

Jake Vasilakes

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

Education

In Progress | **PhD Natural Language Processing**

University of Manchester - National Centre for Text Mining (NaCTeM)

Topic: Natural language processing for scientific text mining

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 - | **Natural Language Processing Research Programmer**

Aug 2020 **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 | **Research Assistant in Speech Processing**

2016 **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: TensorFlow, PyTorch, scikit-learn, NumPy/SciPy/Pandas, NLTK

Biomedical Informatics tools: UMLS, SNOMED-CT, ICD, MetaMap, SemRep

Other tools: Neo4j, Jupyter, Git, PBS, LaTeX

Publications

Vasilakes, J., Zhou, S., Zhang, R., 2020. 'Natural Language Processing', in Al'Aref, S., Singh, G., Baskaran, L., Metaxas, D. eds. *Machine Learning for Cardiovascular Medicine*. Elsevier.

Vasilakes, J., Bompelli, A., Bishop J.R., Adam, T.J., Bodenreider O., Zhang, R., 2020. *Assessing the Enrichment of Dietary Supplement Coverage in the Unified Medical Language System*. Journal of the American Medical Informatics Association (JAMIA).

Rizvi, R.*, **Vasilakes, J.***, Adam, T.J., Melton, G.B., Bishop, J., Cui, T., Zhang, R. 2020. *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*. In 17th World Congress on Medical and Health Informatics, MEDINFO 2019.

Vasilakes, J., Rizvi, R.F., Zhang, J., Adam, T.J. and Zhang, R., 2019. *Detecting Signals of Dietary Supplement Adverse Events from the CFSAN Adverse Event Reporting System (CAERS)*. AMIA Joint Summits on Translational Science Proceedings.

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.

Bompelli, A., Silverman, G., Finzel, R., **Vasilakes, J.**, Knoll, B., Pakhomov, S. and Zhang, R., 2020. *Comparing NLP Systems to Extract Entities of Eligibility Criteria in Dietary Supplements Clinical Trials Using NLP-ADAPT*. In International Conference on Artificial Intelligence in Medicine.

Rizvi, R.F., Wang, Y., Nguyen, T., **Vasilakes, J.**, Bian, J., He, Z. and Zhang, R., 2019. *Analyzing Social Media Data to Understand Consumer Information Needs on Dietary Supplements*. In 17th World Congress on Medical and Health Informatics, MEDINFO 2019.

He, X., Zhang, R., Rizvi, R., **Vasilakes, J.**, Yang, X., Guo, Y., He, Z., Prosperi, M., Huo, J., Alpert, J. and Bian, J., 2019. *ALOHA: Developing an Interactive Graph-based Visualization for Dietary Supplement Knowledge Graph through User-centered Design*. BMC medical informatics and decision making, 19(4).

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

Rizvi, R.F., Adam, T.J., Lindemann, E.A., **Vasilakes, J.**, Pakhomov, S.V., Bishop, J.R., Melton, G.B. and Zhang, R., 2018. *Comparing Existing Resources to Represent Dietary Supplements*. AMIA Summits on Translational Science Proceedings.

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).

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).

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).