University of Nebraska - Lincoln
DigitalCommons@University of Nebraska - Lincoln

Managing Vertebrate Invasive Species | USDA National Wildlife Research Center Symposia

8-1-2007

MONITORING THE SALE AND TRAFFICKING OF INVASIVE VERTEBRATE SPECIES USING AUTOMATED INTERNET SEARCH AND SURVEILLANCE TOOLS

Karl Suiter
NSF Center for Integrated Pest Management, North Carolina State University, Raleigh, North Carolina, USA

Sanda Sferrazza
NSF Center for Integrated Pest Management, North Carolina State University, Raleigh, North Carolina, USA

Follow this and additional works at: http://digitalcommons.unl.edu/nwrcinvasive

Part of the Environmental Indicators and Impact Assessment Commons

Suiter, Karl and Sferrazza, Sanda, "MONITORING THE SALE AND TRAFFICKING OF INVASIVE VERTEBRATE SPECIES USING AUTOMATED INTERNET SEARCH AND SURVEILLANCE TOOLS" (2007). *Managing Vertebrate Invasive Species.* Paper 51.
http://digitalcommons.unl.edu/nwrcinvasive/51

This Article is brought to you for free and open access by the USDA National Wildlife Research Center Symposia at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Managing Vertebrate Invasive Species by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

MONITORING THE SALE AND TRAFFICKING OF INVASIVE VERTEBRATE SPECIES USING AUTOMATED INTERNET SEARCH AND SURVEILLANCE TOOLS

KARL SUITER AND SANDRA SFERRAZZA, NSF Center for Integrated Pest Management, North Carolina State University, Raleigh, North Carolina, USA

Abstract: The internet is often an unregulated pathway that can lead to the introduction of invasive species. Surveillance of internet-related sale or trade activities can reduce the movement of potentially invasive species and serve to lessen the risk of an introduction. The Invasive Species Internet Monitoring System automates the process of searching the internet for suspect sites (e.g., internet storefronts, chat rooms, etc.) involved in the sale or trade of targeted species. Additionally, this software system incorporates data management features that permit the archiving of suspect website content, assists subject matter experts in evaluating sites for risk, send informational emails to website owners and can track sites for compliance. This software system is currently being used by United States Department of Agriculture, Animal and Plant Health Inspection Service, Plant Protection and Quarantine and the Smuggling, Interdiction and Trade Compliance Program to search for websites selling regulated insects, weeds, mollusks, fruits and vegetables, and for animals and animal products that have the potential to carry Highly Pathogenic Avian Influenza.

Key Words: automated search, internet, invasive species, search engine, surveillance.

Managing Vertebrate Invasive Species: Proceedings of an International Symposium (G. W. Witmer, W. C. Pitt, K. A. Fagerstone, Eds). USDA/APHIS/WS, National Wildlife Research Center, Fort Collins, CO. 2007.

INTRODUCTION

The introduction of alien invasive species, either accidentally or purposefully, into non-native habitats poses serious threats to native ecosystems and to human and animal health. The establishment of invasive species has been directly linked to reduced native biodiversity and habitat loss (e.g., D'Antonio and Vitousek 1992). Additionally, the introduction of monkey pox virus in native prairie dog populations from imported Gambian giant pouch rats ultimately led to the infection of humans in the United States (US, Pacelle 2003). Identifying the risks and pathways that can lead to the introduction of invasive species and transmittable zoonotic pathogens can serve to reduce the likelihood of their establishment in new habitats and populations (Kasari 2006).

As early as 1999, it was clear that the US Department of Agriculture (USDA), Animal and Plant Health Inspection Service's (APHIS) regulated organisms were being sold on the internet. Ebay.com, internet discussion and interest groups, chat rooms, etc., all play a part in these sale and trade activities. This mode of illegal plant and animal trade was identified as a new pathway, that unless closed, could lead to the introduction of new alien pests and pathogens into the US. In 2002, the Invasive Species Internet Monitoring System (ISIMS) was developed as a cooperative effort between the USDA and the National Science Foundation (NSF) Center for Integrated Pest Management (CIPM). The goal of the project was to develop techniques and tools that would semi-automate the process of searching for invasive pests being sold on the internet. The integration of automated search tools (e.g., FAST Search and Transfer, Inc.) in combination with custom analysis and databasing tools (e.g., ISIMS), can aid in monitoring the sale and trafficking of invasive vertebrate species. FAST Search is a web crawler that can locate websites selling or trafficking federally regulated invasive species or other organism or animal products of interest. ISIMS is a web application that aids analysts in evaluating, investigating and monitoring websites that may offer invasive vertebrate species for sale or trade.

METHODS

Search Engine

The core component is the FAST Search and Transfer Web Search Engine. This application automates the task of searching the web (all published web pages, news feeds, etc.) using key terms that define distinct query parameters in order to return relevant search results. CIPM has highly customized this application and search pipeline in

order to facilitate searching for invasive species information.

Input Parameters

The list of key search terms typically consists of a list of invasive species (scientific names and common name synonyms). Connector terms that define a sale action on a website are identified (e.g., "for sale", "Mastercard", "shopping cart"), while a list of exclusion terms (e.g., "quarantine", "toy", "www.kfc.com") that remove sites from consideration is also generated. Lists of websites that should always be searched (e.g., "www.invasivespeciesforsale.com") or those that should be excluded from the automated search (e.g., "discovery.com") are created, and the user has the ability to weight or load specific search criteria in order to boost search relevancy. The data is then stored in tables in a Structured Query Language Server database.

Search Details

Scripts automate the creation of the search query strings based on the input parameters provided. Once the query strings are generated and stored in the database, the FAST search engine initiates the search of the indexed web universe for documents matching the search criteria. Once a search is initiated, the FAST search engine continually queries the web for new documents that match the search criteria.

The CIPM utilization of FAST Search is configured to search through web documents, news feeds, blogs, in addition to chat rooms, if accessible. Over sixty languages can be queried in order to search for colloquial words and phases on foreign websites.

Identifying Websites of Interest

A customized web interface is used to display the results of each search. Subject Matter Experts (SME's) review the results (website pages) to determine if a site is selling an invasive species of interest. Sites identified by SME's as relevant are saved to a database for further analysis. The system incorporates a Site Evaluation Tool that permits an evaluator to examine a website in detail. Integrated tools such as a "WHOIS" search assist evaluators in identifying the location and owner of websites that may be selling invasive species. Enforcement officers have the ability to send informative letters via email to potential violators. All information pertaining to a potential violator, in addition to any letters and associated documents related to a site, are stored for reference.

Site Analysis

Summary statistics and reposts can be generated in order to estimate the efficacy of and evaluate the parameters used in each search. These tools permit refining of search parameters in order to generate more relevant search results.

RESULTS

Query results are returned to a custom data view (Figure 1). The results shown are a subset of the species listed on the U.S. Fish and Wildlife Service Threatened and Endangered Animals and Plants list.

Results are ranked in order of their relevancy to the number of search terms matched in the query. Selecting the page title link will bring up the original web page. The contact information of the web page owner or host site can be viewed by selecting the "getWhoIS" link. Sites can be marked as relevant and saved for further analysis or irrelevant. If sites are returned that are not of interest, they can be excluded from the search result list. Additional site filters can be used to return only those sites that correspond to a particular location, language, or indexed keyword.

Utilizing this methodology and the website, we located the GiantPouchedRatz Yahoo discussion group. This demonstrated that Gambian giant pouched rats are still being bred and offered for sale in the US even though now illegal (Figure 2).

DISCUSSION

By automating the search process, the ISIMS software can identify websites selling suspect invasive vertebrate species and other species of interest. However, in order to return relevant search results, the query parameters for each specific collection of organisms must be well defined. The input of subject matter experts when developing query parameters is essential to the success of any search. Additionally, the commitment of resources towards reviewing query results and determining if sales or trade activity warrants further action is also a crucial component of this system.

The focus of much of the work we have performed to date has been to identify websites offering regulated organisms for sale or trade via the internet. However, the types of information

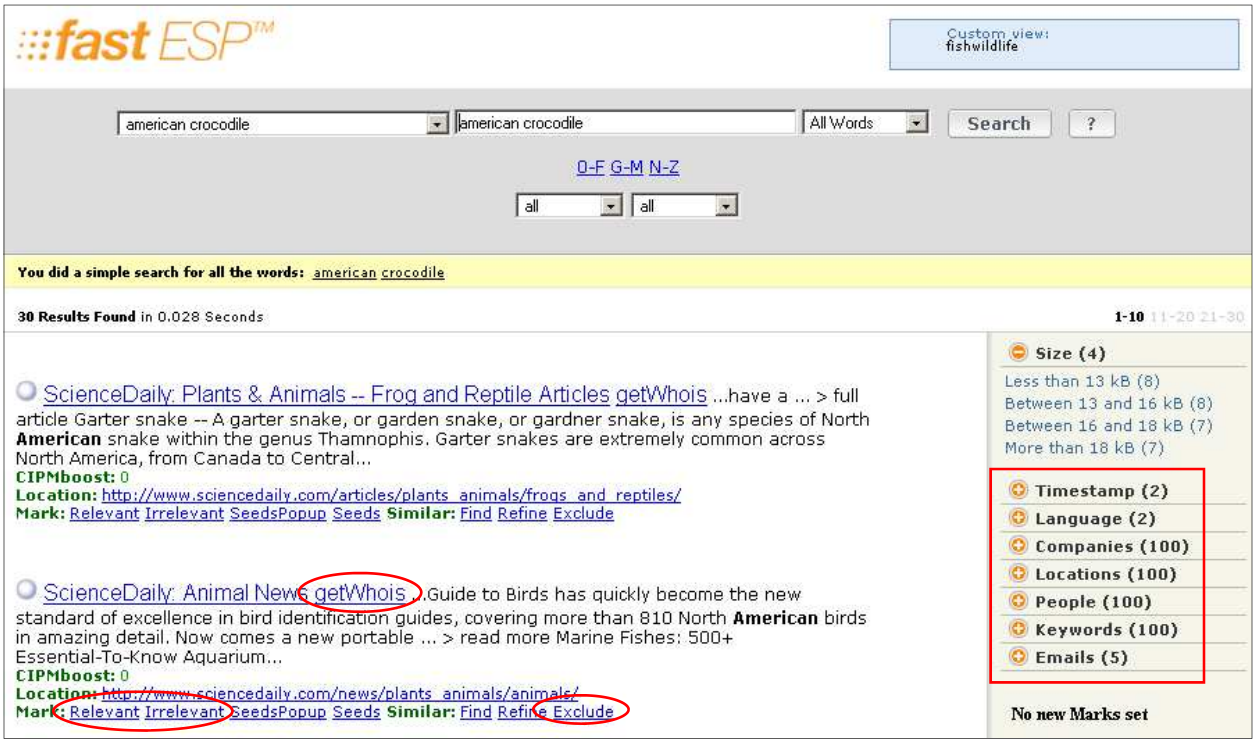

Figure 1. An example of data output from the FAST search engine in a custom data view.

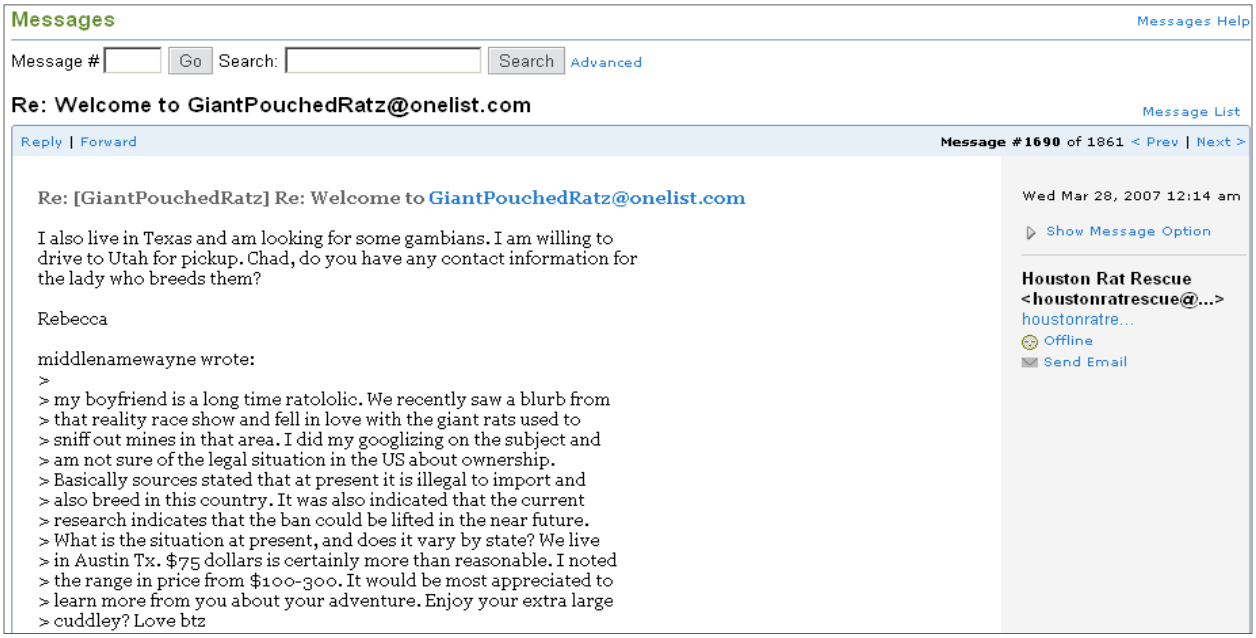

Figure 2. Chat group message indicating that Gambian giant pouched rats are currently being bred and sold in the US.

returned from ISIMS queries can be adapted for many different needs and product groups:

- Identifying sites selling regulated organisms or species of interest,
- Sites selling only within a state or political region,
- Foreign websites selling invasive species,
- Provide a list of growers/producers within a state or region, and
- Internet sales of Convention on International Trade in Endangered Species listed species

One example of this alternate use was identified during a search for websites selling avian products that could serve as a vector for Highly Pathogenic Avian Influenza. After numerous searches and refinements of search terms, it was concluded that few producers and suppliers of avian products, especially those from southeast Asian markets and the US, advertised their products via internet commerce sites. Thus, focus was shifted from identifying sites selling avian products to identifying those overseas suppliers that advertise that they are willing to ship avian products worldwide. The APHIS' Smuggling, Interdiction and Trade Compliance Program is working with other groups to determine if data from these sites can be used to assist in targeting suspect shipments from overseas suppliers.

LITERAURE CITED

D'ANTONIO, C. M. AND VITOUSEK, P. M. 1992. Biological invasions by exotic grasses, the grass/fire cycle, and global change. Annual Review of Ecology and Systematics 23:63–87.

KASARI, T. 2006. Risk analysis team helps guide surveillance plans through pathways analysis. Risk Assessment. NAHSS Outlook Quarter Two. (www.aphis.usda.gov/vs/ceah/ncahs/nsu/outlook).

PACELLE, W. 2003. Importation of exotic species. U.S. Senate Committee on Environment & Public Works Hearing Statements. (http://epw.senate.gov/hearing_statements.cfm?id=212888).