



[HOME](#) • [ABOUT EMEDIA MAGAZINE](#) • [SUBSCRIBE](#) • [ADVERTISE](#) • [CONTACT US](#)



DVD today
news

**FREE
Subscription**

go to
www.emedialive.com

information insider

New Millennium Search Systems: Does X Mark the Sweet Spot?

Robert J. Boeri
Martin Hensel

EMedia Magazine, July 2000
[Copyright © Online Inc.](#)

Searching the Internet has been compared to looking through a massive library after an earthquake. As in a bricks-and-mortar library, information now exists, not only as text, but in other formats and multiple media. Are search systems coping with this approaching information glut? And what's happening with non-Web search systems? Are search systems' recall improving? (That is, do they search and retrieve every document relevant to your search?) Are systems' recall and precision increasing, and are they getting information to you more quickly and easily? In a word: Barely.

The difficulties continue to be the growing amount of information and the heterogeneity of resources to searched. We see four main trends: Emphasis on Web (not local or networked) content, commoditization, multimedia support, and XML.

web versus local or networked media

Web-based ecommerce keeps extending its reach into the unlikeliest places, even search engines. Do you think that if you are looking to buy a "lawn tractor" you'll find all of them, in a relevant order, or just the ones that have marketing arrangements with vendors willing to pay for preferential positioning on search results? Unfortunately,

search systems and their use are seldom "free," and you're sometimes unlikely to get an unbiased ranking via Web search portals. Clicky, debuting last year, randomly sequences the results of queries to show an unbiased list of results. Whatever happened to relevancy?

commoditization

Commodities are usually low-priced necessities (petroleum products being both example and exception). Two commodity search systems are free: Verity's Acrobat search plug-in to Reader and the Atomz Web system (provided your site does not have more than 500 pages). We were fans of Acrobat searching even when you had to pay for the indexing system and search client. Atomz, recognizing the pervasiveness of Acrobat, includes it in its search types.

Commoditization suggests that established search systems are unlikely to re-invent themselves with native support for newer Web-based standards like XML. Although Atomz is new enough to perhaps escape this curse, our inquiry about plans for XML support went unanswered.

multimedia

Searching non-textual media directly is still largely a dream, but search vendors are cleverly doing the next best thing: Searching metadata available in each object (e.g., title in a WAV file) or converting non-textual objects to text. Virage, whose clients include The New York Times for interactive video search, uses captions included with those files. Virage can also use its pattern-recognition technology to go beyond mere caption-based search and can actually index speech.

Streaming audio-video vendor RealNetworks has purchased an interest in Virage for use in RealPlayer. Excalibur's "Screening Room" performs similar searches via metadata included with video files. Even Atomz searches animated Flash files. One of the most comprehensive and fast search systems we've seen is from Fast Search & Transfer ASA, available at <http://www.alltheweb.com>. Fast now searches MP3 files for Lycos and has developed its own technologies for accelerating image and video via the Web, presumably to speed the delivery of multimedia search results. Fast also has made arrangements with Albert Inc., a leading provider of natural language processing software for the Internet. Fast's capabilities in multimedia, natural language, and even XML make it a system for your short list.

XML

As ever, XML's presence continues to grow in search systems, albeit slowly since the population of documents or Web sites using XML is also growing slowly. VoiceXML, the Voice eXtensible Markup Language based on XML, can provide a natural interface to search systems by providing voice access to speech or telephony resources. Voice is the ultimate natural language interface, and this is being pursued vigorously by the likes of AT&T, Lucent, and Motorola.

As for XML-based searching, Fast has developed its own XML vocabulary for searching database content. GOXML .com offers native XML searching, and Xdex, from Sequoia Software Corporation, will index, search, and retrieve XML in any file system or database.

Without a doubt, the most recognizable name in the XML search space is Infoseek, with its release of Ultraseek Server 3.1. This system recognizes XML tags and then searches specific fields. The result of XML document searches promises the precision of database queries, since XML imposes a database-like structure on documents. Combine this with Ultraseek's natural language facility, and you've got an emerging winner for Web or enterprise searching.

How to choose between commodity and leading-edge search systems? As ever, it depends on your present and future needs. If these extend no further than to established information types, and you want inexpensive searching, go with the commodities. But if you're on the XML or multimedia bandwagons, get acquainted with some of the new players. You can only find what you can search and recognize.

Robert J. Boeri (bboeri@world.std.com) and **Martin Hensel** (mhensel@texterity.com) are co-columnists for *Information Insider*. Boeri is an Information Systems Publishing consultant at a Boston-area insurance company. Hensel is president of Texterity, Inc., a Newton, Massachusetts-based consulting firm that builds SGML-based editorial and production systems for publishers, corporations, ecommerce services, and type-setters.

Comments? Email us at letters@onlineinc.com.



[\[EMedia Home\]](#) [\[Subscriptions\]](#)

Copyright © 2000, [Online Inc.](#) All rights reserved.
info@onlineinc.com