John W. Hall

halljw@seas.upenn.edu | 954-632-3507
www.halljw.com

EDUCATION

M.S. in Computer and Information Technology, **University of Pennsylvania**   *May 2019*
M.A. in Linguistics, **University of Illinois**   *May 2017*
    Major area: Computational Linguistics
B.S. in Biomedical Engineering, **Vanderbilt University**   *May 2012*

M.A. THESIS

"Examination of Machine Learning Methods for Multi-label Classification of Intellectual Property Docs"
- Applied variety of machine learning techniques to classification of a corpus of US patent grants
- Built corpus by bulk download of data, cleaning xml-tags, tokenization, and storing of text data
- Implemented multi-class document classifiers using the python sci-kit learn library including Naïve-Bayes, neural network, decision tree, support-vector-machine, and logistic regression
- Examined effectiveness of classifiers in multi-label output using a one-vs-all implementation

TECHNICAL SKILLS

Languages – Python, Java, C/C++, R/RStudio, HTML/CSS, Shell Scripting
Tools/Frameworks – NLTK, Flask, Sci-kit Learn, Bootstrap, LaTeX, Git, Linux, Windows, Blackboard
Skills – Natural Language Processing, Machine Learning, Web Development, Web Scraping

PROJECTS

BlockheadExchange.com, **University of Pennsylvania**, Philadelphia, PA   *2017*
- Collaborated in deployment of site promoting interest and intelligent investing into cryptocurrencies
- Constructed webscraper to gather historical data on 100+ cryptocurrencies, and stored and analyzed market data with the Pandas python library
- Built user interface for conducting basic analysis and visualization of cryptocurrency data

Yup`ik Spellchecker and Transcriber, **University of Illinois**, Champaign, IL   *2017*
- Spellchecker for Central Alaskan Yup`ik, a less commonly studied Eskimo-Aleut language of the native population of Western and Southwestern Alaska
- Transcribes texts in Yup`ik orthography (primarily using Latin Alphabet) to phonetic representations (in the International Phonetic Alphabet) reflecting phonological constraints of the language

IBM Model 1 Statistical Translator, **University of Illinois**, Champaign, IL   *2016*
- Model serving as a foundation for the majority of modern statistical machine translation systems
- Implements a translation model and alignment model from sample corpora of two languages using an expectation-maximization algorithm to establish lexical translation probabilities.

WORK AND TEACHING EXPERIENCE

TA, **University of Illinois**, Champaign, IL   *Aug. to Dec. 2016*
Ling 402: *"Tools and Technology for Speech and Language Processing"*
- TA and grader for introductory computing course of 40+ graduate/undergraduate linguistics students focusing on computing tools for natural language processing using Python in a Linux environment
- Constructed shell scripts to facilitate and automate grading of coding assignments
- Topics included the Linux command line, shell scripting, Python programming, the Natural Language Toolkit library, proper documentation, data structures, and recursion

Biomedical Patent Examiner, **Patent and Trademark Office**, Alexandria, VA   *Jun. 2012 to Oct. 2013*
- Represented official positions of the USPTO in corresponding with inventors and representative entities regarding the technical and legal merits of innovative technologies
- Determined whether patent applications satisfied legal requirements in accordance with US patent code and presented arguments in favor of or against the legality of applications