

special conference issue

information

June 2006

www.sla.org

outlook

vol. 10, no. 6

How to Write an REP for Information Products



Terror
Online

page 57

Extremist Social Movement Groups

TRACKING SUSPECT



And Their Online Digital Libraries

By Edna Reid and
Hsinchun Chen

Terrorist/extremist groups' use of the Web has expanded beyond routine communication and propaganda operations to include training, recruitment, logistical support, and development of virtual communities (Reid et al., 2002; Tsftati and Weimann, 2002). Their Web sites have increased in number, technical sophistication, and multimedia content. How then does one identify, organize, analyze, and provide access to extremist social movement groups' Web-based cultural artifacts (e.g., videos, Web sites, discussion forum messages, weblogs)?

The extensive use of communication networks is empowering social movement struggles internation-

ally and opening new spaces—called “virtual publics”—that move beyond the mere exchange of information to facilitate shared cultures, coordination, and solidarity (Garrido and Halavais, 2003; Jones and Rafael, 2000, pp. 214–23).

For example, terrorist/extremist groups are building dynamic online libraries of multilingual training materials (e.g., monthly magazines, manuals, instructional videos, reports, and speeches), and they are doing so with some support from experts who answer questions and share experiences on message boards or in chat rooms (Glasser and Coll, 2005). Anyone supporting or conducting research on social movement organizations today needs to understand: (a) the wealth and variety of cul-

LEARN MORE AT SLA 2006

Terrorism Informatics: Developing the Skills Needed to Support Homeland Security

8 a.m. – 5 p.m., Saturday, June 10

Baltimore Convention Center, Room 304

Ticketed Event #110—Price: \$299 member/\$399 non-member

How do you identify, organize, analyze, and access terrorist/extremist groups' artifacts, such as videos and Web sites? Anyone supporting or creating research in this area needs to understand the wealth of digital artifacts available, the challenges in using them, and the relevant tools. This workshop will provide an overview of the emerging field of terrorism informatics, an analysis of terrorist/extremist groups' information operations and data sources, an opportunity to use terrorism research tools, and a chance to create a community of practices.

Speaking: Edna Reid, senior research scientist and adjunct lecturer, University of Arizona

Presented by: SLA Professional Development Center

tural artifacts that they produce; (b) the organizations that provide access to them; and (c) advanced research tools that support access and post-retrieval analysis of social movement groups' cultural artifacts, such as

the Dark Web Research Portal testbed being developed by the AI (Artificial Intelligence) Lab, University of Arizona.

This article describes terrorist/extremist groups' cultural artifacts and the ongoing

ing efforts at the AI Lab to build a digital library that enables the communities of librarians/information managers, researchers, and practitioners to effectively respond to information management challenges posed by extremist social movement groups.

Extremist Social Groups' Cultural Artifacts

Extremist social movement groups' cultural artifacts could serve as original and unique information resources to support evidence-based analysis of the terrorism phenomenon. This collection represents the alternate side of the Web, referred to as the "Dark Web" (the use of the Web by terrorist/extremist groups to spread their ideas); it has received extensive government and media attention (Reid et al., 2004). However, librarians/information managers, researchers, and practitioners face major challenges in identifying, collecting, organizing, and using the artifacts because of the dearth of efforts expended for the systematic collection building, organization, and preservation of these ephemeral multilingual and multimedia resources.

Furthermore, some of the extremist Web sites are extremely dynamic, and their contents have a short shelf life. Their dynamic nature poses a major challenge because they can emerge overnight, post messages including videos (e.g., the gruesome beheading tapes), and then swiftly disappear or, in many cases, seem to disappear by changing their URLs or ISPs but retaining much of the content (Weimann, 2005). The new URLs or ISPs are

Table 1

Organizations That Collect and Analyze Terrorist/Extremist Web Artifacts

Organization	Description	Collection	Access
Archive			
Internet Archive	Spiders Web every two months to collect open-access HTML pages.	Created 1996. Provides historical and current snapshots of Web sites.	www.archive.org
Research Community			
AI Lab, University of Arizona	Is a Dark Web Portal test bed—a component of the terrorism DL project. Hsinchun Chen, director.	Created 2003. Has databases of thousands of Web sites, multimedia, and forum messages. Supports post-retrieval analysis.	ai.eller.arizona.edu/research/terror/index.htm
Middle East Media Research Institute	Jihad and Terrorism Studies Project monitors and analyzes militant Islamic groups.	Created 2003. Analyzes and translates its collection of groups' Web sites and multimedia to produce reports, etc.	www.memri.org/jihad.html
Project for the Study of Islamist Movements (Interdisciplinary Center Herzilya, Israel)	Researches Islamist movements. Reuven Paz, director.	Created 2002. Analyzes and translates its collection of groups' Web sites and multimedia to produce reports, etc.	gloria.idc.ac.il/islam
Search for International Terrorist Entities (SITE) Institute	Analyzes artifacts, public records, etc. Spiders Web sites, etc., every 24 hrs. Rita Katz, director.	Created 2003. Analyzes and translates groups' Web sites, forum messages, and multimedia to produce reports.	www.siteinstitute.org Scroll to the bottom of the home page to see a list of video clips. Some clips are available free. Translations are available to subscribers.
Weimann, G. Univ. Haifa, Israel	Spiders Web sites, etc., every 24 hours.	Created 1998. Extensive collection of Web sites and multimedia for research.	Closed
Vigilante Community			
Internet Haganah	Confronts the Global Jihad Project. Is a grass-roots activist organization that monitors Web sites and ISPs (Internet service providers)	Created 2001. Has database of snapshots, ISP data, and links to terrorist/extremist groups' Web sites.	haganah.org.il; www.internet-haganah.us
Johnathanrgalt (geocities)	Monitors groups' Web sites and ISPs. Maintains archival database.	Created 2001. Has database of 60 to 70 closed jihad Web sites.	www.geocities.com/johnathanrgalt/

announced in online discussion forums, weblogs, chat sessions, etc. These activities are examples of what Preece (2000) describes as the "dark sides of online communities"—those that do not necessarily support the traditional concept of positive social interactions.

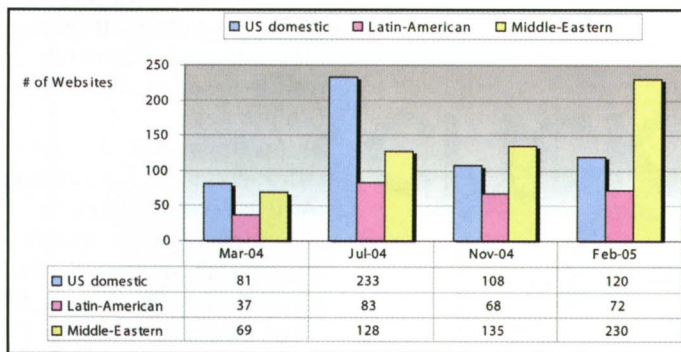
The ISTS (Institute for Security Technology Studies) at Dartmouth College reported terrorist/extremist groups' online social interactions as falling into five distinct areas: propaganda, recruitment and training, fundraising, communications, and targeting (ISTS, November

2003, 11). An example of propaganda is the Alneda Web site, which was considered a primary outlet for "official statements," reports, and videos from senior members of the al-Qaeda movement. Beginning in 2002, the Alneda site used the domain www.alneda.com; the site has been kept alive by "parasiting" itself onto legitimate domains (ISTS, November 2003, 11). Its webmasters buried its file structure deep in subdirectories of legitimate and unsuspecting business sites. Despite being thrown off from various Web hosting

services and being hacked repeatedly by pro-American activists, the Alneda site has endured, retaining the same look, structure, and style. (ISTS, November 2003, 11, and Robbins, 2002).

Another challenge that often overwhelms research, knowledge management, and library communities is the sheer volume of diversely formatted (e.g., multimedia, dynamic files such as php and gci) artifacts available on the Web. Language barriers also pose challenges because sites are often written in languages unknown to most information managers and analysts.

Figure 1
Summary of the Web Sites Collected for the Dark Web Portal Test Bed



Organizations That Collect The Artifacts

Several organizations have taken measures to ensure the continued but controlled availability of terrorist/ extremist groups' cultural artifacts. They conduct research projects (e.g., Project for Research of Islamist Movements) and offer fee-based investigative services (e.g., SITE—Search for International Terrorist Entities—Institute). Some research libraries subscribe to the services provided by the SITE Institute. For example, the SITE Institute and the Project for Research of Islamist Movements monitor Web sites of militant Islamic groups in their native languages and provide access to translated information and metadata about the groups' Web sites and forums. Table 1 contains a list of organizations and subdivides them according to whether they are archives, research communities, or vigilante communities.

Most of the organizations listed in the table use spidering software to collect the artifacts and conduct manual analysis. Of the organizations listed, only the AI Lab provides tools to support post-retrieval analy-

sis, such as automatic indexing, summarization, Web mining, linguistics analysis, visualization, and hyperlink analysis for finding connections among groups' Web sites (see Figure 4).

Dark Web Research Portal Testbed

The AI Lab applied its DL (digital library) toolset to the challenges of terrorism research and designed a collection-building methodology for harvesting terrorist/ extremist groups' cultural artifacts for U.S. Domestic, Middle Eastern, and Latin American groups (Chen et al., 2004). Figure 1 provides a summary of the thousands of Web sites that were collected for the Dark Web Portal testbed in 2004 and early 2005. New batches of

data are collected every two to three months.

Although many researchers, practitioners, and librarians/information managers may not currently analyze terrorist/extremist groups' Web sites, it is reported that Web sites are "the way for everybody in the world to listen to the mujahedin" (Glasser and Coll, 2005). Therefore, a key purpose of the Dark Web Portal is to help researchers and practitioners easily access and analyze multilingual information created by major terrorist/extremist groups and their sympathizers. In addition to Web sites, the Dark Web collection includes multimedia files such as video clips and messages from online discussion forums. Another organization that collects video clips is SITE Institute; it also provides Arabic language translations and analysis. Some of SITE's publications are available free on its Web site.

Terrorist/extremist groups host discussion forums about news topics, share tips on their trade-craft, and provide advice on distribution of viruses. Figure 2 provides an example of information identified from a discussion forum collected for the Dark Web.

An analysis of the Dark Web collection indicates that some groups from the Middle East have a pattern of embedding discussion forums within their Web sites while U.S. domestic and Latin American groups rely on using Yahoo! and Google groups. Figure 3 provides a summary of the number of forums collected.

The Dark Web Research Portal testbed is being designed to support information retrieval and post-retrieval analysis. The portal consists of various components: database searching, keyword suggestion, Web page summarization, categorization, visualization, and advanced information retrieval. Because of the different languages, separate databases for U.S. domestic, Latin American, and Middle Eastern collections have been created. Figure 4 displays the components of the Dark Web Research Portal testbed. Using the U.S. domestic terrorist/extremist groups' collection, they are described as follows:

- **Search Page** for performing a keyword search of the U.S. domestic database provides a sample search of the "Ku Klux Klan." Figure 4a shows that the user can find results (keyword searching), organize results (categorization), and map results

Figure 2
Summary of Content Identified from an Online Discussion Forum

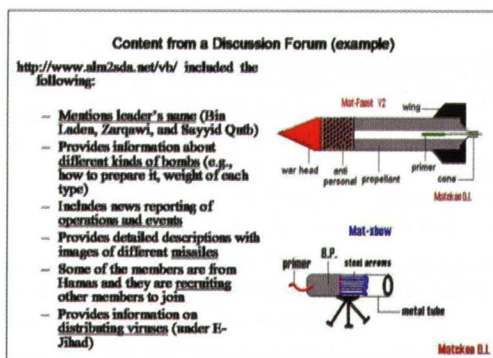


Figure 2
Summary of Forums

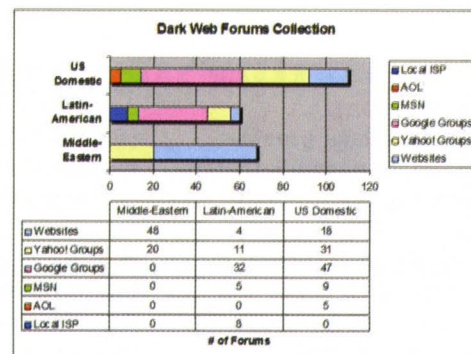
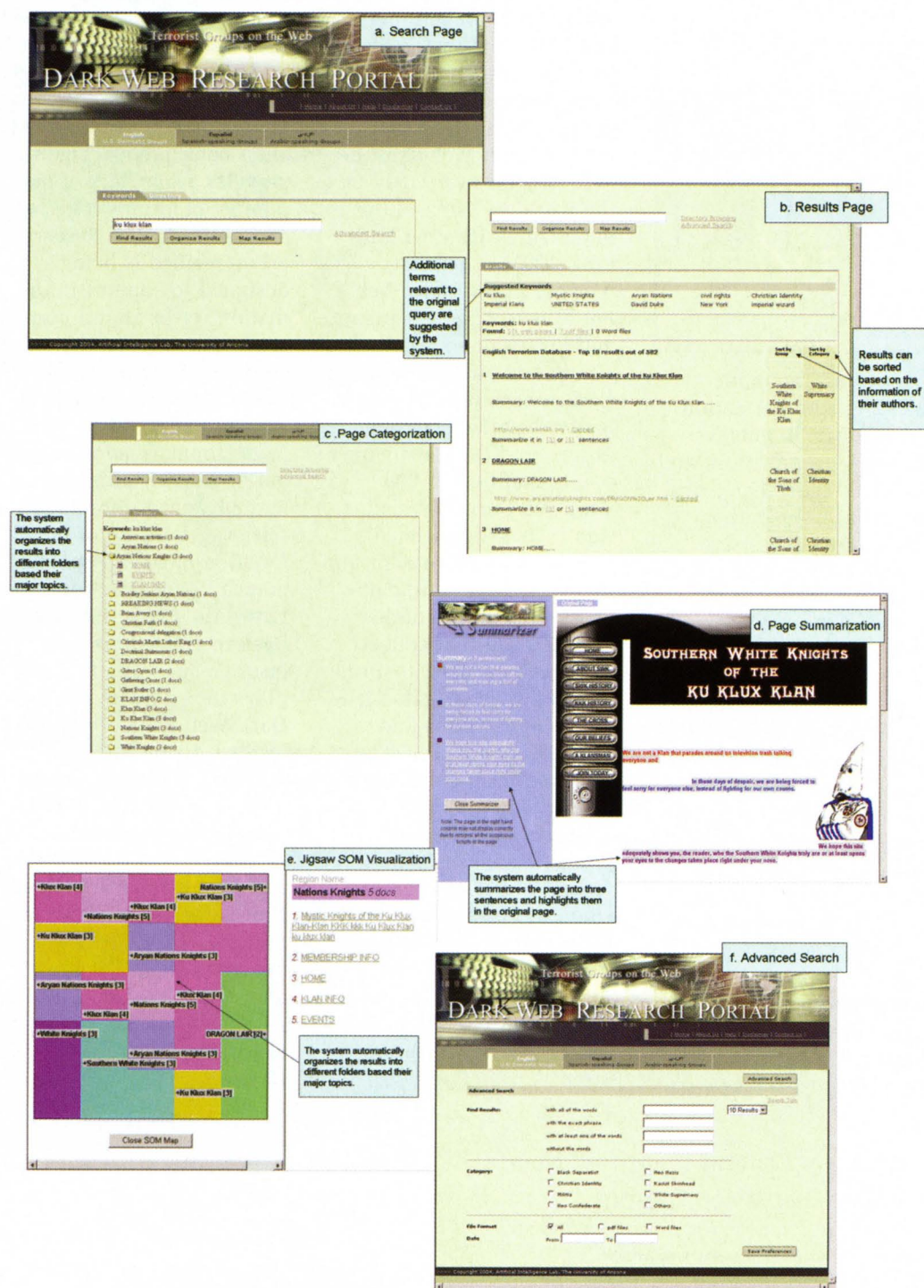


Figure 4
Dark Web Research Portal Testbed



work that automatically learns from the input Web pages and clusters them into different naturally occurring groups. It generates a map of the search results based on keywords (Figure 4e).

- **Advanced search** provides advanced information retrieval options such as search by Boolean operators, group categories, file types, and dates (Figure 4f).

Future Directions

The Dark Web Portal test bed has yet to utilize the multilingual information in the testbed. In this regard, the portal will be expanded to include a CLIR (cross-language information retrieval) component. CLIR is the study of retrieval of information in one language through queries expressed in another language (Reid et al., 2004, p.138). Another component that will be added is a machine translation, which will translate the multilingual information retrieved by the CLIR component back into the researchers' native languages.

The completed Dark Web Portal will make it possible to completely respond to questions like: How do you identify, organize, analyze, and provide access to extremist social movement groups' cultural artifacts? As we continue using the AI Lab's toolkit to develop the Dark Web Portal testbed, we envision extending an invitation to researchers, librarians/information managers, and practitioners who are willing to be early adopters and partners to participate in the evaluation and further design of the test bed. Data from the Dark Web Portal are available to libraries, research centers, and scholars.

(visualization).

- **Results Page** provides additional keywords that are relevant to the original query. It identifies pairs of keywords co-occurring on the same Web pages and extracts them for use as the-saurus terms in the U.S.

domestic database collection (Figure 4b).

- **Page Categorization** organizes the search results into various folders labeled by the key phrases appearing in the page summaries. A program calculates the frequently occurring phras-

es to index the results (Figure 4c).

- **Page Summarization** summarizes Web pages using three to five sentences (Figure 4d).

- **Jigsaw SOM** (self-organizing map) Visualization is a two-layered neural net-

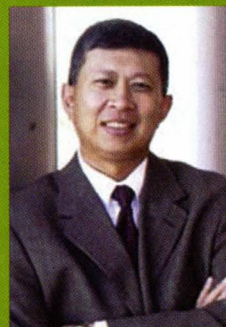
Acknowledgments

This research has been supported in part by the following grants:

- NSF, "COPLINK Center: Information and Knowledge Management for Law Enforcement," July 2000–September 2005.
- NSF/ITR, "COPLINK Center for Intelligence and Security Informatics Research—A Crime Data Mining Approach to Developing Border Safe Research," September 2003–August 2005.
- DHS/CNRI, "BorderSafe Initiative," October 2003–March 2005.

References

- Chen, H., Qin, J., Reid, E., Chung, W., Zhou, Y., Xi, W. 2004. "Dark Web Portal: Collecting and analyzing the presence of domestic and international terrorist groups on the Web." *IEEE Intelligent Transportation Conference*, Washington, D.C. Available at: <http://ai.eller.arizona.edu/research/terror/index.htm>
- Garrido, M., and A. Halavais. 2003. "Mapping networks of support for the Zapatista movement." In *Cyberactivism: Online activism in theory and practice*, edited by M. McCaughey and M. D. Ayers. Routledge: 168–84.
- Glasser, S. B. and Coll, S. 2005. "Terrorists turn to the Web as base of operations." *Washington Post*, August 7, 2005. Also available: http://www.washingtonpost.com/wp-dyn/content/article/2005/08/05/AR2005080501138_pf.html.
- ISTS (Institute for Security Technology Studies) Technical Analysis Group. November 2003. "Examining the cyber capabilities of Islamic terrorist groups." Retrieved August 8, 2005, from <http://csstc.inel.gov/press-room/reports/dartmouth.pdf>.
- Jones, Q., and S. Rafael. 2000. *EM-Electronic Markets* 10 (4). "Time to split, virtually: Discourse architecture and community building create vibrant virtual publics."
- Preece, J. 2000. *Online communities: Designing usability, supporting sociability*. Wiley.
- Reid, E., Qin, J., Chung, W., Xu, Jennifer, Zhou, Y., Schumaker, R., Sageman, M., Chen, H. 2004. "Terrorism knowledge discovery project: A knowledge discovery approach to addressing the threats of terrorism," edited by H. Chen, R. Moore, and D. Zeng. In *Intelligence and security informatics: Second Symposium on Intelligence and Security Informatics*. Springer: 125-145.
- Reid, E., Qin, J., Zhou, Y., Lai, G., Sageman, M. Weimann, G., Chen, H. 2005. "Collecting and analyzing the presence of terrorists on the Web: A case study of Jihad Web sites." In *IEEE International Conference on Intelligence and Security Informatics*, edited by P. Kantor et al. Springer: 402–11. Available at: http://ai.eller.arizona.edu/research/terror/publications/ISI_AILab_submission_final.pdf
- Robbins, J. S. 2002. *National Review Online*. "The jihad online," 30 July. Retrieved August 20, 2005, from www.freerepublic.com/focus/news/724763/posts.
- Tsfati, Y., and G. Weimann. 2002. *Studies in Conflict and Terrorism*, 25. (5) "www.terrorism.com: Terror on the Internet." 317–32.
- Above citation is a journal article.
- Weimann, G. 2004. "www.terrorism.net: How Modern Terrorism Uses the Internet." Special Report 116, U.S. Institute of Peace. Retrieved August 8, 2005, from <http://usip.org/pubs/>.



Professor Hsinchun Chen is director, Artificial Intelligence Lab and McClelland Professor of Management Information Systems at the University of Arizona. He was Andersen Consulting Professor of the Year (1999) and is founder, Knowledge Computing Corporation; founding director, Mark & Susan Hoffman E-Commerce Lab, Department of MIS, Eller College of

Management at the University of Arizona. He is author/editor of 10 books and more than 130 scientific journal articles covering intelligence analysis, biomedical informatics, data/text/web mining, digital library, knowledge management, and Web computing. His recent books include: *Medical Informatics: Knowledge Management and Data Mining in Biomedicine and Intelligence and Security Informatics for National Security: Information Sharing and Data Mining*, both published by Springer. He may be reached at hchen@eller.arizona.edu.



Edna Reid is senior research scientist, Artificial Intelligence Laboratory, and adjunct lecturer, Department of Management Information Systems, Eller College of Management, University of Arizona. She is author of journal articles and conference papers on mapping the terrorism research domain and Web mining terrorist groups' cultural artifacts. She is

one of the editors of a forthcoming book on terrorism informatics. She has more than 20 years of experience in the areas of counterterrorism, information sciences, competitive intelligence, systems analysis, and project management including experience as a counterterrorism specialist in the intelligence community and a senior analyst with several government contractors in Northern Virginia. She was an associate professor with Nanyang Business School, NTU, Singapore, and the president of the Society of Competitive Intelligence Professionals Singapore (SCIPSgp). Formerly, she was an entrepreneur with an Internet start-up in Malaysia. In 1993, she joined the School of Computer Engineering, Division of Information Studies, NTU. She was a member of the National Library Board (Singapore). Her experience includes over thirteen years in Singapore, Malaysia, and Germany. Formerly, she was an Assistant Professor at SCLIS, Rutgers University. Her areas of specialization and research are terrorism informatics, online communities, web mining, and competitive intelligence.