

ĐỀ CƯƠNG CHI TIẾT HỌC PHẦN

1. Tên học phần: Xử lý ngôn ngữ tự nhiên (Natural Language Processing)
- Mã số học phần: CT219

- Số tín chỉ học phần: 3 tín chỉ

- Số tiết học phần: 30 tiết lý thuyết, 30 tiết thực hành và 75 tiết tự học.
2. Đơn vị phụ trách học phần:
- Bộ môn: Khoa học máy tính

- Khoa: Công nghệ Thông tin và Truyền thông
3. Điều kiện:
- Điều kiện tiên quyết: không có

- Điều kiện song hành: không có
4. Mục tiêu của học phần:

Mục tiêu	Nội dung mục tiêu	CDR CTĐT
4.1	- Nắm vững các kiến thức chuyên sâu về xử lý ngôn ngữ tự nhiên và có khả năng tổng hợp, phát triển mở rộng và áp dụng các kiến thức này để giải quyết những bài toán cụ thể, phù hợp yêu cầu thực tế.	2.1.3.c
4.2	- Hiểu biết và có khả năng giải thích, phân tích các mô hình xử lý ngôn ngữ tự nhiên - Có khả năng thiết kế, cài đặt, triển khai và đánh giá mô hình xử lý ngôn ngữ tự nhiên để giải quyết bài toán thực tiễn	2.2.1.b

5. Chuẩn đầu ra của học phần:

CDR HP	Nội dung chuẩn đầu ra	Mục tiêu	CDR CTĐT
	Kiến thức		
CO1	Trình bày được các kiến thức cơ bản về lĩnh vực xử lý ngôn ngữ tự nhiên và ứng dụng của xử lý ngôn ngữ tự nhiên trong thực tế	4.1	2.1.3.c
CO2	Phân biệt và giải thích được các mô hình sử dụng trong xử lý ngôn ngữ	4.1	2.1.3.c
CO3	Trình bày được các phương pháp biểu diễn từ và các phương pháp đo lường sự tương đồng ngữ nghĩa giữa	4.1	2.1.3.c

	các từ		
CO4	Mô tả và giải thích được phương pháp gán nhãn từ loại	4.1	2.1.3.c
CO5	Mô tả và giải thích được phương pháp phân tích cú pháp	4.1	2.1.3.c
CO6	Vận dụng các kiến thức về mô hình ngôn ngữ và ngôn ngữ để rút trích thông tin	4.1	2.1.3.c
CO7	Biết cách vận dụng, tổng hợp các kiến thức đã học và tự nghiên cứu để thiết kế và xây dựng hệ thống giải quyết bài toán trong lĩnh vực xử lý ngôn ngữ tự nhiên đáp ứng yêu cầu thực tế	4.1	2.1.3.c
	Kỹ năng		
CO8	Trình bày, giải thích, phân tích và đánh giá được phương pháp giải quyết bài toán trong lĩnh vực xử lý ngôn ngữ tự nhiên	4.2	2.2.1.b
CO9	Thiết kế, lập trình và cài đặt thực nghiệm phương pháp giải quyết bài toán trong lĩnh vực xử lý ngôn ngữ tự nhiên	4.2	2.2.1.b

6. Mô tả tóm tắt nội dung học phần:

Học phần xử lý ngôn ngữ tự nhiên cung cấp kiến thức nền tảng về xử lý ngôn ngữ tự nhiên, tập trung vào các ứng dụng trên ngôn ngữ của con người. Nội dung của học phần chỉ tập trung xử lý ngôn ngữ ở dạng văn bản mà không bàn đến xử lý ngôn ngữ ở dạng âm thanh (tiếng nói). Trong học phần, sinh viên cũng được cung cấp thông tin về tình hình xử lý ngôn ngữ tiếng Việt. Học phần này cũng trang bị kiến thức và kỹ năng cần thiết giúp người học có khả năng tự nghiên cứu, thích ứng với sự thay đổi của công nghệ, hướng phát triển trong tương lai của các mô hình xử lý ngôn ngữ tự nhiên.

7. Cấu trúc nội dung học phần:

7.1. Lý thuyết

	Nội dung	Số tiết	CDR HP
Chương 1.	Tổng quan xử lý ngôn ngữ tự nhiên	3	
1.1.	Định nghĩa và ứng dụng của xử lý ngôn ngữ tự nhiên	1	CO1
1.2.	Kiến thức cơ bản về xử lý ngôn ngữ tự nhiên	0,5	CO1
1.3.	Các mức độ phân tích ngôn ngữ	1	CO1
1.3.	Cơ sở ngôn ngữ học	0,5	CO1
Chương 2.	Mô hình ngôn ngữ N-gram	3	
2.1.	Mô hình ngôn ngữ N-gram	1	CO2
2.2.	Các phương pháp làm mịn	1	CO2
2.3.	Đánh giá mô hình ngôn ngữ	1	CO2
Chương 3.	Vector ngữ nghĩa (vector semantics)	6	
3.1.	Khái niệm ngữ nghĩa	1	CO3
3.2.	WordNet, VerbNet và FrameNet	1	CO3
3.3.	Vector ngữ nghĩa	1	CO3
3.4.	Các phương pháp biểu diễn từ	2	CO3
3.5.	Các phương pháp đo lường sự tương đồng ngữ	1	CO3

	Nội dung	Số tiết	CDR HP
	nghĩa (measuring semantic similarity)		
Chương 4.	Gán nhãn từ loại (Part-of-Speech tagging)	6	
4.1.	Các khái niệm có liên quan	1	CO4
4.2.	Penn Treebank POS tagset	1	CO4
4.3.	Sử dụng mô hình Markov ẩn (HMM)	2	CO4
4.4.	Giải thuật Viterbi	2	CO4
Chương 5.	Phân tích cú pháp	6	
5.1.	Văn phạm phi ngữ cảnh (context-free grammars)	1	CO5
5.2.	Treebanks	1	CO5
5.3.	Nhập nhằng trong phân tích cú pháp	1	CO5
5.4.	Thuật toán CKY	1	CO5
5.5.	Mô hình xác suất PCFG	1	CO5
5.6.	Phân tích cú pháp phụ thuộc (dependency parsing)	1	CO5
Chương 6.	Rút trích thông tin	6	
6.1.	Các khái niệm có liên quan	1	CO6
6.2.	Rút trích thông tin tên riêng	1	CO6
6.3.	Rút trích mối quan hệ	1	CO6
6.4.	Rút trích thời gian	1	CO6
6.5.	Rút trích sự kiện và thời gian diễn ra sự kiện	2	CO6

7.2. Thực hành

	Nội dung	Số tiết	CDR HP
Bài 1.	Truy cập và sử dụng tài nguyên từ các kho ngữ liệu ngôn ngữ	5	CO1
Bài 2.	Xử lý dữ liệu thô và rút trích thông tin	5	CO1; CO6
Bài 3.	Phân tích cấu trúc câu	5	CO5
Bài 4.	Giải quyết bài toán thực tiễn ứng dụng xử lý ngôn ngữ tự nhiên	15	CO7; CO8; CO9
5.1	Chia nhóm tìm hiểu, đề xuất bài toán cụ thể trong thực tiễn sử dụng xử lý ngôn ngữ tự nhiên	1	
5.2.	Tìm hiểu và báo cáo phương pháp giải quyết bài toán đã đề xuất	2	
5.3.	Lập trình, thực nghiệm và đánh giá kết quả đạt được	6	
5.4.	Báo cáo kết quả thực hiện	1	

8. Phương pháp giảng dạy:

- Lý thuyết:
 - o Giảng viên diễn giảng, đặt vấn đề trao đổi với sinh viên.
 - o Giảng viên tổ chức thảo luận nhóm: sinh viên trao đổi, giảng viên hướng dẫn thảo luận để giải quyết bài toán
- Thực hành:
 - o Giảng viên giao bài thực hành để sinh viên chuẩn bị ở nhà và hướng dẫn sinh viên thực hành trong phòng máy tính.
 - o Giảng viên gợi ý đề tài để sinh viên tự nghiên cứu, thực hành và viết báo cáo.

9. Nhiệm vụ của sinh viên:

Sinh viên phải thực hiện các nhiệm vụ như sau:

- Tham dự tối thiểu 80% số tiết học lý thuyết.
- Tham gia đầy đủ 100% giờ thực hành và có báo cáo kết quả.
- Thực hiện đầy đủ các bài tập nhóm, đề tài và được đánh giá kết quả thực hiện đúng thời gian qui định.
- Tham dự kiểm tra giữa học kỳ.
- Tham dự thi kết thúc học phần.
- Chủ động tổ chức thực hiện giờ tự học.

10. Đánh giá kết quả học tập của sinh viên:

10.1. Cách đánh giá

Sinh viên được đánh giá tích lũy học phần như sau:

TT	Điểm thành phần	Quy định	Trọng số	CDR HP
1	Điểm giữa kỳ	- Thi giữa kỳ (có thể tính điểm chuyên cần, kiểm tra đột xuất)	20%	CO1; CO2
2	Điểm thi kết thúc học phần	Bắt buộc dự thi Có thể chọn các hình thức đánh giá sau: - Thi lý thuyết và thi thực hành - Làm dự án và báo cáo	80% (40% LT + 40% TH)	CO1; CO2; CO3; CO4; CO5; CO6; CO7; CO8; CO9

10.2. Cách tính điểm

- Điểm đánh giá thành phần và điểm thi kết thúc học phần được chấm theo thang điểm 10 (từ 0 đến 10), làm tròn đến một chữ số thập phân.
- Điểm học phần là tổng điểm của tất cả các điểm đánh giá thành phần của học phần nhân với trọng số tương ứng. Điểm học phần theo thang điểm 10 làm tròn đến một chữ số thập phân, sau đó được quy đổi sang điểm chữ và điểm số theo thang điểm 4 theo quy định về công tác học vụ của Trường.

11. Tài liệu học tập:

Thông tin về tài liệu	Số đăng ký cá biệt
[1] Natural language understanding / James Allen, 9780805303346.- 006.35/ A425	CNTT.002943
[2] Natural language processing - IJCNLP 2005: Second international joint conference, Jeju Island, Korea, October 11-13, 2005: proceedings / Robert Dale ... [et al.] (eds.). (Lecture notes in computer science) Lecture notes in artificial intelligence), 9783540291725.- 006.35/ N285	CNTT.001474
[3] Speech and language processing / Daniel Jurafsky and James H. Martin, 0 13 095069 6.- 006.35/ J96	CNTT.000643
[4] Foundations of statistical natural language processing / Christopher D. Manning, Hinrich Schutze.- Cambridge, Mass.: MIT Press, 1999.- xxxvii, 680 p.; 24 cm, 9780262133609.- 410.285/ M283	CNTT.001487

[5] Speech and language processing : An introduction to natural language processing, computational linguistics, and speech recognition / Daniel Jurafsky and James H. Martin,	CNTT.001541
---	-------------

12. Hướng dẫn sinh viên tự học:

Tuần	Nội dung	Lý thuyết (tiết)	Thực hành (tiết)	Nhiệm vụ của sinh viên
1	Chương 1: Tổng quan xử lý ngôn ngữ tự nhiên 1.1. Định nghĩa và ứng dụng của xử lý ngôn ngữ tự nhiên 1.2. Kiến thức cơ bản về xử lý ngôn ngữ tự nhiên 1.3. Các mức độ phân tích ngôn ngữ 1.4. Cơ sở ngôn ngữ học	6	0	-Nghiên cứu trước: +Tài liệu [3]: Chương 1,2 + Tài liệu [1,2,4,5]
2	Chương 2: Mô hình ngôn ngữ N-gram 2.1. Mô hình ngôn ngữ N-gram 2.2. Các phương pháp làm mịn 2.3. Đánh giá mô hình ngôn ngữ	6	0	-Nghiên cứu trước: +Tài liệu [3]: Chương 3 +Tài liệu [4]: Chương 1, 5 + Tài liệu [1,2,5]
3	Chương 3: Vector ngữ nghĩa (vector semantics) 3.1. Khái niệm ngữ nghĩa 3.2. WordNet, VerbNet và FrameNet 3.3. Vector ngữ nghĩa 3.4. Các phương pháp biểu diễn từ 3.5. Các phương pháp đo lường sự tương đồng ngữ nghĩa (measuring semantic similarity)	12	0	-Nghiên cứu trước: +Tài liệu [3]: Chương 6 +Tài liệu [4]: Chương 2, 3, 10 + Tài liệu [1,2,5]
4	Chương 4: Gán nhãn từ loại (Part-of-Speech tagging) 4.1. Các khái niệm có liên quan 4.2. Penn Treebank POS tagset 4.3. Sử dụng mô hình Markov ẩn (HMM) 4.4. Giải thuật Viterbi	12	0	-Nghiên cứu trước: +Tài liệu [3]: Chương 8 +Tài liệu [4]: Chương 5 + Tài liệu [1,2,5]
5	Chương 5: Phân tích cú pháp 5.1. Văn phạm phi ngữ cảnh (context-free grammars) 5.2. Treebanks 5.3. Nhập nhằng trong phân tích cú pháp 5.4. Thuật toán CKY 5.5. Mô hình xác suất PCFG 5.6. Phân tích cú pháp phụ thuộc	12	0	-Nghiên cứu trước: +Tài liệu [3]: Chương 10, 11, 12, 13 +Tài liệu [4]: Chương 8, 9 + Tài liệu [1,2,5]

Tuần	Nội dung	Lý thuyết (tiết)	Thực hành (tiết)	Nhiệm vụ của sinh viên
	(dependency parsing)			
6	Chương 6: Rút trích thông tin (information extraction) 6.1.Các khái niệm có liên quan 6.2.Rút trích thông tin tên riêng (NER) 6.3.Rút trích mối quan hệ (relation extraction) 6.4.Rút trích thời gian 6.5.Rút trích sự kiện và thời gian diễn ra sự kiện	12	0	-Nghiên cứu trước: +Tài liệu [3]: Chương 17 +Tài liệu [4]: Chương 7 + Tài liệu [1,2,5]
7	Bài thực hành 1	0	3	-Nghiên cứu trước: + Tài liệu [1,2,3,4,5]
8	Bài thực hành 2	0	3	-Nghiên cứu trước: + Tài liệu [1,2,3,4,5]
9	Bài thực hành 3	0	3	-Nghiên cứu trước: + Tài liệu [1,2,3,4,5]
10	Bài thực hành 4	0	6	Sinh viên báo cáo tổng quan về các hệ thống hỏi đáp dựa trên tài liệu được cung cấp. Tham khảo tổng hợp lại tất cả bài giảng lý thuyết và tự nghiên cứu từ Internet

Cần Thơ, ngày 09 tháng 5 năm 2019

TRƯỞNG BỘ MÔN



**TL. HIỆU TRƯỞNG
TRƯỜNG KHOA**

Nguyễn Hữu Hòa

Trần Nguyễn Minh Thư