

# Jake Vasilakes

jvasilakes@gmail.com   jvasilakes.github.io   github.com/jvasilakes

## Education

---

- Expected | **PhD Natural Language Processing**  
2023   University of Manchester, National Centre for Text Mining (NaCTeM)  
    *Research Topic:* Contextualizing information extraction  
    *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 2020 - | **Post-Graduate Researcher**  
Present   University of Manchester, National Centre for Text Mining - Manchester, UK
- Developed a variational autoencoder to learn disentangled representations of textual negation and speculation, which was published at ACL 2022.
  - Participated in the n2c2 2022 shared task on contextualized event extraction from clinical text, using transformers and multi-task learning.
- Oct 2017 - | **Natural Language Processing Research Programmer**  
Aug 2020   University of Minnesota, Institute for Health Informatics - Minneapolis, MN
- 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.
  - Teaching Assistant for UMN HINF 5610 Biomedical Natural Language Processing.
  - Organized an introductory tutorial on natural language processing for 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
- 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%.
  - Estimated n-gram language models from web and morphologically decomposed text.
  - 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

**ML tools:** PyTorch, TensorFlow, scikit-learn, NumPy, SciPy, Pandas, NLTK

**Biomedical informatics tools:** UMLS, SNOMED-CT, ICD, MetaMap, SemRep

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

## Publications

---

**Vasilakes, J.\***, Georgiadis\*, P., Nguyen, N., Miwa, M., & Ananiadou, S. (2023). Contextualized Medication Event Extraction with Levitated Markers. To appear in the *Journal of Biomedical Informatics (JBI)*. \* Equal contribution

**Vasilakes, J.**, Zerva, C., Miwa, M., & Ananiadou, S. (2022). Learning Disentangled Representations of Negation and Uncertainty. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT)*.

**Vasilakes, J.**, Zhou, S. & Zhang, R., (2021). Natural language processing. In *Machine Learning in Cardiovascular Medicine* (pp. 123-148). Academic Press.

**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)*, 27(10), 1547-1555.

Rizvi, R. F.\* , **Vasilakes, J.\***, Adam, T. J., Melton, G. B., Bishop, J. R., Bian, J., ... & Zhang, R. (2020). iDISK: the integrated Dietary Supplements Knowledge base. *Journal of the American Medical Informatics Association (JAMIA)*, 27(4), 539-548. \* Equal contribution

**Vasilakes, J.**, Fan, Y., Rizvi, R., Bompelli, A., Bodenreider, O., & Zhang, R. (2019). Normalizing Dietary Supplement Product Names Using the RxNorm Model. *Studies in health technology and informatics*, 264, 408-412.

**Vasilakes, J.**, A., Rizvi, R. F., Zhang, J., Adam, T. J., & 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. AMIA Joint Summits on Translational Science, 2019*, 258-266.

**Vasilakes, J.**, Rizvi, R., Melton, G. B., Pakhomov, S., & Zhang, R. (2018). Evaluating active learning methods for annotating semantic predications. *JAMIA open*, 1(2), 275-282.

**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.

---

Schutte, D., **Vasilakes, J.**, Bompelli, A., Zhou, Y., Fiszman, M., Xu, H., Kilicoglu, H., Bishop, J.R., Adam, T. and Zhang, R. (2022). Discovering Novel Drug-Supplement Interactions using SuppKG Generated from the Biomedical Literature. *Journal of biomedical informatics*, 131, p.104120.

Silverman, G. M., Finzel, R. L., Heinz, M. V., **Vasilakes, J.**, Solinsky, J. C., McEwan, R., ... & Pakhomov, S. V. (2021). An Empirical Study of UMLS Concept Extraction from Clinical Notes using Boolean Combination Ensembles. *arXiv preprint arXiv:2108.02255*.

Bompelli, A., Silverman, G., Finzel, R., **Vasilakes, J.**, Knoll, B., Pakhomov, S., & Zhang, R. (2020, August). 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* (pp. 67-77). Springer, Cham.

Rizvi, R. F., Wang, Y., Nguyen, T., **Vasilakes, J.**, Bian, J., He, Z., & Zhang, R. (2019). Analyzing social media data to understand consumer information needs on dietary supplements. *Studies in health technology and informatics*, 264, 323.

- He, X., Zhang, R., Rizvi, R., **Vasilakes, J.**, Yang, X., Guo, Y., ... & 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), 1-18.
- He, X., Zhang, R., Rizvi, R., **Vasilakes, J.**, Yang, X., Guo, Y., ... & Bian, J. (2018, December). Prototyping an interactive visualization of dietary supplement knowledge graph. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1649-1652). IEEE.
- Rizvi, R. F., Adam, T. J., Lindemann, E. A., **Vasilakes, J.**, Pakhomov, S. V., Bishop, J. R., ... & Zhang, R. (2018). Comparing existing resources to represent dietary supplements. *AMIA Summits on Translational Science Proceedings*, 2018, 207.
- Ragni, A., Wu, C., Gales, M. J., **Vasilakes, J.**, & Knill, K. M. (2017, March). Stimulated training for automatic speech recognition and keyword search in limited resource conditions. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4830-4834). IEEE.
- Ragni, A., Saunders, D., Zahemszky, P., **Vasilakes, J.**, Gales, M. J. F., & Knill, K. M. (2017, March). Morph-to-word transduction for accurate and efficient automatic speech recognition and keyword search. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5770-5774). IEEE.
- Chen, X., Ragni, A., **Vasilakes, J.**, Liu, X., Knill, K., & Gales, M. J. (2017, March). Recurrent neural network language models for keyword search. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 5775-5779). IEEE.