

ĐỀ CƯƠNG MÔN HỌC

XỬ LÝ NGÔN NGỮ TỰ NHIÊN (Natural Language Processing)

Mã số MH : XXXXXXXXX

- Số tín chỉ	:	Tc (LT.BT&TH.TH&H&c): 3					TCHP:					
- Số tiết	:	Tg: 60 LT:45 BT: TH: ĐA: BTL/TL:15										
Số tiết	- Tổng:	60	LT:	45	BT:		TH:		ĐA		BTL/TL	15
								:		:		
- Đánh giá MH:		TT	Phương pháp đánh giá				Số lần		Trọng số (%)			
		1	Bài tập				3		30%			
		2	Kiểm tra giữa học kỳ				1		20			
		3	Thực hành, thí nghiệm									
		4	Tiểu luận, thuyết trình									
		5	Thi cuối học kỳ				1		50			
Thang điểm đánh giá		10/10										
- Môn tiên quyết	:	- Lý thuyết automat và ngôn ngữ hình thức; - Chương trình dịch								MS:		
- Môn học trước	:	- Trí tuệ nhân tạo								MS:		
- Môn song hành	:	-								MS:		
- CTĐT ngành		Khoa học máy tính										
Mã ngành	:	60.48.01										
- Ghi chú khác	:											

❖ Mục tiêu của môn học:

Mục tiêu quan trọng nhất của môn học này là cung cấp sự mô tả toàn diện, sâu về lý thuyết và kỹ thuật, được dùng trong lĩnh vực xử lý ngôn ngữ tự nhiên. Môn học không đi vào tiếp cận cụ thể những vấn đề phức tạp, chuyên biệt mà tập trung vào xác định những khái niệm cơ bản của lĩnh vực này.

Môn học nhằm cung cấp kiến thức cho sinh viên bậc đại học và sau đại học thuộc chuyên ngành Khoa học máy tính và chuyên ngành Ngôn ngữ học tính toán.

Đối với những người có kiến thức lập trình có thể tiếp thu những ý tưởng quan trọng của môn học. Các kiến thức môn học có thể giúp người học viết được chương trình cho hệ thống hiểu ngôn ngữ tự nhiên

Aims: The primary goal of the course is to provide a comprehensive and depth description of the theories and techniques in the Natural Language Processing area. This course will not study the idiosyncratic complexities of particular approaches but concentrates the underlying concepts of the field as a whole.

This course can be used both in undergraduate and graduate level in Computer science or Computational linguistics.

As a result, any student having some basic notions of programming will be able to understanding the principal ideas and techniques used. The knowledge in the course, this is enough to allow studenies to produce working systems for natural language understanding.

❖ Nội dung tóm tắt môn học:

Nội dung môn học có hai phần gồm 12 chương. Mỗi chương sẽ trình bày những vấn đề cốt lõi, yêu cầu sinh viên phải nắm vững để có thể học tiếp ở chương sau.

Phần I Xử lý cú pháp. Phần này tập trung xác định cấu trúc thành phần của câu; xác định cách thức câu được “bẻ gãy” thành các cụm từ như thế nào. Các cụm từ được tách thành các cụm từ nhỏ hơn như thế nào và cuối cùng là cấu trúc của từ.

Phần II Diễn dịch ngữ nghĩa. Phần này tập trung tính toán để tạo ngữ nghĩa độc lập ngữ cảnh được gọi là dạng luận lý – diễn dịch ngữ nghĩa. Các chương của phần này trình bày dạng luận lý, mối liên quan giữa dạng luận lý và cấu trúc cú pháp; cũng như các vấn đề giải quyết nhập nhằng cho nghĩa của từ và cấu trúc ngữ nghĩa.]

Course outline:

Content of the course have two parts, including 12 chapters. Each chapter is concentrated into core materials. The student should cover the core materials in the earlier chapters to best understand the later chapters.

Part I Syntactic processing. The goal of syntactic processing is to determine the structural components of sentences. It determines, how a sentence is broken down into subphrases, and so on, all way down to the actual structure of the words used.

Part II Semantic interpretation. The goal of this part is computing a context- independent notion of meaning, called the logical form – semantic interpretation. It discusses the distinction between logical form and final meaning representation in more detail and develops a logical form language. Then it dress the issue of how the logical form relates to syntactic structure and discusses the important problem of ambiguity resolution to identify the most plausible word senses and semantic structure. Finally, it discusses some alternate methods of semantic interpretation.

❖ Tài liệu học tập

- [1] James Allen, 1995, “Natural Language Processing”, The Benjaming/Cumming Publishing Company, Inc.
- [2] Chrstopher D.Manning and Hinrich Schutze, 2001, “Foundation of Statistical Natural Language Processing”, The MIT Press Cambridge, Massachusetts, London, England.
- [3] Patrich Henry Winston, 1992, “Artificial Intelligence”, Addison – Wesley Publishing Company.
- [4] W.John Hutchins, 1992, “An Introduction to Machine Translation”, Academic Press Harcount Barce Jovanovich Publisher

❖ Các hiểu biết, các kỹ năng cần đạt được sau khi học môn học

Môn học này cung cấp các kiến thức về khoa học máy tính, ngôn ngữ học tính toán cho sinh viên theo hướng phân tích các vấn đề liên quan đến xử lý ngôn ngữ tự nhiên và các kỹ thuật giải quyết các vấn đề đó. Với cách trình bày này của môn học sẽ giúp cho sinh viên khi học xong có khả năng xây dựng được hệ thống của máy tính hiểu ngôn ngữ tự nhiên trong nhiều lĩnh vực.

Learning outcomes:

This course provides knowledge in computer science and computational linguistics.

A structure of the course is organized by problem area rather than by technique, related to natural language. Then, there are techniques for solving the problem in each chapter.

This course is well detailed to allow student to build up a system of natural language understanding.

❖ Hướng dẫn cách học - chi tiết cách đánh giá môn học:

Ngoài các sách tham khảo đã liệt kê trong đề cương, sinh viên nên xem thêm các giáo trình văn phạm tiếng Anh, Việt giảng dạy ở các trường đại học để hiểu sâu môn học và có thể vận dụng xây dựng hệ thống hiểu ngôn ngữ tự nhiên.

Tham dự bài giảng và làm bài tập sẽ được đánh giá điểm chuyên cần 15%.

Bài tập lớn sẽ thực hiện theo nhóm từ 2 đến 3 người, được giao từ tuần thứ 7; được đánh giá 20%.

Kiểm tra giữa học kỳ được đánh giá 25%.

Kiểm tra cuối học kỳ được đánh giá 40%; hình thức kiểm tra: viết, thời lượng 90 phút, theo hình thức mở.

Đối với học viên là NCS, HV được coi là hoàn tất môn học nếu điểm môn học trên 7 điểm

Learning strategies & Assessment scheme:

In addition to the mentioned documents above, student should read documents of English and Vietnamese grammar, which are lectured in undergraduate level. These know ledges are useful that will help student understand deeply this course and apply this in building working system of natural language understanding.

Evaluation of student learning:

Class attention and assignments: 15%.

Course project: 20%

It is assigned to a group with maximum of 3 members. The project will start from 7th week.

Midterm exam: 25%.

Final exam: 40% ; Exam type: Open book and writing ; Exam time: 1.5 hours

❖ Nội dung chi tiết:

▪ **PHẦN GIẢNG DẠY TRÊN LỚP:** 45 tiết

Tuần	Nội dung	Ti liệu	Ghi ch
1	<p><i>Chương 1: Giới thiệu về xử lý ngôn ngữ tự nhiên</i></p> <p>1.1 Nghiên cứu ngôn ngữ tự nhiên</p> <p>1.2 Ứng dụng của hiểu ngôn ngữ tự nhiên</p> <p>1.3 Đánh giá hệ thống hiểu ngôn ngữ tự nhiên</p> <p>1.4 Các mức khác nhau của phân tích ngôn ngữ</p> <p>1.5 Biểu diễn và hiểu ngôn ngữ</p> <p>1.6 Tổ chức của hệ thống hiểu ngôn ngữ tự nhiên</p> <p><i>Phần I Xử lý cú pháp</i></p> <p><i>Chương 2: Cơ sở ngôn ngữ học: Sơ lược cú pháp tiếng Anh</i></p> <p>2.1 Từ</p> <p>2.2 Thành phần của cụm danh từ đơn giản</p> <p>2.3 Cụm động từ và câu đơn giản</p> <p>2.4 Ôn lại cụm danh từ</p> <p>2.5 Cụm tính từ</p> <p>2.6 Cụm trạng từ</p>	[1,4]	Hiểu
2	<p><i>Chương 3: văn phạm và phân tích cú pháp</i></p> <p>3.1 Văn phạm và cấu trúc câu</p> <p>3.2 Cách tạo một văn phạm tốt</p> <p>3.3 Bộ phân tích cú pháp từ trên xuống</p> <p>3.4 Bộ phân tích cú pháp từ dưới lên theo biểu đồ</p> <p>3.5 Văn phạm mạng truyền</p> <p>3.6 Phân tích cú pháp từ trên xuống theo biểu đồ</p>	[1]	Hiểu, nắm vững thức, kiến vận dụng
3,4	<p><i>Chương 4: Nt v văn phạm gia tổ</i></p> <p>4.1 Hệ thống nét và các văn phạm gia tổ</p> <p>4.2 Một số hệ thống nét cơ bản cho tiếng Anh</p> <p>4.3 Phân tích hình vị và Lexicon</p> <p>4.4 Văn phạm đơn giản dùng hệ thống nét</p> <p>4.5 Phân tích cú pháp với hệ thống nét</p> <p>4.6 Mạng truyền gia tổ</p> <p>4.7 Hệ thống nét tạo sinh và văn phạm đồng nhất</p>	[1,2]	Hiểu, vận dụng , tổng hợp

Tuần	Nội dung	Ti liệu	Ghi ch
4	<i>Chương 5: Văn phạm và ngôn ngữ tự nhiên</i> 5.1 Trợ động từ và cụm động từ 5.2 Điều khiển câu hỏi trong văn phạm phi ngữ cảnh 5.3 Mệnh đề quan hệ	[1]	Hiểu, vận dụng, tổng hợp
5	<i>Chương 6: Các phương pháp phân tích cú pháp hiệu quả</i> 6.1 Tham chiếu của con người trong phân tích cú pháp 6.2 Mã hóa không chắc chắn: Bộ phân tích cú pháp đẩy– thu giảm 6.3 Bộ phân tích cú pháp đơn định 6.4 Kỹ thuật mã hóa sự nhập nhằng 6.5 Phân tích cú pháp từng phần	[1]	Hiểu, nắm vững
6,7	<i>Chương 7: Giải quyết nhập nhằng: các phương pháp thống kê</i> 7.1 Lý thuyết cơ bản về xác suất 7.2 Xác suất ước lượng 7.3 Gán nhãn từ loại (POST) 7.4 Xác suất từ vựng 7.5 Văn phạm phi ngữ cảnh có xác suất 7.6 Phân tích cú pháp trước tốt nhất (best – first parsing) 7.7 Bộ phân tích cú pháp trước tốt nhất đơn giản ngữ cảnh phụ thuộc	[1,2]	Hiểu, vận dụng, tổng hợp
7,8	<i>Phần II Diễn dịch ngữ nghĩa</i> <i>Chương 8 : Ngữ nghĩa và dạng luận lý</i> 8.1 Ngữ nghĩa và dạng luận lý 8.2 Nghĩa của từ và nhập nhằng 8.3 Ngôn ngữ của dạng luận lý cơ bản 8.4 Mã hóa nhập nhằng trong dạng luận lý 8.5 Các động từ, trạng thái trong dạng luận lý 8.6 Vai trò chủ đề	[1]	Hiểu, nắm vững, vận dụng
9, 10	<i>Chương 9 : Liên kết cú pháp và ngữ nghĩa</i> 9.1 Diễn dịch ngữ nghĩa và sự kết hợp 9.2 Văn phạm đơn giản và từ vựng với diễn dịch	[1]	Hiểu, nắm vững, vận dụng

Tuần	Nội dung	Ti liệu	Ghi ch
	<p>ngữ nghĩa</p> <p>9.3 Cụm giới từ và cụm động từ</p> <p>9.4 Diễn dịch ngữ nghĩa từ vựng hóa và vai trò ngữ nghĩa</p> <p>9.5 Điều khiển các câu hỏi đơn giản</p> <p>9.6 Diễn dịch ngữ nghĩa bằng nét hợp nhất</p> <p>9.7 Tạo sinh câu từ dạng luận lý</p>		
11, 12	<p><i>Chương 10 : Giải quyết nhập nhằng ngữ nghĩa</i></p> <p>10.1 Giới hạn lựa chọn</p> <p>10.2 Lọc ngữ nghĩa bằng việc dùng giới hạn lựa chọn</p> <p>10.3 Mạng ngữ nghĩa</p> <p>10.4 Giải quyết nhập nhằng nghĩa từ bằng phương pháp thống kê</p> <p>10.5 Hướng đến ngữ nghĩa bằng phương pháp thống kê</p> <p>10.6 Các hướng tổng hợp đến giải quyết nhập nhằng</p>	[1, 2, 3]	Hiểu, nắm vững, vận dụng
13	<p><i>Chương 11 : Các chiến thuật khác cho sự diễn dịch ngữ nghĩa</i></p> <p>11.1 Văn phạm quan hệ</p> <p>11.2 Văn phạm ngữ nghĩa</p> <p>11.3 So trùng mẫu</p> <p>11.4 Kỹ thuật phân tích hướng đến ngữ nghĩa</p>	[1]	Hiểu, nắm vững kiến thức
13,14	<p><i>Chương 12 : Tầm vực và biên dịch cụm danh từ</i></p> <p>12.1. Hiện tượng tầm vực đặc biệt</p> <p>12.2. Định nghĩa về sự mô tả và tầm vực</p> <p>12.3. Phương pháp tầm vực khi phân tích cú pháp</p> <p>12.4. Đồng tham chiếu và ràng buộc</p> <p>12.5. Cụm tính từ</p> <p>12.6. Danh từ quan hệ và định danh</p> <p>12.7. Những vấn đề khác trong ngữ nghĩa</p>	[1]	Hiểu, nắm vững kiến thức
14,15	<p><i>Phần III Ngữ cảnh và tri thức thế giới</i></p> <p><i>Chương 13: Biểu diễn tri thức và suy diễn</i></p> <p>13.1 Biểu diễn tri thức</p> <p>13.2 Biểu diễn trên cơ sở FOPC</p>	[1]	Hiểu, nắm vững, vận dụng

Tuần	Nội dung	Ti liệu	Ghi ch
	13.3 Biểu diễn thông tin khuôn mẫu 13.4 Xử lý định lượng ngôn ngữ tự nhiên 13.5 Thời gian và phân lớp các động từ 13.6 Ngữ nghĩa thủ tục và hỏi đáp		

▪ **PHẦN BÀI TẬP, TIỂU LUẬN NGOẠI KHÓA, HV ĐI NGHIÊN CỨU THỰC TIỄN NGOÀI TRƯỜNG: 15 tiết**

TT	Nội dung	Số tiết	Địa điểm	TLTK
1	Nội dung các chương trên	15	ĐHBK	[1], [2] và các bài báo

Tp. Hồ Chí Minh, ngày tháng năm 2013

**KT. TRƯỞNG KHOA
PHÓ TRƯỞNG KHOA**

CB PHỤ TRÁCH LẬP ĐỀ CƯƠNG

TS. PHẠM TRẦN VŨ

GS. TS. PHAN THỊ TƯƠI