

ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC BÁCH KHOA
KHOA KHOA HỌC & KỸ THUẬT MÁY TÍNH



LUẬN VĂN TỐT NGHIỆP ĐẠI HỌC

HỘI ĐỒNG: Khoa học máy tính
GVHD: PGS. TS. Phạm Hồng Luân
GVPB: TS. Phạm Văn Hai
SVTH: ^{—o0o—}Võ Thị Thuy

TP. HỒ CHÍ MINH, 1/2016

Lời cam đoan

Tôi xin cam đoan...

Lời cảm ơn sdasdasd / Lời ngỏ

Tôi xin chân thành cảm ơn ...

Tóm tắt nội dung

Tóm tắt luận văn ...

Mục lục

1	Giới thiệu	1
1.1	Mở đầu	1
1.2	Yêu cầu và mục tiêu của đề tài	1
1.2.1	Yêu cầu	1
1.2.2	Mục tiêu	2
1.3	Bố cục của luận văn	2

Danh sách bảng

Danh sách hình vẽ

Chương 1

Giới thiệu

1.1 Mở đầu

Ngày nay sự phát triển không ngừng của công nghệ thông tin và truyền thông thông tin, điện thoại di động đã trở thành vật dụng không thể thiếu của hầu như tất cả mọi người.

Theo số liệu thống kê được công bố tại Ngày Internet Việt Nam 2015 do Hiệp hội Internet Việt Nam tổ chức tại Hà Nội 19/11/2015. Hiện nay, Việt Nam có 120 triệu thuê bao di động và theo số liệu thu thập được từ công ty Appota liên quan đến lĩnh vực di động tại Việt Nam, hiện nay nước ta có khoảng 22 triệu người sử dụng smartphone tức là cứ 4 người Việt lại có 1 người sử dụng điện thoại thông minh, chiếm khoảng 19% thuê bao di động.

Với một chiếc điện thoại di động, người dùng chẳng những có thể gọi điện, nhắn tin, phục vụ từ giải trí như: đọc tin tức (báo điện tử), nghe nhạc, xem phim, chơi game; cho đến đọc mail, viết ghi chú, và rất rất nhiều ứng dụng khác. Chính vì thế, chiếc điện thoại thông minh gần như là vật bất li thân của một bộ phận lớn người dùng.

Đi kèm với sự phát triển đó, một vấn nạn nảy sinh theo đó là tin nhắn rác [1].

Với giá thành cho mỗi tin nhắn khá rẻ, chỉ khoảng 200vnd cho mỗi tin nhắn nội mạng và 250vnd cho mỗi tin nhắn ngoại mạng (giá thuê bao trả trước của nhà mạng mobifone).

Ngoài ra thì việc phát tán tin nhắn rác hàng loạt quá dễ dàng: có thể thông qua các đơn vị trung gian cung cấp dịch vụ phát tán tin nhắn rác dưới những cái tên như SMS marketing, tin nhắn marketing. Cước phí của dịch vụ này khá "phải chăng" như mạng Viettel - 45 đồng/tin, mạng Mobifone - 40 đồng/tin, mạng VinaPhone - 27 đồng/tin. (Theo laodong.com.vn). Hoặc với các phần mềm gửi tin nhắn hàng loạt phổ biến như SMS Caster, TOP SMS Marketing, một cá nhân bình thường cũng có thể sử dụng và phát tán hàng nghìn tin nhắn chỉ với một cú click chuột.

Hơn nữa, hình phạt cho việc phát tán tin nhắn rác là rất nhẹ và sự bắt lực của các nhà mạng trong vấn đề này khiến tin nhắn rác ngày một nhiều. Gây rất nhiều phiền toái cho người sử dụng.

Bắt đầu từ những vấn đề nêu trên, nhóm đã quyết định chọn đề tài: “Xây dựng bộ lọc tin nhắn rác cho điện thoại thông minh”. Đề tài nhằm mục đích xây dựng bộ lọc tin nhắn rác cho các tin nhắn tiếng Việt. Đáp ứng nhu cầu bức thiết của người sử dụng dịch vụ di động.

1.2 Yêu cầu và mục tiêu của đề tài

1.2.1 Yêu cầu

Nghiên cứu thuật toán phân loại tin nhắn rác.

Tạo ứng dụng chặn tin nhắn rác trên điện thoại thông minh (Android)

1.2.2 Mục tiêu

Về kiến thức

Phân tích, giải quyết yêu cầu bài toán phân loại tin nhắn rác
Nắm vững các thuật toán phân loại sử dụng
Nắm các kỹ thuật xử lý văn bản: tách token, tính xác suất các token, ...
Nắm các kiến thức lập trình di động (Android)

Về sản phẩm

Tối ưu hóa bộ lọc, ứng dụng
Nắm được quy trình phát triển sản phẩm: phân tích - thiết kế - hiện thực - kiểm tra.
Phát triển ứng dụng thực tế, hướng sử dụng, tương tác.

1.3 Bố cục của luận văn

Tài liệu tham khảo

- [1] J. Ma, Y. Zhang, J. Liu, K. Yu, and X. Wang. Intelligent sms spam filtering using topic model. In *2016 International Conference on Intelligent Networking and Collaborative Systems (INCoS)*, pages 380–383, Sep. 2016.