Bài tập về Language Models:

Bài 1:

Cho 1 file dữ liệu văn bản gồm 1 file train và 1 file test

1. Xây dựng mô hình ngôn ngữ 1-gram, 2-gram, 3-gram

* Nhập vào 1 câu, tính xác suất của câu đó theo từng mô hình

1. Tính Perplexity theo tập Test trên từng mô hình ở câu 1

Bài 2:

Tiếp tục bài 1, bổ sung các chức năng

1. Bổ sung các phương pháp Smoothing:

* Laplace
* Stupid Backoff
* Good Turing

Tính lại Perplexity trên tập test theo phương pháp mới với mô hình 2-gram

1. Nhập vào 1 chuỗi từ “w1 w2 …. wn”, dự đoán **k** từ tiếp theo