# Quantum Natural Language Processing (QNLP)

Lee O'Riordan
Irish Centre for High-End Computing, Ireland
lee.oriordan@ichec.ie

Myles Doyle
Irish Centre for High-End Computing, Ireland
myles.doyle@ichec.ie

Venkatesh Kannan
Irish Centre for High-End Computing, Ireland
venkatesh.kannan@ichec.ie

Fabio Baruffa
Intel Corporation, Germany
fabio.baruffa@intel.com

## ABSTRACT

The objective of this project is to implement the two quantum algorithms ("closest vector problem" and "sentence similarity") of the DisCo model on the Intel Quantum Simulator. Given a corpus, the implemented solution aims to compute the meanings of two sentences (built from words in the corpus) and decide if their meanings match. In this poster, we present the control workflow of the software implementation along with some preliminary results.

## KEYWORDS

Quantum computing, Natural language processing, Sentence similarity

## 1 Introduction

Natural language processing (NLP) is often used to perform tasks such as machine translation, sentiment analysis, relationship extraction, word sense disambiguation and automatic summary generation. Most traditional NLP algorithms for these problems are defined to operate over strings of words, and are commonly referred to as the "bag of words" approach. The challenge, and thus limitation, of this approach is that the algorithms analyse sentences in a corpus based on meanings of the component words and lack information from the grammatical rules and nuances of the language. Consequently, the qualities of results of these traditional algorithms are often unsatisfactory when the complexity of the problem increases.

On the other hand, an alternate approach called "compositional semantics" incorporates the grammatical structure of sentences in a language into the analysis algorithms. Compositional semantics algorithms include the information flows between words in a sentence to determine the meaning of the whole sentence. One such model is "distributional compositional semantics" (DisCo) [1], which is based on tensor product composition to give a grammatically informed algorithm that computes the meaning of sentences and phrases. This algorithm has been noted to potentially offer improvements to the quality of results, particularly for more complex sentences, in terms of memory and computational requirements. However, the main challenge in its implementation is the need for large classical computational resources.

## 2 Distributional Compositional Semantics

The distributional compositional semantics (DisCo) model was originally developed with direct inspiration from quantum theory, and present the quantum version of the DisCo model based on two algorithms [1]:

- Closest vector problem: An algorithm for the "closest vector problem" is used to determine the word/phrase out of a set of words/phrases that has the closest relation (for instance, meaning) to a given word/phrase. This finds application in many computational linguistic tasks such as text classification, word/phrase similarity, test classification and sentiment analysis.
- CSC sentence similarity: This algorithm is an adaptation of the "closest vector problem" quantum algorithm to perform sentence similarity calculations in the distributional compositional framework. This algorithm is based on tensor product composition that gives a grammatically informed algorithm to compute meaning of sentences/phrases and stores the meanings in quantum systems.

## 3 QNLP Implementation

The design of this QNLP software solution is based on a two-level approach: (i) the corpus analysis and formatting layer, and (ii) the quantum processing layer. Layer (i) will be written in Python, to avail of ease-of-use in text processing and analysis. Since this layer will be called at the initial outset of the program run, the performance impact will be negligible compared to the DisCo methods. Layer (ii) will be written entirely in C++, and will directly leverage the Intel Quantum Simulator. The pre-processed data from layer (i) will be loaded, and encoded into the Intel QS to generate the full meaning-space of the analysed corpus. This is illustrated in Figure 1.

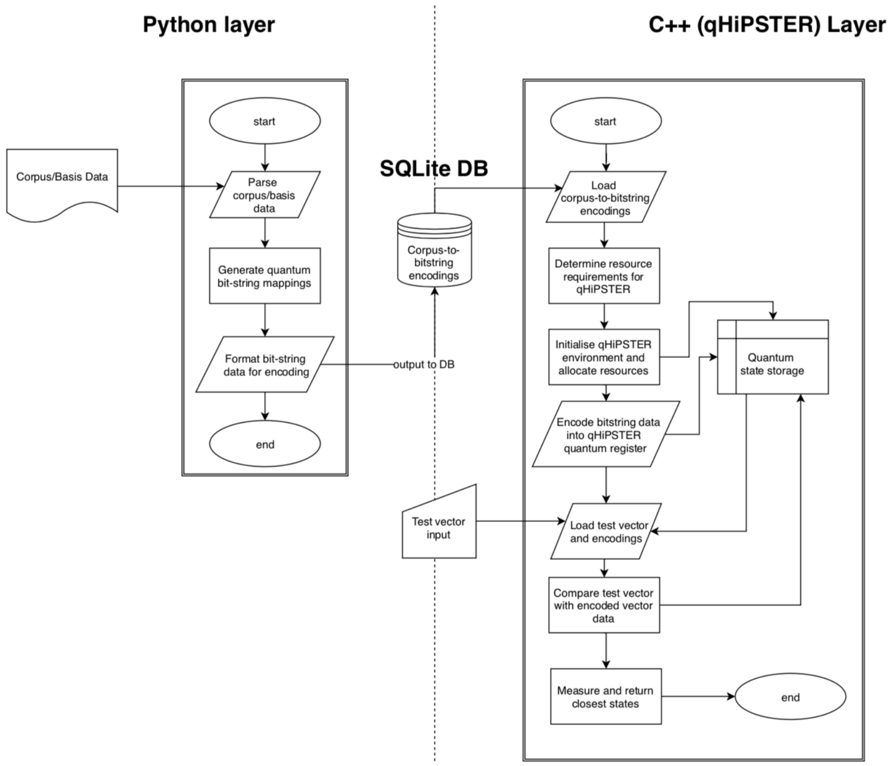


**Figure 1: Control flow of QNLP implementation**

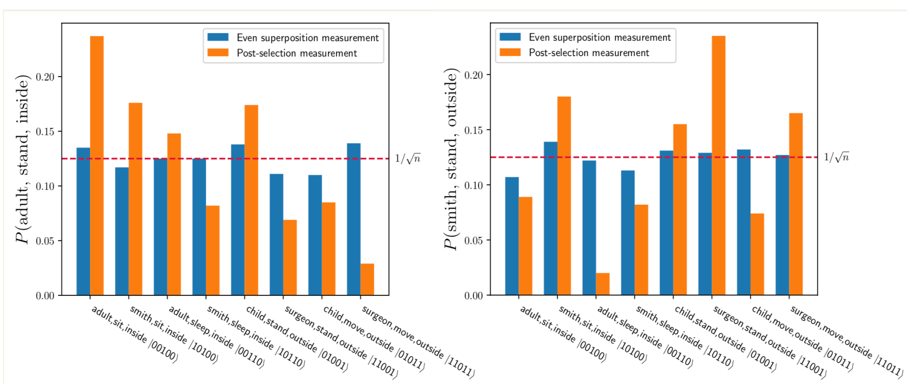


**Figure 2: Preliminary results of sentence similarity tests**

## 4 Summary

In this poster, we have summarized the results of evaluating the scalability of the Intel Quantum Simulator deployed on the Irish national supercomputer Kay. We also present the preliminary results of evaluating the QNLP sentence similarity implementation. This is illustrated in Figure 2. The left figure presents he closest encoded quantum-state vector to the data "adult(s) stand inside". The smallest Hamming distance calculated is "adult(s) sit inside". The right figure, similarly, illustrates the closest state for "smith(s) stand outside" gives "surgeon(s) stand outside".


**REFERNCES**

[1] William Zeng and Bob Coecke,"Quantum Algorithms for Compositional Natural Language Processing", EPTCS 221, 2016.