Dark Web Forums Portal: Searching and Analyzing Jihadist Forums

Yulei Zhang, Shuo Zeng, Li Fan, Yan Dang, Catherine A. Larson, and Hsinchun Chen, *Fellow, IEEE*
Department of Management Information Systems
The University of Arizona
Tucson, AZ 85721, USA
{ylzhang, shuozeng, fanli, ydang}@email.arizona.edu, {hchen, cal}@eller.arizona.edu

***Abstract*—With the advent of Web 2.0, the Web is acting as a platform which enables end-user content generation. As a major type of social media in Web 2.0, Web forums facilitate intensive interactions among participants. International Jihadist groups often use Web forums to promote violence and distribute propaganda materials. These Dark Web forums are heterogeneous and widely distributed. Therefore, how to access and analyze the forum messages and interactions among participants is becoming an issue. This paper presents a general framework for Web forum data integration. Specifically, a Web-based knowledge portal, the Dark Web Forums Portal, is built based on the framework. The portal incorporates the data collected from different international Jihadist forums and provides several important analysis functions, including forum browsing and searching (in single forum and across multiple forums), forum statistics analysis, multilingual translation, and social network visualization. Preliminary results of our user study show that the Dark Web Forums Portal helps users locate information quickly and effectively. Users found the forum statistics analysis, multilingual translation, and social network visualization functions of the portal to be particularly valuable.**

Keywords- Dark Web forums, Web-based knowledge portal, multilingual translation, social network visualization

I. INTRODUCTION

The fast development of the Web has increased the popularity of computer mediated communication (CMC) [1]. Web forum is an open discussion platform as well as a knowledge database for specific topics [2]. It is also a significant source of user-generated content which covers a wide range of topics (e.g., presidential elections, environmental issues, and consumer concerns). With the explosive growth of Internet usage, the Web has become an essential communication media for international Jihadist groups [3] who are increasingly using the Internet to promulgate their agendas. This problematic side of the Web is often referred to as the Dark Web [4]. Due to the high anonymity, easy access, and huge audience of Web forums, international Jihadist groups often use them to promote violence and distribute propaganda materials [5].

Since there exist a large number of heterogeneous and widely distributed Dark Web forums, data integration and retrieval become critical issues for researchers interested in monitoring Dark Web content [6]. With Dark Web forum data growing exponentially, forum information searching has become increasingly difficult, especially searching across multiple data sources. Without a centralized system, it is difficult to search and analyze Dark Web forum data [5].

In addition, due to the dynamic and multilingual nature of Dark Web forums, how to visualize the interactions among participants and properly address the language barrier have become important but challenging questions [3]. Although most Dark Web forum participants tried to minimize their interactions in order to avoid unwanted attention, visualization is a promising way to reveal connectivity and relationships hidden behind their online activities [7]. However, because of the large data size and complicated connections among users, effective visualization is still a challenge.

In this paper, we present a general framework for Web forum data integration, search, and analysis. Specifically, we built a Web-based knowledge portal, Dark Web Forums Portal, which focuses on Jihadist related content. The portal integrates several important functions, including single and multiple forums searching and browsing, statistics analysis, multilingual translation, and social network visualization.

II. LITERRATURE REVIEW

A. *Current Web Forum Research*

With the fast development of Web 2.0, there is a rapid growth of research focusing on user-generated content. Since Web forums facilitate intensive interactions among participants, a number of studies have been done on them from various perspectives, such as online users' behavior analysis [8], sentiment analysis [9], and social interaction analysis [10]. Analyzing Web forums enables new knowledge discovery in many different domains such as health care, education, and politics.

The Dark Web phenomenon has received extensive attention from both the government and the academia [3]. Although Dark Web participants' actions cannot be easily detected, the hidden information in Dark Web forums represents a significant source of knowledge for security and intelligence organizations [11]. Accordingly, a great number of

978-1-4244-4173-0/09/$25.00 ©2009 IEEE

research studies analyzing Dark Web forum data have been done. For example, Abbasi et al. [9] proposed a sentiment analysis framework to classify Dark Web forum opinions in both English and Arabic. Fu et al. [2] utilized a Hybrid Interactional Coherence algorithm to identify Dark Web forum interactions. However, few efforts have been made on Dark Web forum data integration and searching [6]. It has become imperative that a systematic and integrated approach to search, browse, and analyze international Jihadist groups' Web forum data be provided.

B. *Domain Specific Web Portals*

Web portals have been developed for different disciplines, such as infectious diseases, medication and terrorism related fields. For example, the BioPortal system [12] is an integrated, cross-jurisdictional data sharing and analysis portal for infectious diseases. Zhou et al. [13] built a Chinese Medical Web portal (CMedPort) which provided data integration, searching and nature language processing functions. Chen et al. [3] established a Dark Web portal of terrorism Web sites which provided advanced analysis and visualization functionality. However, most of the previous studies were not based on user-generated content. With more and more online, user-generated content appearing on the Web, it is becoming a trend for Web portals to take advantage of online social media data such as web forums [14]. Glance et al. [15] developed a Web portal that gathered and annotated online discussions about consumers and products.

To date, the majority of a Dark Web analyst's time is still spent on collecting data [16]. Interested users must manually gather data they need from Web forums. A few Web portals aiming to study terrorism related Web sites have been developed [3, 6]; however, little attention has been paid to building Web portals for Dark Web forum data.

C. *Multilingual Translation*

From the latest statistics for Internet user by language, non-English speaking users are more than 70% of the whole [17]. For the Dark Web, the multilingual issue is critical because much of Jihadist content is written in various languages such as Arabic, Dutch, French, etc. In order to process multilingual content, different methods have been explored to execute translation tasks [18]. Considering the easy accessibility for some popular languages, a machine translation-based approach is the most available translation resource as many companies have provided their own translation services [19]. The machine translation-based approach uses existing machine translation techniques to provide automatic translation. Google Translation is one of the most popular machine translation tools (http://code.google.com/apis/ajaxlanguage/documentation/#Translation). The Google Translation Service provides translation functions for more than 80 languages. The language detection and language translation APIs it provides can be integrated into Web pages using Javascript. With this service, sentences in languages other than English can be translated to English automatically.

D. *Social Network Analysis on Web Forums*

Social Network Analysis (SNA) is a graph-based method to analyze the network structure of a group or population and its impact on social interactions [20]. SNA has been widely used to study various real-world networks [21]. Web forums are often chosen for SNA research because by default it can record almost all the participants' communication information and the messages themselves can easily be retrieved [22]. Previous studies have employed social network analysis on Web forums. Glance et al. [14] utilized SNA on business-related Web forums to identify hot topics. Yeung et al. [22] established a collaborative learning network cross four open-source forums to study user interactions in single forum and between two forums.

The social networks formed in illegal organizations are referred to as "Dark Networks" [23]. A Dark Network is rich in interaction information, because terrorists or criminals tend to work together or collaborate under an organization. As SNA can extract linkage information from forum data quickly and effectively, adopting SNA for the study of Dark Web forums can facilitate the process of identifying interactions among forum participants [6]. Although some prior studies included SNA analysis on Dark Web forums, to the best of our knowledge, few have tried to incorporate SNA into a real-time system.

III. Research Gaps & Research Questions

A systematic and integrated approach to search, browse, and analyze international Jihadist forums is important and in demand. However, no previous research has developed an integrated framework to gather and store the heterogeneous Dark Web forum data from different sources. In addition, little research attention has been paid to information searching across multiple Dark Web forums. Furthermore, few Dark Web forum analysis portals have incorporated real-time multilingual translation functionality to facilitate users' information searching and browsing. Although a lot of work has been done on Dark Network analysis, few studies have incorporated social network analysis into a real-time, Web-based Dark Web forum analysis system.

Based on the research gaps discussed above, we present the following research questions:

- Q1: How can we develop a Web portal for Dark Web forums which integrate data from multiple forum data sources?
- Q2: How can we develop effective and efficient search and browse methods across multiple forum data sources in our portal?
- Q3: How can we incorporate real-time multilingual translation functionality into our portal to enable automatic forum data translation from non-English (e.g., Arabic) to English?
- Q4: How can we incorporate real-time, user-interactive social network analysis into our portal to analyze and visualize the interactions among forum participants?

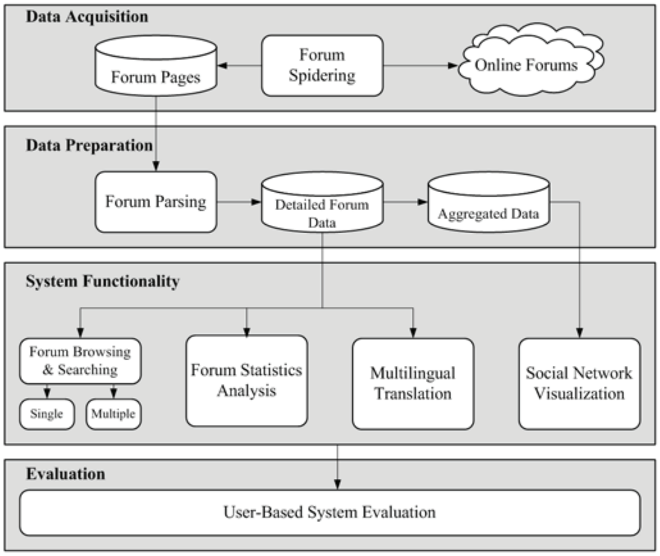


Figure 1. System Design of the Dark Web Forums Portal

TABLE I. Statistics of the Forums Incorporated in the Dark Web Forums Portal

Name	Language	Time Span	Number of Members	Number of Threads	Number of Messages
Alokab	Arabic	04/10/2005 -03/19/2008	1,232	3,699	30,480
Al-Firdaws	Arabic	01/02/2005 -12/06/2007	2,189	9,359	39,775
Montada	Arabic	09/28/2000 -07/01/2007	31,654	93,548	866,693
Hdrmut	Arabic	11/26/2000-05/18/2008	1,707	9,030	45,937
Alsayra	Arabic	04/05/2001-06/03/2008	39,230	42,329	348,933
Hawaaworld	Arabic	03/27/2001-07/01/2008	59,842	20,278	975,695
Islamic Network	English	06/09/2004 -05/07/2008	1,578	12,003	87,769

IV. System Design

As shown in Fig. 1, the Dark Web Forums Portal contains four components: Data Acquisition, Data Preparation, System Functionality, and Evaluation. We detail each component in the following sections.

A. *Data Acquisition*

In this component, spidering programs are developed to collect the Web pages from online forums that contain Jihadist related content identified by domain experts. Forum spidering is performed periodically to keep the collection updated.

B. *Data Preparation*

In this component, forum parsing programs are developed to extract the detailed forum data from the raw HTML Web pages and store it in a local database. For each forum, the structured, detailed forum data extracted include thread names, main message bodies, member names, and post dates. Another set of data (i.e., the data needed to determine the significance of each forum member and the link weight between two different members) used for social network analysis is created by aggregating the detailed forum data.

C. *System Functionality*

The Dark Web Forums portal is implemented using Apache Tomcat and the database is implemented using Microsoft SQL Server 2000. Different functions are developed and incorporated into the system as real-time services, including single and multiple forums browsing and searching, forum statistics analysis, multilingual translation, and social network visualization. For forum statistics analysis, Java applet-based charts are created to show the trends based on the numbers of messages produced over time. The multilingual translation function is implemented using Google Translation API (http://code.google.com/apis/ajaxlanguage/documentation/#Translation). The social network visualization function provides dynamic, user-interactive networks implemented using JUNG (http://jung.sourceforge.net/) to visualize the interactions among forum members.

D. *System Evaluation*

User-based system evaluation is conducted to assess the performance of the Dark Forums Portal.

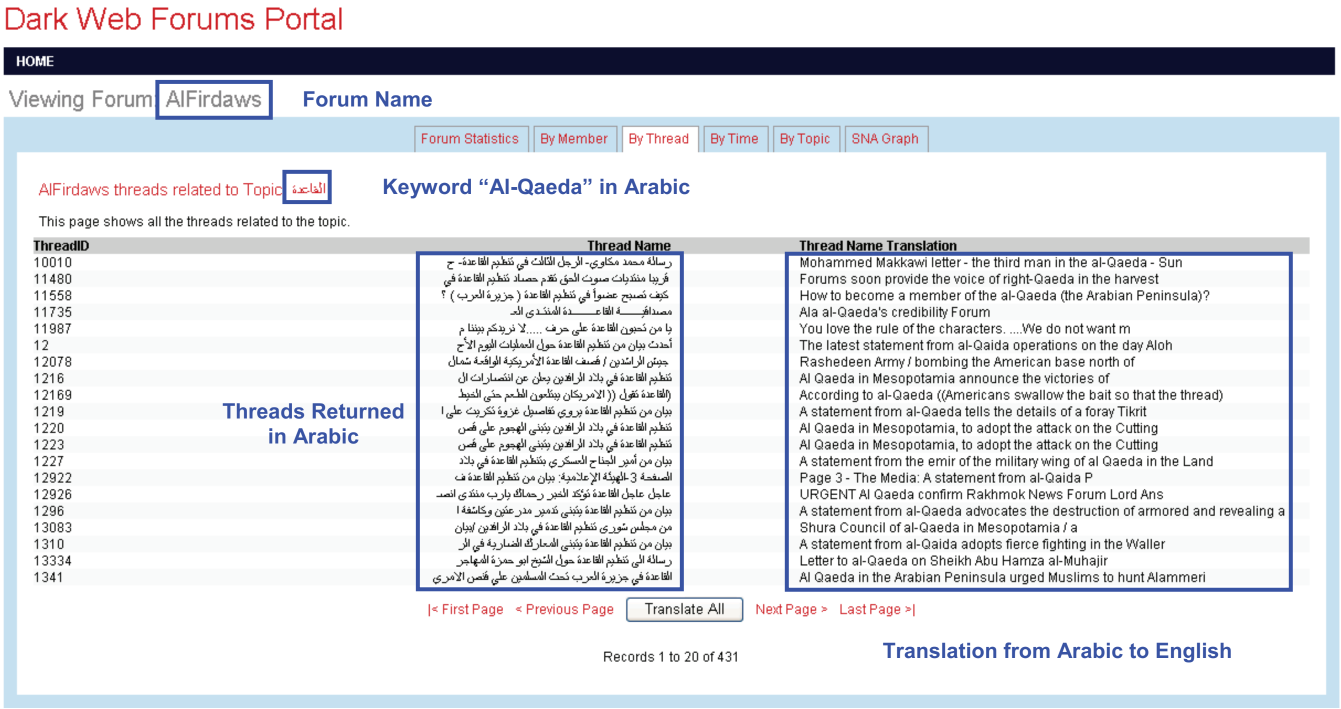
V. Data Set

Table 1 lists the forums incorporated into our system. Currently, the portal contains seven Jihadist forums, among which six are Arabic forums, and one (i.e., “Islamic Network”) is an English forum. All these forums were selected by domain experts as major active Jihadist forums. The total number of messages is about 2M.

Forum “AlFirdaws” is a general forum but has subsections containing discussions of radical Islamic ideologies and supporting Salaf-Jihadi organizations. Forum “Alokab” is dedicated to Islamic theology with some radical content. Forum “Hawaa” is dedicated to Muslim women. Some of its members have shown their sympathy to certain radical groups. Among all seven forums, “Hawaaworld,” “Montada,” and “Alsayra” have many more registered members than the other four forums. Some forums are extremely popular and have close to a million messages.

VI. System Functionality

As we described before, the system has four types of functions: single and multiple forums browsing and searching, forum statistics analysis, multilingual translation, and social network visualization. In this section, we describe each function and present some examples.



AlFirdaws threads related to Topic القاعدة

This page shows all the threads related to the topic.

ThreadID	Thread Name	Thread Name Translation
10010	رسالة محمد مكاوي- الرجل الثالث في تنظيم القاعدة- ح	Mohammed Makkawi letter - the third man in the al-Qaeda - Sun
11480	قريبا منتديات صوت الحق تقدم حصاد تنظيم القاعدة في	Forums soon provide the voice of right-Qaeda in the harvest
11558	كيف تصبح عضواً في تنظيم القاعدة (جزيرة العرب) ؟	How to become a member of the al-Qaeda (the Arabian Peninsula)?
11735	مصداقيـــــة القاعــــــدة المنتدى الع	Ala al-Qaeda's credibility Forum
11987	يا من تحبون القاعدة على حرف .....لا نريدكم بيننا م	You love the rule of the characters. ....We do not want m
12	أحدث بيان من تنظيم القاعدة حول العمليات اليوم الأح	The latest statement from al-Qaida operations on the day Aloh
12078	جيش الراشدين / قصف القاعدة الأمريكية الواقعة شمال	Rashedeen Army / bombing the American base north of
1216	تنظيم القاعدة في بلاد الرافدين يعلن عن انتصارات ال	Al Qaeda in Mesopotamia announce the victories of
12169	(القاعدة تقول ((الامريكان يبتلعون الطعم حتى الخيط	According to al-Qaeda ((Americans swallow the bait so that the thread)
1219	بيان من تنظيم القاعدة يروي تفاصيل غزوة تكريت على ا	A statement from al-Qaeda tells the details of a foray Tikrit
1220	تنظيم القاعدة في بلاد الرافدين يتبنى الهجوم على قص	Al Qaeda in Mesopotamia, to adopt the attack on the Cutting
1223	تنظيم القاعدة في بلاد الرافدين يتبنى الهجوم على قص	Al Qaeda in Mesopotamia, to adopt the attack on the Cutting
1227	بيان من أمير الجناح العسكري بتنظيم القاعدة في بلاد	A statement from the emir of the military wing of al Qaeda in the Land
12922	الصفحة 3 -الهيئة الإعلامية: بيان من تنظيم القاعدة ف	Page 3 - The Media: A statement from al-Qaida P
12926	عاجل عاجل القاعدة تؤكد الخبر رحماك يارب منتدى انصـ	URGENT Al Qaeda confirm Rakhmok News Forum Lord Ans
1296	بيان من تنظيم القاعدة يتبنى تدمير مدرعتين وكاشفة ا	A statement from al-Qaeda advocates the destruction of armored and revealing a
13083	من مجلس شورى تنظيم القاعدة في بلاد الرافدين /بيان	Shura Council of al-Qaeda in Mesopotamia / a
1310	بيان من تنظيم القاعدة يتبنى المعارك الضارية في الر	A statement from al-Qaida adopts fierce fighting in the Waller
13334	رسالة الى تنظيم القاعدة حول الشيخ ابو حمزة المهاجر	Letter to al-Qaeda on Sheikh Abu Hamza al-Muhajir
1341	القاعدة في جزيرة العرب تحث المسلمين على قنص الامري	Al Qaeda in the Arabian Peninsula urged Muslims to hunt Alammeri

Records 1 to 20 of 431

Figure 2. The Screenshot of Single Forum Search Using the Keyword “Al-Qaeda” in the Forum “AlFirdaws”

A. *Forum Browsing & Searching*

1) *Single Forum Browsing & Searching*

Figure 2 shows an example of single forum searching using the keyword “Al-Qaeda” (in Arabic) in forum “AlFirdaws.” In total, 431 threads are returned, each of which has the keyword “Al-Qaeda” (in Arabic) in the thread title. The multilingual translation function automatically translates the returned results from Arabic into English. Some interesting discussion threads related to “Al Qaeda” can be easily identified from the English translations, such as “How to become a member of the al-Qaeda (the Arabian Peninsula)?” “The latest statement from al-Qaida operations on day Aloh” and “Al-Qaeda in Arabian Peninsula urged Muslims to hunt Alammeri.”

2) *Multiple Forums Browsing & Searching*

Besides browsing and searching information in a particular forum, our portal also supports browsing and searching information across all the forums in the portal. For example, when searching across all seven forums using the keyword “bomb,” 136 threads are returned in total, each of which has the keyword “bomb” in the thread title. Among these threads, 108 are from forum “Montada,” 12 from forum “AlFirdaws,” 5 from forum “Alsayra,” 5 from forum “IslamicNetwork,” 3 from forum “Alokab,” 2 from forum “Hdrmut,” and 1 from forum “Hawaa.” The forum “Montada” has many more discussions on this topic than the other forums.

The multilingual translation function automatically translates the threads returned from Arabic forums into English. Some interesting discussion threads related to “bomb” can be easily identified from the English translations, such as “8 bombs destroyed fuel tanks near the U.S. base,” “Investigators reveal the ease of smuggling a dirty bomb to America,” and “US fears of Islamist control of the nuclear bomb.”

B. *Forum Statistics Analysis*

For each forum, statistical data such as the number of members, number of threads, number of messages, start date and end date of the forum is provided. In addition, a Java applet-based chart is created to show the trend based on the number of messages produced in different time periods, which can help users to understand the traffic of discussions over time. For example, the forum “AlFirdaws” experienced intensive discussions from September 2006 to August 2007. After that, the number of messages per month dropped significantly. The peak discussion traffic achieved during September 2006 and August 2007 might be caused by certain events which occurred during that time period.

C. *Multilingual Translation*

The multilingual translation function can automatically translate the returned browsing and searching results from non-English to English. The function is implemented using Google Translation API (http://code.google.com/apis/ajaxlanguage/documentation/#Translation).

Although there are other multilingual translation services available online, we choose the Google Translation API for the following reasons. First, as one of the most widely used online translation services, it provides relatively stable, fast, and accurate translation performance in general. Second, it is free of charge and easy to access. No registration or validation is required. In addition, it can be easily integrated into Java based Web applications.

To conduct automatic translation, the multilingual translation function first checks whether a returned browsing or searching result is in English or not. If not, it will then send the textual data to the remote server to conduct the translation and receive the translated data once the server is done.

D. Social Network Visualization

The social network visualization function analyzes and visualizes the interactions among different members in a given forum according to different topics of interest (such as "bomb"). The social networks are generated based on two assumptions. First, forum members who have posted in the same thread are considered to have interacted. Second, the more messages the members posted in the same thread, the more intensive their interactions are.

As shown in Fig. 3, a social network consists of a set of nodes and links. Each node represents a forum member. The size of a node is proportional to the number of messages he/she has posted on a given topic (such as "bomb"). The more messages posted, the bigger the node. The link between two members indicates that they have posted messages in the same thread(s). The weights of the links are normalized to measure the intensity of the interactions among two members. The thicker the link, the more intensively the two members have interacted with each other. The weight between two members is calculated based on the number of messages they posted in the same thread(s).

The network is dynamic and user-interactive. Any node and link can be dragged and moved. Users can modify three parameters (i.e., forum name, minimum number of important posts, and minimum link weight) to view the social network analysis results dynamically. The "forum name" parameter allows users to select a particular forum. The "minimum number of important posts" parameter allows users to set up the threshold value to restrict the numbers of members displayed in the network by choosing the ones who have posted more messages related to the given topic. Similarly, the "minimum link weight" parameter allows users to set up the threshold value to restrict the number of links displayed in the network by choosing the ones with higher link weights (meaning more interactions).

Fig. 3 is an example of the social network for the forum "IslamicNetwork" on topic "bomb" by specifying the "minimum number of important posts" parameter to be 10 and the "minimum link weight" parameter to be 200. The most active member is "Abu Muqatil" who has posted 74 messages related to "bomb," followed by "Abu Dujanah" with 24 messages and "Helper" with 18 messages. We can also identify that "Abu Muqatil" and "Abu Dujanah" have the most interactions about the topic "bomb."

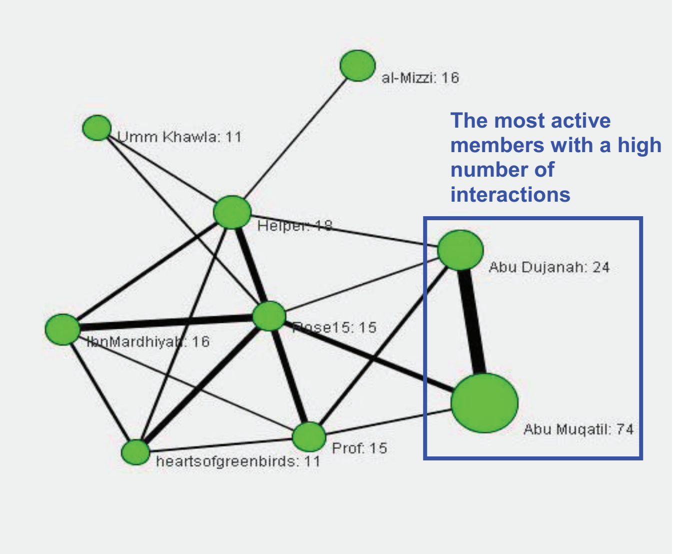


Figure 3. An Example of Social Network Visualization of Members Posting "Bomb" Related Content

VII. Preliminary User Study

We conducted a pilot study with 5 users to evaluate the Dark Web Forums Portal. Each user performed all the tasks related to different functions (i.e., single forum searching, multiple forums searching, statistics analysis, multilingual translation, and social network visualization) provided by the system. For single forum searching, users were asked to compare the function provided by our system with the search function offered by the original "Alokab" forum website, since the forum search function provided by the original "Alokab" forum website was similar to ours. The 5 users completed all the searching tasks successfully using both systems. However, on average, users completed their searching tasks faster when using Dark Web Forums Portal (2.6 minutes) than using the original "Alokab" forum website (5.5 minutes).

For the other functions, no benchmark systems were included, because none of the existing systems or websites offered similar functions on the same data set. Therefore, we used users' subjective assessments of the usefulness, ease of use, intention to use, and satisfaction to evaluate the other functions provided by our system. Seven-point Likert scales were used, with 7 being "strongly positive," 4 being "neutral," and 1 being "strongly negative." For searching multiple forums, users gave average ratings of 6.4, 6.7, 6.3, and 6.5 on usefulness, ease of use, intention to use, and satisfaction, respectively. All scores are much higher than the midpoint (i.e., score of 4). Users' average ratings on the statistics analysis function and multilingual translation function are all very positive, being around 6 and above. During the experiment, users showed strong interest in the automatic multilingual translation function provided by our system. Examples of users' comments include: "I really feel surprised using the imbedded translation module," "the translation function is very handy for searching non-English forums," and "the translations were great." Users also provided quite positive average ratings on the social network visualization function, being 6.6, 6.6, 6.7, and 6.0 on the four dimensions respectively. Users commented that the function provided "a very good tool to understand the interaction among forum members," and "it looks really nice."

VIII. Conclusions and Future Directions

In this study, we presented an integrated approach to search and analyze international Jihadist forums. A Web-based multilingual portal, the Dark Web Forums Portal, was developed based on the approach. The portal integrated forum data from seven major active international Jihadist forums identified by domain experts. Different functions provided by the portal include single and multiple forums browsing & searching, forum statistics analysis, multilingual translation, and social network visualization. These functions were designed to help users locate and understand and eventually utilize the information they want quickly and easily. The Dark Web Forums Portal is an infrastructure to integrate heterogeneous forum data, and will serve as a strong complement to the current databases, news reports and other sources available to the research community.

In the future, we plan to incorporate more Jihadist forums into our portal. We will make the portal available to the research and intelligence communities to seek their feedback.

Acknowledgments

This work is supported by the NSF Computer and Network Systems (CNS) Program, "(CRI: CRD) Developing a Dark Web Collection and Infrastructure for Computational and Social Sciences," (CNS-0709338), September 2007 - August 2010.

References

[1] A. Abbasi, and H. Chen, "Affect Intensity Analysis of Dark Web Forums," IEEE Intelligence and Security Informatics, pp. 282-288, 2007.

[2] T. Fu, A. Abbasi, and H. Chen, "A Hybrid Approach to Web Forum Interactional Coherence Analysis," Journal of the American Society for Information Science and Technology (JASIST), vol. 59, no. 8, 2008.

[3] H. Chen, "Exploring Extremism and Terrorism on the Web: The Dark Web Project," Lecture Notes in Computer Science, 2007.

[4] H. Chen, Intelligence and Security Informatics for International Security: Information Sharing and Data Mining, London: Springer Press, 2006.

[5] Y. Zhou, J. Qin, G. Lai et al., "Collection of U.S. Extremist Online Forums: A Web Mining Approach," in Annual Hawaii International Conference on System Science, 2007.

[6] E. Reid, J. Qin, W. Chung et al., Terrorism Knowledge Discovery Project: A Knowledge Discovery Approach to Addressing the Threats of Terrorism, pp. 125-145, 2004.

[7] J. Xu, and H. Chen, "The Topology of Dark Networks," Communications of the ACM, vol. 51, no. 10, pp. 58-65, 2008.

[8] E. Im, W. Chee, H. Lim et al., "An Online Forum Exploring Needs for Help of Patients With Cancer: Gender and Ethnic Differences," in Oncology Nursing Forum, 2008, pp. 653-660.

[9] A. Abbasi, H. Chen, and S. Arab, "Sentiment Analysis in Multiple Languages: Feature Selection for Opinion Classification in Web Forums," ACM Transactions on Information Systems, vol. 26, no. 3, pp. 1-34, 2008.

[10] C. Yang, and T. D. Ng, "Analyzing Content Development and Visualizing Social Interactions in Web Forum," in IEEE International Conference on Intelligence and Security Informatics (ISI'2008), 2008, pp. 25-30.

[11] H. Chen, W. Chung, J. Qin et al., "Uncovering the DarkWeb: A Case Study of Jihad on the Web," Journal of the American Society for Information Science and Technology (JASIST), vol. 59, no. 8, 2008.

[12] D. Zeng, H. Chen, D. Tseng et al., "BioPortal: A Case Study in Infectious Disease Informatics," in Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL'2005), 2005.

[13] Y. Zhou, J. Qin, and H. Chen, "CMedPort: An Integrated Approach to Facilitating Chinese Medical Information Seeking," Decision Support Systems, vol. 42, no. 3, pp. 1431-1448, 2006.

[14] S. Shim, and B. Lee, "Evolution of Portals and Stability of Information Ecology on the Web," in Proceedings of the 8th international conference on Electronic commerce, 2006.

[15] N. Glance, M. Hurst, K. Nigam et al., "Deriving Marketing Intelligence from Online Discussion," in Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining(KDD'2005), 2005.

[16] R. Popp, T. Armour, T. Senator et al., "Countering terrorism through information technology," Communications of the ACM, vol. 47, no. 3, pp. 36-43, 2004.

[17] InternetWorldStats. "Internet World Stats," http://www.internetworldstats.com/stats7.htm.

[18] Y. Zhou, J. Qin, H. Chen et al., "Multilingual Web Retrieval: An Experiment on a Multilingual Business Intelligence Portal," in Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'2005), 2005.

[19] D. Wu, D. He, H. Ji et al., "A Study of Using an Out-of-box Commercial MT System for Query Translation in CLIR," in Proceedings of the 2nd ACM workshop on Improving Non English Web Searching (iNEWS '2008), 2008.

[20] D. Liben-Nowel, "The Link-prediction Problem for Social Networks," Journal of the American Society for Information Science and Technology (JASIST), vol. 58, no. 7, pp. 1019-1031, 2007.

[21] G. Kossinets, and D. J.Watts, "Empirical Analysis of an Evolving Social Network," Science, vol. 311, pp. 88-90, 2006.

[22] Y. Yeung, "Macroscopic Study of the Social Networks Formed in Web-based Discussion Forums," in Proceedings of the 2005 Conference on Computer Support for Collaborative Learning, 2005, pp. 727-731.

[23] J. Raab, and H. B. Milward, "Dark Networks as Problems," Journal of Public Administration Research and Theory, vol. 13, pp. 413-439, 2003.