

IT4772 Xử lý ngôn ngữ tự nhiên

(Nội dung tóm tắt):

Môn học này cung cấp các kiến thức cơ sở về các phương pháp xử lý ngôn ngữ tự nhiên thông qua máy tính như phân tích hình thái từ, phân tích từ loại, phân tích cú pháp, phân tích ngữ nghĩa. Đó là các kiến thức cơ bản, là phương tiện giúp học viên có thể xây dựng các ứng dụng thực tế hơn như kiểm tra tính chính xác của văn bản, hiểu và tóm tắt văn bản, phân loại văn bản, trích rút thông tin, dịch máy, sinh văn bản, chuyển các văn bản thành dữ liệu có cấu trúc, giao diện ngôn ngữ tự nhiên để truy vấn CSDL...

1. Tên học phần: **Xử lý ngôn ngữ tự nhiên**

2. Mã số: **IT4772**

3. Khối lượng: 3 (3-1-0-6)

- Lý thuyết: 45 tiết
- Bài tập lớn: 15 tiết
- Thí nghiệm: 0 bài

4. Học phần thay thế:

5. Đối tượng tham dự: Sinh viên đại học các ngành kỹ thuật từ học kỳ 5.

6. Điều kiện học phần:

- Học phần tiên quyết: Sinh viên cần phải có kinh nghiệm lập trình trên một trong các ngôn ngữ C, C++, Java.
- Học phần học trước: Lý thuyết tính toán
- Học phần song hành: -

7. Mục tiêu học phần: Sinh viên sau khi học xong phải nắm bắt được các kiến thức sau:

- Cách thức biểu diễn các cấu trúc (từ, cú pháp, ngữ nghĩa) của văn bản.
- Các thuật toán để xây dựng các cấu trúc của văn bản.
- Các xây dựng các mô hình xác suất để lựa chọn kết quả tốt nhất.
- Phương pháp xây dựng các ứng dụng của xử lý ngôn ngữ trên cơ sở các kiến thức nền tảng đã học.

8. Nội dung vắn tắt học phần:

Học phần cung cấp các kiến thức về:

- Các khái niệm cơ bản về xử lý ngôn ngữ tự nhiên
- Các phương pháp phân tích từ
- Các phương pháp phân tích từ loại
- Một số cách tiếp cận cơ bản trong phân tích cú pháp
- Phương pháp biểu diễn và phân tích ngữ nghĩa
- Một số hướng nghiên cứu mới, tiêu biểu trong xử lý ngôn ngữ

9. Nhiệm vụ của sinh viên:

- Dự lớp: đầy đủ theo quy chế
- Bài tập: hoàn thành các bài tập của học phần

10. Đánh giá kết quả: KT/BT(0.3)-T(TN/TL:0.7)

- Điểm quá trình: trọng số 0.3

- Làm đầy đủ bài tập
- Dự lớp đầy đủ
- Thi cuối kỳ: tự luận, trọng số 0.7

11. Tài liệu học tập

- Bài giảng bằng tiếng Việt
- Sách và tài liệu tham khảo bằng tiếng Việt và tiếng Anh: xem đề cương chi tiết

12. Nội dung chi tiết học phần:

1. Mở đầu (3 tiết)

- 1.1 Giới thiệu chung
- 1.2 Các hướng nghiên cứu trong xử lý ngôn ngữ tự nhiên
- 1.3 Các mức phân tích trong xử lý ngôn ngữ tự nhiên
- 1.4 Các ứng dụng của xử lý ngôn ngữ tự nhiên

2 Hình thái từ (4 tiết)

- 2.1 Định nghĩa
- 2.2 Các dạng hình thái từ
- 2.3 Phân tích hình thái từ dựa trên ôtomat hữu hạn

3 Từ loại (6 tiết)

- 3.1 Định nghĩa
- 3.2 Các loại từ trong tiếng Anh và tiếng Việt
- 3.3 Các cách tiếp cận trong phân tích từ loại
 - 3.3.1 Phương pháp Hidden Markov Model
 - 3.3.2 Phương pháp dựa trên ràng buộc ngữ pháp
 - 3.3.3 Phương pháp chuyển đổi

4 Cú pháp (14 tiết)

- 4.1 Một số khái niệm
- 4.3 Phân tích cú pháp sử dụng văn phạm phi ngữ cảnh
 - 4.3.1 Giới thiệu
 - 4.3.2 Phân tích kiểu trên xuống
 - 4.3.3 Phân tích kiểu dưới lên
 - 4.3.4 Một số vấn đề trong phân tích cú pháp
- 4.4 Phân tích cú pháp theo cách tiếp cận xác suất
 - 4.4.1 Một số khái niệm về xác suất
 - 4.4.2 Phân tích cú pháp sử dụng xác suất
- 4.5 Một số cách tiếp cận hiện đại trong phân tích cú pháp

5 Ngữ nghĩa (6 tiết)

- 5.1 Một số khái niệm
- 5.2 Phương pháp biểu diễn ngữ nghĩa
- 5.3 Phân tích ngữ nghĩa
- 5.4 WordNet – cơ sở dữ liệu các mối quan hệ từ vựng
- 5.5 Một số ứng dụng của phân tích ngữ nghĩa

6. Các hướng nghiên cứu mới trong XLNNTN (12 tiết)

- 6.1 Dịch máy
 - 6.1.1 Các mô hình dịch máy cổ điển
 - 6.1.2 Cách tiếp cận dịch máy thống kê
- 6.2 Phân loại văn bản

- 6.2.1 Bài toán phân loại văn bản
- 6.2.2 Một số cách tiếp cận trong phân loại văn bản
- 6.3 Phản hồi thông tin
 - 6.3.1 Định nghĩa và các ứng dụng của phản hồi thông tin
 - 6.3.2 Một số cách tiếp cận trong phản hồi thông tin

13. Nội dung các bài thí nghiệm (thực hành, tiểu luận, bài tập lớn)

Phần thực hành tạo cho học viên có khả năng chủ động sáng tạo trong nghiên cứu và hiểu sâu hơn về môn học. Học viên sẽ có một bài tập lớn để thực hành các kiến thức đã học theo hai hướng:

1. Viết một tiểu luận so sánh các cách tiếp cận mới trong một lĩnh vực nghiên cứu cụ thể của xử lý ngôn ngữ tự nhiên.
2. Cài đặt một số thuật toán đã học và phân tích các ưu nhược điểm của chúng.

14. Tài liệu tham khảo

1. Christopher Manning and Hinrich Schütze. 1999. **Foundations of Statistical Natural Language Processing**. The MIT Press.
2. Dan Jurafsky and James Martin. 2000. **Speech and Language Processing**. PrenticeHall.
3. James Allen. 1994. **Natural Language Understanding**. The Benajmins/Cummings Publishing Company Inc.
4. Barton, E., Berwick, R., & Ristad, E.. 1987. **Computational Complexity and Natural Language**. The MIT Press. ISBN 0-26-02266-4.