No One Best Search Engine: One-Size Fits All Won't Work

At Enterprise Search East in May 2007 and again at Enterprise Search West in November 2007, I asserted that "enterprise search is dead". I will probably say the same thing in 2008.

The remark underscores the crisis in search and retrieval behind an organization's firewall. The phrase "enterprise search" is misleading but ubiquitous. Dissatisfaction with enterprise search hovers in the 60 percent range, a figure that has been increasing since I started keeping records in 2002. The reason is simple. Key word retrieval is useful some of the time, but most users need suggestions, tools to find previously-accessed documents, or answers to questions like "What's the telephone number of Fujiwara Deliveries?" Most enterprise search systems do a lousy job of delivering immediately useful information. Reviewing a laundry list of results is not helpful. In fact, sticky notes on monitors are proof that an organization's enterprise search system are not cutting it.

I'VE TALLIED MORE THAN 150 VENDORS OF SEARCH SYSTEMS. THEY RANGE FROM STARTUPS LIKE POLYSPOT (WWW.POLYSPOT.COM) IN PARIS TO VENDORS WITH TECHNOLOGY THAT IS A QUARTER CENTURY OLD SUCH AS OPENTEXT'S BRS ENGINE (WWW.OPENTEXT.COM) OR AUTONOMY-VERITY (WWW.AUTONOMY.COM). IN BETWEEN THESE TWO EXTREMES ARE KEY WORD RETRIEVAL VENDORS (DTSEARCH IN BETHESDA, MARYLAND WWW.DTSEARCH.COM), RICH TEXT PROCESSING VENDORS (COGNITION TECHNOLOGIES, INC. IN SANTA MONICA, CALIFORNIA WWW.COGNITIONTECH.COM), FREE SEARCH SYSTEMS (LUCENE AT HTTP://LUCENE.APACHE.ORG), SPECIALIST VENDORS (BLOSSOM SOFTWARE, INC. WWW.BLOSSOM.COM), UTILITY VENDORS (DATA HARMONY IN ALBUQUERQUE, NEW MEXICO WWW.DATAHARMONY.COM), SEMANTIC ENGINES (SIDEREAN IN EL SEGUNDO, CALIFORNIA WWW.SIDEREAN.COM), SUPER-PLATFORM VENDORS (IBM, MICROSOFT, AND ORACLE), AND THE BIG THREE (AUTONOMY PLC, ENDECA INC WWW.ENDECA.COM., AND FAST SEARCH & RETRIEVAL WWW.FASTSEARCH.COM). THE ENTERPRISE SEARCH SPECTRUM IS MORE OVERLOADED THAN WIRELESS BANDWIDTH.

Into this crowded field comes is Google's Appliance. Does Google's enterprise division have a recipe that is superior to the other 150+ vendors? Recent pronouncements from experts, pundits, and industry observers assert that the Google Appliance is a weak sister in the behind-the-firewall search market. In a horse race of 150 stallions, Google's running dead last caught my attention.

The Appliance is a search toaster. It's similar to other search appliance offerings from EPI Thunderstone, Index Engines, and Planet Technologies. Each of these products is designed to be unpacked, plugged in, configured, and made available to users. Each is a hardware device, and the customer is relieved of some of the muss and fuss associated with procuring hardware, installing software, and getting the system up and running. The Appliance vendors offer convenience, but the trade offs for some organizations may be unattractive.

Appliances can be customized, but their purpose is fast deployment of search. A licensee can customize any appliance because the vendors know that some customers will "learn" what their users need after getting the system up and running. Thunderstone offers a wide range of configuration options plus a software development kit, a driven API, and fast indexing. But a novice is likely to spend some time figuring out this system.

Google takes a similar approach. The basic Appliance can be serving queries in less than a day, longer if there are hundreds of millions of documents and hundreds of servers to configure. Index Engines and Planet Technology take a similar approach.

Comparing these Appliance solutions to a hosted service where the customer does nothing except open a port is useful. Hosted services are easier to deploy, faster to get up and running, and less hassle than any on-premises installation.

Comparing these appliance solutions to the box-of-parts delivered, customized, tuned, and deployed is also helpful. The search systems available from such companies as Autonomy, Endeca, Fast Search & Transfer, IBM, Microsoft, Oracle, and many other firms require considerable care and feeding.

Isolating a single vendor's system and identifying it as lacking this or that feature, or suggesting one vendor's approach is better than another's is unhelpful. The reason is that each system does some things well and others things differently. Depending on the specific situation, certain systems may be more or less well suited.

Consider the Google Appliance. Each Appliance can be extended using the Google OneBox API

(http://code.google.com/enterprise/documentation/oneboxguide.html). This API makes it possible for a savvy licensee or an authorized Google partner to make the Appliance perform most of the tricks associated with state-of-the-art search systems. The Appliance can make content available from different systems and repositories in a single interface. Analysts describe this function as federation or federated search. Users don't know this term, but they want available information searchable from one interface. The OneBox API can hook third-party applications into the Appliance and display reports, graphics, and facts pulled from enterprise systems. The Google Appliance can generate maps that "mash up" information from different sources. For licensees who want to add natural language processing to the Appliance, Madrid, Spain based Bitext provides an easy-to-install NLP subsystem.

The Google Appliance, therefore, is not a so-so search engine. The Appliance is simply a hardware device that makes it possible for motivated licensees and resellers to craft access solutions needed by a particular organization's users. In a sense, Google's Appliance is neither a Trojan horse purpose-built to get Google into an organization nor a limited-function search "toaster" appliance.The Google Appliance is a trainable work horse, handling certain functions that enhance key word queries. As surprising as it seems, the Google Appliance is more like the customizable systems from Autonomy, Endeca, and Fast Search & Transfer than like systems that offer few features and cannot be extended. Users want entity extraction, summarization, and assisted navigation or point-and-click interfaces..

What the Google Appliance does well is make it almost painless to deploy search in a very short time. Google's engineers have also made it possible for those who want to customize to build a tailored system. The extensibility comes from the Google OneBox API, and from the increasing array of widgets, third-party integrators, and engineers who are using the Appliance as way to reduce the time required for hardware procurement and configuration in order to free up time to concentrate of delivering services users demand.

Recall that enterprise search is dead. The vendors are offering products that, by definition, leave key word retrieval, unstable systems, and budget-breaking costs behind. Most of the companies offering enterprise search systems will find the market's expectations changing rapidly.

Google—and its 11,500 Appliance installations at the end of 2007—will have to change as well. Search is a work in progress. Giants like Microsoft is providing a software search appliance and including enterprise search in SharePoint and its other server products. Even IBM has teamed with Yahoo and gives away a crippled version of Lucene. Fence-straddler Oracle offers its Secure Enterprise Search (SES 11g) solution and also peddles Google's Appliance. For tough bargainers, most search vendors will throw away the price list and ask "How much do you want to spend?"

The net net is that neither Google nor any other vendor is flawless. Judge for yourself in head-to-head comparisons. Take analysts' observations under advisement. Gather your requirements, collect your facts first hand, then decide. Yes, search is dead. But we're moving beyond search.

http://www.eset.com/