

# Octavia Maria Şulea

## PERSONAL INFORMATION

### Contact

+1 929 227 5083  
+40 743 760587  
[mary.octavia@gmail.com](mailto:mary.octavia@gmail.com)

### Links

[Google Scholar](#)

## WORK EXPERIENCE

<b>NLP Data Scientist @ GumGum, Santa Monica, CA, USA</b> Improved neural language model fine-tuning time through near-duplicate detection using LSH (deployed); Improved text extraction models by redesigning their data collection pipeline (deployed)	Sep 2018 – Jul 2019
<b>NLP Data Scientist @ Lark Health, Mountain View, CA, USA</b> Improved their food parser by implementing a neural sequence labeling model to split user text into meals (deployed)	May 2018 – Aug 2018
<b>NLP Engineer @ Agolo, New York, NY, USA</b> Improved multi-document extractive summarization of news articles through Bayesian optimization (deployed); Lead research into query and theme based extractive summarization of financial documents using clustering (PoC)	Apr 2017 – Jan 2018
<b>Language Technology Researcher @ DFKI, Saarbrücken, Germany</b> Lead pre-grant research implementing a system to predict the date, law area, and ruling of legal cases in the French Supreme Court (published); Co-created a system to identify the language variety/ location of news articles (published)	Mar 2016 – Mar 2017
<b>Computational Linguistics Researcher @ Saarland University, Saarbrücken, Germany</b> Researched neural systems for textual entailment in Twitter as part of the PHEME EU research grant (published)	May 2016 – Dec 2016
<b>Machine Learning Researcher @ Bitdefender, Bucharest, Romania</b> Lead research implementing 1) a highly accurate web page classifier using only URLs (published); 2) a highly accurate predictor of gender, age, and personality for Twitter users (published); 3) a novel way to train bilingual word embeddings which translates named entities from French to English better than Bing (published)	Mar 2015 – Nov 2015

## EDUCATION

<b>Ph.D. in Computer Science (Natural Language Processing)</b> University of Bucharest, Doctoral School of Computer Science, Romania	Oct 2014 – Oct 2019
<b>M.Sc. In Artificial Intelligence</b> University of Bucharest, Faculty of Mathematics and Computer Science, Romania	Oct 2011 – June 2013
<b>M.A. in Theoretical Linguistics</b> University of Bucharest, Faculty of Foreign Languages and Literatures, Romania	Oct 2010 – June 2012
<b>B.Sc. in Computer Science</b> University of Bucharest, Faculty of Mathematics and Computer Science, Romania	Oct 2007 – June 2010

## SKILLS

### Programming & Scripting Languages

Python for ML/NLP (TensorFlow, Keras, NLTK, Gensim, Scikit-learn, SpaCy, SciPy, NumPy, XGBoost), text processing (Pandas, PyPDF2, PySpark, DataSketch), visualization (Matplotlib), web scraping (Scrapy, Selenium, Requests); MongoDB (PyMongo); SQL (PyMySQL); C, C#

### Environments, Frameworks & Tools

AWS (EC2, S3, Athena, CloudWatch), Jupyter Notebook, PyCharm, Sublime Text, IntelliJ, Git, Mercurial, LaTeX, JIRA, Breeze, Scrum

### Human Languages

Romanian (native), English (near-native), French (intermediate), German (intermediate)

## ADDITIONAL INFORMATION

### Scholarships, Grants, Awards

Doctoral Scholarship, EU strategic grant POSDRU/187/1.5/S/155559, 2015  
The Doctoral Researcher Gala 2015, strategic grant POSDRU/187/1.5/S/155559  
Scholarship for Scientific Achievements, University of Bucharest, 2011-2012  
Scholarship for Scientific Achievements, Romanian Ministry of Education, 2011-2012

### References

Available upon request

# Octavia Maria Şulea

## LIST OF PUBLICATIONS

**Octavia-Maria Şulea**, Steve Young. *Unsupervised inflection generation using neural language modelling*. Advances in Computational Intelligence - 15th International Work-Conference on Artificial Neural Networks (IWANN 2019). June 2019

**Octavia-Maria Şulea**, Steve Young, Liviu P. Dinu. *MorphoGen: Full Inflection Generation Using Recurrent Neural Networks*. 20th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2019). April 2019

**Octavia-Maria Şulea**, Liviu P. Dinu, Bogdan Dumitru. *Full Inflection Learning Using Deep Neural Networks*. 19th International Conference on Computational Linguistics and Intelligent Text Processing (CICLing 2018). March 2018

**Octavia-Maria Şulea**. *Recognizing Textual Entailment in Twitter Using Word Embeddings*. Proceedings of the 2nd Workshop on Evaluating Vector Space Representations for NLP (RepEval@EMNLP 2017). September 2017

**Octavia-Maria Sulea**, Marcos Zampieri, Mihaela Vela, Josef van Genabith. *Predicting the Law Area and Decisions of French Supreme Court Cases*. Proceedings of the International Conference Recent Advances in Natural Language Processing (RANLP 2017). September 2017

**Octavia-Maria Sulea**, Marcos Zampieri, Shervin Malmasi, Mihaela Vela, Liviu P. Dinu, Josef van Genabith. *Exploring the use of text classification in the legal domain*, Proceedings of the Second Workshop on Automated Semantic Analysis of Information in Legal Texts co-located with the 16th International Conference on Artificial Intelligence and Law (ASAIL@ICAIL 2017). June 2017

**Octavia-Maria Şulea**. *Semi-supervised approach to Romanian noun declension*. Procedia Computer Science, Volume 96 Proceedings of Knowledge-Based and Intelligent Information and Engineering Systems: Proceedings of the 20th International Conference (KES 2016). September 2016

Marcos Zampieri, Shervin Malmasi, **Octavia-Maria Şulea**, Liviu P. Dinu. *A Computational Approach to the Study of Portuguese Newspapers Published in Macau*. Natural Language Processing meets Journalism (NLP4J@IJCAI 2016). July 2016

**Octavia-Maria Şulea**, Sergiu Nisioi, Liviu P. Dinu. *Using Word Embeddings to Translate Named Entities*. Proceedings of the 10<sup>th</sup> International Conference on Language Resources and Evaluation (LREC 2016). May 2016

**Octavia-Maria Şulea**, Liviu P. Dinu, Alexandra Peşte. *Using NLP-Specific Features for Non-NLP Specific Tasks. A Web Security Application*. Proceedings of the 22nd International Conference of Neural Information Processing (ICONIP 2015). November 2015

**Octavia-Maria Şulea**, Daniel Dichiu. *Automatic Profiling of Twitter Users based on their Tweets*. CLEF 2015 Working Notes. September 2015

Liviu P. Dinu, Vlad Niculae, **Octavia-Maria Şulea**. *Romanian syllabication using machine learning*. Proceedings of Text, Speech and Dialogue (TSD 2013). September 2013

Liviu P. Dinu, Vlad Niculae, **Octavia-Maria Şulea**. *Sequence tagging for verb conjugation in Romanian*. Proceedings of Recent Advances in Natural Language Processing (RANLP 2013). September 2013

Alina Maria Ciobanu, Anca Dinu, Liviu P. Dinu, Vlad Niculae, **Octavia-Maria Şulea**. *Temporal text*

## Octavia Maria Şulea

*classification for Romanian novels set in the past*. Proceedings of *Recent Advances in Natural Language Processing* (RANLP 2013). September 2013

Alina Maria Ciobanu, Anca Dinu, Liviu P. Dinu, Vlad Niculae, **Octavia-Maria Şulea**. *Temporal classification for historical Romanian texts*. Proceedings of the 7th Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities (LaTeCH@ACL 2013). August 2013

Liviu P. Dinu, Vlad Niculae, **Octavia-Maria Şulea**. *Dealing with the Grey Sheep of the Romanian Gender System, the Neuter*. 24th International Conference on Computational Linguistics, Proceedings of the Conference: Demonstration Papers (COLING 2012). December 2012

Iulia Dănăilă, Liviu P. Dinu, Vlad Niculae, **Octavia-Maria Şulea**. *String Distances for Near-duplicate Detection*. Polibits, Research journal on Computer science and computer engineering with applications. June 2012

Liviu P. Dinu, Vlad Niculae, **Octavia-Maria Şulea**. *The Romanian Neuter Examined through a Two-Gender N-Gram Classification System*. Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 2012). May 2012

Liviu P. Dinu, Vlad Niculae, **Octavia-Maria Şulea**. *Pastiche detection based on stopword rankings. Exposing impersonators of a Romanian writer*. EACL 2012 Workshop on Computational Approaches to Deception Detection. April 2012

Liviu P. Dinu, Vlad Niculae, **Octavia-Maria Şulea**. *Learning How to Conjugate the Romanian Verb. Rules for Regular and Partially Irregular Verbs*. 3th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2012). April 2012

Liviu P. Dinu, Emil Ionescu, Vlad Niculae, **Octavia-Maria Şulea**. *Can Alternations be Learned? A Machine Learning Approach to Verb Alternations*. Recent Advances in Natural Language Processing (RANLP 2011). September 2011