

# Information Today

A REPRINT FROM  
Information Today

THE NEWSPAPER FOR USERS AND PRODUCERS OF ELECTRONIC INFORMATION SERVICES

Volume 19, Issue 4 • April 2002

<http://www.infotoday.com>

## *IT Interview*

# Nstein Targets the E-Publishing Industry

*President Randall Marcinko leads the focus on automated indexing solutions*



*Nstein president Randall Marcinko*

by Paula J. Hane

**R**andall Marcinko has been an entrepreneur in the library and information technology field for over 20 years. Having forged a solid network of business relationships during that time, he is well-known by database producers, online services, and e-publishers. Marcinko founded Dynamic Information Corp., a full-service information brokerage that has specialized in document delivery since 1986. After selling Dynamic Information to EBSCO Industries in 1994, Marcinko headed up EBSCOdoc, its new document delivery division. He remained at EBSCO through March 1996 as he implemented the first complete document delivery interface on the Web. Since then, Marcinko has led Marcinko Enterprises, Inc. into consulting and back-office development for primary publishers, secondary publishers, and other information industry players.

Marcinko has divested his other businesses and is now working full time as the new president of Nstein Technologies, Inc., a Montreal-based company that offers categorizing and indexing products. He talked with me recently about the company's technology and its new focus on providing solutions to the e-publishing market.

**Q** What attracted you to work for Nstein?

**A** My former company, Marcinko Enterprises, Inc., had worked with various clients that needed a solution for indexing. As we worked with clients that had larger and larger problems and greater backfile sizes, there became a real need for computer-aided indexing (CAI) that wasn't going to be filled by manual indexing. In late 2000, we started working with UPI (United Press International)—there was no way to make it work for them with manual indexing. We

started doing surveys of companies in the CAI field—companies like Nstein—and to my surprise, in every case, Nstein came out on top. It was a company I hadn't heard of. They brought me up to Montreal to look at it. Because I was there as a consultant working for a client, we really kicked the tires. I had no intention of working for them or of even using them at first. At the beginning they were the new kid on the block, but I was really fascinated by what they were doing. Eventually, they were providing solutions that were very successful for my clients and made me look good. Ultimately I did a little consulting for [Nstein] and in the fall of 2001 we started talking about what else was possible. They offered me the opportunity to become president and see what I could do to make their operations grow.

**Q** I can see why they were drawn to you, given all your contacts working in the industry over a span of many years. You



have worked with, consulted for, or been involved with—in licensing or some operation—almost every major player in the information industry.

**A** Correct. I think that was one of the major reasons why there was synergy. They had a technology that had grown well before I came, to the point that it was ready to be commercialized. My coming on board is at a time when the technology is quite mature and ready to be seen in prime time. What's needed now is to take the technology and do something with it that is definable, with a focus, and that meets the needs of clients in an industry that is ready to buy. I think my background—both with relationships and deals I've done and my familiarity with indexing software—is a good match for what they need.

**Q** What do you see that sets Nstein's technology apart from that of competitors?

**A** Let me give you a sketch of what Nstein does so we can get to the point of where indexing is possible. Nstein takes textual documents and runs them through the Nstein engine. In the first phase we extract concepts out of the document. We feel we use the three leading technologies for doing this. We use linguistic analysis, statistical analysis, and lexicon-based techniques. Using all three approaches lets us extract the richest and most comprehensive group of concepts within a document. The concepts also have a ranking of which are most important. We then couple this with the next step, which is to create a training set of documents that have index terms attached to them already. Algorithmically the program is then able to create what we call a Linguistic DNA, or a linguistic fingerprint of that document. That information is then stored in what we term the Knowledge Base. Now, when new documents are submitted to the engine, we can automatically associate index terms with them.

So it's a three-step process: concept extraction, training and Knowledge Base creation, and then indexing. The concepts can also be extracted from languages other than English and these are mapped to a taxonomy so that a search in English will find the foreign-language documents with those concepts as well. We can also provide document summarization for clients, using our Nsummarizer specialty module.

We also do something called "Nfinding"—some people call it "entity extraction." We can identify what those concepts are. For example, we can find company names within a document. We can do it for geographical locations, drug names, people's names—for many different types of entities.

**Q** So you're capable of handling the need for "field" searching? I ask this because I covered a content-extraction product last fall that couldn't identify information by field or type.

**A** Yes. That's the other way of looking at it. In a traditional database approach, for a field called "drug name" we can create an Nfinder to populate that field. This is becoming more important today, as people want to link between databases. They need to have a solution that handles disambiguation of concepts.

**Q** Tell us about Nstein's indexing work for the APA (American Psychological Association).

**A** They are holding our entire engine in-house, as all our clients do. We send people in and help them train the Knowledge Base. It's an ongoing process. There are various ways people are choosing to use our technology. In some cases, the results, precision, and recall are great, and they'll use it as the final product. In other cases, they have an indexing staff, and they may either need to trim the staff or to put through a much more significant volume than before. They use us to save anywhere from 10 or 20 percent to 80 or 90 percent of the labor per record. Some users will say, "Suggest to us, Nstein, the terms that our indexers should look at." There can then be a component that's manual, added to the automated indexing by the Nstein software.

We work with customers to get the level of accuracy that they need. We recognize that, in the world of indexing, there are still some concepts that are incredibly intangible. In the psychological area, for example, there are probably areas where one would want to infer from the article, based on one's prior knowledge, some concepts that should link to an index term. That would be something they [APA] would add on top of the automated indexing to maintain the quality that everyone respects them for.

**Q** How long did it take to get this process operational for the APA?

**A** The typical on-site installation is anywhere from 1 to 10 days of labor that our people spend. It depends on how much customization the client wants. One of the things we pride ourselves on is not delivering a piece of software. We deliver a solution.

**Q** So a company like APA licenses your software and then pays for a certain amount of support and customization consulting time?

**A** Yes. In our current model, clients license perpetually a current version of the software. If they purchase a maintenance agreement, it allows them to upgrade the software throughout the term of the agreement.

**Q** Your agreement with UPI represented a big entree into the American market for CAI. Hadn't UPI been using any kind of CAI system?

**A** No they weren't, and they had been working with a taxonomy that had been in the news industry for a long time and that was really inadequate for today's news. The challenge with UPI was to assist with the growth of a good representative taxonomy—not only good in terms of lexicographic standards but also one that would meet the needs of high-power searching, so that people could do cross-file searching. So we built the taxonomy and then interposed it within their production system. UPI also decided to purchase our Archive Server, which will

allow them to work back and index the many million archival records they have.

**Q** On your site you list a number of server products for different markets: Nportal server, Ncrm server, Nmarketplace server, Nserver e-publisher suite. Are these all basically the same or do they provide different capabilities?

**A** One of the big changes we've realized since I came on board is that we really want to lead in the space that we're in. We've had to make some definitions and define our focus. We are very useful in all the areas men-

---

“  
*When you  
go to sell to  
someone in CRM,  
for example,  
you first have to  
explain the need.  
When you sell  
in the e-publishing  
space and say  
'indexing,' they  
all nod.*  
”

---

tioned, but you'll see in the next version of our Web site and in my explanation that the territory we're choosing to focus on now is what we call e-publishing. This means anybody producing substantially textual material that can be stored in a database and to which they need to add index terms in order to be more competitive and provide a better product. Given that focus, that opens up all the primary publishers, secondary

publishers, syndicators—people that have databases used in the corporate world and in CRM [customer relationship management]. We're not going to be selling portal software or CRM software.

**Q** So this represents a change to the previously stated company strategy to target multiple markets and be a “content management software company.” And, in fact, all of your announced customers have been in the e-publishing market space.

**A** We realized we couldn't do a good job and be a leader in all of them. We want to be proud of understanding our current clients and become a leader in this market space and dominate it. I'm very comfortable with the e-publishing space. There's a need there that is very different from any other space. When you go to sell to someone in CRM, for example, you first have to explain the need. When you sell in the e-publishing space and say “indexing,” they all nod. You say it's expensive, it takes manual labor, and you'd like to save money and improve quality, and they nod. So all that's left is to explain our solution and let them do testing. It's an incredibly different sales process. We don't have to educate the potential client on what indexing means.

**Q** When companies like APA use your indexing solution, they can then offer their customers a better quality product in a timelier manner and one where costs can be kept down. Correct?

**A** Yes, and I would add that our customers also gain the ability to put through a much larger volume of content in order to get a more fully rounded product.

**Q** In our tough economic times, many companies have greatly reduced budgets and aren't buying new products. However, when I interviewed Flip Filipowski, the CEO of divine, recently, he said that current conditions actually have increased the need for content management products that cut costs and engage customers more closely. How do you feel about this?

**A** I absolutely agree with that. Throughout my career, I've always had businesses that support the information community, and at times of economic stress, people need to look outside to save costs or to replace functions that they can't maintain in-house. When you're in a service company and the economy is in a stressful time, as long as you can mind your own budget carefully you are presented with a good opportunity to work with clients who desperately need your help.

**Q** Do you think the provision of solutions such as classification, tagging, and indexing represents a major trend? Could this be the next killer app that saves content producers?

**A** It could be, but I tend to look at it differently. I see it as the oldest killer app that's finally making it in prime time. For 20 years we've been saying that you can't do good searching without good indexing under the hood. We need that solution. What has made it hit prime time, I believe, is the flailing of some dot-com companies that didn't understand the history.

**Q** It's funny. For a while, people in the industry were concerned that abstracting-and-indexing (A&I) companies were dinosaurs. We worried that they would be pushed out by the advent of full-text searching and the Web. Now, those A&I services seem to be even more needed and the future looks bright.

**A** The smart secondaries could go far. They can't simply index though. They need to provide added value.

**Q** What are your goals for Nstein for the next year and then longer term?

**A** Right now we are all behind the mission of making Nstein the undisputed leader in the provision of services in the e-publishing marketplace. That's our short-term goal, say within a year. In the longer term, we would like to be very dominant in the e-publishing space but we'd like to grow our tools to meet the needs of the e-publishers that

are going to grow and continue to add. In the 2- to 3-year time frame we hope to branch out into other areas. We'll evaluate which areas we are best suited to and can get to most quickly. We're not going to try to do everything. We'll pick spaces where we know we could be strong. We're going to remember basic business and Economics 101, and we're going to bring money to the bottom line, because otherwise we won't be here for our customers.

We have an incredible advantage that I hope our clients realize. We're doing much of our development work in Canada and many of our clients are in the U.S. There's an obvious advantage there in the costs on the dollar and the difference in the price for labor and production. I think that's why we're able to price our products affordably. The price of the dollar makes a difference in development costs to the tune of about 40 percent.

**Q** Do you still have an agreement with a Quebec university for research?

**A** We licensed some linguistic technology from Laval University in Quebec. In addition, we have an ongoing relationship with their linguistics lab that allows us to learn about new technologies that they are working with and to be a potential employer for people in their program. It's a very healthy business relationship with the academic community in Quebec.

**Q** What kind of sales and marketing operation do you have for reaching customers?

**A** We have an office in Montreal and one in Chicago, and people in various locations in the U.S. We're adamant that we get out there at conferences and trade shows and let people know what we're doing. We're involved with many of the professional associations within our industry. We're going to do more advertising. We're also going to work on seminars and other educational opportunities for our clients. We're already ramping up our sales staff.

There will be a significant number of people added in the next few months in computational linguistics and in programming, as well as sales. We do see significant growth in order to maintain and stay ahead of the curve. There's some R&D on an ongoing basis to improve our products and there's a lot of support for ongoing clients. You can expect to see announcements of new customers over the next few months that will indicate we are moving rapidly in this industry.

For more information on Nstein Technologies, visit <http://www.nstein.com> or call 877/678-3461.

---

*Paula J. Hane is editor of NewsBreaks, contributing editor of Information Today, a former reference librarian, and a long-time online searcher. Her e-mail address is [phane@infotoday.com](mailto:phane@infotoday.com).*