

ĐỀ CƯƠNG MÔN HỌC

1. THÔNG TIN CHUNG

Tên môn học (tiếng Việt): **Ngôn ngữ học máy tính**

Tên môn học (tiếng Anh): **Computational Linguistics**

Danh mã môn học:

Thuộc khối kiến thức: Cơ sở (bắt buộc)

Khoa/Bộ môn quản lý: Ngoại ngữ

Giảng viên phụ trách: PGS.TS. Nguyễn Ngọc Vũ

Điện thoại: 0918025951

Email: vunn@huflit.edu.vn

Giảng viên khác: TS. Đinh Điền

PGS.TS. Nguyễn Tuấn Đăng

Môn học trước: Không

2. MÔ TẢ MÔN HỌC

Khóa học giới thiệu những ứng dụng thực tế của ngôn ngữ học máy tính. Người học sẽ học các nguyên tắc cơ bản về cách máy tính được sử dụng để biểu diễn và xử lý thông tin dạng văn bản và lời nói. Người học cũng sẽ được giới thiệu về những thách thức của các vấn đề kỹ thuật ngôn ngữ trong thế giới thực và tìm hiểu cách chúng được xử lý bằng ngôn ngữ lập trình Python. Các chủ đề chính bao gồm: ngữ nghĩa trong ngôn ngữ học máy tính, cú pháp trong ngôn ngữ học máy tính, xây dựng ngân hàng ngữ liệu cú pháp (treebank), phân loại tài liệu, ứng dụng của ngôn ngữ học máy tính. Sinh viên sẽ được đào tạo thực hành về những

kiến thức cơ bản về xử lý văn bản bằng Python và sẽ có cơ hội làm việc với NLTK, một bộ ứng dụng xử lý ngôn ngữ tự nhiên phổ biến.

3. MỤC TIÊU MÔN HỌC

Mục tiêu	Mô tả
1	Hiểu các nguyên lý và khái niệm cơ bản của ngôn ngữ học máy tính
2	Giải thích được các mô hình/thuật toán/phương pháp để phân tích cấu trúc ngôn ngữ
3	Đánh giá, phân tích được những khó khăn của các hệ thống xử lý ngôn ngữ tự nhiên
4	Vận dụng được kiến thức đã học để xây dựng một số ứng dụng xử lý ngữ liệu đơn giản

4. CHUẨN ĐẦU RA CỦA MÔN HỌC

CĐR-MH	CĐR của CTĐT	Mô tả CĐR-MH	Mức độ giảng dạy
CLO1	PLO1	Hiểu các hệ thống ngôn ngữ tự nhiên từ góc nhìn của ngôn ngữ học máy tính	I
CLO2	PLO4	Ứng dụng các mô hình mô tả dữ liệu ngữ nghĩa, cú pháp tiếng Anh và tiếng Việt trong ngôn ngữ lập trình Python	T&U
CLO3	PLO3	Xây dựng được một số ứng dụng ngôn ngữ học máy tính đơn giản bằng Python	T&U
CLO4	PLO12	Xây dựng được ngân hàng khối liệu đơn giản phục vụ công tác nghiên cứu ngôn ngữ	T&U
CLO5	PLO6, PLO10	Tham gia tích cực và độc lập vào các cuộc thảo luận hội thảo bằng cách trình bày các bài tập nghiên cứu về ngôn ngữ học máy tính	U

I – Giới thiệu (Introduction); T – Giảng dạy (Teach); U - Ứng dụng (Utilize)

CLO (Course Learning Outcomes) - PLO (Program Learning Outcomes)

5. NỘI DUNG, KẾ HOẠCH GIẢNG DẠY

Buổi	Nội dung	CĐRM H	Hoạt động (giảng, trao đổi, thảo luận, thuyết trình)	Thành phần Đánh giá
Buổi 1: Số tiết: 5	Introduction to CL - History of CL - Essentials of CL - Areas of CL - Applications of CL	CLO1, CLO2	- Giảng dạy, thảo luận với các ví dụ. - Thực hành với Python	A2
Buổi: 2&3 Số tiết: 5	Word meaning & Ambiguation - Wordnet Introduction - Grice's theory of meaning - Word sense disambiguation - Machine translation approaches: Issues and Challenges	CLO1, CLO3	- Giảng dạy, thảo luận với các ví dụ. - Thực hành với Python	A2
Buổi: 4 Số tiết: 5	Part of Speech Tagging - Open vs Closed POS - Hidden Markov Model - POS tagging in Python	CLO1, CLO5	- Giảng dạy, thảo luận với các ví dụ. - Thảo luận, thuyết trình, phản biện theo nhóm. - Thực hành với Python	A2
Buổi: 5 Số tiết: 5	Text classification - Classification methods - Bag of words - Naive Bayes model	CLO2, CLO4	- Giảng dạy, thảo luận với các ví dụ. - Thực hành với Python	A2
Buổi 6&7: Số tiết: 5	Syntax and Grammar in Computational Linguistics - Constituency relations - Dependency relations - Syntax parsing - Treebanks	CLO4, CLO5	- Giảng dạy, thảo luận với các ví dụ. - Thảo luận, thuyết trình, phản biện theo nhóm.	A2

Buổi	Nội dung	CĐRM H	Hoạt động (giảng, trao đổi, thảo luận, thuyết trình)	Thành phần Đánh giá
			- Thực hành với Python	
Buổi 8: Số tiết: 5	Corpus Linguistics - Text collection - Corpus Annotation - Corpus Exploitation	CLO3, CLO5	- Giảng dạy, thảo luận với các ví dụ. - Thảo luận, thuyết trình, phản biện theo nhóm.	A2
Buổi: 9 Số tiết: 5	Applications of CL - Language Teaching - Contrastive Linguistics - Computer-Aided Translation - Computational Lexicography	CLO1, CLO2	- Giảng dạy, thảo luận với các ví dụ.	A2

6. ĐÁNH GIÁ MÔN HỌC

Thành phần đánh giá	CĐR-MH	Tỷ lệ (%)
A1. Điểm chuyên cần	CLO5	10
A2. Điểm quá trình (tham gia hoạt động nhóm, bài tập nhóm, thuyết trình, phản biện)	CLO5	30
A3. Điểm đánh giá tiểu luận kết thúc học phần	CLO1, CLO2, CLO3, CLO4	60
Tổng cộng		100

7. QUY ĐỊNH CỦA MÔN HỌC

Học viên có trách nhiệm tham dự lớp học, tham gia phát biểu, trao đổi và thảo luận các vấn đề, tình huống quản trị được đưa ra trong giờ học và nghiên cứu các tài liệu theo sự hướng dẫn của giảng viên. Học viên chỉ được đánh giá kết thúc học phần khi đảm bảo ít nhất 75% số giờ lên lớp theo quy định của học phần.

Khi điều kiện về học liệu trên hệ thống LMS được đáp ứng, giảng viên có thể tổ chức dạy trực tuyến môn học này với tỉ lệ học online là 30%.

8. TÀI LIỆU HỌC TẬP, THAM KHẢO

8.1. Giáo trình

Roland Hausser (2001), *Foundation of Computational Linguistics*. Berlin: Springer-Verlag Hedelberg Publisher. ISBN: 978-3-642-07626-8

Đinh Điền (2019). *Ngôn ngữ học Tính toán*. Tp HCM: NXB. ĐHQG-HCM.

8.2. Tài liệu tham khảo

Đinh Điền (2018). *Ngôn ngữ học Ngữ liệu*. Tp HCM: NXB. ĐHQG-HCM. ISBN: 978-604-73-6609-5

Đinh Điền (2019). *Từ điển học Tính toán*. Tp HCM: NXB. ĐHQG-HCM. ISBN: 978-604-73-6874-7

Adam Przepiórkowski, Maciej Piasecki, Krzysztof Jassem, and Piotr Fuglewicz (2013), *Computational Linguistics: Applications*, Berlin: Springer-Verlag Hedelberg Publisher. ISBN: 978-3-642-34399-5

Igor Bolshakov, Alexander Gelbukh (2004), *Computational Linguistics: Models, Resources, Applications*. Printed in Mexican. ISBN: 970-36-0147-2

Alexander Clark, Chris Fox, and Shalom Lappin (2010), *The Handbook of Computational Linguistics and Natural Language Processing*, UK: Blackwell Publishing. ISBN: 978-1-4051-5581-6

Website: www.clc.hcmus.edu.vn

Website: <https://www.aclweb.org/portal/>

GIẢNG VIÊN

TRƯỞNG KHOA/ BỘ MÔN