

## **Building on Basis for Multilingual Digital Forensics**

By Mark Terry - May 2006 Issue, Posted May 16, 2006

<http://www.econtentmag.com/?ArticleID=15478>

All Content Copyright © 1998-2006 EContentmag.com - All Rights Reserved

Basis Technology recently announced an initiative to create the next generation of digital forensics products. Basis Technology specializes in multilingual information retrieval, focusing on the problem of searching, sorting, classifying, and organizing information in many different languages. The company's clients include Google, Microsoft, MSN Search, Yahoo!, AOL, and numerous others. Carl Hoffman, CEO of Basis Technology, says, "Our software enables these search engines to index and retrieve Web pages in languages other than English. So if you go to Yahoo! and enter a query in Japanese, it will give back to you Japanese Web pages. It's our software under the hood that makes that work."

Basis believes it can leverage its analytic multilingual search approach to enhance the field of digital forensics. To support this endeavor, the company hired Dr. Brian Carrier, PhD, to serve as the director of its Digital Forensics Group. Carrier is the author of the textbook *File System Forensic Analysis* and of two open source forensics tools, *Sleuth Kit* and *Autopsy*. Basis also hired Dr. Simson Garfinkel, PhD, as a consultant with the title of consulting scientist and forensics software architect. He is the author of the open source *Advanced Forensics Format (AFF)* software library and the author or co-author of 14 books on computing.

Hoffman says, "They each have unique strengths and experiences: Dr. Carrier is one of the world's experts on the forensic analysis of file systems. Dr. Garfinkel is in many ways the visionary of this system. He has been cooking up a system like this one in his head for many years now, and he actually first began working for us about a year ago in a consulting capacity, writing our early design papers. He is also an expert on what is referred to as drive sanitization practices, and is the author of a fairly landmark study on sanitization practices."

Hoffman sees Basis' digital forensics market target as law enforcement, defense, the military, the intelligence community, and litigation support. "These tools are especially important in law enforcement because if you look at any criminal enterprise today, it runs like a corporation and uses computers extensively."

Basis Technology's primary product is called the *Rosette Linguistics Platform*, which focuses on documents and text, similar to the current generation of

forensic processes. This has two components: acquisition of information on the computer into a specific format, and analysis of the information.

To leverage the company's background in multilingual search, Carrier will work to integrate Basis Technology's foreign language support with forensic tools. Carrier says, "The industry has some basic searching capabilities for foreign languages, but they don't really have the high-powered analysis that Basis currently delivers. The software will now go through and extract out files and data from the hard disk and then import it into the analysis software and perform additional analysis techniques beyond just the foreign language analysis."

United States operating systems typically use the American Standard Code for Information Interchange (ASCII) for formatting, which has 128 symbols. European countries use a 256-character set, which can accommodate Latin figures and diacritical marks in a variety of languages including French, German, Spanish, and Italian. Once you start bringing in Russian, Middle Eastern, or Asian languages, the subject of data retrieval and forensic recovery becomes significantly more complex. Hoffman says, "Let's say you have a hard drive and you're searching for any document on a hard drive that contains keywords that you're looking for. Most forensic products today will do a pretty good job of finding documents containing those keywords, just like your desktop search utility would, provided that the keywords you're searching for are in English." Clearly, Basis feels it can build upon its expertise to help organizations glean information from digital media in any language.

Carrier says, "What we're looking at is much more automation in the large scale investigations. There currently is a lot of manual processing involved. So we're working on automating part of this process to make it more efficient and allow people to correlate and link computers together."

([www.basistech.com](http://www.basistech.com))