

Dark Web has been supported in part by the National Science Foundation (NSF CRI CRD), #CNS-0709338 and NSF (EXP-LA), #CBET-0730908

Dr. Hsinchun Chen, Director, Artificial Intelligence Lab, University of Arizona, <http://ai.arizona.edu>

Artificial Intelligence Lab Management Information Systems The University of Arizona, Tucson, AZ

Introduction

- Dark Web forums offer members of extremist and terrorist groups a medium to communicate
- Information contained within Dark Web forums represents a significant source of knowledge for security and intelligence organizations.
- We have developed systems supporting the large-scale **collection**, **search**, and **analysis** of Dark Web forums, specifically addressing the needs of security analysts.



1. Collection | AZ Forum Spider

Overview

The goal of the Arizona Forum Spider for Dark Web forums is to assist in the automated collection of forum communications and other content of interest.

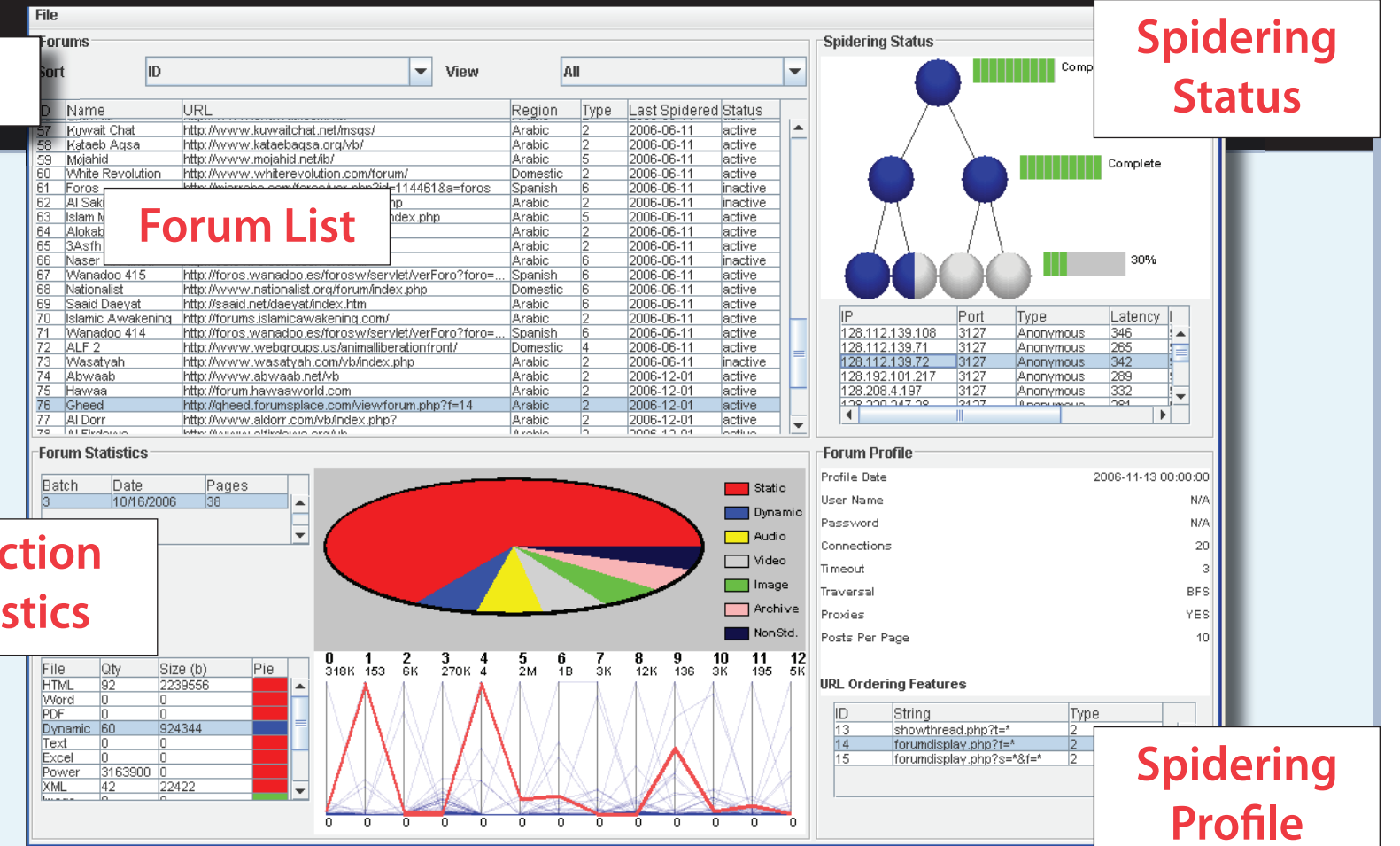
Major Functionality

- The Arizona Forum Spider uses breadth and depth first (BFS and

DFS) traversal for crawl space URL ordering based on URL tokens, anchor text, and link levels.

- Automated discovery of proxy servers to distribute collection and increase reliability.
- Website wrappers are utilized to identify updated content for collection, allowing for incremental spidering.

AZ Forum Spider



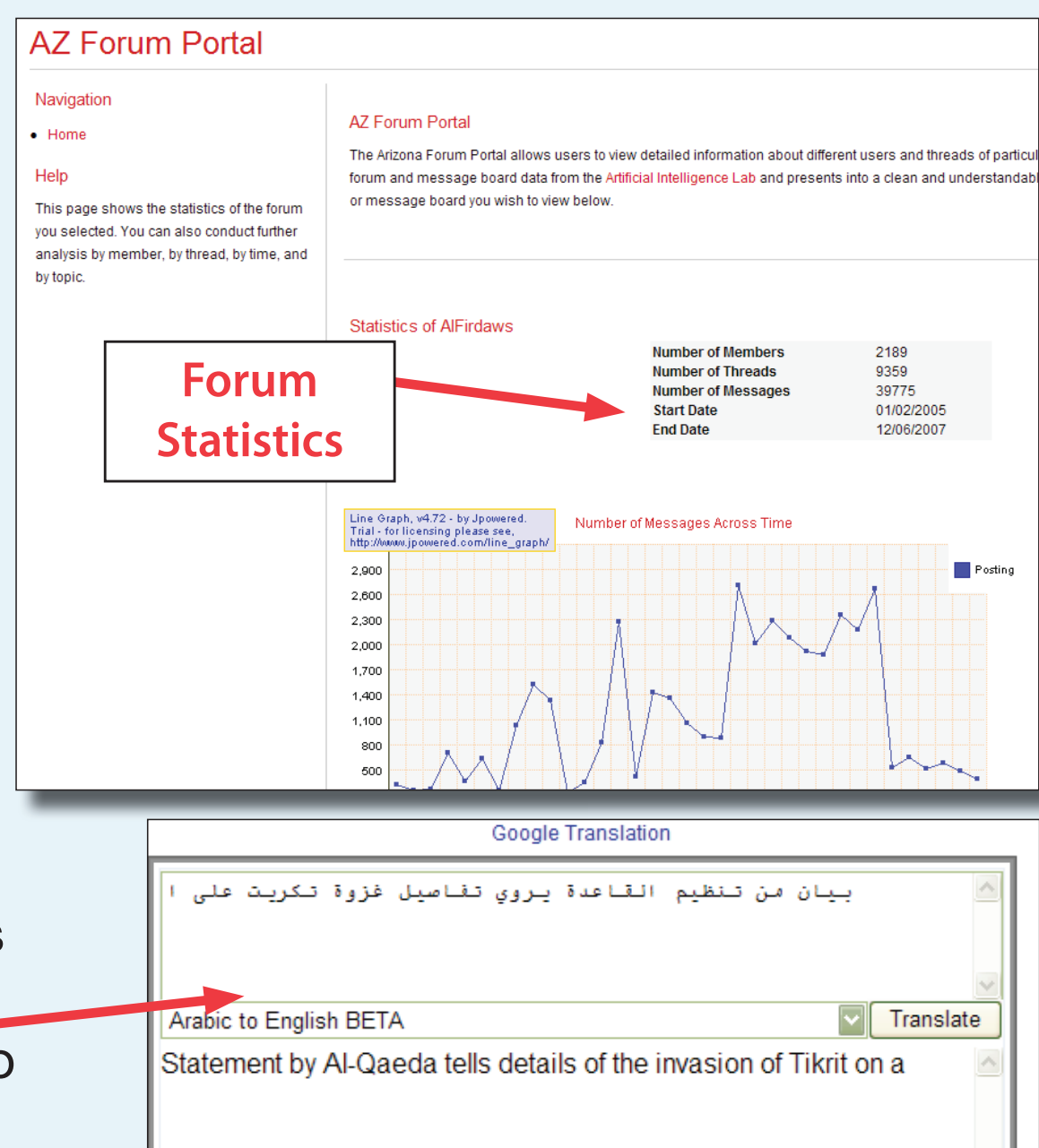
2. Search | AZ Forum Portal, Sentiment Analyzer

AZ Forum Portal Overview

The Arizona Dark Web Forum Portal is a centralized repository of Dark Web forum communications. The goal of the portal is to provide the functionality to browse, search, and analyze the Dark Web forum collections. The current portal contains more than 2.5 million messages across 7 forums.

Major Functionality

- The Arizona Dark Web Forum Portal provides functionality to aid in the analysis of Dark Web forums within a centralized repository.
 - Forum-level statistics and graphical depiction of posting activity over the forum lifespan
 - Facility to browse forum communications by member, thread, or time period
 - Keyword search of thread titles or message bodies
 - Embedded **Google Translation** to facilitate multilingual analysis



AZ Sentiment Analyzer Overview

The Arizona Sentiment Analyzer is a web-based portal for the analysis of opinions and emotions expressed in Dark Web forum communications, to enhance the understanding of forum communities and participants. For the security analyst, the portal eases the acquisition and consumption of knowledge discovered in sentiment and affect analyses of Dark Web forum communications.

Major Functionality

- Forum-level statistics characterize sentiment and affect expression of the community, for easy comparison among forums
- Analysis across various dimensions, featuring impactful visualizations
- Temporal analysis by thread or member to identify sentiments and affects expressed at specific time periods
- Search functionality for threads titles or members
- Sentiment and affect regression analysis



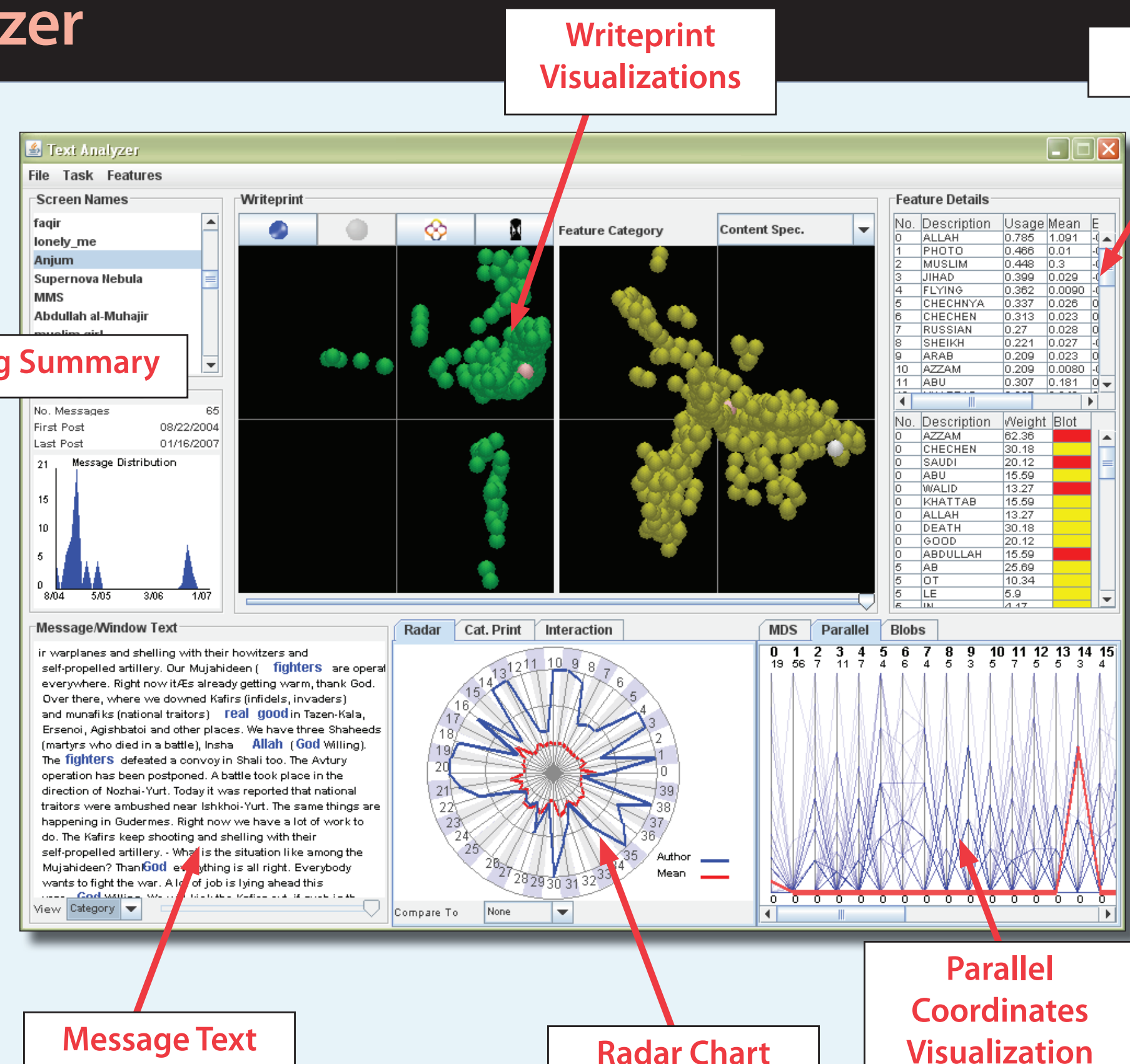
3. Analysis | AZ CyberGate Text Analyzer

Overview

The Arizona CyberGate Text Analyzer is a comprehensive system for the analysis of forum communications, utilizing automated text mining and visualization methods.

Major Functionality

- CyberGate can aid in Dark Web forum analysis by providing the facility to characterize the forum and compare its participants utilizing text mining approaches. Rich visualization offers new perspectives of the data to analysts.
- CyberGate Writeprint and Ink Blot techniques utilizes various linguistic feature categories for text categorization and analysis
 - Syntactic, structural, lexical, content-specific, and lexicons of interest
- Incorporates feature selection methods based upon principle component analysis and decision tree models
- Presents rich text visualization using multi-dimensional scaling, parallel coordinates, and ink blots



Features

Literature

- Abbasi, A., Fu, T., and Chen, H. 2006. "A Focused Crawler for Dark Web Forums," paper submitted to the *Journal of the American Society for Information Science and Technology*.
- Abbasi, A. and Chen, H. 2008. Writeprints: A stylometric approach to identity-level identification and similarity detection in cyberspace. *ACM Trans. Inf. Syst.* 26, 2 (Mar. 2008), 1-29.
- Abbasi, A. and Chen, H. 2007. Categorization and analysis of text in computer mediated communication archives using visualization. *In Proceedings of the 7th ACM/IEEE-CS Joint Conference on Digital Libraries (Vancouver, BC, Canada, June 18-23, 2007)*. JCDL '07. ACM, New York, NY, 11-18.
- Chen, H. 2008. Sentiment and affect analysis of Dark Web forums: Measuring radicalization on the internet. *IEEE International Conference on Intelligence and Security Informatics, 2008 (Taipei, June 17-20, 2008)*. ISI 2008, 104-109.