

Real-Time Search and Discovery of the Social Web



By David Spark
Founder, Spark Media Solutions
December 7th, 2009

Twitter: [@dspark](#)

Updated: December 10th, 2009

Table of contents

Executive summary	3
Real-time interaction on the Internet	4
Traditional search engines and the real-time web	
Publishing for a few, implications for all	
The Michael Jackson wake-up call	
Looking for real-time answers	
What real-time search delivers	
The technologies and behaviors behind real-time search	9
The APIs	
The protocols	
The function of a real-time search engine	
Market dynamics and ecosystem	11
Indexing real-time data: explicit vs. implicit	
Twitter ecosystem	
Traditional search players	
Real-time search startups	
Web users rely on each other for a successful real-time web	
Industry economics and business models	16
Emerging business models for real-time search	
What should we expect from real-time search	18
Endnotes	19
Author	20

Executive summary

Each day, users of the social web post more than 230 million pieces of content. They're sharing all types of published media — videos, pictures, tweets, Facebook status updates, blog posts, comments, reviews, and more — with their immediate community and the greater social web. Given the short lifespan of much of this information, traditional search engines have historically ignored publishing a significant portion of this content. This gap in search engine reporting has opened up new opportunities for real-time search and discovery engines. On a minute-by-minute or even second-by-second basis, these new search engines digest and make sense of what people are talking about across the social web. But their autonomy from the traditional search engine players is about to dissipate. The real-time search landscape is changing as both Google and Microsoft's Bing form real-time search partnerships, expanding their definition of searchable content to include the publication of dynamic social web content.

People's use of search has evolved with the growth of social media. It's estimated that as much as 40 percent of all searches have a real-time component. That means new competitors who aim to serve this market by providing answers specifically for real-time queries could take away up to \$40 billion dollars from traditional search engines. That's just the beginning revenue potential as real-time search can deliver additional money making services such as real-time conversations, discovery, programming, and trend tracking.

The future of the real-time search market is being shaped by emerging technologies, new players, the evolving incumbent search players, and the changing ways web users interact with each other.

NOTE: This report focuses mainly on user-generated real-time content. It purposely omits machine-to-machine real-time content and long explanations of location-based information.

Real-time interaction on the Internet

Traditional search engines and the real-time web

Google has proven to everyone that there's money in search. Competitors Microsoft Bing, Yahoo (now partnered with Microsoft) and Ask have all made strong efforts to steal even the smallest market share away from Google. The fight has been worth it because it's estimated that capturing a single percent of the search market can increase market capitalization by \$1 billion¹. For years they've been fighting over the same pie of long-term indexed content. They haven't been paying attention to the ever widening and shifting pie that is the real-time web.

Traditional search's foundation is built on durability over time. Google and its traditional search competitors operate like a well-catalogued encyclopedia, indexing data and its popularity over time to find content which proves to be the most valuable. Only information that's been culled by a community of other long-term web publishers appears at the top of search engine results. However, much of the content being created via social media tools such as Facebook and Twitter is being created for a specific moment, rather than its long-term archival value. Status updates and tweets are published to be consumed as they're being created. A single tweet's value can come and go in less than a minute.

Therefore, real-time web content's lifespan is antithetical to a traditional search engine's algorithms. Since content doesn't appear or get referenced in the ways traditional search engines discover and evaluate content, it has largely been ignored. Prior to its deal with Twitter, Google said it had been indexing real-time content such as Twitter for years. Yet it also admits it returns very few results from Twitter and doesn't offer the same experience as real-time search engines do.

Real-time interaction on the Internet

Publishing for a few, implications for all

Each day the real-time web generates more than 230 million pieces of contentⁱⁱ. That includes everything from tweets, Facebook status updates, bookmarks, article recommendations, blog posts, comments, videos, pictures, and more. Most of that content is being produced for an already connected social networking audience such as online friends, followers, or fans.

While real-time content may initially be created for an existing audience, when it's aggregated and analyzed, it has larger implications for everyone. Real-time search engines can uncover revealing in-the-moment trends, watch stories unfold, and connect users across the world as they experience and comment on a single event.

Real-time web content is being created by users that were traditionally seen only as search consumers, but their reactions, now published, have value. Any topic or issue that is evolving (or otherwise timely) benefits from real-time search. Examples include:

- Concerned citizens tracking tragedies or political unrest, such as the Iran election protests.
- MTV fans collectively commenting on Kanye West's interruption at the MTV Video Awards.
- Obsessive football fans getting up-to-the-minute results on their fantasy teams.
- Companies providing customer service at the moment people express interest or frustration with their product.

During any hot news story, consumers have repeatedly proven their hunger for up-to-date information. That behavior has been expressed repeatedly through real-time web searches. A single user can enter the same search query as many as ten times a day, according to real-time search sources.

Our behavior is changing as we're seeing what the real-time web can provide. People are getting used to watching stories unfold on Twitter. "In the age of real time, Twitter is Walter Cronkite," argued Techcrunch. News organizations are even using the real-time web as a resource to track and source information. Cable news networks are constantly struggling to find stories and images to fill up hours of programming. Especially during a crisis, viewers take comfort in the constant availability of information, sometimes to the sacrifice of its quality. "Emotions are high during a crisis and people find comfort in information — regardless of the quality of that information," said Keri Stephens, assistant professor of communication studies at The University of Texas at Austin.

Real-time interaction on the Internet

The Michael Jackson wake-up call

During live events with unpredictable stories that unfold over time — disasters and sporting events, for example — the demand for real-time information skyrockets. People are eager for up to date information. The most notorious example was the news of Michael Jackson's death in June, 2009.

As the story broke, there were so many queries for "Michael Jackson" that for a full [25 minutes Google News mistook the surge for an automated attack](#) and some people received a "We're Sorry" page for their results. It is believed that Google's response was the result of an automatic trigger set to go off when there's such a dramatic surge in searches on a single term. While some people may have seen the TMZ or Perez Hilton articles that broke the story, Google failed to reveal the public chatter users were looking for. Google News indexes a select number of sources which includes TMZ, but not the entire blogosphere. Still, even during this dramatic episode, whether they caught the news stories or not, many saw at the top of the page, the same two results as they do today: michaeljackson.com and Jackson's Wikipedia page.

As bits of information on the real-time web appeared, the story of Michael Jackson's death started to unfold, most notably on blogs and via Twitter and Facebook status updates. The story wasn't being explained in a pre-digested methodical standard published format, but rather via a stream of community published and distributed information. Over time, the story revealed itself as real-time web users shared and pointed to information gleaned from other real-time web users.

Many web users believe that Google can provide the best information for search needs, but when it comes to real-time information generated and distributed via the social web, that assumption was wrong. The response to Jackson's death demonstrated that by ignoring the real-time web, Google and its search engine competitors were providing a poor experience to a real-time query.

Real-time interaction on the Internet

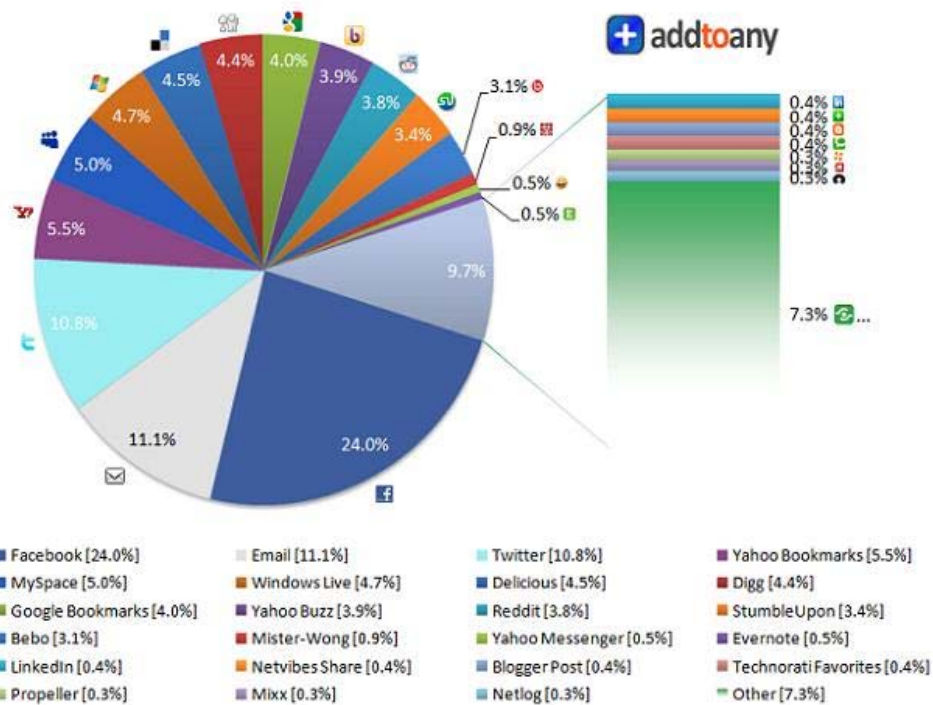
Looking for real-time answers

Those users conducting searches for personalities currently in the news are inevitably trying to find the latest information. Wanting real-time results is a common search behavior, and it comprises an estimated 15-40 percent of all searchesⁱⁱⁱ. In addition to consumer interest in trends, breaking news and the latest meme, businesses are using real-time search engines, also sometimes known as social media monitoring tools. They're looking for greater insight into public sentiment about their products and operations and want to be alerted to relevant industry conversations.

Google has prided itself on incorporating different kinds of media into its search results, as it has started to fold video, image, and blog searches into its overall universal search. But when it came to the real-time web, Google, Yahoo, and Microsoft Bing were all struggling to satisfy users' demand for real-time user generated information.

As of October 2009, however, both Google and Microsoft have made search deals with Twitter. Microsoft also has a search deal with Facebook, in which it is an investor. The details of these search deals are becoming evident now. On the date this report was published, Google announced its rollout of real-time search incorporated into general search with results from social services such as Twitter, blogs, MySpace, FriendFeed, and even Facebook, Techcrunch reported.

Even with this announcement, two of the biggest real-time content providers, Twitter and Facebook, only comprise a small portion of the real-time content being shared on the social web. According to social media data shared via the [AddToAny](#) widget, Twitter only comprises 10.8 percent of the user-generated real-time web and Facebook 24 percent^{iv}, indicating that there's still plenty more real-time content to index.



Real-time interaction on the Internet

What real-time search delivers

Real-time search engines have multiple ways of collecting and indexing data as well as different approaches to displaying results, which often represent one or a combination of the following two methods:

1. **The fire hose** - Results just stream in as the content is being produced and indexed from a particular source or set of sources.
2. **Relevancy now** - When multiple people are currently linking to a published story, video, or other, indexing this sharing via social tools such as Twitter and Digg can show current and collective interest. Relevancy gives rise to discovery, allowing searchers to see what's hot now whether universally, within a specific locale, or within a specific topic.

Traditional search engines are most concerned with relevance over time, but real-time search engines must factor in both recency and popularity. Finding the appropriate mix to calculate spikes and changes in public interest of terms, topics, and stories is an ongoing effort for real-time search engines.

The technologies and behaviors behind real-time search

One of the initial core issues for traditional search engines has been the time it takes to index newly published information. Ten years ago, a recently published piece of web content could take weeks to get indexed. The problem stemmed from the constant need to poll all web sites to see if there was any new information to index and wait for a response. This need for two-way communication for site updates, even if nothing had changed, was horribly inefficient.

About a year ago, social media sites realized that, given the volume of content being created, they would have to gather this information more quickly than with traditional polling. They began using long polling which still sent out a request, but the receiver would just sit and wait and only respond if content had changed, thereby eliminating “we have nothing” responses. This dramatically increased the speed with which content could be indexed.

The APIs

Twitter Search API — Of all the real-time search tools, Twitter has the best combination of valuable real-time dialogue, open access to indexing, and a critical mass of usage. This entire discussion of real-time search can be traced back to the introduction of Summize, an early popular real-time Twitter search engine that was so good that Twitter bought it and renamed it as Twitter Search.

Even though, by one estimation, Twitter comprises less than 11 percent of the user-generated real-time web, indexing the tweets and the sites they point to can provide seemingly accurate real-time search results. Some real-time search engines are fully based on this API and just index Twitter conversations.

Today, real-time search tools are based on utilizing a variety of different APIs (application program interfaces) and technologies that allow them to gather information very rapidly, sometimes as fast as two to three times a second. Below is a review of the technologies, and how they’re being used.

Facebook’s API — Creating similar and more dynamic content than Twitter, Facebook has been developing its own real-time search engine and will soon be offering results for Microsoft Bing. After that, Facebook’s API is closed for outside search engines. Facebook application developers can scan user content only if each individual Facebook user approves it. In addition, not all of Facebook’s content is open, even within Facebook. Users have control over what content they want indexed and viewable by the public and their friends.

The technologies and behaviors behind real-time search

The protocols

Simple Update Protocol (SUP) - A "ping feed" that any web service, usually some type of publishing application, can use to alert subscribers of a feed update. It was developed by Friendfeed, a service that alerts subscribers on many different kinds of content feeds. Simple and compact, SUP eliminates frequent polling, thereby making it far more efficient.

PubSubHubBub and RSSCloud - Blogging protocols for distributing content in real time. The first, developed by Google, is an extension of the ATOM syndication feed. The second is an element of RSS 2.0 and it's being used by WordPress.com. Rather than sending updates directly from publisher to subscriber, these protocols leverage a "hub" that manages the complexities of distribution, taking the workload away from the publishers. This can cut the time needed to deliver content from blog publisher to RSS reader to just seconds, from what used to take 15-60 minutes using a direct publisher to subscriber interface.

Extensible Messaging and Presence Protocol (XMPP) - This is not a feed management protocol, but rather a 10-year-old series of open XML technologies for presence and real-time communication, such as instant messaging, VoIP, machine-to-machine communications, and now social networking. Some real-time search engines, such as Collecta, are utilizing the historical breadth of real-time communications tools built on top of XMPP to create their real-time search engines.

The function of a real-time search engine

Real-time search engines perform the following functions:

Surfacing - As the seemingly endless stream of data comes in, the real-time search engine indexes and finds trends, bringing them to our attention, dynamically. This is sometimes not a function of "search" but rather of "discovery."

Rank real-time results - Algorithm that integrates a variety of factors: freshness of content, authority of domain that published content, authority of people recommending the content, and the increasing volume of people recommending the content.

Manage volume and spam - Not only do real-time search engines need to collect information in real time, they also need to get rid of garbage in real time.

Market dynamics and ecosystem

Indexing real-time data: explicit vs. implicit

The major gatekeepers of the explicit real-time web are currently Twitter, which is open, and Facebook, which is closed. The explicit real-time web refers to content that users choose to publish publicly or for their select community. Content can be either a self-contained original statement or it can point to some other web content (e.g. “check out this blog post <http://tinyurl.com/...>”).

The implicit web refers to what people don’t expressly publish, yet is very measurable behavior. Traditional search engines index the implicit web, monitoring user activity such as what users search for and what links they click.

The overwhelming majority of real-time search engines only index the explicit web. This approach has its limitations. In particular, the only content that is seen and relevant is what people publish and others choose to share. Harvard Business Review estimates that 90 percent of the real-time web is being published by just 10 percent of users^v. On Twitter, only 5 percent of the users create 74 percent of the content^{vi}. Depending how you look at it, those sub-10 percent users are either a representative sample or highly skewing the results.

The only way to gauge broader user activity and interest is to measure online behavior, or the implicit web. Searching the implicit web requires tons of data, and Twitter search is probably the only real-time search engine that’s got enough activity to track passive behavior, currently. However, Twitter Search doesn’t seem to be factoring any implicit behavior into its results.

Conversely, OneRiot, which doesn’t get nearly as many queries as Twitter search, is still indexing implicit behavior through search partnerships and offering a free toolbar. They currently have partnerships with 80 search engines, such as [Zibb](#) and [Scour](#), plus users can install a toolbar that monitors their online behavior anonymously, and in return they get better real-time insights into their interests.

Using its search engine, its toolbar, and data from its partners’ search engines, OneRiot tracks the links users click on and their “dwell time.” For example, if a user clicks on a link and immediately bounces back, spending no time, then that’s a signal that the link may not be valid. Conversely, if a user clicks on a six-minute video and they watch for five and a half minutes, then it’s probably valuable.

OneRiot’s search partner network has increased its branding and opened up the number of queries it is able to ingest and analyze — from 5 million in June without search partners to now more than 40 million monthly searches across its entire partner network, which includes the 3 million registered opt-in users of its toolbar.

For implicit search to be valuable, you need feedback. More importantly, you need a lot of feedback. Achieving that critical mass was only possible when OneRiot syndicated its search to its partner network thus enabling them to improve the quality of search results more quickly. Explicit search alone doesn’t give you a full picture. In addition, their ad inventory jumped dramatically which allowed them to work with large advertisers and that resulted in increased revenue.

(cont’d)

Market dynamics and ecosystem

Indexing real-time data: explicit vs. implicit (cont'd)

Newly launched real-time search engine [Wowd](#) also has an elective add-on so it can track implicit search behavior through its own peer-to-peer network. However, given Wowd's limited audience and lack of a search partner network, it simply doesn't have enough traffic to create valuable implicit real-time search data. In initial tests, head to head, OneRiot's results are more timely and accurate than Wowd.

Adding implicit behavior is necessary for valuable real-time search results as it can show interest in the overwhelming cases when people choose not to publish their interest over the real-time web.

Twitter ecosystem

For years, "YouTube" was synonymous with "online video." Three years ago if you asked someone where they watch video online, they'd probably say YouTube. Today, YouTube is still dominant, but it's got well over 1,000 competitors, and it no longer owns the term "online video." Ironically, a search for "video" or "online video" on Google (YouTube's owner), places YouTube in the middle of the first page of results. Other brands, such as Hulu and Metacafe, have emerged, taking a bite out of YouTube's online video dominance.

Like YouTube did three years ago, "Twitter" is currently enjoying that public mindshare of being synonymous with "real-time web." Facebook has plenty of real-time conversation as well, but other than itself and Microsoft Bing it's not open for full public digestion like Twitter.

By opening its data through an open API, Twitter has become a catalyst for the real-time web. Its content and intellectual property is making this discussion possible. An entire ecosystem of Twitter applications and search engines have been built

completely using the Twitter APIs. Many have argued that Twitter could make money if it licensed its API. If Twitter did charge for its API, however, the sizeable ecosystem of applications using it probably wouldn't have developed.

Most real-time search competitors are simply digesting mostly Twitter data in easy to consume chunks, or making the content more discoverable. There are two standouts, however. OneRiot and Collecta have their own intellectual property and established relationships that position them ahead of other players in the market. OneRiot has relationships with browsers to distribute its toolbar, relationships with other search engines, and it's measuring both explicit and implicit real-time behavior. Collecta's engine is using the real-time communications tool XMPP to constantly retrieve search queries and present them in a very fast viewable stream of real-time programming. Its true market value is yet to be seen. That's why Collecta has opened up its API up to developers in hopes that one will create a killer application to raise awareness of the value of its search data.

Market dynamics and ecosystem

Traditional search players

With the introduction in August 2009 of a new indexing methodology, code named "Caffeine," Google has dramatically increased the speed at which it indexes web content. There's now an option to see latest results in the past hour. Unsanctioned by Google, an outside developer introduced a [hack](#) that can show search results within the past second. Such results indicated that Google was poised to make its own play in the real-time web, beyond its partnership deal with Twitter.

As mentioned earlier, Microsoft's Bing has partnership deals with Facebook and Twitter, and at the time of writing this report, rumors have started circling that Yahoo will form a partnership deal with OneRiot.

These initial real-time search partnerships appear to be Band-Aid solutions, however, aimed at stemming the flow of people from traditional search engines to real-time search engines. In early implementations, we're seeing real-time results branded within the display of the traditional search engine's results. This is unlike all previous specialized search integration, which have been integrated into overall search results.

Real-time search startups

Real-time search is a true competitor to incumbent search players, in that it can answer questions traditional search algorithms cannot. Because of incidents like Michael Jackson's death, traditional search engines have realized that they must begin returning real-time results or they will lose traffic, and subsequently revenue, to real-time search engines.

Like specialized searches that have come before (e.g. video, pictures, blogs), custom search companies have been the first to tackle the unique search problem. There are dozens of real-time search engines that simply index Twitter content. Below, we highlight select companies that either have a significant presence in real-time search or are providing a unique service.

[almost.at](#) - Focused on tracking discussion in real time around specific live events. Tracks tweets, tweets linking to blog posts, and photos. Unique feature is it records live discussions like a DVR. Want to know what everyone was saying exactly at half-time? You can "rewind" to that point and see.

[TweetMeme](#) - Excellent branding as it's slowly becoming the ubiquitous tool to use on blogs posts. Theory is more tweets for a single piece of content indicates more people are interested in it. A single click of their retweet button automatically launches Twitter with the content of the post and a shortened URL already filled in. Search engine only looks at time stamps and number of times a specific article or post is retweeted. Each retweet is equally weighted.

(cont'd)

Market dynamics and ecosystem

Real-time search startups (cont'd)

OneRiot - Combines both explicit (what people are posting) and implicit (what people click on and consume) real-time web behavior. Also tracking Digg, Delicious, and YouTube. Branded algorithm PulseRank is a real-time version of Google's PageRank.

Topsy - Indexes only Twitter. Ranks results by timestamp and tweeter's influence. For example, a single retweet from Digg founder and popular tweeter Kevin Rose is worth 341 retweets from the least influential tweeter, said CEO Vipul Prakash.

Twingly - About.com for the real-time web. Users pick a topic they want to curate and create the ongoing searches on that topic. Stories are revealed based on those searches and people can add comments to the stories.

Twendz - Fire hose results of Twitter search, but attempts to show sentiment (via color codes) and reveals trending subtopics to your search. Information is constantly being refreshed and updated.

YourVersion - A discovery engine based on individuals' unique interests. Set up a profile with all collected interests and YourVersion delivers updates. Very impressive UI makes it easy to navigate and share.

Collecta - Combining real-time search with real-time communications to deliver a constant stream of fast results. Appears like a channel of programming. No aggregation or digesting. Saves queries allowing users to "change channels" between streams of live web programming.

Scooper - A hybrid between Collecta and OneRiot. Offers a channel of streaming programming, but not as fast. Delayed 45 seconds to one minute. Scans multiple additional feeds such as YouTube, blogs, Twitpic, Digg, Delicious, and Flickr. Aggregates popular results and includes discovery by maintaining a top ten list of hot topics.

CrowdEye - A Twitter-only aggregator. Search results provide related terms allowing for deeper discovery on search query. Simple and nice interface for search and discovery of hot news stories.

Twazzup - Very similar to CrowdEye and Scooper. Twitter-only aggregator with an updated stream of content. Provides not only related terms but also tweeters that are active on the subject.

Surchur - A discovery engine that tracks the existing trend reporting on Twitter, Google, Yahoo, CNN, Technorati, and Bing.

Instant - Visually stunning real-time search aggregator. Only indexing Twitter, this program creates a snapshot view of online discussion and tries to determine a sentiment score. Are people talking positively or negatively about the subject? Sentiment engine is far from perfect admits CEO Joe Langevin, as it is difficult to identify sarcasm. Instant has plans for a freemium business model. They'll offer the attractive graphical snapshot for free and then charge for detailed analytics.

(cont'd)

Market dynamics and ecosystem

Real-time search startups (cont'd)

We'll definitely see some level of consolidation among real-time search engines as well as some key acquisitions by larger players. Real-time search startups best hope is to impress the search incumbents with their offerings whether it be breadth of indexing, quality of results, usage base, usability, or new discoverable content.

Impressing Google and Microsoft may not be strictly necessary as real-time search delivers a variety of different experiences not normally seen in traditional search. Most notably those new experiences include discovery and a channel of programming content that will be both competitive and compatible with traditional media

programming. Like TV, radio, print, and online, real-time web programming can be produced (a.k.a. curated) and provide a source of news and entertainment for which advertising can be sold against it.

While it's estimated that real-time search could constitute as much as \$40 billion of the traditional search market, it could also represent a lot more new revenue, as real-time search opens up new avenues for search behavior and related monetization that haven't been appropriate for the traditional search market.

Web users rely on each other for a successful real-time web

The real-time web is more than just content to be indexed and searched against. The content that's being provided is revealing sentiments, ideas, and opinions of anyone connected to the Internet. With real-time search, the traditional behavior of using a search engine to find what you want then leaving is changing. We're seeing the following search behavior changes with real-time search.

Real-time search is becoming a programming channel — Initially, real time search has been an auxiliary channel adding more value to an existing experience (e.g. "I'm watching 'American Idol' right now, what are other people saying about it?"). More recently, it's become a primary channel of content, especially at conferences where there is no other programming (e.g. a TV show) and only those attending can create the content.

Real-time searchers are entering the same query multiple times a day - This behavior doesn't usually happen with traditional search engines. A repeated query shows a tendency to want to track certain subjects. Recognizing this behavior, Collecta saves searches so users can click and change channels between their queries, allowing them to switch back and forth to see what's changed.

Because of these emerging behaviors, real-time search will affect many industries including major brands, live events such as sports and entertainment, news, and marketing. Real-time search affects many industries beyond just traditional search.

Industry economics and business models

Emerging business models for real-time search

Expect to see several new business models evolve from real-time search. Some are already in their nascent stages of development.

Paid search – Advertising is the core revenue source for traditional search engines. For marketers, real-time search adds another level of relevancy, which is time, and in some cases, location. It's an even more qualified version of search engine marketing (SEM). If users are searching multiple times a day on the same query, advertisers will have many times in a given day to market to them. They can use that knowledge to build a story and alter their messaging throughout the day.

Licensing APIs - A possible argument for how Twitter could make money. Right now its API is free. Could that possibly move into a paid model? Many technologies may lose audience interest if they force payments on developers. Some search engines, such as Collecta, have already shown intention of licensing their API as a revenue source.

Sponsoring discussion around live events - By tracking hashtags on Twitter, we're already seeing people using a live event's real-time content as a content channel. It will be even more customized like what we've already seen with The Huffington Post and their real-time reporting around The World Series. They created a content page that included real-time programming from YouTube, Twitter hashtag searches, and tweets from a select group of sports reporters. Just like event producers get sponsors to put up signage or a booth at a sporting event or conference. The same type of sponsorship opportunities are also possible for real-time programming.

Curated live searches - Real-time search can help improve the discovery of current discussions, but like with any programming there will be a need to have that content produced by an actual human that has an editorial eye. For example, a 24-hour webcam pointed at the street streaming video to the web will get some traffic, but once a producer looks at that video and edits it down to a shorter program it has more value, attracting a larger audience. The same is true with real-time search results. Turning on a fire hose to spit out real-time search results can be interesting. Add a human editorial production layer on those search results can create exponentially more value. We've seen examples of this already with blog posts touting "best tweets" on a given subject or event (e.g. ["The 10 best tweets \(so far\) from Microsoft's MVP Summit"](#)).

This is another step beyond just sponsoring a discussion around a live event. If managed correctly, a human-powered curated live search becomes another programming channel against which to sell advertising.

(cont'd)

Industry economics and business models

Emerging business models for real-time search (cont'd)

Syndicated content feeds - Once a real-time search is created, it can then be syndicated to multiple locations. Producers can add value layering real-time searches with human-powered curation and a custom interactive display of results. Brands and sites will pay for this content to bring traffic to their site, especially during live events.

Customer service and CRM - Currently, [more than 100 companies offer some type of social media monitoring solution](#). All have some type of real-time search component. Some are free, while some require a monthly subscription that businesses happily pay just to know what people are saying about them. Most just provide alerts and summarized reports. Others are vertically integrated with real-time communications tools (e.g. [TweetDeck](#), [PeopleBrowsr](#), [Seesmic](#)) allowing users to respond back. Some tools (e.g. [CoTweet](#), [HootSuite](#)) allow companies to manage responses across a team of people. When marketers see a real-time discussion, they can make a timely response. To improve the customer service experience, expect to see a consolidation of these three types of applications: real-time search engines, social media monitoring, real-time communications tools.

What should we expect from real-time search

1. Real-time search engines that are only indexing Twitter (not including Twitter search) will have to expand their offerings to scan other real-time content and to include implicit behavior.
2. Real-time search will soon be seen as a valid programming channel that will include human-powered search curators that will produce more valuable real-time productions.
3. Consumers and marketers are slowly realizing the value of using real-time search as a customer relations management tool. Why pick up the phone and complain to Comcast, when you can just send out a public tweet complaining? They'll call you. It's already happening. To facilitate the process, we'll see a consolidation of the tools being used for monitoring and interaction (e.g. Radian6, Tweetdeck, CoTweet) with the real-time web.
4. Search advertising will be a big boon, as marketers will be able to send multiple targeted messages in a given day.

As for all the parties involved, here are some recommendations:

Investors - Avoid investing in real-time search engines that are completely based on Twitter's API. You don't want to invest in a company that's fully based on a startup. Instant is a search engine built on Twitter's API. CEO Joe Langevin admits that his biggest fear is Twitter releasing its own killer application. Act quickly, sources fear that real-time search could become a commodity in three to five years.

Startups - Just being able to digest information in a cool way probably won't be enough. Only one company that's doing it the best, possibly Scoopler, will succeed. The real winners will be presenting a new way to index information (e.g. OneRiot), will be offering another experience like a real-time programming channel (e.g. Collecta), or will offer an application that takes advantage of real-time information in a unique and valuable way.

Incumbents - Through their partnerships with Twitter and Facebook, Microsoft and Google have made a good first step that will save them from initially losing the 15-40 percent of people seeking real-time results. But they'll need to expand their concept of search to include discovery and content programming.

Gatekeepers - Facebook will reveal a real-time search tool that will invite more content and activity into its walled garden. Since they have a critical mass of activity, they won't need to make any other partnerships or bring in other real-time activity. Twitter may be fine making a deal with Microsoft and Google as people will be satisfied with those results and not bother to jump to other "real-time only" search engines.

This is just the beginning. Search is a learned behavior. Ultimately, how people will interact with real-time search will be a symbiotic education for the real-time search engine creators and the consumers. User behavior will evolve and will craft the presentation, application, and redefinition of real-time search. We expect that with the expanded definition of real-time search, we may start referring to it as something else such as real-time programming.

Endnotes

ⁱ ["Why 1% of search market share is worth over \\$1 Billion"](#) by Don Dodge (2007).

ⁱⁱ Combination of content from [GigaTweet](#) and [AddToAny's report as reported by Mashable](#) on distribution of real-time content for its web sharing tool.

ⁱⁱⁱ Analysis extrapolated by Tobias Peggs, CEO of OneRiot and Gerry Campbell, CEO of Collecta from three research reports on the study of intent of when people use a search engine: ["A Taxonomy of Web Search"](#) by Andrei Broder, 2002, ["Understanding User Goals in Web Search"](#) by Rose and Levinson (2004), and ["Determining the informational, navigational, and transactional intent of Web queries"](#) by Jim Jansen at Penn State (2007).

^{iv} ["Sharing on Facebook Now More Popular Than Sharing on Email"](#) by Adam Ostrow of Mashable, sourcing AddToAny (July, 2009)

^v ["New Twitter Research: Men Follow Men and Nobody Tweets"](#) by Bill Heil and Mikolaj Piskorski (June, 2009)

^{vi} ["Do You Know Who's on Twitter?"](#) by eMarketer (August, 2009)

Author: David Spark

David Spark is the founder of [Spark Media Solutions](#), a firm that helps companies build industry voice through storytelling and social media. For more than 14 years, Spark has worked as a journalist and producer reporting on the tech industry in print, radio, TV, and online. His articles and advice have appeared in more than 30 media outlets including eWEEK, Wired News, PC Computing, PC World, Smart Computing, and TechTV. He's worked as an analyst for The Guidewire Group and Gerson Lehrman Group.

In addition, Spark squandered more than a dozen years working as a stand-up comedian and comedy writer for Second City comedy troupe in Chicago.

Today, Spark is a regular contributor to Mashable, Socialmedia.biz, Green 960, ABC Radio in NYC, John C. Dvorak's "Cranky Geeks," and KQED's "This Week in Northern California." For more on Spark, follow his blog the ["Spark Minute."](#)



Contact: david@sparkmediasolutions.com

Spark Media Solutions

Spark Media Solutions helps businesses become media networks as an arm of its marketing, sales, and public relations effort. We are a custom publishing organization that provides editorial guidance and production across all forms of media. Our goal is to turn our clients into the number one voice in their respective industries through storytelling and social media. The content and the conversation that surrounds our efforts is designed to position our clients as thought leaders in their respective industry.