

## SUMMARY

Motivated science lead with more than 10 years of experience in natural language processing and machine learning. Team player dedicated to promoting culture of trust and personal growth. Proven track record of bringing innovation into a product-focused environment.

## PROJECTS

### Clinical Information Extraction System

*Architect and contributor, Babylon Health*

A clinical information system built using the spaCy architecture and leveraging a clinical ontology, bespoke relation extraction rules and cutting edge SRL and coreference resolution models.

### Consultation Notes Generation

*Science lead, Babylon Health*

Generation of clinical consultation notes based on an automatically transcribed consultations.

### Data-Driven Differential Diagnosis

*Science lead, Babylon Health*

Predicting possible diagnosis or treatment based on co-occurrence data using Bayesian models.

### News Analytics

*Co-Architect and contributor, Datamaran*

A news analytics system that ingests and processes news articles as they are published and from a backlog.

## TECHNOLOGY

**GitHub:** <https://github.com/savkov>

- tensorflow, pytorch, transformers
- spaCy, scikit-learn, numpy, pandas
- Kafka, Redis
- continuous integration workflow
- micro-service architecture

## PUBLICATIONS

**Google Scholar:** <http://scholar.sasho.io>

- *Estimating Mutual Information Between Dense Word Embeddings*, Zhelezniak et al. 2020, ACL
- *Correlations between Word Vector Sets*, Zhelezniak et al. 2019, EMNLP
- *Correlation coefficients and semantic textual similarity*, Zhelezniak et al. 2019, NAACL
- *Don't Settle for Average, Go for the Max: Fuzzy Sets and Max-Pooled Word Vectors*, Zhelezniak et al. 2019, ICLR
- *Annotating patient clinical records with syntactic chunks and named entities: the Harvey Corpus*, Savkov et al. 2016, LREC Journal
- *Chunking clinical text containing non-canonical language*, Savkov et al. 2014, BioNLP
- *Linguistic Processing Pipeline for Bulgarian*, Savkov et al. 2012, LREC
- *HPSG-based Bulgarian-English statistical machine translation*, Simov et al. 2012, Littera et Lingua

## LEADERSHIP

- Great at promoting innovation by aligning product goals with science capabilities.
- Proven record of building an NLP team from the ground up and establishing a culture of trust, innovation and growth.

## EXPERIENCE

### Lead NLP Scientist

*Babylon Health, June 2019 to present*

I led the team through a couple of company reorganisations that shifted our focus closer to the product. I am currently leading two strategic innovation initiatives focusing on generation of consultation notes and data-driven triage.

### Senior NLP Scientist

*Babylon Health, Dec 2018 to May 2019*

I led the newly established NLP squad as part of the Research Tribe. We focused on speech-to-text processing, sentence similarity metrics, and maturing and expanding our information extraction stack.

### NLP Scientist

*Babylon Health, Feb 2017 to Nov 2018*

I led the transition of the Information Extraction system from a GATE architecture to a spaCy architecture. In addition, I worked on word sense disambiguation, sentence-level textual similarity and a custom rule engine for annotation processing. I was also coordinating the hiring process for NLP.

### Data Scientist

*eRevalue (now Datamaran), Sep 2015 to Jan 2017*

I was part of a small team that designed and delivered a live news analytics module and a Kafka data ingestion and processing architecture for the Datamaran platform.

### Research Assistant

*Bulgarian Academy of Sciences, Apr 2010 to Dec 2011*

I was part of a group working on BG-EN translation pair as part of an EU-funded science project. I built a morphosyntactic part-of-speech tagger and a dependency parser, and integrated them in a machine translation pipeline.

## EDUCATION

### University of Sussex

*PhD in Clinical NLP, 2012-2015*

My thesis was about information extraction from primary care notes, focusing on data annotation and machine learning applications.

### University of Tübingen

*MA in Computational Linguistics, 2006-2009*

### University of Tübingen

*BA in Computational Linguistics, 2003-2006*