

ĐẠI HỌC BÁCH KHOA HÀ NỘI

ĐỒ ÁN TỐT NGHIỆP

Thiết kế xây dựng ứng dụng lắng nghe mạng xã hội

NGUYỄN QUANG HUY
huy.nq183554@sis.hust.edu.vn

Ngành Khoa học máy tính

Giảng viên hướng dẫn: ThS. Nguyễn Tiến Thành

Chữ kí GVHD

Khoa: Khoa học máy tính

Trường: Công nghệ Thông tin và Truyền thông

HÀ NỘI, 12/2023

LỜI CẢM ƠN

Đầu tiên, em xin được cảm ơn gia đình của em. Gia đình đã luôn ở bên cạnh động viên, hỗ trợ cả về vật chất lẫn tinh thần để em được học tập một cách tốt nhất trong năm năm vừa qua.

Em xin được gửi lời cảm ơn tới các thầy cô giảng viên trường Công nghệ thông tin và Truyền thông - Đại học Bách Khoa Hà Nội. Các thầy cô đã luôn tận tâm chỉ bảo, dạy cho em những nền tảng kiến thức để em có thể thực hiện đồ án tốt nghiệp và là hành trang việc phát triển sự nghiệp trên đoạn đường phía trước. Đặc biệt, em xin được gửi lời cảm ơn sâu sắc đến ThS. Nguyễn Tiến Thành. Thầy là người trực tiếp hướng dẫn đưa ra các định hướng để em có thể hoàn thành đồ án một cách tốt nhất.

Em cũng xin được gửi lời cảm ơn tới tất cả bạn bè của em. Họ là người cùng em trải qua tất cả mọi khoảnh khắc trong quãng thời gian sinh viên, động viên, giúp đỡ trong mọi thời điểm khó khăn.

Cuối cùng, xin cảm ơn chính bản thân mình trong năm năm học đã nỗ lực không ngừng để có thể đi đến được ngày hôm nay.

Tuy nhiên, trong quá trình thực hiện đồ án do hạn chế về thời gian, kiến thức và kinh nghiệm nên đồ án của em không tránh khỏi những sai sót. Em rất mong nhận được sự góp ý của các thầy cô và hội đồng để đề tài của em được hoàn thiện hơn.

Em xin chân thành cảm ơn.

TÓM TẮT NỘI DUNG ĐỒ ÁN

Marketing đóng vai trò quan trọng trong hoạt động kinh doanh của một tổ chức. Một chiến dịch marketing hiệu quả thường đi kèm với kết quả như xây dựng được thương hiệu và sự tin tưởng từ khách hàng, thúc đẩy doanh số bán hàng, tạo lập mối quan hệ với khách hàng, tối ưu hóa chi phí marketing.

Có rất nhiều các phương pháp để tăng hiệu quả của một chiến dịch marketing trong đó việc theo dõi, phân tích, đánh giá phản hồi từ mạng xã hội là một phương pháp tiềm năng trong thời kỳ công nghệ và mạng xã hội đóng vai trò ngày càng quan trọng trong cuộc sống, tóm gọn lại phương pháp lại được gọi là lắng nghe mạng xã hội. Lắng nghe mạng xã hội đóng vai trò quan trọng trong lĩnh vực marketing, một ứng dụng lắng nghe mạng xã hội giúp doanh nghiệp đánh giá tình hình thị trường, cung cấp thông tin thời gian thực về cảm nhận của người dùng với sản phẩm, dịch vụ của doanh nghiệp, lắng nghe ý kiến phản hồi của khách hàng trên các nền tảng mạng xã hội, tạo chiến lược tiếp thị dựa trên các thông tin thu thập được, đo lường hiệu suất chiến lược tiếp thị dựa trên tương tác, phản hồi tích cực và tiêu cực và đánh giá ảnh hưởng trên thương hiệu.

Để giải quyết vấn đề này, đồ án xây dựng ứng dụng thu thập dữ liệu sử dụng ngôn ngữ lập trình Java, framework Spring Boot xây dựng server, thư viện Selenium xây dựng ứng dụng thu thập dữ liệu, ngôn ngữ lập trình Javascript và framework ReactJs để xây dựng website trực quan hóa, hệ quản trị cơ sở dữ liệu Postgresql với các ứng dụng nguồn mở Kafka, Logstash, Elasticsearch.

Sinh viên thực hiện
(*Nguyễn Quang Huy*)

MỤC LỤC

CHƯƠNG 1. GIỚI THIỆU ĐỀ TÀI.....	1
1.1 Đặt vấn đề.....	1
1.2 Mục tiêu và phạm vi đề tài.....	1
1.3 Định hướng giải pháp.....	1
1.4 Bố cục đồ án	2
CHƯƠNG 2. KHẢO SÁT VÀ PHÂN TÍCH YÊU CẦU.....	3
2.1 Khảo sát hiện trạng	3
2.1.1 Người dùng	3
2.1.2 Các hệ thống đã có.....	3
2.2 Tổng quan chức năng	4
2.2.1 Biểu đồ use case tổng quát	4
2.2.2 Biểu đồ use case phân rã quản lý nhóm	5
2.2.3 Biểu đồ use case phân rã quản lý chủ đề.....	6
2.2.4 Biểu đồ use case phân rã quản lý tin bài	7
2.2.5 Biểu đồ use case phân rã chức năng thống kê.....	8
2.2.6 Biểu đồ use case phân rã quản lý email	9
2.2.7 Quy trình nghiệp vụ	9
2.3 Đặc tả chức năng	11
2.3.1 Đặc tả use case đăng nhập.....	11
2.3.2 Đặc tả use case thêm chủ đề.....	13
2.3.3 Đặc tả use case thêm mới nhóm.....	14
2.4 Yêu cầu phi chức năng	16
CHƯƠNG 3. CÔNG NGHỆ SỬ DỤNG.....	17
3.1 Java và Spring Boot	17

3.2 Thư viện Selenium.....	17
3.3 Javascript và ReactJs.....	17
3.4 Hệ quản trị cơ sở dữ liệu Postgresql.....	18
3.5 Kafka	18
3.6 Logstash	19
3.7 Elasticsearch	19
3.8 Hệ thống Kafka, Logstash, Elasticsearch.....	20
CHƯƠNG 4. THIẾT KẾ, TRIỂN KHAI VÀ ĐÁNH GIÁ HỆ THỐNG	21
4.1 Thiết kế kiến trúc.....	21
4.1.1 Lựa chọn kiến trúc phần mềm	21
4.1.2 Thiết kế tổng quan.....	23
4.1.3 Thiết kế chi tiết gói	24
4.2 Thiết kế chi tiết.....	30
4.2.1 Thiết kế giao diện	30
4.2.2 Thiết kế lớp	35
4.2.3 Thiết kế cơ sở dữ liệu	36
4.3 Xây dựng ứng dụng.....	41
4.3.1 Thư viện và công cụ sử dụng	41
4.3.2 Kết quả đạt được	41
4.3.3 Minh họa các chức năng chính	41
4.4 Kiểm thử.....	48
4.5 Triển khai	50
CHƯƠNG 5. CÁC GIẢI PHÁP VÀ ĐÓNG GÓP NỔI BẬT.....	51
5.1 Tổng quan về hệ thống	51
5.2 Tổng quan về hệ thống thu thập dữ liệu.....	51

5.3 Giải pháp thu thập dữ liệu từ Facebook.....	52
5.3.1 Cách các website sử dụng để tránh việc bị thu thập dữ liệu.....	52
5.3.2 Giải pháp thu thập dữ liệu.....	53
5.3.3 Tìm hiểu về các website của mạng xã hội Facebook	54
5.4 Xây dựng hệ thống thu thập dữ liệu	54
5.4.1 Hệ thống lệnh được lập lịch.....	55
5.4.2 Hàng đợi lệnh	55
5.4.3 Bot thu thập dữ liệu	55
5.5 Tích hợp hệ thống tìm kiếm tin bài với Elasticsearch.....	55
5.5.1 Tổ chức dữ liệu trên Elastichsearch	56
5.5.2 Tích hợp Elasticsearch vào tìm kiếm tin bài theo chủ đề.....	57
5.5.3 Tích hợp Elasticsearch vào tìm kiếm tin bài theo nhóm.....	58
CHƯƠNG 6. KẾT LUẬN VÀ HƯỚNG PHÁT TRIỂN	59
6.1 Kết luận.....	59
6.2 Hướng phát triển.....	59
7. ĐẶC TẢ USE CASE	61
7.0.1 Đặc tả use case đăng ký	61
7.1 Đặc tả use case “xóa chủ đề”	62
7.2 Đặc tả use case “xóa nhóm”	63
7.3 Đặc tả use case “xóa email”	64
7.4 Đặc tả use case “xem thống kê theo nhóm”	66
7.5 Đặc tả use case “xem thống kê theo chủ đề”	66
7.6 Đặc tả use case “xem danh sách tin bài”	68
7.7 Đặc tả use case “xem chi tiết tin bài”	69

DANH MỤC HÌNH VẼ

Hình 2.1	Biểu đồ usecase tổng quát.	4
Hình 2.2	Biểu đồ usecase phân rã quản lý nhóm.	5
Hình 2.3	Biểu đồ usecase phân rã quản lý chủ đề.	6
Hình 2.4	Biểu đồ usecase phân rã quản lý tin bài.	7
Hình 2.5	Biểu đồ usecase phân rã chức năng thống kê.	8
Hình 2.6	Biểu đồ usecase phân rã chức năng thống kê.	9
Hình 2.7	Biểu đồ hoạt động quy trình nghiệp vụ tạo nhóm mới.	10
Hình 2.8	Biểu đồ hoạt động quy trình nghiệp vụ tạo chủ đề mới.	11
Hình 4.1	Kiến trúc 3 tầng	22
Hình 4.2	Biểu đồ phụ thuộc gói phần server	23
Hình 4.3	Biểu đồ phụ thuộc gói hệ thống thu thập dữ liệu	24
Hình 4.4	Thiết kế gói controller	25
Hình 4.5	Thiết kế gói service	26
Hình 4.6	Thiết kế gói entity	27
Hình 4.7	Thiết kế gói scheduler	28
Hình 4.8	Thiết kế gói service	29
Hình 4.9	Thiết kế gói worker	30
Hình 4.10	Thiết kế giao diện đăng nhập	31
Hình 4.11	Thiết kế giao diện đăng ký	31
Hình 4.12	Thiết kế giao diện thống kê	32
Hình 4.13	Thiết kế giao diện danh sách chủ đề	32
Hình 4.14	Thiết kế giao diện thêm mới chủ đề	33
Hình 4.15	Thiết kế giao diện danh sách tin bài	33
Hình 4.16	Thiết kế giao diện danh sách nhóm	34
Hình 4.17	Thiết kế giao diện thêm mới nhóm	34
Hình 4.18	Thiết kế giao diện danh sách email	35
Hình 4.19	Thiết kế giao diện thêm mới sách email	35
Hình 4.20	Biểu đồ thực thể liên kết	36
Hình 4.21	Hình ảnh giao diện danh sách chủ đề	42
Hình 4.22	Hình ảnh giao diện thêm mới chủ đề	43
Hình 4.23	Hình ảnh giao diện chỉnh sửa chủ đề	43
Hình 4.24	Hình ảnh giao diện đăng nhập	44
Hình 4.25	Hình ảnh giao diện đăng ký	44
Hình 4.26	Hình ảnh giao diện thống kê	45

Hình 4.27	Hình ảnh giao diện thống kê	45
Hình 4.28	Hình ảnh giao diện danh sách tin bài	46
Hình 4.29	Hình ảnh giao diện chi tiết tin bài	46
Hình 4.30	Hình ảnh giao diện danh sách nhóm	47
Hình 4.31	Hình ảnh giao diện chi tiết nhóm	47
Hình 4.32	Hình ảnh giao diện thêm mới nhóm	48
Hình 4.33	Hình ảnh giao diện danh sách email	48
Hình 5.1	Tổng quan hệ thống lắng nghe mạng xã hội fbsprompt	51
Hình 5.2	Tổng quan hệ thống thu thập dữ liệu	52
Hình 5.3	Tổng quan hệ thống tìm kiếm tin bài	58

DANH MỤC BẢNG BIỂU

Bảng 2.1	Luồng sự kiện thay thế use case đăng nhập.	12
Bảng 2.2	Dữ liệu đầu vào đăng nhập.	12
Bảng 2.3	Luồng sự kiện thay thế use case thêm mới chủ đề.	13
Bảng 2.4	Dữ liệu đầu vào thêm chủ đề mới.	14
Bảng 2.5	Luồng sự kiện thay thế use case thêm nhóm.	15
Bảng 2.6	Dữ liệu đầu vào thêm mới nhóm.	15
Bảng 2.7	Dữ liệu đầu ra thêm mới nhóm.	16
Bảng 4.1	Bảng mô tả chi tiết table user.	37
Bảng 4.2	Bảng mô tả chi tiết table role.	37
Bảng 4.3	Bảng mô tả chi tiết table user_role.	37
Bảng 4.4	Bảng mô tả chi tiết table topic.	38
Bảng 4.5	Bảng mô tả chi tiết table email_notification.	38
Bảng 4.6	Bảng mô tả chi tiết table facebook_group.	39
Bảng 4.7	Bảng mô tả chi tiết table facebook_group_user.	39
Bảng 4.8	Bảng mô tả chi tiết table smtp_email.	40
Bảng 4.9	Bảng mô tả chi tiết table crawled_facebook_post.	40
Bảng 4.10	Danh sách thư viện và công cụ sử dụng.	41
Bảng 4.11	Thống kê thông tin về hệ thống.	42
Bảng 5.1	Bảng mô tả chi tiết tin bài.	56
Bảng 7.1	Luồng sự kiện thay thế use case đăng ký.	61
Bảng 7.2	Dữ liệu đầu vào đăng ký.	62
Bảng 7.3	Luồng sự kiện thay thế use case xóa chủ đề.	63
Bảng 7.4	Dữ liệu đầu vào xóa chủ đề.	63
Bảng 7.5	Luồng sự kiện thay thế use case xóa nhóm.	64
Bảng 7.6	Dữ liệu đầu vào xóa nhóm.	64
Bảng 7.7	Luồng sự kiện thay thế use case xóa email.	65
Bảng 7.8	Dữ liệu đầu vào xóa email.	66
Bảng 7.9	Dữ liệu đầu vào xem thống kê tin bài theo nhóm.	67
Bảng 7.10	Dữ liệu đầu vào xem thống kê tin bài theo chủ đề.	68