Jake Vasilakes

jvasilakes@gmail.com +1 763-339-9216 • jvasilakes.github.io github.com/jvasilakes

Education

Currently | **PhD Natural Language Processing**
University of Manchester - National Centre for Text Mining (NaCTeM)
Advisor: Prof. Sophia Ananiadou

Aug 2015 | **MS Speech and Language Processing, *distinction***
University of Edinburgh
Thesis: "Automatic Generation of Wide-scale Semantic Representations in NLTK"
Advisor: Dr. Ewan Klein

June 2013 | **BA Philosophy with Honors, *magna cum laude***
Loyola University - Chicago
Thesis: "The World of Speech"
Advisor: Dr. Hanne Jacobs

Experience

Oct 2017 - Aug 2020 | **Natural Language Processing Research Programmer**
University of Minnesota, Institute for Health Informatics - Minneapolis, MN

Research
- Created iDISK, an open-source Neo4j knowledge base of dietary supplements using data automatically integrated from multiple semi-structured sources.
- Researched active learning and core-set selection methods to reduce the amount of labeled data required to build machine learning models.
- Deployed and managed annotation projects to support new research directions.

Service
- Teaching Assistant for UMN HINF 5610 Biomedical Natural Language Processing.
- Gave a talk "Introduction to Natural Language Processing" and associated tutorial at a workshop organized by the University of Minnesota Carlson School of Management.

Feb - Nov 2016 | **Research Assistant in Speech Processing**
University of Cambridge - Cambridge, UK

Research
- Trained and evaluated machine learning systems for multilingual speech recognition on datasets containing over 80 hours of audio data.
- Developed a statistical model to predict speech recognition performance on unseen languages to within 5%.
- Built n-gram language models from web and morphologically decomposed text.

Service
- Supervised an undergraduate student's research project on optimizing a search graph, which was published in IEEE ICASSP 2017.

Skills

Programming Languages: Python, Julia, R, C, *nix shell, SQL, Cypher
AI & NLP tools: NumPy/SciPy/Pandas, scikit-learn, NLTK, TensorFlow, PyTorch
Health Informatics tools: UMLS, SNOMED-CT, ICD, MetaMap, SemRep
Other tools: Neo4j, Jupyter, Git, PBS, LaTeX

Projects

iDISK: The integrated DIetary Supplements Knowledge base
iDISK is a standardized knowledge base of clinically relevant information concerning dietary supplements.
Code and data releases available at `github.com/zhang-informatics/iDISK`.

Y'ALL: Yet another Active Learning Library
Built on top of scikit-learn, Y'ALL implements a variety of active learning and core-set selection methods.
Code available at `github.com/jvasilakes/yall`.

Automatic Generation of Wide-Scale Semantic Representations in NLTK
M.S. thesis project. Extends the CCG parsing framework to encode sentence semantics as logical forms.
Code available at `https://github.com/jvasilakes/nltk/tree/develop/nltk/semparse`.

Publications

Rizvi, R.*, **Vasilakes, J.***, Adam, T.J., Melton, G.B., Bishop, J., Cui, T., Zhang, R. (2019). *iDISK: The integrated DIetary Supplements Knowledge base.* Journal of the American Medical Informatics Association (JAMIA).
* Equal contribution

Vasilakes, J., Fan, Y., Rizvi, R., Bompelli, A., Bodenreider, O., Zhang, R. (2019). *Normalizing Dietary Supplement Product Names using the RxNorm Model.* MedInfo, Lyon, France

Vasilakes, J., Rizvi, R., Zhang, J., Adam, T.J., Zhang R. (2019). *Detecting Signals of Dietary Supplement Adverse Events from the CFSAN Adverse Event Reporting System (CAERS).* American Medical Informatics Association (AMIA) Informatics Summit, San Francisco, CA

Vasilakes, J., Rizvi, R., Melton, G.B., Pakhomov, S., Zhang, R. (2018). *Evaluating Active Learning Methods for Annotating Semantic Predications.* Journal of the American Medical Informatics Association (JAMIA) Open

Vasilakes, J., Wang, H., Ragni, A., Gales, M.J.F. & Knill, K.M. (2016). *Speech Recognition and Keyword Spotting Performance Analysis Across Languages.* Poster presented at UK Speech Conference, Sheffield, UK

Rizvi, R., Wang, Y., Nguyen, T., **Vasilakes, J.**, Bian, J., He, Z., Zhang, R. (2019). *Analyzing Social Media Data to Understand Consumers Information Needs on Dietary Supplements.* MedInfo, Lyon, France

He X., Zhang R., Rizvi R., **Vasilakes, J.**, et al. (2019) *ALOHA: developing an interactive graph-based visualization for dietary supplement knowledge graph through user-centered design.* BMC Medical Informatics and Decision Making.

Xing, H., Zhang, R., Rizvi, R., **Vasilakes, J.**, Yang, X., Guo, Y., He, Z., Prosperi, M., Bian, J. (2018). *Prototyping an Interactive Visualization of Dietary Supplement Knowledge Graph.* IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Madrid, Spain

Rizvi, R., Adam, T.J., Lindemann, E., **Vasilakes, J.**, Pakhomov, S., Bishop, J., Meltion, G.B., Zhang, R. (2018). *Comparing Existing Resources to Represent Dietary Supplements.* American Medical Informatics Association (AMIA) Summits on Translational Science, San Francisco, CA

Ragni, A., Wu, C., Gales, M.J.F., **Vasilakes, J.**, Knill, K.M. (2017). *Stimulated training for automatic speech recognition and keyword search in limited resource conditions.* IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA

Ragni, A., Saunders, D., Zahemszky, P., **Vasilakes, J.**, Gales, M.J.F., Knill, K.M. (2017). *Morph-to-word transduction for accurate and efficient automatic speech recognition and keyword search.* IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA

Chen, X., Ragni, A., **Vasilakes, J.**, Liu, X., Knill, K.M., Gales, M.J.F. (2017). *Recurrent neural network language models for keyword search.* IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA