

# Request

Harito ID

2025-10-14

## Thông báo: Bài Lab về Word Embeddings

### 1. Lý thuyết Tổng quan

Các bạn vui lòng xem lại lý thuyết nền tảng về **Embedding** (nhúng từ) trong Xử lý Ngôn ngữ Tự nhiên.

- **Tài liệu tham khảo chính:** `lecture/week3/Embedding.pdf`
- **Trọng tâm lý thuyết:**
  - Nghiên cứu kỹ các mô hình: **Word2Vec**, **GloVe**, và **fastText**.
  - Hiểu rõ cách các mô hình này biểu diễn ngữ nghĩa của từ trong không gian vector.

### 2. Bài Lab: Triển khai và Trực quan hóa

#### Phần 1: Giảm chiều và Trực quan hóa Vector

**Mục tiêu:** - **Giảm chiều:** Sử dụng các kỹ thuật như PCA hoặc t-SNE để giảm chiều các word vector (từ Word2Vec, GloVe, hoặc fastText) xuống còn 2 hoặc 3 chiều. - **Trực quan hóa:** Vẽ biểu đồ (scatter plot) để hiển thị các từ trong không gian 2D/3D, qua đó quan sát mối quan hệ ngữ nghĩa giữa chúng.

#### Phần 2: Yêu cầu chi tiết

Nội dung và yêu cầu cụ thể cho bài lab được mô tả trong file `lab/lab4_word_embeddings_part2`.

### 3. Yêu cầu Nộp Bài

- **Định dạng:** Nộp 01 file PDF được xuất từ Google Colab hoặc Jupyter Notebook.
- **Nội dung file nộp:**
  - **Hiển thị đầy đủ kết quả:** File PDF phải chứa toàn bộ code và kết quả chạy (outputs) của các cell. Giảng viên sẽ không chạy lại code.
  - **Bình luận và Phân tích:** Cần có các phần diễn giải (markdown cells) để giải thích code, phân tích kết quả và đưa ra nhận xét, đánh giá về các phương pháp đã sử dụng. Đây là phần quan trọng để đánh giá mức độ hiểu bài.