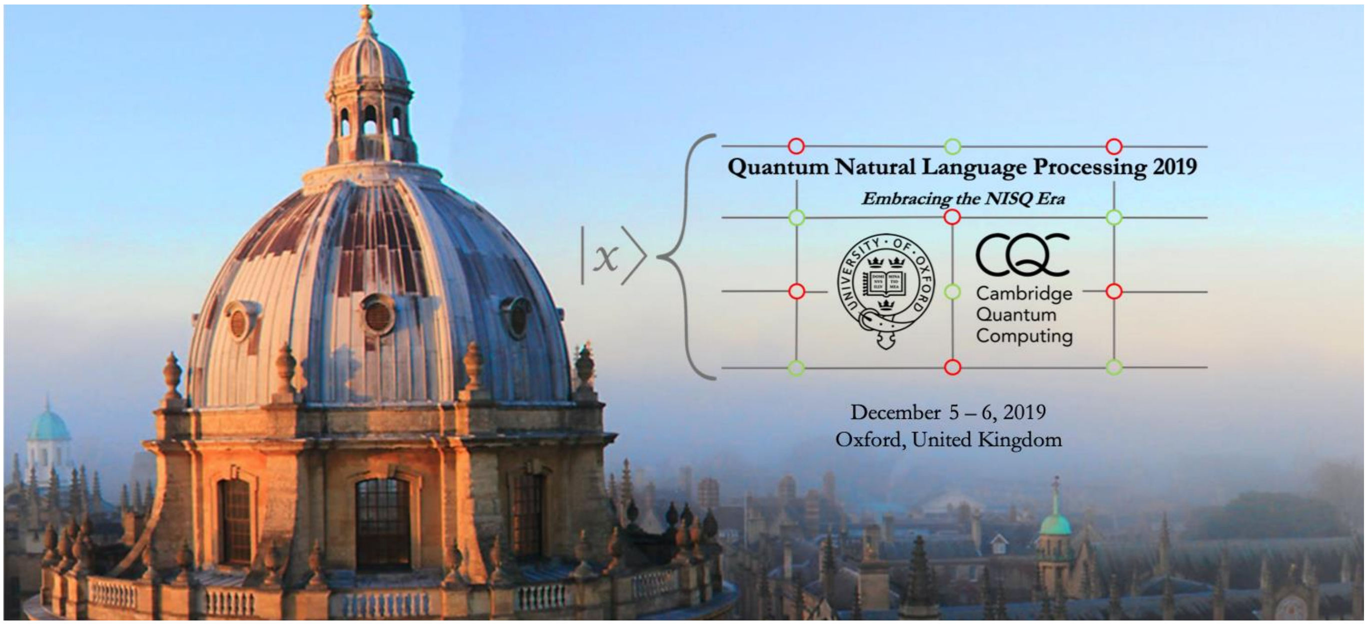

Speaker:

Johannes Bausch

Talk Title:

A Quantum Search Decoder for Natural Language Processing

Talk Abstract:

Probabilistic language models, e.g. those based on an LSTM, often face the problem of finding a high probability prediction from a sequence of random variables over a set of words. This is commonly addressed using a form of greedy decoding such as beam search, where a limited number of highest-likelihood paths (the beam width) of the decoder are kept, and at the end the maximum-likelihood path is chosen. The resulting algorithm has linear runtime in the beam width. However, the input is not necessarily distributed such that a high-likelihood input symbol at any given time step also leads to the global optimum. Limiting the beam width can thus result in a failure to recognise long-range dependencies.

In practice, only an exponentially large beam width can guarantee that the global optimum is found: for an input of length n and average parser branching ratio R, the baseline classical algorithm needs to query the input on average Rn times.

In this work, we construct a quantum algorithm to find the globally optimal parse with high constant success probability. Given the input to the decoder is distributed like a power-law with exponent k>0, our algorithm yields a runtime R^f(R,k), where f≤1/2, and f→0 exponentially quickly for growing k. This implies that our algorithm always yields a super-Grover type speedup, i.e. it is more than quadratically faster than its classical counterpart. We further modify our procedure to recover a quantum beam search variant, which enables an even stronger empirical speedup, while sacrificing accuracy. Finally, we apply this quantum beam search decoder to Mozilla's implementation of Baidu's DeepSpeech neural net, which we show to exhibit such a power law word rank frequency, underpinning the applicability of our model.

Based on:

https://arxiv.org/abs/1909.05023

QNLP Website: http://www.cs.ox.ac.uk/QNLP2019/