

# Một phương pháp nhận dạng văn bản tiếng

Việt : Luận văn ThS / Lê Minh Hoàng ; Nghd.: TS.

Lương Chi Mai, TS. Ngô Quốc Tạo . - H. : Khoa Công  
Nghệ, 2001 . - 86 tr.

Tóm tắt: Trình bày cơ sở lý thuyết toán học của bài toán nhận dạng chữ Việt (chữ in và chữ viết tay). Xây dựng các mô hình nhận dạng cấu trúc chữ và mô hình nhận dạng từ tiếng Việt. Xử lý văn bản và nhận dạng văn bản. Các kết quả cài đặt thử nghiệm

## MỤC LỤC

|                                                                        |    |
|------------------------------------------------------------------------|----|
| Chương 1: Tổng quan.....                                               | 6  |
| 1. Mục đích nghiên cứu của đề tài .....                                | 6  |
| 2. Một số khái niệm .....                                              | 7  |
| IU. Vài nét lịch sử và hướng nghiên cứu trong nhận dạng chữ viết ..... | 12 |
| IV. Đối tượng và phạm vi nghiên cứu của đề tài.....                    | 21 |
| Chương 2: Cơ sở lý thuyết.....                                         | 22 |
| 1. Tính Markov .....                                                   | 22 |
| 2. Xích Markov rời rạc và thuần nhất .....                             | 23 |
| HI. Mô hình xích Markov .....                                          | 25 |
| IV. Mô hình Markov ẩn (Hidden Markov Model - HMM).....                 | 26 |
| V. Ba bài toán đặt ra đối với mô hình Markov ẩn.....                   | 30 |
| VI. Giải pháp cho ba bài toán .....                                    | 31 |
| VU. Nhận xét.....                                                      | 42 |
| Chương III: Xây dựng các mô hình nhận dạng cấu trúc chữ Việt.....      | 44 |
| 1. Mô tả cấu trúc chữ.....                                             | 45 |
| 2. Nhận dạng chữ .....                                                 | 49 |
| HI. Phân lớp các mẫu huấn luyện và mô hình nhận dạng.....              | 58 |
| IV. Xử lý nhập nhằng trong quá trình nhận dạng.....                    | 60 |
| Chương IV: Xây dựng mô hình nhận dạng từ.....                          | 62 |

|                                                                         |    |
|-------------------------------------------------------------------------|----|
| 1. Xác định miền liên thông.....                                        | 62 |
| 2. Tách dấu ra khỏi ảnh .....                                           | 64 |
| III. Xử lý nét đứt và nét dính .....                                    | 64 |
| IV. Kết luận.....                                                       | 68 |
| Chương V: Xử lý và nhận dạng văn bản .....                              | 70 |
| 1. Tiền xử lý văn bản .....                                             | 70 |
| 2. Tách dòng.....                                                       | 73 |
| IU. Tách từ.....                                                        | 73 |
| IV. Nhận dạng.....                                                      | 74 |
| V. Hậu xử lý .....                                                      | 74 |
| Kết luận.....                                                           | 75 |
| 1. Đánh giá tóm tắt kết quả đạt được.....                               | 75 |
| 2. Những vấn đề còn chưa khắc phục được và đề xuất cách giải quyết..... | 76 |
| 3. Hướng nghiên cứu tiếp theo của đề tài .....                          | 78 |
| Phụ lục: Cài đặt thử nghiệm .....                                       | 80 |
| 1. Các chức năng của hệ thống thử nghiệm.....                           | 80 |
| 2. Một số kết quả thử nghiệm.....                                       | 83 |
| Tài liệu tham khảo.....                                                 | 85 |