

Information Today

May 2008
Vol. 25 | Issue 5

The Newspaper for Users and Producers of Electronic Information Services | www.infotoday.com



Insider's Perspective

The Search Gap

by

JEROME PESENTI

From MAY 2008 Issue
Volume 25 • Number 5

AUTHOR COPY:
All Rights Reserved



Copyright © 2007
by Information Today, Inc.

This article originally appeared in
Information Today, Inc.
Volume 25, Issue 5, May 2008, Page 20

Information Today is published monthly (combined issue July/August) and is copyrighted © 2008 by Information Today, Inc., 143 Old Marlton Pike, Medford, New Jersey 08055. Registered in U.S. Patent and Trademark Office (ISSN: 8755-6286).

Subscriptions from U.S.: \$82.95; Canada and Mexico: \$109 U.S. funds; outside North America: \$119 U.S. funds; order from Information Today, Inc. at above address.

Information Today is available in full-text format for online searching through the information services of DataStar, Dialog, EBSCO, Factiva, FT Profile, Gale, Intelligence Data, Newsbank, NEXIS, Profound, Reuters, STN, and Westlaw. Full-text and full-page versions of articles are now available through ITI InfoCentral. Individuals can go to www.iti-infocentral.com. *Information Today* is abstracted in the Internet and Personal Computing Abstracts (IPCA) database, which is available through Dialog (File #233) and in OCLC's FirstSearch and EPIC services. It is also available in the Information Science & Technology Abstracts (ISTA) database, which is available through Dialog (File #202) and SilverPlatter Information. Microfilm, microfiche, and article and issue photocopies are available from ProQuest (formerly UMI), Ann Arbor, Michigan, (800) 521-0600.

Authorization to photocopy items herein is granted by Information Today, Inc., provided that the base fee of \$3.50 per copy, plus 50¢ per page is paid directly to Copyright Clearance Center, 222 Rosewood Street, Danvers, MA 01923. For those organizations participating in the Transactional Reporting Service, the fee code is 8755-6286/88 \$2.50 + 50¢ pp.

[Insider's Perspective is a column written by an expert who shares insights about challenges and trends in his or her specific niche in the information industry.]

Do a simple experiment: Ask yourself or a colleague who works in a large organization this question: "How often do you use a web search engine versus a search engine that is 'behind the firewall' (i.e., to



Jerome Pesenti

find information only available inside your organization)?" The two answers are typically very different. In fact, most internet-savvy people use a web search engine every day. The same people rarely or never use a search engine behind the firewall. But maybe they just don't need it?

Then ask a second question: "How do you exchange and/or refer to information inside versus outside the firewall?" If I want to learn about a certain subject on the internet, my first step will not be to call or email the most knowledgeable person in my entourage. It will likely be a web search. Only after doing my homework (my own research), will I interact with people (and often find the right people to interact with). Inside the firewall, the picture is completely different. Finding information on your own is almost impossible, so I will likely call or email one of my colleagues (who may not necessarily be the most knowledgeable), starting a chain of emails, interactions, and disturbances to find the right person.

Similarly, once I find a knowledgeable person, the exchange of information is completely different. On the web, there is an expectation that this information will be found in the common repository. "Just do a search for X" is a popular answer to many forum posts. Inside the firewall, the

standard is, "Email me the document," duplicating the information and making it unnecessary to share it more broadly.

Outside the firewall, search has allowed efficient collaboration and the exchange of valuable information between people with absolutely no prior connection. This allows anybody to mine and leverage collective wisdom and creates an incentive to augment it. Inside the firewall, where interests are much closer and exchanges could be even more valuable, such interactions through search have not happened. People need search, yet they don't use it at all.

A third question is "Why?" Why aren't people using search behind the firewall? This question usually produces one of the following answers:

- "There is no search engine," or, more often than not, people don't know that there is one (or even more than one).
- "Searches return junk," meaning that bad quality makes the search engines useless.
- "I don't know where to search," meaning the information is in too many silos.

How can this search gap be filled? To answer this question, let's go back in history.

Remember 1998? At that time, search was known, but it was a dispensable tool, not a big deal. AltaVista was moving into the portal business because of the lukewarm prospects of search. Enter Google. The service was clean, easy to use, and had an intuitive interface that gave the impression of being able to find the needle in the haystack. Google was game changing. It made

search an indispensable tool of everyday internet life.

In other words, what you need is a "Google behind the firewall." Now, a "Google behind the firewall" is not actually Google behind the firewall. The Google enterprise appliance has been around for many years, but it gives a false sense of solving the issue. Sure, it was better than Verity and the like, but it did not significantly increase the usage of search. The reason? The appliance is designed to grab the easily accessible information, not the critical information. It relies on a static, web-based relevance model, and it relies on only relevance with low emphasis on more advanced user experience. It also does not come with the 1,000 Ph.D.s to constantly increase coverage, relevance, and usability that you see on Google.

So what is a "Google behind the firewall"? First, you obviously need a search engine, but you need it to be good enough to start a virtuous circle. Like Google in 1998, good enough means a leap in search experience: a leap in coverage, findability, and usability.

Coverage

Search needs to aim at being universal. Like Google outside the firewall, you should aim to organize and make accessible through search all information in your organization. You cannot expect your users to figure out where to search before actually doing a search. This is also key to raising expectations. When users start trusting the engine and can't find information, they will request that the information be made accessible and searchable, creating an incentive to share information.

In most organizations, increasing coverage means making secure

content searchable across multiple repositories; secured content is more often than not the most valuable. You'll need a search technology that's able to tackle difficult and diverse connectivity, security, and data normalization issues in a minimum deployment time.

Findability

Findability is not just synonymous with relevance. Behind the firewall, there is no magical page rank or popularity metric. Clustering, metadata navigation, spotlighting, personalization, and source-dependent relevance formulas are key to raising findability across a wide range of information sources to an acceptable level.

Look for search technology that provides these advanced tools out of the box and makes them easy to deploy (this is tied closely to the connec-

tivity issue, as it requires the ability to leverage implicit and explicit meta-data in the different repositories).

Usability

The web has created high standards for usability; enterprise applications need to match and even surpass that standard. Pay attention to little details: breadcrumbs, tabs, default syntax, overall look and feel, and UI personalization.

Look for a search technology that provides a simple, weblike look and feel with well-integrated and well-thought-out advanced features.

Only after resolving coverage, findability, and usability issues will you be able to break the vicious circle—low quality, low usage, low IT priority, low quality—and turn it into a virtuous circle—higher usage, raised user expectations, higher IT focus, and better search.

Jerome Pesenti, chief scientist and co-founder of Vivísimo, acts as the visionary for the company, driving development and delivery of the company's products. He also plays a crucial role in its overall strategic vision and growth. Before Vivísimo, he was a visiting scientist at Carnegie Mellon University's computer science department working on document clustering, data mining, and artificial intelligence. He is a frequent presenter at industry conferences, including Microsoft's U.S. Public CIO Summit, the ICIC 2006 Conference, the Life Sciences Conference, and ASIDIC. He is an alumnus of the École Normale Supérieure in Paris. His academic degrees consist of a B.S. in philosophy from the Sorbonne, an M.S. in cognitive science from the University of Paris IV, and an M.S. and Ph.D. in pure mathematics from the University of Paris-Sud.

Information Today

with an
EYE ON
TOMORROW



- Online Services
- The Internet
- Electronic Publishing
- Online Media
- Copyright



Vice President, Content
Richard Kaser

Editor-in-Chief
Barbara Brynko

Editorial Services Manager—Publications
Deborah Poulson

Copy Editor
Cindy Martine

Assitant Editor
John Carr

Proofreader
Celeste Perterson-Sloss

Contributors Jim Ashling, Jamie Babbitt, Phillip Britt, Steven M. Cohen,

Paula J. Hane, Donald T. Hawkins, Shirley Duglin Kennedy, David King, Thonmas Krivak, Marji McClure, Thomas Pack, Lauree Padgett, Loraine Page, Robin Peek, George H. Pike, Barbara quint, Gib Robbie

Vice President, Graphics and Production
M. Heide Dengler

Senior Graphic Designer
Lisa Boccadutre

Production Assistant
Elizabeth Federici

Ad Trafficking Coordinator
Michael Hardwick

Advertising Director
Michael V. Zarrello

(U.S. and International)
609-654-6266 x132
mzarrello@infotoday.com

Classified Sales
Chuck Fiorello
cfiorello@infotoday.com

Vice President, Marketing & Business Development
Thomas H. Hogan, Jr.

Circulation Manager
Janeen Welsh

President & CEO, Information Today, Inc.
Thomas H. Hogan

Chairman of the Board
Roger Bilboul

Vice President, Administration
John Yersak

Main and Editorial Offices
(609) 654-6266
Fax: (609) 654-4309

Website
www.infotoday.com

 **Information Today, Inc.**
Website: www.infotoday.com
Copyright © 2008 by Information Today, Inc.
All rights reserved.