ACID Scans Web for Pirated Multimedia

April 5 2007

The Virage division of Autonomy has developed search technology that can scan the Web for pirated video and multimedia clips.

Broadcasters, movie producers and media publishers of all types have access to a new search technology that can scan the Internet to discover Web sites that are illegally distributing copyrighted video and images.

ACID (Automatic Copyright Infringement Detection), developed by the Virage division of Autonomy, can detect illegally posted rich media in any format wherever it is posted, according to Autonomy founder and CEO Michael Lynch. Autonomy officially introduced the ACID technology in a press conference here on April 4.

"Acid watches very large amounts of video and it can spot video that is owned by someone else," in a highly automated process, Lynch said.

It could be used to detect movies, television clips or any copyrighted media posted, for example, on YouTube, on personal Web sites or on any other Web site, according to company officials.

The technology is equally useful to these file-sharing sites, because it gives them the means to scan their sites for infringing media before it's posted, or at least before it results in a lawsuit. The technology allows these scans to be performed quickly to avoid lengthy delays in posting new content online.

The automated search technology will free copyright owners from

having to spend hours manually searching through files on video-sharing sites, company officials said.

With the current explosion of interest in video distribution and video sharing over the Internet, the technology will give media producers the means to try to stay ahead of the increasing volume of illegally distributed multimedia content, Lynch said.

Furthermore, it doesn't rely on tagging technology or video watermarking to locate copyright-infringing media. These techniques can be thwarted by changing formats or by video codec changes, according to Lynch.

Instead, ACID uses Autonomy's "meaning-based computing" technology, which allows computers to find relationships within many different types of unstructured data, including text, word processing documents, e-mails, audio and multimedia. Acid uses patented image and audio analysis technology to look for known examples of copyrighted material no matter what format it's stored in.

ACID is also based on Autonomy's IDOL (Intelligent Data Operating Layer), a basic search platform that can analyze information in more than 1,000 formats including text, voice and video.

Since ACID works with all media formats, it can detect whether a portion of a copyrighted video or audio tract has been overlaid or stored as part of a new and original media file.

Meaning-based computing is Autonomy's overall strategy for helping organizations mine useful information from unstructured data. It is based on search technology that goes beyond keyword-based search engines to enable organizations to locate data within the organization that would otherwise go undiscovered, according to Lynch.

Autonomy's search technology uses automatic hyperlinking and link clustering that the company claims isn't built into keyword search engines. According to the company, this technology allows computers to perform searches with greater context, so it finds a wider range of related documents or research citations than is possible from keyword searches.

The computer industry, Lynch maintains, has to turn to technology like meaning-based computing because, "We are going to see less and less structure in data because there is going to be more and more data and it is going to be unstructured."

The computer industry's current "obsession with structure and - keyword - tagging is fundamentally wrong" as a way to bring order to unstructured data, Lynch said.

Global defense and aerospace company BAE Systems is implementing IDOL to make it easier for the company to find documents and information scattered across a multitude of offices and information systems, according to Scott Petrie, a knowledge engineer with BAE's San Diego-based National Security Solutions group.

Petrie said he was recruited by the company to help find ways for the company to mine more value from its unstructured data resources. BAE chose IDOL because it provided the best technology for finding relationships within the vast store of unstructured data in repositories of all types within the company, Petrie said.