



XỬ LÝ NGÔN NGỮ TỰ NHIÊN

GIỚI THIỆU MÔN HỌC

NGUYỄN TRỌNG CHÍNH



TRÌNH BÀY

1. MỤC TIÊU MÔN HỌC
2. NỘI DUNG MÔN HỌC
3. TÀI LIỆU HỌC TẬP
4. TRIỂN KHAI LỚP HỌC
5. PHƯƠNG PHÁP HỌC
6. ĐÁNH GIÁ KẾT QUẢ HỌC TẬP
7. CÁC QUY ĐỊNH



MỤC TIÊU MÔN HỌC



MỤC TIÊU MÔN HỌC

- Hiểu và nắm vững các khái niệm cơ bản trong xử lý ngôn ngữ tự nhiên.
- Hiểu các bài toán phân tích cảm xúc, gán nhãn từ loại và phân tích cú pháp .
- Hiểu một số phương pháp phân lớp cơ bản và một số kiến trúc mạng nơron để giải quyết các bài toán trong môn học
- Áp dụng các phương pháp và xây dựng mô hình để giải quyết một số bài toán thực tế.



NỘI DUNG MÔN HỌC



NỘI DUNG MÔN HỌC

- Chương 1. Phân tích cảm xúc.
- Chương 2. Mô hình ngôn ngữ.
- Chương 3. Ngữ nghĩa vector.
- Chương 4. Gán nhãn từ loại.
- Chương 5. Phân tích cú pháp.



TÀI LIỆU HỌC TẬP



TÀI LIỆU HỌC TẬP

Sách tham khảo:

- Dan Jurafsky and James H. Martin. Speech and Language Processing. [Online]. <https://web.stanford.edu/~jurafsky/slp3/>
- Natural Language Processing, David Bamman, Spring 2020: <https://people.ischool.berkeley.edu/~dbamman/nlp20.html>
- Applied Natural Language Processing, David Bamman, Fall 2021: <https://people.ischool.berkeley.edu/~dbamman/info256.html>



TRIỂN KHAI LỚP HỌC



TRIỂN KHAI LỚP HỌC

1. Lớp lý thuyết:

- Giới thiệu các bài toán cơ bản trong xử lý ngôn ngữ tự nhiên.
- Trình bày các phương pháp giải quyết cho mỗi bài toán.
- Trình bày phương pháp đánh giá.
- Thảo luận về đề án môn học.



TRIỂN KHAI LỚP HỌC (tt)

2. Lớp thực hành:

- Hướng dẫn cài đặt và huấn luyện các mô hình.
- Hướng dẫn sử dụng các thư viện lập trình.
- Công cụ thực hành: Python, Google Colab.

3. Nhóm thuyết trình:

- Áp dụng lý thuyết vào vấn đề cụ thể
- Thử nghiệm, đánh giá
- Trình bày các kết quả của nhóm.



PHƯƠNG PHÁP HỌC



PHƯƠNG PHÁP HỌC

1. Nghe giảng:

- Hiểu vấn đề.
- Hiểu phương pháp giải quyết vấn đề.

2. Đọc tài liệu và làm bài tập:

- củng cố kiến thức và mở rộng vấn đề.
- Tìm hiểu thêm các phương pháp giải quyết khác.



PHƯƠNG PHÁP HỌC (tt)

3. Thực hành:

- Cài đặt chương trình để áp dụng kiến thức.
- Đánh giá kết quả để kiểm chứng và hình thành ý tưởng mới



ĐÁNH GIÁ KẾT QUẢ HỌC TẬP



ĐÁNH GIÁ KẾT QUẢ HỌC TẬP

1. Báo cáo đồ án cuối kỳ (50% điểm tổng kết, theo nhóm)

- Xác định bài toán cần giải quyết của đồ án.
- Các khái niệm, các mô hình và phương pháp đánh giá kết quả đồ án.
- Giải quyết vấn đề lý thuyết liên quan đến đồ án.

2. Các bài tập thực hành (50% điểm tổng kết, cá nhân)

- Cài đặt chương trình cho 3 bài tập thực hành tính điểm.
- Hiệu quả của chương trình phải trên baseline.



CÁC QUY ĐỊNH



CÁC QUY ĐỊNH

1. Lập nhóm

- Số lượng 2 – 3 sinh viên.
- Lập nhóm ngay từ đầu để phân chia công việc rõ ràng.



CÁC QUY ĐỊNH

2. Các bài nộp

- Không copy..
- Được phép sử dụng AI, nhưng phải trình bày lại.
- Chỉ nộp trên courses.uit.edu.vn.
- Phải nộp đúng hạn.



CÁC QUY ĐỊNH

3. Chấm đồ án

- Trình bày (50% điểm đồ án). Thành viên không tham gia sẽ không có điểm này
- Cuốn báo cáo (50% điểm đồ án). Điểm chung cho nhóm.
- Source code: nếu thiếu hoặc không đúng như trong cuốn báo cáo sẽ trừ vào các điểm cài đặt đánh giá (2đ) và nhận xét (2đ).
- Data: nếu thiếu hoặc không đúng như trong cuốn báo cáo sẽ trừ vào điểm ngữ liệu (2đ) và nhận xét (2đ).



CÁC QUY ĐỊNH

4. Khung điểm đồ án:

1. Giới thiệu bài toán. (giải quyết vấn đề gì? thách thức gì?) (2đ).
2. Trình bày ngữ liệu của bài toán. Phân tích thủ công 40-50 mẫu (nhóm 2 SV) hoặc 60-70 mẫu (nhóm 3 SV) (2đ).
3. Phương pháp (mô hình) kèm ví dụ minh họa từ ngữ liệu của đồ án (2đ).



CÁC QUY ĐỊNH

4. Khung điểm đồ án (tt):

5. Cài đặt thử nghiệm và đánh giá. (2đ)

6. Nhận xét kết quả đánh giá và diễn giải một số mẫu đúng và sai. (2đ)



CÁC QUY ĐỊNH

5. Tham dự lớp học:

- Khuyến khích tham dự đầy đủ các buổi học để hiểu bài và thảo luận về đề án.
- Diễn đàn môn học cũng là kênh tham gia lớp học chính thức.
- Thường xuyên xem thông báo để biết thông tin về các bài thực hành tính điểm, thời gian trình bày đề án và thời hạn nộp file báo cáo đề án.



CÁC QUY ĐỊNH

6. Đề tài:

- Phải thông báo trước khi thực hiện những đề án làm cho nhiều môn học khác nhau để tránh bị xác định là đạo văn khi chấm.
- Các đề tài theo các vấn đề đã học, gồm: phân tích cảm xúc, word embeddings, mô hình ngôn ngữ, gán nhãn từ loại, trích chọn thông tin, phân tích cú pháp thành tố và phân tích cú pháp phụ thuộc.



CÁC QUY ĐỊNH

6. Đề tài:

- Đề tài RAG: giải quyết vấn đề tạo sinh câu trả lời.
- Đề tài QA: Giải quyết vấn đề xác định câu trả lời.
- Đề tài IR: giải quyết vấn đề chọn term hoặc biểu diễn văn bản theo ngữ nghĩa vector. Không áp dụng cho ảnh, âm thanh và phim.
- Các đề tài khác: cần thảo luận với giảng viên về tính phù hợp với môn học trước khi tiến hành.