QuanText: Text Data Visualization in Quantum Computing

Abu Kaisar Mohammad Masum, and Naveed Mahmud
Department of Electrical Engineering and Computer Science, Florida Institute of Technology
{amasum2022}@my.fit.edu
{nmahmud}@fit.edu

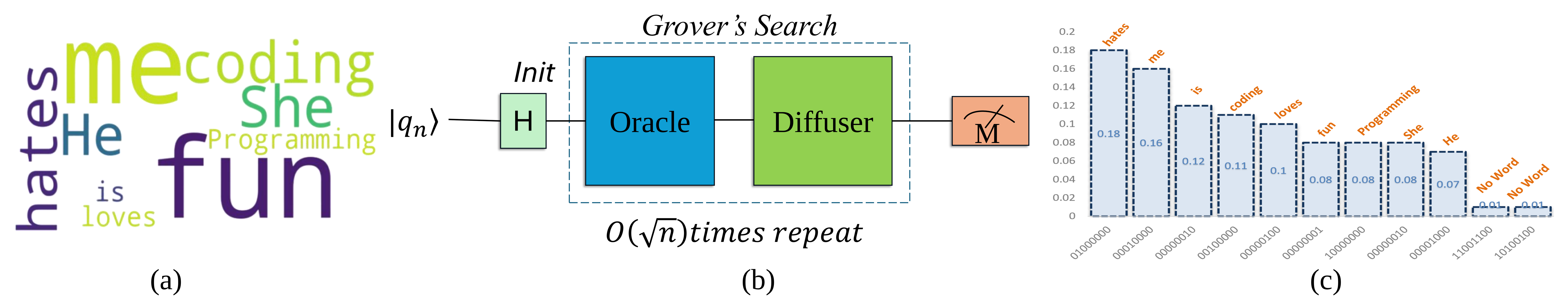

Figure 1: A simple explanation of Quantum Text Data Visualization is as follows: (a) creating a word cloud from sentences, (b) encoding words as basis states using Grover's circuit, which includes initialization, oracle, and diffuser stages, and repeats operations $O(\sqrt{n})$ times, and (c) visualizing the output probabilities with a bar chart, where the final measurement gate provides bitstrings whose cumulative probability sums to 1.

❑ Introduction:

- Qubits can be in states |0⟩, |1⟩, or a superposition of both, providing more computational power than classical bits [1]. Qubits further extend their capabilities by existing in states beyond |0⟩ and |1⟩, forming a linear combination known as superposition [2].
- Efficient text data visualization in quantum computing enhances Quantum Machine Learning (QML) and Quantum Natural Language Processing (QNLP).
- Quantum Data Visualization (QDV) for images exists, but text data visualization is underexplored. Uses Grover's algorithm to visualize text data by encoding text, generating bitstrings, and visualizing word probabilities.

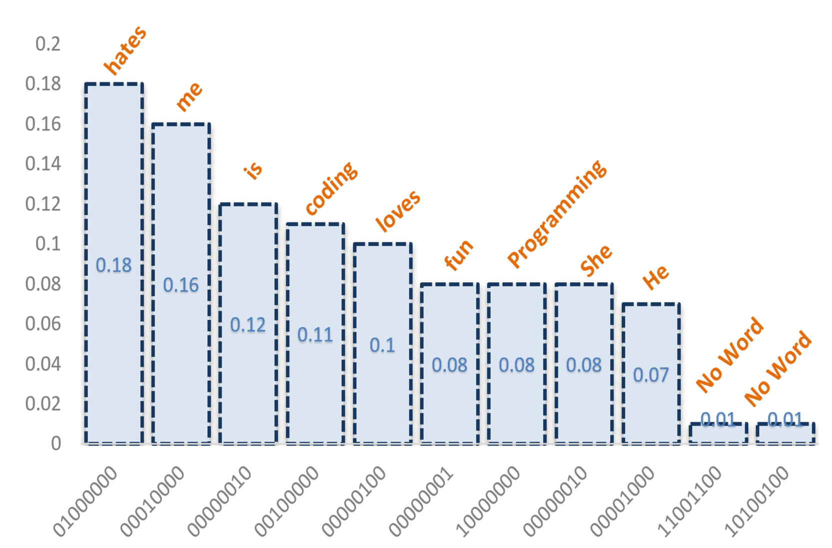

Figure 2: Bar plot of binary pattern probabilities with associated words.

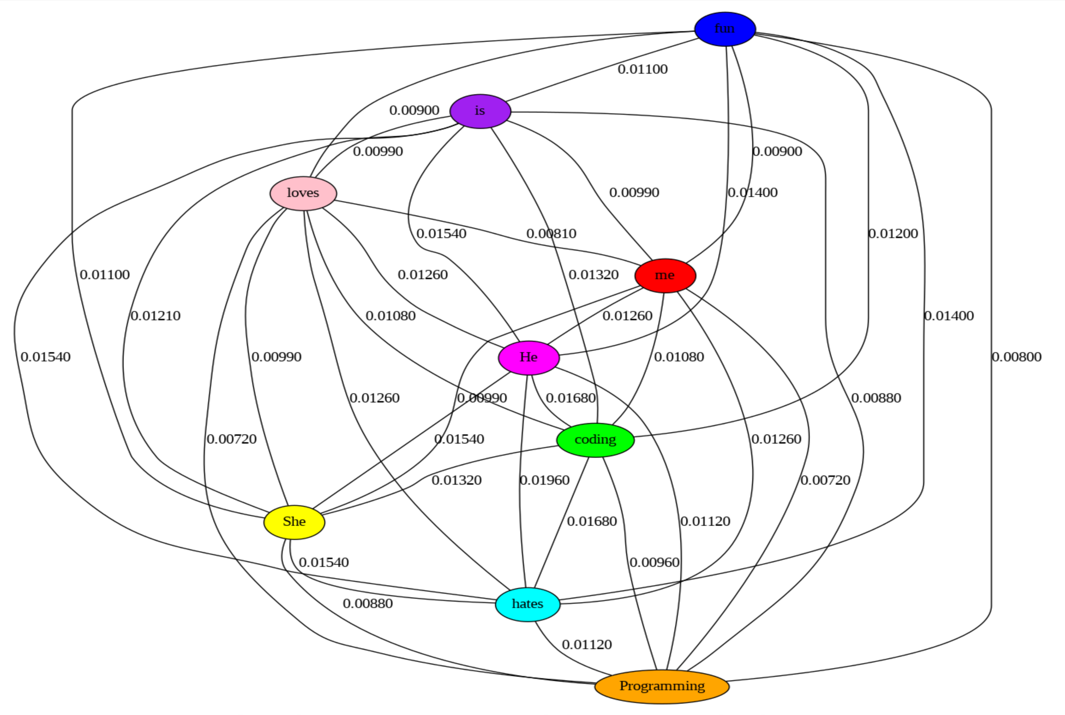

Figure 3: Graphical representation of the traversal of Grover's circuit output binary sequences. Nodes represent binary sequences, and edges represent transition probabilities.

❑ Method:

We implemented an algorithm to traverse the output binary sequences from Grover's circuit to analyze our approach further. In this algorithm, w_i represents a word or node in the binary sequence, and binary_dict is a dictionary mapping binary sequences to their probabilities.

Algorithm 1 Probabilistic Binary Sequence Traversal Algorithm

1: **if** w_i is a leaf node **then**
2: **return** binary_dict[seq_i]
3: **else**
4: Choose w_j randomly from out(w_i) with probability binary_dict[seq_j]
5: Continue this process until reaching a leaf node.
6: **end if**

❑ Analysis:

- **Co-occurrence:** Measures the frequency with which pairs of words appear together.
- **Contextual Similarity:** Uses cosine similarity to determine how similar the contexts of different word pairings are.
- **Semantic Similarity:** Measures the level of meaning relatedness between words or phrases.

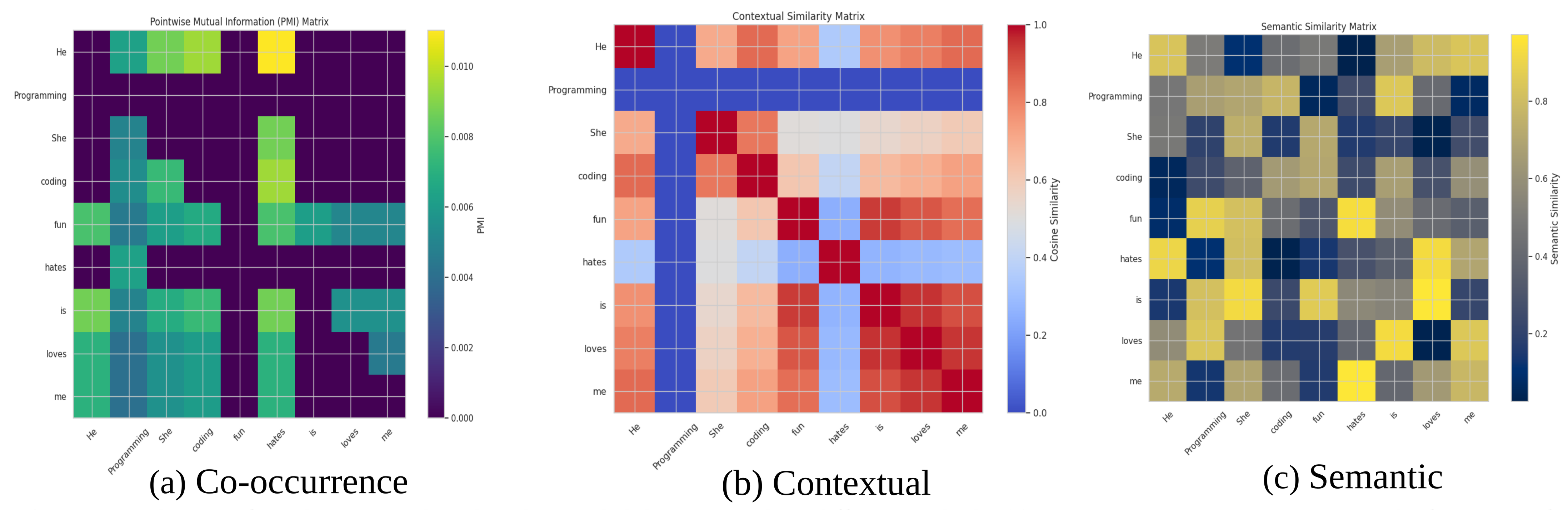

Figure 4: Evaluation of word sequences based on their probabilities using three different methods: (a) Co-occurrence matrix representing the frequency of word pairs appearing together, (b) Contextual similarity matrix using cosine similarity to measure context similarity, and (c) Semantic similarity matrix indicating the degree of meaning-relatedness between words or phrases.

❑ Reference:

[1] H. Arya, Z. Kaul, R. Wadhwa, K. Taira, T. Hirano, and S. C. Kaul. Quantum dots in bio-imaging: revolution by the small. Biochemical and biophysical research communications, 329(4):1173–1177, 2005. 1

[2] N. Linden, S. Popescu, and J. A. Smolin. Entanglement of superpositions. Physical review letters, 97(10):100502, 2006. 1

❑ Conclusion and Future Work:

The transition from classical to quantum data visualization stands out for its utilization of quantum principles to accelerate computations, enhance parallelism, and unlock new applications. This research approach introduces the potential of quantum data representation across various domains, underscoring the significance of quantum computing as an innovative field with far-reaching implications. Future work includes exploring quantum-based text analysis, text generation applications, and language modeling to further advance this field.

Acknowledgement

We dedicate this poster to the memory of our innocent students and general peoples who lost their lives in Bangladesh quota reform movement 2024. May they find eternal peace and the highest blessings in the afterlife.

VIS2024
IEEE VIS: Visualization & Visual Analytics
St. Pete Beach, Florida, SA, October 13-18