ĐẠI HỌC THÁI NGUYÊN
TRƯỜNG ĐẠI HỌC CNTT & TT

HOÀNG NGỌC PHAN

XÂY DỰNG CÔNG CỤ LỌC NỘI DUNG DỊCH VỤ WEB

CHUYÊN NGÀNH: KHOA HỌC MÁY TÍNH
MÃ SỐ: 60. 48. 01

LUẬN VĂN THẠC SĨ KHOA HỌC CÔNG NGHỆ THÔNG TIN

NGƯỜI HƯỚNG DẪN KHOA HỌC
TS. NGUYÊN NGỌC CƯƠNG

Thái Nguyên, 2010

LỜI CAM ĐOAN

Tôi xin cam đoan, kết quả của luận văn hoàn toàn là kết quả của tự bản thân tôi tìm hiểu, nghiên cứu. Các tài liệu tham khảo được trích dẫn và chú thích đầy đủ.

Tác giả

Hoàng Ngọc Phan

LỜI CẢM ƠN

Tôi xin được bày tỏ lòng biết ơn chân thành và sâu sắc nhất đến thầy giáo hướng dẫn, Tiến sĩ Nguyễn Ngọc Cương, người đã tận tình dẫn dắt và tạo mọi điều kiện tốt nhất để tôi có thể hoàn thành luận văn này.

Tôi cũng xin chân thành cảm ơn các thầy cô giáo trường Đại học Công Nghệ Thông Tin & Truyền Thông Thái Nguyên, Viện Công nghệ Thông tin đã giúp đỡ và tạo mọi điều kiện thuận lợi trong quá trình học tập và nghiên cứu.

Xin chân thành cám ơn các anh chị lớp cao học Khoa học máy tính khoá 2012 và các thầy cô giáo, các bạn đồng nghiệp đã luôn bên cạnh, động viên, khuyến khích tôi trong suốt thời gian học tập và thực hiện đề tài.

Xin chân thành cám ơn!

Học viên

Hoàng Ngọc Phan

MỤC LỤC

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

ADSL : Asymmetric Digital Subscriber Line

CMAE : Content Management in Adversarial Environments

COSIM : Cosine Simarility

DNS : Domain Name Service

DWK : Depraved Web Killer

FTP : File Transfer Protocol

HTTP : Hypertext Transfer Protocol

IP : Internet Protocol (nghi thức mạng)

IR : Information Retrieve

ISP : Internet Service Provider

SIM : Simarility

TCP : Transmission Control Protocol

URL : Uniform Resource Locator

PICS : Platform for Internet Content Selection

SMTP :

ICMP : Internet control message protocol

AUP : Acceptable-Use Policy

VPN : Virtual Private Network

VLAN : Virtual Local Area Network

DTD : Document Type Definitions

ISS : Internet Information Server

ASP : *Active Server Pages*

MTA : Mail Transfe Agent

LỜI MỞ ĐẦU

Ngày nay, số người dùng Intenet và các dịch vụ chạy trên Internet ngày càng nhiều và Internet được xem như là một phương tiện để tiếp nhận và truyền tải thông tin. Đặc biệt là Web và Mail, số người truy cập và sử dụng dịch vụ này nhiều nhất.

Tuy nhiên, cũng có những người sử dụng phương tiện Internet để truyền bá những thông tin không lành mạnh và cũng có những đối tượng tham gia vào việc truy cập những thông tin này.

Chính vì lý do đó, công việc hỗ trợ quản lý và đảm bảo an ninh - an toàn thông tin trên mạng Internet đã trở thành mối quan tâm của mỗi gia đình, mỗi tổ chức, mỗi quốc gia. Về phương diện gia đình, mối quan tâm của các bậc phụ huynh là ngăn ngừa việc thâm nhập các trang Web độc hại đối với con em mình. Về phương diện quốc gia, với đặc thù về chính trị và kinh tế ở nước ta, là một trong số ít nước xã hội chủ nghĩa, vừa mới đấu tranh thống nhất và đưa đất nước hoàn toàn thoát khỏi ách đô hộ của đế quốc trong một thời gian ngắn; các thế lực phản động cả ở trong nước và ngoài nước tận dụng triệt để những thuận lợi của mạng Internet để phục vụ cho mục đích tuyên truyền, phát tán tài liệu phản động và thực hiện các hành vi phản động khác chống phá nhà nước Cộng hòa Xã hội Chủ nghĩa Việt Nam. Do mạng Internet không có giới hạn về địa lý, lại có các phương tiện thuận lợi như thư điện tử, diễn đàn, các trang web,… nên các thế lực thù địch rất dễ dàng phát tán thông tin đến số đông người dùng mạng chỉ trong một thời gian ngắn mà hầu như không mất phí tổn gì. Đây là một vấn đề nhức nhối không chỉ ở Việt Nam, mà ở bất kỳ quốc gia nào khác trên thế giới.

Cùng với việc tăng cường năng lực cơ sở hạ tầng thiết bị, hệ thống phần mềm, nhân lực nhằm phát triển Internet, Đảng và Nhà nước ta cũng đã ban hành các hệ thống pháp lý đối với việc khai thác Internet.

Nghị định số 55/2001/NĐ-CP ngày 23-8-2001 của Chính phủ về *Quản lý, cung cấp và sử dụng dịch vụ Internet* đã đề cập về vấn đề này (Điều 2; Điều 6; Điều 11; Điều 18; Điều 28; Điều 33; Điều 35; Điều 41 và Điều 45). Một số nội dung chi tiết hơn được thể hiện trong *Quy định về biện pháp và trang thiết bị kiểm tra, kiểm soát đảm bảo an ninh quốc gia trong hoạt động Internet ở Việt Nam* của Bộ Nội vụ được ban hành kèm theo Quyết định số 848/1997/QĐ-BNV(A11) ngày 23.10.1997 (Mục 2 Khoản 3 Điều 5, Khoản 3 Điều 6). *Quy định về Đảm bảo an toàn, an ninh trong hoạt động quản lý, cung cấp, sử dụng Internet tại Việt Nam* được ban hành kèm theo Quyết định số 71/2004/QĐ-BCA (A11) ngày 29 tháng 1 năm 2004 của Bộ trưởng Bộ Công an quy định toàn diện và chi tiết về các nội dung đảm bảo an toàn an ninh trên Internet của Nhà nước ta, Thông tư 02 (02/2005/TTLT-BCVT-VHTT-CA-KHĐT), có quy định "*Quyền và nghĩa vụ của đại lý Internet*":

Nhận thức được tầm quan trọng cũng như yêu cầu cấp thiết của vấn đề này, được sự đồng ý của giáo viên hướng dẫn và của Trường Đại học CNTT & TT Thái Nguyên, em đã chọn đề tài ***: « Xây dựng công cụ lọc nội dung thông tin dịch vụ Web »***

Nội dung Đề tài gồm 3 chương :

Chương 1 : Tổng quan về bảo đảm an ninh an toàn thông tin

Chương này nghiên cứu, Đánh giá tình hình quản lý Nhà nước về lọc nội dung trên Internet của các quốc gia và Việt Nam, phân tích các khái niệm về an toàn thông tin và an ninh nội dung thông tin, các giải pháp đảm bảo an ninh an toàn thông tin.

Chương 2 : Tìm hiểu về các phương thức trao đổi thông tin qua giao dịch web và các kỹ thuật lọc nội dung thông tin đối với dịch vụ web

Chương này nghiên cứu về các phương thức trao đổi thông tin qua giao dịch web, mô hình hệ thống lọc nội dung thông tin đối với giao dịch web và một số kỹ thuật lọc nội dung web

Chương 3 : Xây dựng công cụ lọc nội dung Web độc hại

Nghiên cứu, ứng dụng xây dựng công cụ lọc nội dung web