



U.S. Department of Justice

Criminal Division

Washington, D.C. 20530

CRM-200400930F

JAN 27 2006

Chris Hoofnagle
Electronic Privacy Information Center
1718 Connecticut Avenue, NW
Suite 200
Washington, DC 20009

Dear Chris Hoofnagle:

While processing your request dated January 15, 2004, for records related to the Acxiom Corporation and General Wesley Clark, the Office of Information and Privacy located three records (items 1-3) of interest to the Criminal Division of the Department of Justice and referred them to us for review and direct response to you. We received this referral on June 7, 2004.

In light of our review, we have determined to release items 1 and 2 in full and item 3 in part. We are withholding the portions indicated of the item pursuant to the following Freedom of Information Act (FOIA) exemption set forth in 5 U.S.C. 552(b):

- (6) which permits the withholding of personnel and medical files and similar files the disclosure of which would constitute a clearly unwarranted invasion of personal privacy.

Copies of the records are attached.

You have a right to an administrative appeal of this partial denial of your request. Department regulations provide that such appeals must be filed within sixty days of your receipt of this letter. 28 C.F.R. 16.9. Your appeal should be addressed to: The Office of Information and Privacy, United States Department of Justice, Flag Building, Suite 570, Washington, D.C. 20530. Both the envelope and the letter should be clearly marked with the legend "FOIA Appeal." If you exercise this right and your appeal is denied, you also have the right to seek judicial review of this action in the federal judicial district (1) in which you reside, (2) in which you have your principal place of business, (3) in which the records denied are located, or (4) for the District of Columbia. If you elect to file an appeal, please

include, in your letter to the Office of Information and Privacy, the Criminal Division file number that appears above your name in this letter.

Sincerely,

Thomas J. McIntyre

Thomas J. McIntyre, Chief
Freedom of Information/Privacy Act Unit
Office of Enforcement Operations
Criminal Division

Department of Justice
EXECUTIVE SECRETARIAT
CONTROL SHEET

DATE OF DOCUMENT: 11/26/2001
DATE RECEIVED: 01/28/2002

WORKFLOW ID: 108160
DUE DATE:

FROM: The Honorable Vic Snyder
U.S. House of Representatives

Washington, DC 20515

TO: AG

MAIL TYPE: Congressional Constituent

SUBJECT: Ltr on behalf of the University of Arkansas and the Axiom Corporation, Little Rock, AR, enclosing a proposal to apply advanced techniques of information retrieval to automatically find and identify Web sites belonging to terrorists and others who advocate violence in support of extremist views.

DATE ASSIGNED
03/04/2002

ACTION COMPONENT & ACTION REQUESTED
Criminal Division
For response by component.

INFO COMPONENT: OLA

COMMENTS: 6/13/02: CRM replied by ltr dated 6/3/02. cc: AG files. Forward copies of incoming letter and draft response to OLA for review and approval prior to mailing. After OLA approval, return control sheet with signed and dated copy of response to ES.

FILE CODE:

EXECSEC POC: Paula Stephens: 202-616-0074



U.S. Department of Justice

Criminal Division

*Assistant Attorney General**Washington, D.C. 20530*

June 3, 2002

The Honorable Vic Snyder
U.S. House of Representatives
Washington, DC 20515

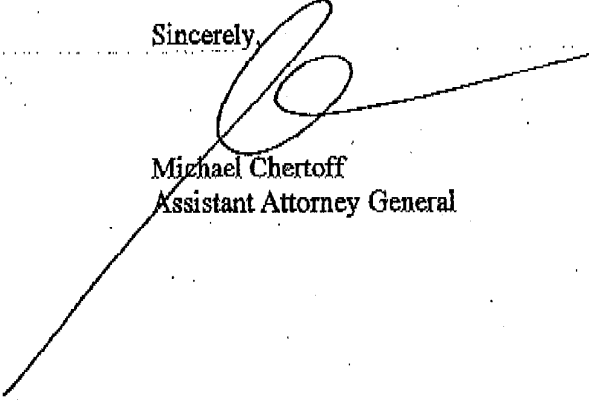
Dear Congressman Snyder:

Thank you for your November 26, 2001, letter concerning the advanced techniques that may be used in uncovering terrorist financing over the Internet. We apologize for the delay in our response.

Since September 11, we have been involved in efforts to enlist the private sector partners into our financial war on terrorism. We have been in contact with a variety of computer companies, including Axiom Corporation, and are well aware of its impressive technical capabilities. Please be assured that we will use every tool at our disposal – both public and private – in our law enforcement counterterrorism efforts. Companies such as Axiom are a key part of this campaign.

I hope this information is helpful. If we can be of further assistance in this or any other matter, please do not hesitate to contact this office.

Sincerely,



Michael Chertoff
Assistant Attorney General

VS:mc
Enclosure

PRINTED ON RECYCLED PAPER

3

Research Project Proposal

Automatic Identification of Extremist Internet Websites

Department of Computer science
University of Arkansas at Little Rock
2801 S. University Avenue
Little Rock, AR 72204

²Axiom Corporation
1000 Technology Drive
Little Rock, AR 72212

Overview

The authors propose to apply advanced techniques of information retrieval, such as pattern analysis and feature extraction, to automatically find and identify websites belonging to advocates of extremist views and actions that may pose threats to peace and security. The proposed research will focus on adapting techniques already successfully developed by the principal investigators under the auspices of the Axiom Data Engineering Laboratory housed within the Cyber College at the University of Arkansas at Little Rock (UALR) under a grant from ACXIOM/ASTA/UALR. These techniques are currently being used for extracting business contact information from .COM websites. From the knowledge and experiences gained in the process of designing such techniques, a model for identifying sites that pose potential security risks can be developed and implemented.

Introduction

There are individuals and groups in every society who hold extreme views on many issues, such as, abortion, racial superiority, politics, religion, immigration, and foreign affairs, to name a few. They often promote their views and attempt to recruit new adherents through Internet websites. Whether intentionally or unintentionally, these sites can sometimes create a culture of hatred that some of their adherents commit criminal or terrorist acts against other individuals or groups, such as local and or federal governments agencies, politicians, immigrants, people of certain race or faith, or celebrities. These unlawful acts might include bombing, assassinations, kidnapping, hijacking, or sabotage, and may result in death or injury to innocent bystanders, as well as, to their intended targets.

A system could be modeled and developed that automatically and constantly searches the World Wide Web and identify such sites. Once identified, law enforcement and security officials will be able to monitor and cross reference these sites. As a result, some of the criminal and terrorist activities that they engender can be prevented. In addition, the system could be enhanced to extract features, such as, individual names and titles, telephone numbers, e-mail addresses, and mailing addresses. Once properly identified,

Version 1, 11/4/2001

ble

indexed, and stored in a database, this information could be made available to the proper authorities for further action.

Proposed System Functions

- 1) The system would continuously "crawl" the Internet searching for extremist websites.
- 2) Once a potential extremist site is identified, the URL and other features of the site would be recorded in a database for later follow-up.
- 3) Other personal identification features associated with the site, such as, names, email, or telephone numbers would also be extracted and recorded in the database.
- 4) Personal information extracted from sites confirmed as extremist could be
 - a) used for cross-reference analysis to establish possible connections between extremist groups,
 - b) collected for an Identity Verification System to be used by airlines, rental car agencies, and other business and governmental agencies.

System Architecture

The architecture for the proposed system is based in part on the knowledge and experiences gained in the process of designing and developing a system for extracting business contact information from .COM websites.

The system is composed of four components, namely:

- Crawler
- Grabber
- Extractor
- Database.

The Crawler constantly searches through the World Wide Web for all reachable domains.

The Grabber downloads the eligible HTML pages, removes the HTML tags, and then delivers the textual contents as a text file.

The Extractor checks the text file for special types of patterns and features, and attempts to determine if these patterns and features meet the criteria of an extremist site. If the threshold of suspicion is met, then the personal information are extracted and stored in the Database along with the patterns and features that were previously obtained.

To explain further, the Extractor has three sub-components namely:

- Inspector
- Feature Extractor
- Sifter

Inspector is in charge of determining whether a web site is suspicious. This process is completed in two steps:

- b4
1. **Marker Words and Phrases:** Search for marker words in the text file, and identify their location and type within the text. Examples of possible marker words are
 - Verbs, e.g., bomb, kill, burn, kidnap, highjack
 - Buildings, e.g., Embassy, Congress, Mosque, Church
 - Places, e.g., Congress, Israel, Kashmir, Ireland, Embassy
 - People, e.g., Bush, Chaney, Saddam Hussein, Bin Laden, Tony Blair
 - Organizations, e.g., Al-Qaidah, Hamas, CIA, FBI, Mosad, IRS, KKK
 - Derogatory Racial Terms
 - Titles, e.g., President, King, Judge, Senator, Mullah, Rebel, Prime Minister
 - Other Suspicious Terms, e.g., explosives, bombs, Jihad, Kamikaze
 2. **Pattern Analysis:** When marker words of a particular type appear in specific patterns, then the site may be classified as extremist.

The Feature Extractor identifies and extracts the features of the extremist site. In addition to the site's URL, other features of interest would be individual names and titles, telephone numbers, numbers, fax numbers, e-mail addresses, and mailing addresses.

The Sifter component takes two actions on the extracted features, which are:

- Parsing and removing invalid characters,
- Validating each feature against a repository of the knowledge (facts and rules).

Finally, the Feature Extractor writes the validated features into a database for use by the proper authorities in other applications.

Areas of Research

- The development of algorithms that can check the contents of the website against a very large list of marker words are a great challenges. These algorithms must be fast, and the marker word tables must be comprehensive and easy to maintain.
- The design, development, and testing of pattern analysis algorithms that can effectively and efficiently locate key features of the site.
- The development of a knowledge base for validation of the extracted features.
- Design and development of an appropriate schema for storing the extracted features in a secure database in such a way that they can be readily accessed and cross-referenced.

Automatic Identification of Extremist Internet Websites

[]
¹Department of Computer science and
Axiom Data Engineering Lab.
University of Arkansas at Little Rock
Little Rock, AR 72204

] b6

²Axiom Corporation
Axiom Data Engineering Lab.
1000 Technology Drive
Little Rock, AR 72212

Goals

- Finding websites that may pose threats to peace and security by advocating the extremist views and actions.
- Extracting the information from the websites and store them in a database for further use.

Proposed System Functions

- 1- **The system would continuously "crawl" the Internet searching for extremist websites.**
- 2- **Once a potential extremist site is identified, the URL and other features of the site would be recorded in a database for later follow-up.**
- 3- **Other personal identification features associated with the site, such as, names, email, or telephone numbers would also be extracted and recorded in the database.**

Proposed System Functions

- 4- **Personal information extracted from sites confirmed as extremist could be**
 - **used for cross-reference analysis to establish possible connections between extremist groups,**
 - **collected for an Identity Verification System to be used by airlines, rental car agencies, and other business and governmental agencies.**

Web Grabber Technology

Web Grabber Technology already successfully developed by the principal investigators and it is currently being used for extracting business contact information from .COM websites.

Sponsored by: the Acxiom Data Engineering Laboratory
housed within the Cyber College at the University
of Arkansas at Little Rock

A grant from: ACXIOM/ASTA/UALR.

Budget

Item	1st Year	2nd Year
PIs' Salaries	\$200,000	\$200,000
Students' Salaries	\$100,000	\$100,000
Research Expenses	\$200,000	\$100,000
Sub-Total	\$500,000	\$500,000
Total		\$1,000,000