Trong bài báo của chúng tôi, chúng tôi đưa ra giải pháp đo độ tương đồng trên câu dựa trên thuật toán Levenshtein và dùng mô hình học sâu Siamese để cải thiện cho tiếng việt.

* Tiền xử lý dữ liệu tách văn bản thành các câu và lưu trữ vào database của chương trình.
* Khi so sánh thì sẽ tìm tập các văn bản có khả năng cao xảy ra tương đồng từ database.
* Dùng thuật toán Levenshtein để đo độ tương đồng giữa các câu được tách ra từ văn bản.
* Cải tiến bằng cách sử dụng mô hình học sâu để huấn luyện bộ dữ liệu tiếng việt để tang khả năng nhận biết ngữ nghĩa của các câu.