Applied NLP - Data, Models, Pipelines, Business Decisions: Lessons from the field

Łukasz Augustyniak

Wrocław University of Science and Technology (Poland), AVAYA (USA)

Abstract:

Natural Language Processing (NLP) projects often fail in their conception (e.g., lack of data, wrong data, wrong annotations), in their delivery, or their business usefulness. It is a good idea to identify what is the proper research problem, what a client is looking for, what are our limitations, and finally, how you would solve the problem. The applied NLP is still rapidly evolving and changing area - I want to present the experience from a several NLP project from social media analysis, news data extraction, call center analytics, legal documents annotation and many more. What are the best practices for data acquisition, creating the first models, packing all into pipelines, and how we can serve these models? In addition, I show how we may communicate our work beyond the client deliverables to guide other applied NLP practitioners, guiding them towards success and pointing them away from many NLP pitfalls.

Biography:

Lukasz is a Data Scientist with 7+ years of experience in various ML projects (social media monitoring, call center's transcriptions analysis, recommendation engines, information extraction from texts, and many more). He is finishing a Ph.D. in Machine Learning, related to aspect-based sentiment analysis at the Wroclaw University of Science and Technology, Poland. He received MSc in Computer Science from the Wroclaw University of Technology in 2013 with distinction. He also received an MA in Law from Wroclaw University in 2014, and he is still actively interested in the legal aspects of IT and analysis of legal documents using NLP.

Publication of speakers:

1. Lukasz Augustyniak, wordnet2vec: Corpora agnostic word vectorization method, Neurocomputing, Volumes 326–327, 31 January 2019, Pages 141-150

2. Lukasz Augustyniak, Evaluation of in vivo graphene oxide toxicity for Acheta domesticus in relation to nanomaterial purity and time passed from the exposure, Journal of Hazardous Materials, Volume 305, 15 March 2016, Pages 30-40

3. Lukasz Augustyniak, Vitellogenin expression, DNA damage, health status of cells and catalase activity in Acheta domesticus selected according to their longevity after graphene oxide treatment, Science of The Total Environment, Volume 737, 1 October 2020, 140274

4. Lukasz Augustyniak, Localization of conduction electrons in hydrothermally reduced graphene oxide: electron paramagnetic resonance studies, Carbon, Volume 168, 30 October 2020, Pages 665-672