

ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG

TRẦN THÁI SƠN

NGHIÊN CỨU MỘT SỐ PHƯƠNG PHÁP TÌM KIẾM XÁP XỈ VÀ XÂY DỰNG ỨNG DỤNG
HỖ TRỢ LỰA CHỌN PHẢN BIỆN CHO TẠP CHÍ KHOA HỌC VÀ CÔNG NGHỆ ĐẠI HỌC
THÁI NGUYÊN

Chuyên ngành: Khoa học máy tính

Mã số: 8480101

LUẬN VĂN THẠC SĨ KHOA HỌC MÁY TÍNH

Người hướng dẫn khoa học: TS. NGUYỄN ĐÌNH DŨNG

THÁI NGUYÊN, 2018

MỤC LỤC

LỜI CẢM ƠN	vi
LỜI CAM ĐOAN.....	vii
DANH MỤC CÁC BẢNG.....	viii
DANH MỤC CÁC HÌNH VẼ.....	ix
LỜI MỞ ĐẦU	xi
CHƯƠNG 1. TỔNG QUAN.....	1
1.1. Tổng quan về hệ thống hỗ trợ lựa chọn phản biện	1
1.1.1. Thuật toán Kalmukov	2
1.1.2. Bài toán lựa chọn phản biện CMACRA	6
1.1.3. Lựa chọn phản biện với thuật toán xấp xỉ 1/3	9
CHƯƠNG 2. CƠ SỞ LÝ THUYẾT	11
2.1. Một số phương pháp so khớp mẫu trong tính toán độ gần tựa ngữ nghĩa	11
2.1.1. Thuật toán Brute Force	11
2.1.2. Thuật toán Knuth-Morris-Pratt	12
2.2. Phương pháp quy hoạch động và ứng dụng trong lựa chọn phản biện	13
2.2.1. Phương pháp quy hoạch động.....	14
2.2.2. Thuật toán lựa chọn phản biện.....	16
2.3. Otomat hữu hạn mờ và ứng dụng trong lựa chọn phản biện bài báo.....	20
2.3.1. Độ gần tựa ngữ nghĩa theo mô hình Otomat hữu hạn mờ	20
2.3.2. Thuật toán lựa chọn phản biện.....	22
CHƯƠNG 3. XÂY DỰNG ỨNG DỤNG.....	25
3.1. Phát biểu bài toán	25
3.2. Mô tả yêu cầu hệ thống	25
3.3. Phân tích hệ thống.....	30
3.3.1. Xác định Actor và Use Case	30
3.3.2. Biểu đồ Use Case	32
3.3.3. Biểu đồ lớp phân tích.....	43
3.3.4. Biểu đồ trạng thái.....	44
3.4. Thiết kế hệ thống.....	44
3.4.1. Biểu đồ tuần tự.....	44
3.4.2. Biểu đồ lớp chi tiết.....	53

3.4.3. Các đối tượng cơ sở dữ liệu và các chức năng của chương trình.....	54
KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	71
TÀI LIỆU THAM KHẢO.....	72

LỜI CẢM ƠN

Luận văn này được hoàn thành tại Trường Đại học Công nghệ Thông tin và Truyền thông dưới sự hướng dẫn của TS. Nguyễn Đình Dũng. Tác giả xin bày tỏ lòng biết ơn tới các thầy cô giáo thuộc Trường Đại học Công nghệ Thông tin và Truyền thông đã tạo điều kiện và giúp đỡ tác giả trong quá trình học tập và làm luận văn tại Trường, đặc biệt tác giả xin bày tỏ lòng biết ơn tới TS. Nguyễn Đình Dũng đã tận tình hướng dẫn và cung cấp nhiều tài liệu cần thiết để tác giả có thể hoàn thành luận văn đúng thời hạn.

Xin chân thành cảm ơn anh chị em học viên cao học và bạn bè đồng nghiệp đã trao đổi, động viên và khích lệ tác giả trong quá trình học tập và làm luận văn tại Trường Đại học Công nghệ Thông tin và Truyền thông – Đại học Thái Nguyên.

LỜI CAM ĐOAN

Tôi xin cam đoan luận văn này do chính tôi thực hiện, dưới sự hướng dẫn khoa học của TS. Nguyễn Đình Dũng, các kết quả lý thuyết được trình bày trong luận văn là sự tổng hợp từ các kết quả đã được công bố và có trích dẫn đầy đủ, Chương trình ứng dụng trong luận văn này được tác giả thực hiện là hoàn toàn trung thực, nếu sai tôi hoàn toàn chịu trách nhiệm.

Thái Nguyên, tháng 4 năm 2018

Học viên cao học

Trần Thái Sơn

DANH MỤC CÁC BẢNG

Bảng 1. Các từ khóa của phản biện (<i>ReviewerKeyword</i>).....	17
Bảng 2. Khoảng cách chuyên môn (<i>ExpertiseDistance</i>).....	17
Bảng 3. Bảng RK minh họa thuật toán.....	19
Bảng 4. Kết quả thực hiện thuật toán	19
Bảng 5. Xác định các tác nhân.....	30
Bảng 6. Xác định các use case	32
Bảng 7. Scenario cho use case Đăng nhập	33
Bảng 8. Scenario cho use case Cập nhật lý lịch sơ lược	35
Bảng 9. Scenario cho use case Tổng hợp bài báo theo phản biện.....	36
Bảng 10. Scenario cho use case Tổng hợp bài báo theo tác giả.....	39
Bảng 11. Scenario cho use case Chọn bài báo phản biện.....	42
Bảng 12. Bảng tblCriticism lưu các thông tin về phản biện.....	55
Bảng 13. Bảng tblJournal lưu các thông tin về bài báo.....	56
Bảng 14. Bảng tblChuyensan lưu trữ danh mục các chuyên san.....	57
Bảng 15. Bảng tblCri_Journal_and_Book	57
Bảng 16. Bảng tblCri_ProcessOfWork	57
Bảng 17. Bảng tblCri_ScientificResearch.....	57
Bảng 18. Bảng tblGender lưu thông tin danh mục giới tính.....	57
Bảng 19. Bảng tblKeywordKHX.....	58
Bảng 20. Bảng tblKeywordNLY.....	58
Bảng 21. Bảng tblKeywordTNK.....	58
Bảng 22. Bảng tblManageRegist.....	58
Bảng 23. Bảng tblMenu lưu thông tin danh mục menu trang chủ.....	58
Bảng 24. Bảng tblPermission lưu thông tin danh mục phân quyền trong hệ thống....	59
Bảng 25. Bảng tblResultOfCriticism	59
Bảng 26. Bảng tblUser lưu thông tin người sử dụng hệ thống	59

DANH MỤC CÁC HÌNH VẼ

Hình 1. Bước 1-Xác định ma trận độ tương tự.....	3
Hình 2. Bước 2 – Xây dựng mảng cấu trúc involvedReviewers; Bước 3 – Xử lý involvedReviewers và phân bài báo cho phản biện tại hàng đầu tiên của ma trận K; Bước 4 – Kết thúc lựa chọn phản biện.....	4
Hình 3. Nhóm Use case Quản trị hệ thống.....	32
Hình 4. Use case Đăng nhập.....	33
Hình 5. Nhóm Use case Quản lý phản biện.....	34
Hình 6. Phân rã Use case Cập nhật thông tin phản biện.....	34
Hình 7. Phân rã Use case Tổng hợp bài báo theo phản biện.....	35
Hình 8. Use case Thống kê và tìm kiếm.....	36
Hình 9. Nhóm Use case Cập nhật bài báo.....	37
Hình 10. Phân rã Use case Cập nhật bài báo khoa học tự nhiên.....	37
Hình 11. Phân rã Use case Cập nhật bài báo khoa học Xã hội và Nhân văn.....	38
Hình 12. Phân rã Use case Cập nhật bài báo Nông Lâm Ngư Y Sinh.....	38
Hình 13. Nhóm Use case Tổng hợp bài báo theo tác giả.....	39
Hình 14. Use case Đổi mật khẩu.....	40
Hình 15. Use case Đăng ký làm phản biện.....	40
Hình 16. Nhóm Use case Tìm kiếm phản biện.....	41
Hình 17. Phân rã Use case chọn bài báo phản biện.....	41
Hình 18. Nhóm Use case Chỉ định phản biện.....	42
Hình 19. Biểu đồ lớp phân tích.....	43
Hình 20. Biểu đồ trạng thái lớp Journal.....	44
Hình 21. Biểu đồ tuần tự cho chức năng đăng nhập.....	44
Hình 22. Biểu đồ tuần tự cho chức năng tìm kiếm phản biện theo từng từ khóa.....	45
Hình 23. Biểu đồ tuần tự cho chức năng tìm kiếm phản biện theo dãy từ khóa.....	46
Hình 24. Biểu đồ tuần tự cho chức năng Chọn bài báo.....	47
Hình 25. Biểu đồ tuần tự cho chức năng Cập nhật lý lịch sơ lược.....	48
Hình 26. Biểu đồ tuần tự cho chức năng Cập nhật từ khóa chuyên môn.....	49

Hình 27. Biểu đồ tuần tự cho chức năng đổi mật khẩu.....	50
Hình 28. Biểu đồ tuần tự cho chức năng Thêm bài báo.....	50
Hình 29. Biểu đồ tuần tự cho chức năng Chi định phản biện 1.....	51
Hình 30. Biểu đồ tuần tự cho chức năng Chi định phản biện 2.....	52
Hình 31. Biểu đồ lớp chi tiết.....	53
Hình 32. Sơ đồ thực thể liên kết.....	59
Hình 33. Các chức năng hệ thống hỗ trợ lựa chọn phản biện.....	60
Hình 34. Form Đăng nhập hệ thống.....	61
Hình 35. Form Chức năng tìm kiếm phản biện.....	61
Hình 36. Form Đăng nhập lại hệ thống.....	62
Hình 37. Form Đổi mật khẩu.....	62
Hình 38. Form Chức năng cập nhật thành viên.....	63
Hình 39. Form Chức năng quản lý đăng ký phản biện.....	63
Hình 40. Form Xem thông tin đăng ký tham gia phản biện.....	64
Hình 41. Form Cập nhật lý lịch của phản biện.....	64
Hình 42. Form Cập nhật quá trình đào tạo.....	65
Hình 43. Form Cập nhật quá trình công tác.....	65
Hình 44. Form Cập nhật đề tài nghiên cứu khoa học.....	66
Hình 45. Form Cập nhật công trình khoa học đã công bố.....	66
Hình 46. Form Cập nhật các từ khóa chuyên môn.....	67
Hình 47. Form Tổng hợp bài báo theo phản biện.....	67
Hình 48. Form Cập nhật bài báo.....	68
Hình 49. Form Thống kê bài báo theo các tiêu chí.....	68
Hình 50. Form Tổng hợp bài báo theo tác giả.....	69
Hình 51. Form Chức năng xem hồ sơ của phản biện.....	69

LỜI MỞ ĐẦU

Trong thời đại ngày nay, cùng với các phương pháp về chế độ tài chính, chính sách nhân sự, công nghệ thông tin (CNTT) được xác định là một trong các yếu tố quan trọng của một tạp chí. Ứng dụng CNTT trong hoạt động tạp chí có thể hiểu là việc nghiên cứu, phát triển, đầu tư đưa các công nghệ tiên tiến trong lĩnh vực CNTT vào ứng dụng thực tiễn trong các công đoạn của quá trình hoạt động của tạp chí, từ khâu nhận bài, phản biện, biên tập và xuất bản. Các hoạt động chủ yếu gồm thiết lập một chiến lược tổng thể về phát triển CNTT, các quy chế, quy trình quản lý, xây dựng hạ tầng phần cứng, đường truyền, cơ sở dữ liệu, phần mềm ứng dụng.

Hiện nay đã có một số tạp chí như Tạp chí Khoa học và Công nghệ Đại học Đà Nẵng [26], Tạp chí Khoa học kỹ thuật Mỏ - Địa chất [27]... đã có giải pháp ứng dụng công nghệ thông tin và hệ thống tiện ích của Internet, tự động hóa một số bước trong quy trình quản lý tạp chí. Nội dung chính của giải pháp là xây dựng hệ thống website tự động hóa các khâu quản lý tạp chí như: quá trình nộp bài của tác giả, mời đọc phản biện, theo dõi quá trình xử lý một công trình từ khi gửi đăng cho đến lúc bài được nhận đăng vào các số của tạp chí. Việc nghiên cứu xây dựng hệ thống phần mềm website tạp chí khoa học đã từng bước hiện đại hóa quá trình quản lý trong thời kỳ hội nhập quốc tế, tạo môi trường trao đổi giữa các nhà khoa học và tận dụng tối đa nguồn tài liệu quý.

Để đạt được và giữ vững các tiêu chí cơ bản của một tạp chí khoa học quốc tế, hiện nay một số các tạp chí quốc tế được xếp hạng bởi ISI ngoài việc đảm bảo phát hành đều đặn, liên tục, đúng thời hạn theo kế hoạch thì tạp chí còn phải tuân thủ các điều kiện chặt chẽ khác như quá trình biên tập, về nội dung bài, quá trình xét duyệt, hình thức xét duyệt. Muốn vậy, việc tác nghiệp của những tạp chí này chủ yếu phải thực hiện qua mạng, qua hệ thống phần mềm, vì qua hệ thống này các công ước biên tập quốc tế được lưu giữ trong hệ thống cơ sở dữ liệu (nội dung bài báo, tác giả và tài liệu tham khảo được trích dẫn), Ban biên tập là ban biên tập quốc tế gồm các nhà khoa học tiêu biểu cho các hướng chính của tạp chí có thể trao đổi tác nghiệp một cách hiệu quả; bài gửi đăng được tiếp nhận và xử lý theo đúng thông lệ quốc tế, được phản biện kín, độc lập và khách quan.

Nhằm nâng cao hiệu quả của việc tiếp nhận, chọn lọc và phản biện, hiện nay một số hội nghị trong nước và quốc tế đã đăng ký sử dụng dịch vụ quản lý và tổ chức hội nghị khoa học của phần mềm Easychair [28], đây là một phần mềm được đặt tại máy chủ của Khoa Khoa học máy tính thuộc Đại học Manchester. Việc nộp báo cáo của các thành viên đều nộp qua mạng tại trang đó. Quá trình lấy bài, nhận xét và cho kết quả phản biện của các thành viên trong Ban chương trình đều thực hiện trực tuyến thông qua phần mềm. Việc cho điểm bài báo được các phản biện thực hiện khách quan theo các tiêu chí và kèm theo nhận xét. Trao đổi thông tin giữa tác giả bài báo và thành viên trong Ban chương trình hay giữa các thành viên trong Ban chương trình đều thực hiện bằng email thông qua sự quản lý của hệ thống Easychair. Đối với kết quả phản biện, những tác giả có bài được chấp nhận sẽ nhận được bản nhận xét và yêu cầu chỉnh sửa, bài đã chỉnh sửa cũng sẽ được gửi lại thông qua phần mềm và cuối cùng là hệ thống sẽ tự động lập danh sách mục lục tác giả, bài được tuyển chọn phục vụ cho việc in ấn.

Tạp chí Khoa học và Công nghệ Đại học Thái Nguyên thành lập theo quyết định số 1571/QĐ-BGD&ĐT ngày 31 tháng 3 năm 2006 của Bộ trưởng Bộ Giáo dục và Đào tạo; Giấy phép hoạt động báo chí số 149/GP-BVHTT ngày 17 tháng 04 năm 2001 do Bộ Văn hóa Thông tin cấp và Giấy phép sửa đổi, bổ sung số 1270/GP-BTTTT ngày 26 tháng 8 năm 2010 do Bộ Thông tin và Truyền thông cấp với mục đích công bố các công trình nghiên cứu khoa học và công nghệ của cán bộ, sinh viên Đại học Thái Nguyên, của các nhà khoa học trong và ngoài nước; các thông tin về những kết quả nghiên cứu và hoạt động khoa học công nghệ nổi bật của các đơn vị thành viên thuộc Đại học Thái Nguyên. Sau hơn 10 năm hoạt động, Tạp chí đã góp một phần không nhỏ trong việc quảng bá, nâng cao vị thế của Đại học Thái Nguyên trong lĩnh vực hoạt động khoa học công nghệ, số lượng bài gửi đăng ngày càng lớn, chất lượng bài ngày càng đòi hỏi được nâng cao. Để tiệm cận với chuẩn quốc tế, Tạp chí Khoa học và Công nghệ Đại học Thái Nguyên cần tập trung nâng cao chất lượng các chuyên san, xử lý linh hoạt và cải tiến hơn nữa trong quy trình phản biện, biên tập; nâng cao điểm số công trình của bài báo; quảng bá rộng rãi hơn nữa các công trình nghiên cứu của tác giả trên mạng.

Đề tài luận văn ” Nghiên cứu một số phương pháp tìm kiếm xấp xỉ và xây dựng ứng dụng hỗ trợ lựa chọn phản biện cho tạp chí khoa học và công nghệ Đại học Thái Nguyên ” tập trung nghiên cứu và thực hiện 2 nội dung chính sau: