

Viện Công nghệ Thông tin và Truyền thông  
Bộ môn Hệ thống Thông tin  
**Thi cuối kỳ 20202**  
Ngày thi: 08/09/2021

Đề thi môn **IT4652 Học sâu và ứng dụng**  
**PHẦN 1**  
Thời gian: 30 phút  
<Đề thi mở, không trao đổi. Viết trên giấy và  
chụp lại gửi lên Microsoft Forms>

<b>Giáo viên ra đề</b> Đã ký Trần Việt Trung	<b>Bộ môn duyệt</b>
--	---------------------

### Bài 1

Xác định các cụm từ đồng nghĩa, trái nghĩa, hay không có quan hệ là bài toán thú vị trong xử lý ngôn ngữ tự nhiên. Đây có thể coi là bài toán phân loại 3 lớp: Đồng nghĩa, Trái nghĩa, Không có quan hệ. Dưới đây là một mẫu dữ liệu:

- + “lập trình viên”, “kỹ sư phần mềm”, Đồng nghĩa
- + “xuất sắc”, “kém cỏi”. Trái nghĩa
- + “Đại học Bách Khoa Hà Nội”, “Chủ nghĩa tư bản”, Không có quan hệ

Bộ dữ liệu cho trước gồm 20.000 cặp đồng nghĩa, 15.000 cặp trái nghĩa và 5.000 cặp không có quan hệ.

s

Để xây dựng mô hình cho bài toán này, các em hãy trình bày các việc sau:

1. Tiền xử lý và làm giàu dữ liệu
2. Xác định đầu vào và đầu ra cho bài toán
3. Xây dựng kiến trúc mạng nơ ron phù hợp
4. Giải thuật tối ưu mạng nơ ron sử dụng
5. Xây dựng hàm mất mát phù hợp
6. Các tham số khởi tạo khi huấn luyện mô hình
7. Trình bày giả mã nguồn (pseudo code) quá trình huấn luyện mô hình
8. Các vấn đề giám sát và thay đổi trong quá trình huấn luyện mô hình