NLP solutions, (deep) machine learning
- **Programming languages**

  Java, Python, C++, C#, C, Matlab, SQL
- **Language proficiency**
  - o **Serbian** (native)
  - o **English** (fluent – *Cambridge Certificate of Proficiency in English (CPE), grade A*)
  - o **French** (limited)