**1- Building a vector space model for Vietnamese**

Thực hiện huấn luyện 2 mô hình Skip-gram và Fasttext.

Mô hình đã huấn luyện được lưu ở folder “model”.

Hai mô hình cùng được huấn luyện trên một bộ dữ liệu lấy từ wiki đã qua tiền xử lý, với vector 150 chiều và context window là 5.

Kết quả thực nghiệm:

* Visualize data: các vector của các từ có quan hệ với nhau, cũng có khoảng cách tương đối với nhau
* In ra top 10 từ gần với một từ cụ thể: mô hình Fasttext cho ra kết quả chính xác hơn mô hinhf Skip-gram.

Link Github: [phuongmt3/Vietnamese\_word2vec (github.com)](https://github.com/phuongmt3/Vietnamese_word2vec)

**2- Finish script for pipeline in slide 28 (Word2vec in the week 2)**

Link Colab: [TruyenKieu\_Word2vec](https://colab.research.google.com/drive/1fv0YNfKmI1K-3YZVS3I_tYJ8rXY-qYbs?usp=sharing)