



## Interview with Hsinchun Chen on the Dark Web

*At the University of Arizona, a group of researchers is working to catalogue the single largest databank of terrorist writing on the web, from social media to chat rooms. This searchable databank is part of a project called the [Dark Web](#), which has been funded by the National Science Foundation and the Defense Department. [Hsinchun Chen](#), a computer scientist who is heading the project, spoke with the Medill National Security Zone about his work, which he believes could provide unique insights into terrorist networks.*

**Medill National Security Zone:** Can you give me an overview of the Dark Web Project?

**Chen:** We started looking at the presence of terrorism and extremism phenomena on the Web. We initially started out in 2002 and 2003 with websites, because it was their public face to the world. Then I hooked up with Dr. [Marc Sageman](#) who is an expert in terrorist networks. He was writing his book at the time I was doing my research.



Since then what started as a static web presence in terms of websites turned into looking at social media, the members that sympathize and how they interact. We look at what kind of interaction patterns, what kind of emotional state and what kinds of opinions are spread in social media. I think we are using a more automated computer science method longer and more comprehensively than any group in the world.

**MNSZ:** How do you deal with the language barrier?

**C:** There is a very significant distinction between what my lab does and what another research center would do. We are very computationally driven in that I don't really use linguists. You don't get most of the high linguist Arabists. They do the manual qualitative analysis.

In our lab we are computer scientists. We develop linguistic computational programs to look at English, Arabic and French postings. We look at a topic. We correct them with a computer program and highlight the important signature of the poster in the original language.

We analyze the emotional state of Arabic writing, which may be anger or violence. Everything that we do is based around computer programs that are carefully developed with feedback from the linguist, Arabist or language expert. Once we do that we process a large amount of content.

A typical human analyst can process 50 or 100 documents in a couple of days. We process millions of documents and pull out the key important signals based on the computation methods that we've developed. This is why it is quite unique.

**MNSZ:** Are people able to log in and search through things?

**C:** There is a link to a system called the [Dark Web Forum Portal](#). You can send your request. We will have a researcher or staff member qualify you. There are about 300 people around the world on it. They are mostly analysts and researchers.

They are using our content to do their investigation or their own analysis. That may be used either for computation of context or in a socio-political context. This portal is one of our more recent contributions to the academic research world.

It consists of about 20 to 30 foreign languages such as English, Arabic, French, German and Russian. It consists of close to 15 million messages posted by a quarter of a million jihadists.

We developed a system that has translation ability, search ability and visualization ability. You can find any relevant topics and time frames. People can find any topics that interest them in any time frame in any forum. They can search across all 29 forums and all 50 million messages to find out the postings of members and their interconnections, meaning their social networks.

**MNSZ:** Let's say I want to research Hamas rocket attacks. Is that something that I could do?

**C:** Definitely. That's one of the thousand examples. You would enter "Hamas comma rocket attack." You can say you want to do a cross-forum search or you can find a particular Hamas forum and go to that forum. It will take about 15 to 20 seconds. It will tell you how many hits. It will measure the keywords by partial or exact match for the forums. It will tell you that maybe there are one to 30 postings in Islamic Awakenings and 10 in this one. Then you click on the posting. If there are more than 30 you can go to that forum.

It will show you all the forums and the threads. You can filter. You can look at a subset of that posting. You can say, "The post is in Arabic. I can't read it." You do a Google translation of what the message is about.

Then you can move the content to a social network visualization tool that can illustrate how many people are major posters of that message and how many messages they post. Then you can create a picture of the member and what they say. You can have everything altogether.

**MNSZ:** Would the Dark Web have helped people understand the Arab Spring?

**C:** The Arab Spring content is more relevant to Twitter and Facebook. Forums can probably capture or reflect on the phenomena during that three to five months or even up to a year.

You can see the sentiment change. I cannot say that forums had as much impact as Twitter and Facebook at that time. Different social media forums allow you to study phenomena in a more longitudinal fashion. It's open. The collection is typically comprehensive. Facebook and Twitter on the other hand, the content is a spurt of ideas. People get together like a flash mob. People come and go.

We have been requested to expand from Dark Web to another project called Geopolitical Web. This Geopolitical Web Project doesn't just look at the terrorist groups. We are looking at regions and countries.

This is not just in forums. We're also collecting Twitter, Facebook and YouTube content. YouTube content can be very insightful and contiguous. We are expanding the kind of content we look at and the different social media types.

We are looking at 15 regions including the typical ones like Syria, Somalia, Yemen, Iraq, Afghanistan and Indonesia. We are looking at the more volatile regions that could have the next Arab Spring or another revolution.

[Those interested in signing up for the Dark Web, should contact the Lab's Associate Director, Ms. Cathy Larson at [cal@eller.arizona.edu](mailto:cal@eller.arizona.edu)]



## A B O U T W A R 2 . 0

War 2.0 presents original research on the intersection of It catalogues the research projects run by various parts connections between the entities funding and performing national security strategy in the years ahead...

Feel free to Tweet or like us either on

 

## R E C E N T P O S T S

- Predicting Tomorrow's News
- Pentagon Funds Social Media Contest
- Hearing on Jihadist Social Network Use Today

- The CIA, Link Analysis and a Silicon Valley Startup
- New Justice Department Proposal Could Make Dating Prof

## ARCHIVES

- February 2012
- December 2011
- November 2011
- October 2011

## BLOGROLL

- Complexity and Social Networks
- Swarm Creativity

---

## WHAT IS WAR 2.0

- Introduction: The Science of Networks
- A Brief History of Network Analysis
- Interview with Hsinchun Chen on the Dark Web
- Interview with Kalev Leetaru, on Culturomics 2.0
- Glossary of Terms
- A Brief Reading List

## OTHER SITE PAGES

- The Blog
- About the Project
- The Network: Introduction

## CASE STUDIES

- Connecting the 9/11 Hijackers
- Finding the IED Cells
- The Capture of Saddam Hussein
- The Hunt for Osama bin Laden

## THE NETWORK: UNIVERSITIES

- University of Arizona
- Carnegie Mellon University
- University of Maryland
- Massachusetts Institute of Technology

## THE NETWORK: PRIVATE COMPANIES AND NONPROFIT

- FMS Advanced Information Group
- GeoEye
- HBGary Federal LLC
- IBM
- INTENSITY Corporation
- Mitre Corporation
- Northrop Grumman
- NTrepid
- Orgnet.com
- Palantir Technologies
- Recorded Future
- SS8
- Visible Technologies

## THE NETWORK GOVERNMENT AGENCIES

- Central Intelligence Agency
- Defense Advanced Research Projects Agency (DARPA)
- Department of Homeland Security
- Federal Bureau of Investigation
- Intelligence Advanced Research Projects Activity (IARPA)
- National Security Agency
- U.S. Special Operations Command