

Andreas Grivas

Natural Language Processing Engineer

18/2 Duncan Street
Edinburgh, Scotland
☎ (0044) 07920178572
✉ andreasgrv@gmail.com
🌐 <http://grv.overfit.xyz>

Summary

I enjoy exploring, developing, testing and understanding the shortcomings of natural language processing (NLP) systems. During my MSc I learned a lot about cutting edge neural network models and implemented them for a variety of NLP tasks, such as language modelling, machine translation and syntactic analysis (dependency parsing). In my dissertation I demonstrated how neural network subword representations can improve syntactic analysis for languages that have morphologically-rich word forms. I have two years of previous work experience extracting entity mentions (Named Entity Recognition) and their relations (Relation Extraction) from text.

Academic Background

Studies

2016–2017 **MSc Artificial Intelligence**, *University of Edinburgh, with distinction – GPA 79/100.*

2010–2014 **BSc Informatics**, *Harokopio University, Athens, 1st honours – GPA 9.1/10.0.*

MSc thesis

title Parsing morphologically-rich languages using neural networks
supervisors Adam Lopez & Clara Vania
description For my MSc project I implemented three syntactic analysis models for 14 languages. The parsing models which constructed word representations compositionally from characters employing recurrent or convolutional neural networks were competitive with state of the art on one of the two datasets they were evaluated on.

Online Courses Completed

2016 **Edx**, *Introduction to Probability.*

2016 **Coursera**, *Machine Learning (Andrew Ng's course).*

2013 **Coursera**, *Digital Signal Processing.*

2013 **Coursera**, *Cryptography I.*

2013 **Coursera**, *Understanding Einstein - The Special Theory of Relativity.*

2012 **Coursera**, *Introduction to Mathematical Thinking.*

Research Interests

- Natural Language Processing
- Dependency Parsing

- Morphologically aware word representations
- Named Entity Recognition
- Relation Extraction

Publications

2015 **PaloPro: A platform for knowledge extraction from big social data and the news**, *IJBDI*.

Publications - other

2015 **Author profiling using stylometric and structural feature groupings**, *Notebook for PAN at CLEF 2015*.

Awards

Pan 2015

2015 **Author Profiling task (text classification)**, *3_{rd} place*.

Wrote submission for Pan 2015 in the Author Profiling task. The task was to predict the gender, age and personality traits of twitter users given a dataset of tweets. Our machine learning approach placed *3_{rd}* overall and *1_{st}* on the gender identification subtask.

Hackathons

2015 **Taxnorris team**, *2_{nd} place*, #apodeiksi Hackathon.

Worked on machine learning OCR approach (dataset creation, preprocessing, use of open OCR library ocopy) that translated images of receipts to text.

2014 **A.g.i.n.a.r.a team**, *1_{st} place*, Open Public Data Hackathon.

Coded part of backend in python and bash scripts.

Work Experience

Reveal Project

2015–2016 **Research Assistant**, *NCSR Demokritos*, Worked on relation extraction and text classification (Author Profiling) on social network data.

- Designed and coded approach for Pan 2015 Author Profiling challenge that came *3_{rd}* place.
- Created tictacs, a python module that creates scikit-learn machine learning pipelines from yaml config files. <http://github.com/andreasgrv/tictacs>
- Became competent with the scikit-learn machine learning library.

ICT4Growth Project

2014–2015 **Research Assistant**, *AUEB group*, Developed named entity recognition models for Greek and Serbian.

- Trained models on automatically annotated data leveraging wikipedia as a source of text and dbpedia as a gazeteer.
- Implemented annotator for marking named entities.
- Deployed web api for the named entity recognition system (written with Flask). Gained experience interfacing Python and Java through jni with the help of the jnius library.

Programming Skills

Programming Languages

python, C++, java, javascript, C, bash, octave, L^AT_EX, tikZ

Markup and Description Languages

yaml, xml, html, css

Operating Systems

- Unix, Linux
- Basic administration skills

Machine Learning Libraries and Tools

- chainer, tensorflow
- numpy, matplotlib
- scikit-learn

Databases and “Big Data”

- postgres, mysql
- mongodb
- hadoop

Web Frameworks

flask, react

Tools

vim, tmux, git, make, maven

Languages

English Mothertongue

Cambridge Proficiency Certification

Greek Mothertongue

Certificates

2015 Cambridge Proficiency Certificate in English - Grade A

2004 Cambridge first Certificate in English - Grade A

Organizing

2013 **Fosscomm 2013**, *member of the organizing committee*, www.fosscomm.gr.

Interests

- Electric & Bass Guitar
- Comedy - stand up comedy