

This is a copy of the article that appeared on  
<http://www.streamingmediaworld.com/gen/reviews/searchassociation/>

## **The Streaming Search Engine That Reads Your Mind**

by [John Townley](#)  
August 10, 2000

We recently had the pleasure of a wonderful lunch at one of North Shore Long Island's four-star restaurants, the landmark Cypriot dining destination [Ayhan's Shish-Kebab](#), in the company of the main spinners of a new search engine that chases streaming media in a new way.

The company is called [Taalee](#), brainchild of Indian semantic savant Amit Sheth, PhD and Pinnacle founder Ajay Chopra. It does something that other search engines do not do, which is infer meaning when you do a search, rather than simply find a given metadata keyword or image. In a way, it's designed to read your mind. And, in the process, it both shrinks Internet data overload and then expands it in quite another way.

Hmmm...you may say. How's that possible? Are we talking search engine AI here?

Well, sort of. When you go to your average search engine, it looks for a tag. A word or a phrase usually, and it brings back everything it has on record that contains it. The results can be staggering and totally daunting. You go looking for Roger Clemens's batting average and you find 100,000+ references to Mark Twain (Samuel Clemens), and surprisingly little about baseball. So you give up, swamped in irrelevant data you haven't time for. Or, you can refine your search to "Roger Clemens" (using quotation marks to restrict it to the full phrase) and find reams of data about English marriages to a variety of Rogers in that clan, long before baseball was invented.

### **Finding by Association**

But suppose you had a search engine that would look for a set of associations, not just a literal tag? Northern Light sort of does this, by spotting commonly associated words to the one you're looking for and sorting them in folders in a kind of flat file system. Taalee claims it's system can do that and go one step better by allowing users to drill down in each folder a number of levels creating a complete hierarchical, or relational system.

Want connections to films or radio reports from baseball "stadiums" (it will know that's where the game's played) or about the Dominican Republic (which is where the balls are made)? Or bats (not the furry kind)? It will line you up on the core and ancillary issues of the subject matter, not just the word or phrase itself, thus focusing in on what you really want to know and not diluting it with irrelevancies.

This sounds like a relief, but it may take some getting used to. This is not the first time an attempt has been made to group specifics with their surrounding information environments. There was a time when the Encyclopedia Britannica read like a dictionary - a simple word search, and then a whole lot of data on the word. It took up two shelves, and you could thumb through it all day before you found what you wanted.

### **Hard Copy Flop**

Then the publishers revamped the whole concept by grouping the information by associated subjects (sort of a print version of hyperlinks), which took years, millions of dollars, and brought us the modern version of the reference shelf classic. It confused us totally, and we have not gone back to those bookshelves since.

### **Interactive Ease**

Taalee has taken something of that approach to Net searching, but has eliminated the confusion by allowing interactive customization, something not possible in hard copy print. It will, depending on the choice of its clients who lease and install it on their sites, chase down your own personal preferences according to your option, making it tailored to you, the Web site it's coming from, and the broader context of the subject you're interested in.

"The simplest-sounding things are sometimes the most difficult to do," said Amit Sheth, founder and chief executive officer of Taalee Inc. "But now, providers of Web-based information can offer their users the power of Taalee's Semantic Web technology to search the Web faster and more efficiently than ever before. Taalee's services, which today include Semantic Search, Directory, Cataloging, Personalization, and Targeting, are powered by its Semantic Engine and WorldModel. We're providing a meaningful context that enables our customers- content owners, publishers, and advertisers, both enterprise and media, to provide the user with the content in which they're interested. Meanwhile the content owner, publisher, and advertiser achieve their objectives with respect to communication and monetization."

Sheth's brilliance can sometimes lose the listener, which may at times be intentional, as the Semantic Engine and WorldModel are proprietary developments. Where is he coming from?

### **Sheth's Credentials**

Dr. Sheth combines experience at Honeywell, Unisys, and Bellcore with that gained as Founder and Director of the University of Georgia's Large Scale Distributed Information Systems Lab. Taalee's Semantic Infrastructure, he told us, has been developed through extensive academic and applied research over the last five years. It combines traditional and multimedia databases, metadata management, context and domain modeling, encyclopedia data, distributed intelligent agents and XML to provide an unprecedented degree of Web searching power.

"Up to now," Sheth said, "Web assets have been found, cataloged, and searched/navigated using two popular approaches. The first is manual editor-based directories, where pages are manually put into categories, similar to a book's table of contents. The second is crawler-based analysis and keyword indexing, which is similar to the index at the end of a book. Human editors can produce high quality search results, but they can't keep up with the Web's growth of content; they have the 'depth' but not the 'breadth.' Crawlers are efficient, but lack the human understanding of what the content means. Without contextual or relationship information, crawlers have the 'breadth' but not the 'depth' ".

## **Semantic Engine**

Taalee's Semantic Engine, Sheth pointed out, overcomes the shortcomings of both these approaches through its use of intelligent agents, thereby collecting a "breadth" of information. It then cross-references this information to a knowledge base, cataloging the information assets in categories and sub-categories. This process creates the "depth" of knowledge necessary to provide precise and fresh results. "This outcome brings cataloging and searching to entirely new levels," he said, "It's the first implementation of what Tim Berners-Lee has called the Semantic Web -- his vision for the next phase of the Web's evolution."

## **Semantic Web**

The Semantic Web, Dr. Sheth noted, is a contextual network built of meaning-related content from many sites. This content is made up of domain-specific data. Individual user desires and relevant content are mapped to each other in meaningful contexts. These meaningful relationships provide Web publishers with opportunities to monetize users' searches through context/intent relevant (i.e., meaningful) advertising/promotional content. "Thus," he said, "a Semantic Web has been created and utilized by various interested parties to exchange various forms of value."

Combining semantics and automated intelligent agents, Sheth told us, produces two key advantages: precision and richness in interacting with content and freshness of the content. These advantages enable a far greater degree of asset cataloging and management, search, and personalization. "When our automated intelligent agents visit a site," he said, "they know what to look for, and they understand what they find. Taalee agents acquire much more meaningful information, in the form of structured relational metadata, than do mere 'page scrapers,' which take all of a page's words indiscriminately and index them as keywords. From this structured relational metadata, Taalee provides a much higher quality of search, and presents more interesting and informative content."

With their trademarked Semantic Search and Semantic Asset Management, Taalee brings this search technique into the real world. Taalee Semantic Search combines a digital media search engine with a service that locates fresh, up-to-the-minute audio and video content. It empowers Internet users' effortless access to rich, fresh and highly relevant Web content. Its Blended Semantic Browsing and Querying combines the best of

directory-like browsing with the precision of a structured, easy-to-use query, and yields, Sheth says, precise and timely results. It employs a proprietary metabase- a database of metadata- of digital media. Semantic Search has an understanding and knowledge of various digital media assets on the Web. This metabase is a database that describes each asset that results from intelligent automated crawlers' gathering of assets and their subsequent post-processing/enhancement. The gathering process is driven by Taalee's WorldModel, which contains the rules, ontology, domain knowledge, and knowledge experts necessary to carry it out.

## **Asset Cataloging**

Taalee Semantic Asset Cataloging is a semantic cataloging and asset management service; it is used by content creators for precise cataloging of their original content. Semantic Asset Cataloging also assists in correctly and automatically cataloging partners' content into Taalee's asset metabase. Once the content has been cataloged, it is searchable from any site using Semantic Search, increasing exposure for the content creators.

"The benefits that Taalee offers its partners," Sheth boasted, "are many and varied; improved customer satisfaction, delighted users and quick customization, deployment and time to market. We're developing infrastructure that will support production, distribution and advertising for digital media (audio and video enhanced with some text and photos for now) for the third generation of the web. And the Semantic Engine is obviously much less expensive than a building full of editors [yikes! - ed.]. It provides opportunities for multiple user impressions through the display of searches, results, related links, and Taalee's Rich Media Reference pages, all on the host site. Taalee is the future of services based on semantic Web searching -- it's just that simple."

Doesn't sound that simple to us - not at all. But if it works as well as Sheth says it does, it really will help you find more meaningful data more easily.

## **Coming to a Site Near You?**

Why is this particularly useful to streamers? Sheth explained:

"The value for streaming content is very high. The cost of streaming content is very high. So the understanding of the content that Taalee has is all the more valuable. You cannot eyeball an audio-video stream. You can eyeball text. While Taalee can also improve simple text, it is all the more relevant in audio-video. Additionally, Taalee has figured out a whole number of different ways - dealing with all types of files and making sure that you have sufficient information about all types of files in content and rights management issues for our partners. So there are a whole bunch of specialized services that Taalee has completed for audio-video."

## **Wrapping up Lunch**

Sheth wound up our pleasant lunch of gyros and Greek salad with a wrap on Taalee:

"There are two basically important things about Taalee," he concluded. "One is that this is the first time we have achieved understanding of the content on a highly scaleable manner. The second is that a comprehensive file structure that can be rendered as services for asset management and cataloging for our customers, for search and personalization of directories. Taalee is not a single technology or a single application company like many of the search engines are. Its core infrastructure eyes every piece of content, user interest, and e-commerce opportunity and relates it to any of the different applications related to usage and monetization of content.

### **Some Examples**

To find out how it's beginning to be used - and it is by no means fully implemented yet -- you can check it out at: [Redband Broadcasting](#) - a spin-off of the LookSmart 'super' portal. Taalee powers the cataloging and search capabilities of Redband.com. Redband's audio programs are delivered over the extensive Internet distribution network of LookSmart, AltaVista and Sony Music Entertainment, as well as third-party web sites. Taalee also powers [AWZ.com](#), a portal community for teens by teens.