

---

## **Real Time Search: Where Retrieval and Discovery Collide**

---

Organizations with bloated and expensive business intelligence systems are vulnerable without real time information from inside and outside the institution.

The word “search” means keying 2-3 terms into a search box and browsing a list of results. Most of the results are for textual information. Much of the information in the results list is days, weeks, even months old.

Most organizations and their business intelligence systems are bodily obese. Result: sluggish, slow, and unfit for the rigors of today's business environment. With potentially significant information available from Twitter “tweets” and videos, information used for decision making may be misleading.

Even worse, specialized business intelligence and content processing systems cannot be used without specialized training. Even when a modern business intelligence system is installed and working, the average manager cannot get instant answers to certain questions. One major American credit card company cannot do horizontal analyses of buying trends because the database system crashes when asked to traverse a decade of data.

At the same time, competitors and innovators are perturbing markets. The volume of data and the pace of change have created a need to know what the "social sphere" thinks, does, and knows quickly. Many organizations are rushing to use Facebook, LinkedIn, and customized social networks built with tools such as those available from Ning.com.

But for a business professional looking for current information about a competitor, consumer comments about a product, or emerging trends, a routine Web query will not provide the needed information.

The intersection of traditional business intelligence systems and technology and new systems such as Facebook, Twitter, and Google Wave creates a knowledge gap. This gap is the location of a collision between traditional information retrieval and discovering new information available via newsfeeds, Web log articles, short text messages from Twitter users, and user-generated videos posted to YouTube.com and Vimeo.com, among other sources.

The fat, lethargic organization has two problems. The first is an organizational obesity or what I call "weight"--the sheer mass and friction of systems from traditional enterprise software and established business methods slow down an organization's information reflexes. The result is missed opportunities and increased risk from not having timely information where and when it is required.

Flash back to the age of clipper ships. Today's organizations cannot operate with the methods or schedules of the 19th century. Most business intelligence systems are poorly equipped to deal with real time information flows. Even Google lags in the indexing of real-time information. Run a query for the status of the Apple - AT&T iPhone deal on Bing.com, Google.com, or Yahoo.com. You will not get the same results that you obtain when you run the query on Collecta.com, IceRocket.com, or Scoopler.com.

Ignoring real time information can lead to some challenging situations. In early 2009, the New York Police Department was caught in the boundary between traditional information and real-time information. Tyra Banks, hostess of "America's Next Top Model", scheduled an open try out or cattle call for model hopefuls less than five feet, seven inches. Hopefuls showed up at a posh Central Park South hotel. Using Twitter, the word flashed around Manhattan. A crowd gathered. Traffic was blocked. The New York Police patrol car arrived to find several hundred male and female models pushing, shoving, and squirting hair spray to cut a path through the crowd to the hotel entrance. A number of municipal law enforcement agencies are now monitoring Twitter messages called "tweets".

In boundary conditions, forces collide and can change an environment quickly and without warning. Traditional search and business intelligence systems cannot cope with this type of data flow.

Traditional business intelligence systems assert that these seven figure installations operate in real time. They do not. The reason is that programmers have to configure the systems to produce reports from data that are not in the system. When the traditional business intelligence system acquires a Tweet stream, RSS (really simple syndication) content from Web logs, or outputs from standing queries for public facing Facebook.com content--the traditional systems, like those in use at pharmaceutical companies and some government agencies, choke and fall over.

Savvy information professionals recognize this "intelligence gap". The present economic conditions create an opportunity to rethink the systems and the type of information processed by an organization. But to obtain payoffs from traditional business intelligence and from real time sources, business processes and infrastructure must adapt.

Once an organization has taken steps to become more nimble by losing "weight" in certain information processes, another benefit is that some decision cycles will be accelerated. In order to lose "wait", organizations need systems that integrate external real time information into the organization's own repositories and message traffic. External information about certain people, organizations, and products keeps the organization's information from becoming stale. The ability of an organizations to monitor information in real time can work like the 19<sup>th</sup> century Gold Rush prospector's panning for gold system. The modern system processes the message stream for "nuggets", key pieces of high value information.

What can you do to tackle the boundary condition between traditional search and real-time information discovery?

The first step is to work through several sample queries on the systems identified in the table below. Typical test queries can be “Apple AT&T”, “Microsoft Bing”, and “Michael Jackson”. You can substitute your own queries drawn from today’s headlines. The key is to compare the results among the identified systems.

System	url	Comment
Collecta	<a href="http://www.collecta.com">www.collecta.com</a>	Displays time for each query
Google	<a href="http://www2.sandbox.google.com/">http://www2.sandbox.google.com/</a>	Google's next generation search system
IceRocket	<a href="http://www.icerocket.com">www.icerocket.com</a>	Strong Web log coverage
itpints	<a href="http://www.itpints.com">www.itpints.com</a>	Independent developer. Indexes Tweets by tag
Microsoft Bing	<a href="http://bingtweets.com/">http://bingtweets.com/</a>	Microsoft's index of Twitter postings
Newsflashr	<a href="http://www.newsflashr.com">www.newsflashr.com</a>	Near real time indexing of news from a variety of sources
OneRiot	<a href="http://www.oneriot.com">www.oneriot.com</a>	Received an infusion of \$4.0 million in funding in mid 2009
Scoopler	<a href="http://www.scoopler.com">www.scoopler.com</a>	Identifies hot topics; one click narrowing by content type
Surchur	<a href="http://www.surchur.com">www.surchur.com</a>	Dashboard display of what is hot; search function
Topsy	<a href="http://www.topsy.com">www.topsy.com</a>	Search Tweet postings
Tweetmeme	<a href="http://www.tweetmeme.com">www.tweetmeme.com</a>	Current topics and a Tweet search function

Let me conclude with two suggestions.

Explore real time and work with your information technology team to implement some real-time content into the existing repositories at your organization. If a business intelligence system is up and running, you can extend your test to that system as well. Second, the future of real time is in creating applications that eliminate the need to search. You can navigate to the public demonstration of the ArnoldIT.com Overflight service at <http://www.arnoldit.com/overflight>.) Click on "Products". You see the most recent Google news from Google. This is an example of a simple “always on” dashboard of information. Real-time content makes the boundary between traditional information retrieval and the data flows outside an organization an exciting and challenge space filled with opportunities.

Stephen Arnold, ArnoldIT.com

[www.arnoldit.com/sitemap.html](http://www.arnoldit.com/sitemap.html)

[www.arnoldit.com/wordpress](http://www.arnoldit.com/wordpress)

Mr. Arnold is an independent consultant residing in Harrod's Creek, Kentucky. He is the author of more than 10 monographs. His most recent is Google: The Digital Gutenberg, available from [www.infonortics.com](http://www.infonortics.com). He writes monthly columns for KMWorld, Information World Review, and the Smart Business Network.

STEPHEN E. ARNOLD