

Jake Vasilakes

Mail: jvasilakes@gmail.com • Phone: +44 7572356508
Homepage: jvasilakes.github.io • Github: github.com/jvasilakes

Experience

Feb 2016 - Present | **Research Assistant in Speech Processing (IARPA Babel project)**
University of Cambridge - Cambridge, UK

- Building state of the art ASR and keyword search systems for 7+ languages using HTK.
- Implemented pipelines for building language models from web and morphologically decomposed data using Python and shell.
- Conducting experiments, reporting and analysing results.

Sept - Nov 2015 | **Algorithm Development Intern**
ICAN Future Star Ltd - Edinburgh, UK

- Formulated a supervised machine learning system for computing how well a given student matches a university given their academic qualifications.
- Designed a system for assisting students in the university application process by clustering students and modeling the per-cluster application process.
- Assisted in writing an ultimately successful grant for funding.

Education

August 2015 | **MSc Speech and Language Processing (Distinction)**
University of Edinburgh

Thesis: *Automatic Generation of Wide-scale Semantic Representations in NLTK*
Advisor: Ewan Klein

Coursework

- | | |
|--|---|
| • Advanced Natural Language Processing | • Introductory Applied Machine Learning |
| • Speech Processing | • Topics in Natural Language Processing |
| • Statistics and Methodology using R | • Automatic Speech Recognition |
| • Natural Language Understanding | • Machine Translation |
| • Automated Reasoning | • Semantic Web Systems |

June 2013 | **B.A. Philosophy (Honors)**
Loyola University of Chicago

GPA: 3.84/4.00 (Equivalent to UK first)

Thesis: *The World of Speech* Advisor: Hanne Jacobs

Minors: Classics, Italian

Honors and Awards: Outstanding Philosophy Senior Award 2013, 2nd place Ancient Greek Translation Contest 2012, Member - Eta Sigma Phi Classical studies honor society

Skills

Programming Languages and Software

Python: Very good knowledge. Python 2 and 3, PEP 8 coding standards, unit testing.

Shell scripting: Good knowledge. `bash` and `tcsh`.

Regular Expressions: Good knowledge.

R: General knowledge (core aspects of the language, applications to machine learning).

C/C++: General knowledge (core aspects of the languages).

Operating Systems: Linux/UNIX (including OS X), Windows XP - 8.

Software and Libraries: NLTK, HTK, NumPy, TensorFlow, WEKA, Festival TTS, Praat, Wavesurfer.

Version Control: Git.

Web and Data technologies: HTML, CSS, XML.

Ontologies: RDF, OWL, SPARQL.

Concepts

Natural Language Processing and Speech Technology:

- Automatic speech recognition (HMM/{GMM, DNN} systems)
- Language modelling (n-grams, neural networks).
- Syntactic parsing (constituent, dependency, CCG).
- Computational semantics (semantic parsing, formal and distributional semantics).
- Statistical machine translation (alignment models, decoding).

Machine Learning and Statistical Modelling:

- Generative and discriminative modelling.
- Supervised and unsupervised methods.
- Dimensionality reduction, feature selection.

Languages

English (native), Italian (conversational), some knowledge of German and Spanish, also 4 years study of Ancient Greek.