


BÁO CÁO

Họ và tên	LƯƠNG MẠNH HÀ (CH2001028)
Ảnh	
Số buổi vắng	0
Bonus	8
Tên đề tài (VN)	NHẬN DẠNG VĂN BẢN ĐỐI VỚI MỘT SỐ NGÔN NGỮ HỆ LA TINH
Tên đề tài (EN)	RESEARCH ON TEXT RECOGNITION FOR SOME LATIN LANGUAGES
Giới thiệu	<ul style="list-style-type: none">Bài toán: nhận dạng ngôn ngữ tự nhiên dựa vào phân hoạch không gian (nhận dạng theo thống kê toán học), trong đó một lớp ngôn ngữ tiêu biểu được nghiên cứu đó là ngôn ngữ La Tinh như tiếng Anh, tiếng Pháp, v.vLí do chọn đề tài: Nhận dạng chữ là bài toán rất hữu ích, quen thuộc và được ứng dụng nhiều trong thực tế, đặc biệt là trong lĩnh vực nhận dạng và phân loại văn bản. Vì thế, chủ đề đã thu hút nhiều tác giả quan tâm nghiên cứu bằng các phương pháp nhận dạng khác nhau: logic mờ, giải thuật di truyền, mô hình xác suất thống kê, mô hình mạng nơ ron. Đã có

	<p>rất nhiều công trình nghiên cứu thực hiện việc nhận dạng, phân loại văn bản La Tinh đã đạt tỷ lệ chính xác cao, tuy nhiên các ứng dụng đó cũng chưa thể đáp ứng hoàn toàn nhu cầu của người sử dụng. Ngày nay, các nghiên cứu về phương pháp nhận dạng tốt hơn áp dụng cho thiết bị di động và các bài toán thời gian thực vẫn đang được thực hiện.</p> <ul style="list-style-type: none"> ● Khả năng ứng dụng thực tế, tính thời sự: Kết quả nghiên cứu góp phần nhận dạng được các loại ngôn ngữ tự nhiên hệ La Tinh, có khả năng mở rộng ứng dụng trong việc xây dựng chương trình kiểm soát e-mail, đặc biệt là chương trình phân tích bản mã tự động. Phương pháp nhận dạng ngôn ngữ tự nhiên sẽ đưa khả năng ứng dụng cao, đặc biệt đối với an ninh quốc phòng, hoặc các công ty, tổ chức có nhu cầu nhận dạng các loại ngôn ngữ tự nhiên. ● Phạm vi nghiên cứu: <ul style="list-style-type: none"> ○ Tìm hiểu tổng quan về nhận dạng; ○ Tìm hiểu các đặc trưng của ngôn ngữ La tinh. ○ Xây dựng, cài đặt một số kỹ thuật nhận dạng ngôn ngữ La tinh. ● Hướng nghiên cứu của đề tài <ul style="list-style-type: none"> ○ Nghiên cứu quá trình Markov hữu hạn trạng thái. ○ Nghiên cứu và xây dựng mô hình Markov ứng với các ngôn ngữ tự nhiên như : Tiếng Anh, Tiếng Pháp, Tiếng Đức, Tiếng Tây Ban Nha.
Mục tiêu	<ul style="list-style-type: none"> ● Nghiên cứu và xây dựng chương trình nhận dạng văn bản ứng với các ngôn ngữ tự nhiên như: Tiếng Anh, Tiếng Pháp, Tiếng Đức, Tiếng Tây Ban Nha. ● Xây dựng các chương trình thử nghiệm.
Nội dung và	<ul style="list-style-type: none"> ● Tìm hiểu và cập nhật các kiến thức và phương pháp cơ bản về nhận dạng ngôn ngữ tự nhiên, trí tuệ nhân tạo, khảo sát lý thuyết các mô hình,

<p>phương pháp thực hiện</p>	<p>công cụ toán học, thiết kế và xây dựng thuật toán, kỹ thuật tổ chức dữ liệu và ngôn ngữ lập trình.</p> <ul style="list-style-type: none"> • Tìm hiểu và xây dựng các đặc trưng ngôn ngữ mà đề tài quan tâm. • Tìm đọc các bài báo, các công trình nghiên cứu khoa học liên quan đến chủ đề nghiên cứu trong nước và trên thế giới. Cụ thể là các tài liệu kỹ thuật thống kê toán học các quá trình Markov; các quy luật ngôn ngữ như là một quá trình ngẫu nhiên dừng, không hậu quả; các kỹ thuật nhận dạng ngôn ngữ tự nhiên. Hình thành được tổng quan tương đối đầy đủ về tình hình nghiên cứu liên quan đến chủ đề hiện nay trên thế giới. • Lập trình cài đặt một số kỹ thuật nhận dạng ngôn ngữ La Tinh và đánh giá kết quả.
<p>Kết quả dự kiến</p>	<ul style="list-style-type: none"> • <i>Sử dụng các ứng dụng kỹ thuật thống kê Toán học để nhận dạng các ngôn ngữ tự nhiên và tìm hiểu đặc trưng của một số ngôn ngữ tự nhiên tiêu biểu.</i> • <i>Thuật toán nhận dạng văn bản La Tinh có thể đưa ra kết quả với một số mẫu ngắn thuộc ngôn ngữ cho trước .</i>
<p>Tài liệu tham khảo</p>	<p>[1] Hoàng Minh Tuấn, Một số vấn đề trong xây dựng siêu máy tính chi phí thay cho các ứng dụng xử lý thông tin và tính toán khoa học kỹ thuật, Luận văn Tiến sĩ kỹ thuật, Mã số 62.52.70.01. tr. 35-55.</p> <p>[2] Hồ Văn Canh, Phạm Quốc Doanh (2002), Thuật toán nhận dạng các ngôn ngữ tự nhiên, tr. 3-20.</p>

- | | |
|--|---|
| | <p>[3] Lương Mạnh Bá, Nguyễn Thanh Thuỷ (1999), Nhập môn xử lý ảnh số, Nhà xuất bản khoa học và kỹ thuật, tr.154-170.</p> <p>[4] Nguyễn Viết Thế, Hồ Văn Canh (2010), Nhập môn Phân tích thông tin có bảo mật, NXB Thông tin và Truyền thông.</p> <p>[5] Trần Duy Hưng, Nguyễn Ngọc Cường (2002), Nhận dạng tự động ngôn ngữ tiếng Anh, Tạp chí "Tin học và điều khiển học", Trung tâm Khoa học tự nhiên và Công nghệ Quốc gia số 3/2002.</p> <p>[6] Andrew R. Webb. 2002. John Wiley & Sons, Ltd. Statistical Pattern Recognition, Second Edition.</p> <p>[7] Richard O Duda, Peter E Hart, David G Stork . Wiley-interscience. "Bayesian decision theory", Pattern Classification, Second Edition: 39-78.</p> <p>[8] Wilks, S. S. 1962. Mathematical Statistics. New York: John Wiley. Mark Stam, Richard M. Low (2007): Applied Cryptanalysis Breaking ciphers in the Real World. A John Wiley & Sons, Inc, publication 2007.</p> |
|--|---|