



CS221(2023 01) - meow

Xử lý ngôn ngữ tự nhiên (Trường Đại học Công nghệ thông tin, Đại học Quốc gia Thành phố Hồ Chí Minh)



Scan to open on Studocu



ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN

ĐỀ CƯƠNG MÔN HỌC

Môn: XỬ LÝ NGÔN NGỮ TỰ NHIÊN
(Tiếng Anh: NATURAL LANGUAGE PROCESSING)

Chương trình đào tạo:

Cử nhân Trí tuệ nhân tạo;

Cử nhân Khoa học máy tính.

(Bản Cập nhật tháng 01 năm 2023)

Người biên soạn: Nguyễn Thị Quý

TP. HCM – 2023

1. THÔNG TIN CHUNG (General information)

1.1 Thông tin về giảng viên:

- Họ và tên: Nguyễn Thị Quý
- Chức danh, học vị: Tiến sĩ.
- Email: quynt@uit.edu.vn

1.2. Thông tin về môn học

Tên môn học (tiếng Việt):	Xử lý ngôn ngữ tự nhiên
Tên môn học (tiếng Anh):	Natural Language Processing
Mã môn học:	CS221
Thuộc khối kiến thức:	Đại cương <input type="checkbox"/> ; Cơ sở nhóm ngành <input type="checkbox"/> ; Cơ sở ngành <input type="checkbox"/> ; Chuyên ngành <input checked="" type="checkbox"/> ; Tốt nghiệp <input type="checkbox"/>
Khoa/Bộ môn phụ trách:	Khoa Khoa học Máy tính
Giảng viên phụ trách biên soạn:	TS. Nguyễn Thị Quý
Số tín chỉ:	4
Lý thuyết:	3
Thực hành:	1
Môn học tiên quyết:	
Môn học trước:	Nhập môn lập trình (IT001), Đại số tuyến tính (MA003), Cấu trúc rời rạc (MA004)

2. MÔ TẢ MÔN HỌC (Course description)

Xử lý ngôn ngữ tự nhiên (NLP) nghiên cứu kết hợp hai lĩnh vực Ngôn ngữ học và Khoa học máy tính nhằm giúp máy tính có thể hiểu ngôn ngữ như con người. Môn học này giới thiệu đến sinh viên những phương thức, thuật toán cơ bản được sử dụng trong nhiều bài toán NLP như phân tích cảm xúc, gán nhãn chuỗi, phân tích cú pháp, ... Sinh viên sẽ thực hành áp dụng các thuật toán này trên dữ liệu để tạo ra các công cụ NLP.

3. MỤC TIÊU VÀ CHUẨN ĐẦU RA MÔN HỌC

3.1 Mục tiêu môn học:

3.1.1. Mục tiêu chung: Trang bị cho người học những kiến thức cơ bản về một số bài toán NLP.

3.1.2. Mục tiêu cụ thể

Sau khi hoàn thành môn học này người học có thể có được các khả năng sau:

Ký hiệu	Mục tiêu môn học
CG1	Hiểu và nắm vững các khái niệm, các thuật toán, kiến thức tổng quan về NLP.
CG2	Áp dụng các thuật toán để xây dựng ứng dụng NLP.
CG3	Sử dụng các thư viện như NLTK, Gensim, Scikit-learn để xây dựng ứng dụng NLP viết bằng ngôn ngữ lập trình Python
CG4	Rèn luyện ngoại ngữ và các kỹ năng nghiên cứu khoa học, làm việc nhóm

3.2 Chuẩn đầu ra.

Sau khi hoàn thành môn học này, sinh viên có thể:

CĐRMH	Mô tả CĐRMH	Ánh xạ CĐR CTĐT	Cấp độ CĐRMH về NT, KN, TD
CLO1	Đánh giá được ưu điểm và nhược điểm của từng thuật toán máy học áp dụng trên bài toán NLP.	LO3.4	KN4
CLO2	Biết cách chọn đặc trưng, thuật toán máy học phù hợp để giải quyết bài toán NLP cụ thể.	LO3.3, LO3.4	KN4
CLO3	Biết cách đánh giá độ chính xác bài toán NLP cụ thể.	LO2.2, LO4.3	NT3, KN4
CLO4	Vận dụng các kỹ năng của nghiên cứu khoa học như đọc bài báo khoa học, thuyết trình, đặt câu hỏi.	LO6.2, LO5.2, LO7.1	KN3
CLO5	Cài đặt một ứng dụng NLP sử dụng các thư viện như NLTK, Gensim, Scikit-learn, viết bằng ngôn ngữ lập trình Python.	LO4.1, LO2.1	KN4
CLO6	Có thái độ, quan điểm và nhận thức đúng đắn về môn học	LO8.2	KN2

3.3 Mối quan hệ giữa Mục tiêu và Chuẩn đầu ra:

Mối quan hệ giữa mục tiêu môn học và chuẩn đầu ra môn học được thể hiện trong ma trận sau:

Các thành phần	Mục tiêu	Chuẩn đầu ra
Kiến thức	CG1	CLO1, CLO2
Kỹ năng	CG2, CG3	CLO5, CLO3, CLO4
Thái độ	CG4	CLO6

4. YÊU CẦU ĐỐI VỚI NGƯỜI HỌC

- Dự lớp đầy đủ;
- Tham gia tích cực, nghiêm túc các hoạt động thảo luận, thực hành học tập trên lớp theo yêu cầu và tích cực tự nghiên cứu nội dung môn học theo hướng dẫn;
- Thực hiện các bài tập kiểm tra đầy đủ và đúng thời gian quy định.

5. NỘI DUNG MÔN HỌC, KẾ HOẠCH GIẢNG DẠY

Buổi	Nội dung giảng dạy	CĐRMH	Hoạt động dạy và học	Thành phần đánh giá
1	Giới thiệu về môn học <ul style="list-style-type: none">- các nội dung giảng dạy- phương pháp học- tài nguyên học tập- đánh giá môn học Giới thiệu tổng quan về một số bài toán NLP cơ bản, NLP applications, ngữ liệu	CLO1, CLO4		A1
2-4	Phân loại văn bản (text classification):	CLO2, CLO3, CLO4, CLO5, CLO6		A1, A2

	<ul style="list-style-type: none"> - Sentiment analysis using Naive Bayes, Logistic regression - Feedforward neural network 			
5, 6	Vector semantics and static word embeddings	CLO2, CLO3, CLO4, CLO5, CLO6		A1, A2
7, 8	Mô hình ngôn ngữ (language modeling): <ul style="list-style-type: none"> - N-gram model - Recurrent neural network (RNN) 	CLO2, CLO3, CLO4, CLO5, CLO6		A1, A2
9-11	Gán nhãn chuỗi (sequence labeling problems): <ul style="list-style-type: none"> - POS tagging, named-entity recognition (HMM, MEMM, CRF) - Neural sequence labeling 	CLO2, CLO3, CLO4, CLO5, CLO6		A1, A2
12-14	Phân tích cú pháp (syntax and parsing) <ul style="list-style-type: none"> - Context-free syntax - Context-free parsing algorithms - Dependency parsing 	CLO2, CLO3, CLO4, CLO5, CLO6		A1, A2
15	Ôn tập	CLO2, CLO3, CLO4, CLO5, CLO6		A1

6. PHƯƠNG PHÁP VÀ PHƯƠNG TIỆN DẠY HỌC

- Phương pháp giảng dạy chủ yếu: trình bày các kiến thức nền tảng, hướng dẫn

nghiên cứu tài liệu; hướng dẫn thực hành và diễn giải theo chủ đề.

- Học liệu: các nguồn tài liệu được sử dụng trên lớp và máy chiếu các slides.

7. ĐÁNH GIÁ MÔN HỌC (Course assessment)

Thành phần	CĐRMH	Tỉ lệ (%)
A1. Quá trình		50%
A2. Đồ án		50%

8. TÀI LIỆU THAM KHẢO

- *Speech and Language Processing (3rd ed. draft)*, Dan Jurafsky and James H. Martin, 2021: <https://web.stanford.edu/~jurafsky/slp3/>
- *Natural Language Processing*, Jacob Eisenstein, 2018
- *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>

9. PHẦN MỀM HAY CÔNG CỤ HỖ TRỢ THỰC HÀNH

- Google Colab, Kaggle
- Ngôn ngữ lập trình Python

10. Ngày phê duyệt:

11. Cấp phê duyệt: Bộ môn Tính toán Đa phương tiện.

Tp.HCM, ngày 12 tháng 01 năm 2023

Trưởng bộ môn
(Ký và ghi rõ họ tên)

Ngô Đức Thành

Giảng viên biên soạn
(Ký và ghi rõ họ tên)

Nguyễn Thị Quý