Gồm 2 câu:

1) Xây dựng mô hình ngôn ngữ tiếng Việt

input: cho một tập các văn bản (sẽ cho trước)

output: một mô hình ngôn ngữ dùng 1-gram, 2-gram, hoặc 3-gram

test:

\* cho vào 1 câu => đưa ra xác suất của câu đó.

\* cho vào một cụm từ (1-gram, hoặc 2-gram, hoặc 3-gram)

hãy cho biết tần suất của cụm đó

2) Ứng dụng cho một trong các bài toán sau

- Sửa lỗi chia động từ (áp dụng cho tiếng Anh)

- Gợi ý từ tiếp theo (dùng cho typing)

- Dịch teencode hoặc từ viết tắt

- Khôi phục dấu cho câu không dấu

- Phát hiện lỗi chính tả

- Tìm từ đồng nghĩa