**Mô hình ngôn ngữ và ứng dụng cho kiểm lỗi chính tả - Bài thực hành trên lớp**

**Mục tiêu: xây dựng chương trình kiểm lỗi cú pháp tiếng Anh đơn giản**

Bài 1:

1. Xây dựng mô hình ngôn ngữ dựa trên n-gram sử dụng phương pháp smoothing là Laplace, cho các mô hình:

* 1-gram
* 2-gram
* 3-gram

1. Tính xác suất của một câu và tính Perplexity của một câu dựa theo 1-gram, 2-gram, 3-gram
2. Phân tích kết quả

Bài 2:

1. Cải tiến mô hình bằng cách sử dụng smoothing là phương pháp interpolation theo “Stupid backoff” (Brants et al. 2007)
2. So sánh với kết quả bài 1
3. Dùng mô hình vừa xây dựng để sinh các từ tiếp theo với một chuỗi từ cho trước.
4. Kết hợp với hàm tính khoảng cách giữa các từ để dự đoán từ đúng cho vị trí từ sai chính tả. (from difflib import get\_close\_matches)