

## Collexis: The Research Fingerprint

The new year ushered in more than champagne and confetti for Collexis Holdings, Inc., a relatively new search system that hit the ground running last year with its one-of-a-kind applications for search, indexing, and aggregation. So far in 2008, the company is off to a good start:

- On Jan. 3, the company announced that it finalized a \$2.2 million private placement of restricted common stock, the proceeds of which were used as the first payment to acquire SyynX Solutions GmbH, a German-based software company that has been Collexis' longtime software development partner.
- On Jan. 10, Dell and Collexis announced plans to launch BioMedExperts, a social networking community designed to promote collaborative medical research and development. Through the BioMedExperts community, healthcare and life science professionals will be able to collaborate and continue research for 1.4 million biomedical experts with 12 million pre-established network connections from more than 120 countries. The site will also allow users to analyze associated professional connections with the network and view scientific publications. Dell will provide the computer hardware and marketing support for the group.
- On Jan. 11, at the start of the ALA Midwinter Meeting in Philadelphia, Collexis announced its intention to collaborate with

The screenshot shows the Collexis website's 'Our Uniqueness' section. It features a navigation bar with links like 'Home', 'Products & Services', 'Industry Solutions', 'Technology', 'News & Events', 'Investor Relations', 'Customers', and 'Partners'. The main heading is 'Our Uniqueness'. Below it, there are three sub-sections: 'The Increasing Demand for Knowledge Management /Discovery Applications', 'The Collexis Difference', and an image of a hand pointing to a fingerprint. The text describes how web-based technologies have led to a huge amount of unstructured information, making it difficult to access. It highlights that Collexis makes significant improvements to standard data and information retrieval capabilities, uncovering unique information and creating potential new hypotheses based on large volumes of unstructured content.

Thomson Scientific on future product offerings that include marrying Web of Science data in the Collexis Knowledge Dashboards for the academic and government markets.

These three events have firmly established Collexis as a key player in the search market, a position that is reinforced by its long list of established clients that include the Mayo Clinic, The Johns Hopkins Uni-

The diagram illustrates the Collexis fingerprinting process. It starts with 'What we know' and explains that finding information using fingerprint technology is powerful. It then shows a flow from 'contents fingerprints' (including Emails, Word, RFPs, Jobs, CVs, Skills, and Articles, books) to 'people fingerprints' and finally to 'organization fingerprints'. The text explains that when one knows who the author of a publication is and where he works, the Collexis system can create profiles of this author and organization based on the publications. At the bottom, it states: 'Using the same approach as when we are looking for documents, we can type in a query (some terms or a whole piece of text) and look for experts and organizations. Aggregated Fingerprints are available via Collexis' patented technology can create fingerprints, or profiles, of individual authors and the organization based on the publications.'

Collexis' patented technology can create fingerprints, or profiles, of individual authors and the organization based on the publications.

versity, The University of California—San Francisco, Merck & Co., Lockheed Martin, the Wellcome Trust, the National Institutes of Health, and the U.S. Department of Defense, to name just a few. By December 2007, Collexis signed 15 new customer contracts in the second half of the year as it emerged from a one-core platform to one offering a suite of products.

The Collexis applications now reach into three basic verticals: legal, government, and STM. Its technology can be customized to just about any discipline, but the company concentrated its major efforts in biomedicine, academia, organizations, and government. Search capabilities extend beyond simple discovery; when armed with solid thesauri, Collexis can spot relationships, point to emerging trends, and even analyze the competition.

Collexis is far from an overnight success. Since its founding in 1999, Collexis was in the prep stage for 6 years. It “spent that time focused on R&D and developing its core engine to make it a success,” according to Bill Kirkland, CEO. The company isn’t in competition with Google either. “We have an industry-specific knowledge platform with a consumer approach,” he says. This search is geared specifically to a select niche in each individual industry.

Collexis simply approaches discovery in a different way. Its patented technology retrieves information, it can discover hidden relationships in multiple content sources, and it can even map trends to help researchers formulate hypotheses from unstructured content. Much like with our own “fingerprints” (those unique whorls, loops, and arches), Collexis not only creates “fingerprints” from all content, but also from search information, whether that search involves a few words, a sentence, or a complete document.

Collexis also uses thesauri to retrieve select data; for those without thesauri, Collexis can build one for the specific organization. Its matching technology also calculates the relative distance between query items so that even partial matches are possible, and many of those turn up

in unexpected places and offer connections that were previously never explored. The partial match capability also means that users can use free-text search instead of a complex Boolean search query.

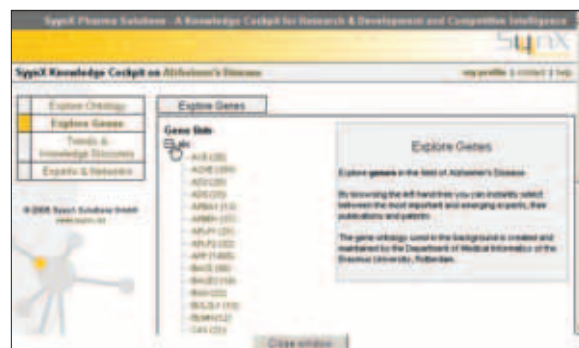
Another unique factor is its ability to manipulate the fingerprints so results can be clustered, aggregated, associated, or sorted into a user-friendly configuration. Its visualization tools set up the connections, an intricate road map of lines and links that can be reconfigured easily in one click. These network connections bridge time and place, offering fresh insights and unexpected results. “It pulls knowledge out of the content and the content out of the repository,” says Kirkland.

Darrell W. Gunter, executive vice president and CMO, joined the company last July and brought with him extensive industry experience, most recently from Elsevier. The buzz surrounding the Collexis booth at the 2007 London Online conference reflected the interest in the technology. Gunter called the show a “huge success” and says the session that Collexis offered was standing room only. After the session, booth traffic was so brisk that Gunter says he completely ran out of brochures near the end of the show.

After London Online in December 2007, Collexis hosted its first Knowledge Discovery Symposium titled *The Integration of Knowledge Discovery and Professional Social Networking*, which Gunter reports will be an annual event. The symposium, which provided overviews from organizations that now use Collexis technology, shared case studies from the Wellcome Trust and Asklepios Hospital Group, among others.

The organizations used the technology to leverage their own rich data, according to Kirkland. In terms of expert profiles in biomedicine, researchers can track a scientist’s work, see the extent of

his expertise and key projects, or scout unknown connections with other experts who may or may not be in the same field. An organization, for example, can also provide real discovery within its four



Users can tap into specialty areas and explore fields of expertise.



Visualization tools let users drill deeply into data.

walls by creating a knowledge guide that can let a researcher browse through its faculty roster and find out about specific expertise in the same institution.

According to Outsell’s Daniel Pollock in *Insights*, “Visualization tools mark the coming together of sophisticated technologies spanning indexing, taxonomies, semantic processing and advanced UI [user interface]. Their value proposition lies in achieving productivity gains and competitive advantage by cutting through the noise of ‘too many search results.’”

At ALA Midwinter, Gunter was busy fielding questions and arranging demos for the crowd surrounding the booth. “We’re off to a great start in 2008. The good news keeps coming.”

—Barbara Brynko