



## DSICCR Tuesday Seminar Series

Tuesday at Noon, Click [Here](#) to Join the Seminar

### Distributed Representations for Natural Language Processing: From Pre-Embeddings to Transformers

**Kirk Roberts, Ph.D.**

Associate Professor

School of Biomedical Informatics

UTHealth Houston

For decades natural language processing (NLP) has relied on the acquisition of knowledge about language directly from its contextual use. As stated by Firth, "you shall know a word by the company it keeps". The advantage of this approach is that it requires no manual knowledge: simply use a large corpus of text documents to derive the meanings of words. While manually-built dictionaries (like WordNet) have been used considerably in NLP, we have constantly relied on automatically-built distributional representations. This talk will provide an overview of the history of these representations, from early word co-occurrence lists to the modern large language model (LLM) pre-training techniques used by transformer models like GPT, BERT, and T5.

Tuesday, November 29<sup>st</sup>, 2022. 12p – 1p. [Webcast](#)

Contact: Xiaohong.Bi@uth.tmc.edu

 #SBMIseminar

