

The background of the cover is an abstract, high-contrast blue image. It features several thick, dark blue lines that radiate from the bottom center towards the top corners, creating a sense of depth and movement. The lines are set against a lighter, almost white, background that is slightly blurred, giving the overall effect a dynamic and modern feel.

**The
Social Web Analytics
eBook 2008**

"...The social web will be the most critical marketing environment around.

"...The social web will become the primary center of activity for whatever you do when you shop, plan, learn, or communicate. It may not take over your entire life (one hopes), but it will be the first place you turn for news, information, entertainment, diversion."

Larry Weber, Chairman, W2 Group, "Marketing to the Social Web"

"We've been liberated! Before the Web came along, there were only two ways to get noticed: buy expensive advertising or beg the mainstream media to tell your story for you. Now we have a better option: publishing interesting content on the Web that your buyers want to consume.

"The tools of the marketing and PR trade have changed.

"The skills that worked offline to help you buy or beg your way in are the skills of interruption and coercion. Success online comes from thinking like a journalist and a thought leader."

David Meerman Scott, author of The New Rules of Marketing & PR,
for the Social Web Analytics eBook 2008

If you could go back to the mid-90s and offer a marketer a little box that could sit on her desk and let her listen in on thousands of customer conversations and participate in those discussions regardless of geography or time zone, it would appear so far-fetched that she'd probably call security. This eBook is about that reality.

ACCOMPANYING WEBSITE

URL: www.socialwebanalytics.com

AUTHOR INFORMATION

Name: Philip Shelldrake
Email: philip@racepointgroup.com
Email 2: me@philipshelldrake.com
Company: www.racepointgroup.com
Twitter: [@shelldrake](https://twitter.com/shelldrake)
FriendFeed: [/shelldrake](https://friendfeed.com/shelldrake)
Blog: www.philipshelldrake.me.uk

LEGAL INFORMATION

Published: 1st July 2008

License: [Creative Commons Attribution-Noncommercial-Share Alike 2.0 UK: England & Wales License](https://creativecommons.org/licenses/by-nc-sa/2.0/uk/)



TABLE OF CONTENTS

About the Social Web Analytics eBook 2008.....	1
About the Social Web.....	4
The Cluetrain Manifesto.....	5
Marketing to the Social Web.....	5
The need for Social Web Analytics.....	9
'New' PR.....	9
Brand.....	13
Measurement & evaluation.....	14
Market research & new product development.....	17
About Social Web Analytics.....	25
What is SWA?.....	25
What are you looking for in a SWA service?.....	25
Indexing.....	25
Spider capability.....	27
Semantic analysis.....	27
Search query structure.....	29
APIs and libraries.....	30
Infrastructure.....	31
Commercial, licensing and terms of use.....	32
The free tools.....	34
Google, Yahoo!, MSN Live, Ask.....	34
Google Alerts.....	34
Google Trends.....	35
Google Blog Search.....	36
Technorati.....	36
Twingly.....	37
IceRocket.....	37
BlogPulse.....	38
News readers.....	39
Alexa.....	39

Del.icio.us.....	39
Digg.....	40
Summize.....	40
The vendors.....	43
Vendor information and your participation.....	44
Attentio.....	46
Biz360.....	50
Brandimensions.....	52
BuzzLogic.....	54
Cision.....	59
CollectiveIntellect.....	63
CyberAlert.....	69
Cymfony.....	70
DNA13.....	74
Dow Jones.....	77
Integrasco.....	81
Kaava.....	84
Magpie.....	88
Nielsen BuzzMetrics.....	93
Radian6.....	95
Vocus.....	98

About the Social Web Analytics eBook

2008

Technology has revolutionised communications, massively and irrevocably, to the benefit of the consumer, the adaptive and agile organisation, and those who cherish an open society.

This ebook gives a brief overview of the characteristics of the Social Web (also known as Social Media), but that's not its primary purpose.

Rather, I review here how all organisations can try and make the most of the unprecedented wealth of information afforded by the Social Web, the incredible facility to 'listen in' on conversations close to their heart, and to initiate and engage in this dialogue. It has been relatively straight forward for PR professionals to work with a few dozen journalists; it has been a means to an end for advertisers to bludgeon brand values into targets; but today, keeping tabs on thousands of conversations is quite another challenge altogether – two-way dialogue between your stakeholders, and between you and your stakeholders.

This ebook is an introduction to Social Web Analytics (SWA, also known as Social Media Analytics), the driver for it, how it can be applied, the key vendors and their services, and considerations for your organisation's procurement of such services.

I stop short of making recommendations of one vendor or one tool over another however; that's for each reader to investigate equipped with the understanding lent them here and married to their insight into their organisation's specific needs.

Readers of [my blog](#)¹ and our [company blog](#)² will see that I have leaned on the content of past posts in compiling this ebook.

Lastly, but importantly, I urge readers to consider "[A Bill of Rights for Users of the Social Web](#)"³ by Joseph Smarr, the irrepressible Marc Canter, Robert Scoble and Michael Arrington.

¹ <http://www.philipsheldrake.me.uk>

² <http://www.racetalkblog.com>

³ <http://opensocialweb.org>



About the Social Web

Fellow Londoner [Sir Timothy John Berners-Lee](#) put the first website online 6th August 1991⁴, and things have moved pretty fast since then. The first consumer Web revolution took us well into the current decade, embodied by companies such as Yahoo!, AOL, Amazon, eBay, PayPal, Ticketmaster and services such as browser based email and online banking. This was the Transactional Web if you like.

The second phase, the Social Web, is catalysed by the so-called Web 2.0 technologies facilitating easy-to-use, engaging and rewarding online social interaction. It's about self-expression, relationships, user-rating, affiliation, trust and user-created content.

The term Social Web was coined, according to [the Wikipedia entry](#)⁵, in 1998, as both a technological and social term. This duality is apt given that our focus here is on the application of technology to infer social meaning.

Interestingly, although possibly only to some readers so I'll keep it very brief, this ebook broaches upon the semantics of Web content and therefore on the prospects of a [Semantic Web](#)⁶; what some pundits refer to as the third phase of the Web, or Web 3.0 for short.

⁴ http://en.wikipedia.org/wiki/Tim_Berners-Lee

⁵ http://en.wikipedia.org/wiki/Social_web

⁶ http://en.wikipedia.org/wiki/Semantic_Web

The Cluetrain Manifesto

The ramifications for organisations of this Social Web reality were first considered and presented by the authors of the seminal [Cluetrain Manifesto](#)⁷ in 1999.

The Cluetrain Manifesto asserts that the Internet allows markets to revert back to the days when a market was defined by people gathering and talking amongst themselves about buyer reputation, seller reputation, product quality and prices. This was lost for a while as the scale of organisations and markets outstripped the facility for consumers to coalesce. The consumers' conversation is now reignited.

Marketing to the Social Web

In his book "[Marketing to the Social Web: How Digital Customer Communities Build Your Business](#)"⁸, Larry Weber describes the opportunity the Social Web presents organisations. I recommend the book (disclosure – Larry is chairman of my organisation), and he has selected these quotes from his book for this ebook:

"The social web is the online place where people with a common interest can gather to share thoughts, comments, and opinions. It includes social networks such as MySpace, Gather, Friendster, Facebook, BlackPlanet, Eons, LinkedIn, and hundreds more. It includes branded web destinations like Amazon, Netflix, and eBay. It includes enterprise sites such as IBM, Circuit City, Cisco, and Oracle. The social web is a new world of unpaid media created by individuals or enterprises on the web.

⁷ Rick Levine, Christopher Locke, Doc Searls, and David Weinberger

http://en.wikipedia.org/wiki/Cluetrain_Manifesto

⁸ <http://www.marketingtothesocialweb.com>

"...The real job of the marketer in the social web is to aggregate customers. You aggregate customers two ways: (1) by providing compelling content on your website and creating retail environments that customers want to visit and (2) by going out and participating in the public arena.

"...The social web will be the most critical marketing environment around.

"...The social web will become the primary center of activity for whatever you do when you shop, plan, learn, or communicate. It may not take over your entire life (one hopes), but it will be the first place you turn for news, information, entertainment, diversion.

"...Marketing therefore has to wrap around that – because what is truly changing in the social web is media, and marketing has always had to shape itself around media."

I also invited [Brian Solis](#)⁹, Founder and Principal of FutureWorks, as a social media thought leader, to contribute a perspective to put this ebook in context:

"Social Media is no longer an option or debatable. It is critically important to all businesses, without prejudice. It represents a powerful, and additional, channel to first listen to customers, stakeholders, media, bloggers, peers, and other influencers, and in turn, build two-way paths of conversations to them. Yes, conversations are taking place about your company, product, and service, right now, with or without you. This represents priceless opportunities to build relationships and shape perceptions at every step. In the process, you become a resource to the very people looking for leadership, expertise, vision, and also solutions. The

⁹ <http://www.briansolis.com>

most important driver for outbound and proactive online relations is that it's measurable and absolutely tied to the bottom line.

"Much in the same way Web marketers integrate calls-to-action and dedicated splash pages to direct responses, successful conversations can also benefit from strategically carved inbound and interconnected paths that can be tracked and measured. From listening, participation, to analytics, social media creates new opportunities to make deep and meaningful connections, forge relationships, and influence without manipulation. And, in the process, we also earn a place within their network as a trustworthy resource."

Want some numbers? Thanks then to Courtney Hughes at SWA vendor BuzzLogic for pointing me to the forecast from eMarketer¹⁰, May 2008, envisaging that two thirds of the US population will read a blog post at least once a month by 2012. Courtney also alerted me to research by Synovate¹¹, 2007, that found that 65% of people who read blogs do so explicitly to get an opinion.

¹⁰ http://www.emarketer.com/Reports/All/Emarketer_2000494.aspx

¹¹ <http://www.synovate.com/news/article/2007/08/new-study-shows-americans-blogging-behaviour.html>



The need for Social Web Analytics

Each and every organisation will have its own specific motivations for adopting SWA, and I have grouped important drivers here under the headings of:

- 'New' PR
- Brand
- Measurement & Evaluation
- Market Research & New Product Development.

The sections on New PR and Brand will be useful to anyone looking to understand the immediate application of SWA or convey its importance to colleagues and clients. The latter two sections are, however, more advanced and will be most relevant to those who have already dabbled in SWA and are looking to go to the next level, so you might chose to skip these bits.

'New' PR

For many years, PR has been defined as journalist relations – a linear relationship between PRs, journalists and the target audience. The industry became increasingly focused on traditional media as the best, if not sole way, to reach the 'public'.

I believe terms such as 'new' PR or "PR 2.0" simply refer to a reversion to the objective of building a dialogue with all your influencers and audiences, and developing content that helps to earn understanding and support. We apply descriptors such as "new" or "2.0" because a new and large swathe of those channels is now digital.

In much the same way the Cluetrain Manifesto anticipated the return to markets as we once knew them, the Social Web has taken us back to the original definition of PR.

The New Rules of Marketing & PR

In his excellent book "[The New Rules of Marketing & PR](#)"¹², David Meerman Scott lists three uses of blogs for marketing and PR:

- “1. To easily monitor what millions of people are saying about you, the market you sell into, your organization, and its products
2. To participate in those conversations by commenting on other people’s blogs
3. To begin and to shape those conversations by creating and writing your own blog.”

I discuss more reasons and uses for getting involved in social media here, or more precisely SWA, but in this quote David succinctly lists the only reasons you should need!

If anything, this eBook drills down on the first part of this quote, “To easily monitor...”.

There is no doubt that you can enrich your insight into your market and its perspective of your company from your personal use of the free tools described later in this eBook, such as Technorati and Twingly. However, David recognises that the word “easily” suitably describes getting going, but that you may need some assistance to go further.

He writes “Text mining technologies extract content from millions of blogs so you can read what people are saying; in a more sophisticated

¹² <http://www.davidmeermanscott.com/books.htm>

use, they also allow the measurement of trends.” We call those mining technologies Social Web Analytics.

Isn't this just a consumer thing?

The Social Web impacts all marketing communications, business-to-consumer, business-to-business, not-for-profit, government. If being an expert or leader in your market is defined as others' regard for your insight, skills or services, then you must participate in the networks where this expertise is being shared, and where the people you want to influence are going to help shape their viewpoint. For many professions, these networks remain predominantly offline, but this balance will tip in favour of online for most if not all professions eventually.

Example social *professional* networks include [MarCom Professional](#)¹³ for marketing communicators, [sermo.com](#)¹⁴ for physicians, [ArtCloud](#)¹⁵ for the art world, and [inmobile](#)¹⁶ for the wireless industry. And almost all professions have other online media dedicated to them – traditional media's online presence, dedicated news and opinion sites, blogs etc.

Distributed conversations

We cannot, however, hang out in four or five virtual places to gain an insight into the zeitgeist of our markets. The conversations relating to your market, to your products and services, to your campaign, don't neatly happen at a handful of websites. You only have to click around the links branching away from a polemical and popular blog post to see

¹³ <http://www.marcomprofessional.com>

¹⁴ <http://www.sermo.com>

¹⁵ <http://www.artworld.com>

¹⁶ <http://www.inmobile.org>

how quickly the conversation seeps out through the equivalent of the backstreets, coalesces again elsewhere, and then fragments once more.

There's a time dimension here too. It's not uncommon, for example, for regular Web users to receive a viral email (viral means simply containing good content so interesting it compels you to pass it on) months and sometimes years after they first saw it.

Ultimately, the World Wide Web is the biggest social network of them all, and it's way too big a place to hope to secure a thorough understanding of the respect your brand commands, the buzz about your competitors, the expectations for the market going forward, simply by meandering around. As I mentioned above in relation to David's recommendations, a meander is better than simply staying out of it, but it isn't sufficient of itself if you intend to 'get serious' here.

myChannel

The user (aka the recipient of news and information, the listener, the viewer, the inter-actor) has been empowered to set the schedule. It's what they want, when they want it and how they want it. Video on demand. Personal video recorders (PVR). Newsfeeds (RSS). Alerts. Lifestreaming. Podcasts. Web radio. Mobile TV.

To all intents and purposes, we're just a short hop away from everyone having their own customised channel, a channel tailored uniquely from your own subscriptions, your friends' subscriptions and recommendations, and automated "if you like that, you'll like this" discovery.

In my presentation at Internet World 2005 I labelled this eventuality myChannel. One billion connected people equals one billion separate "channels".

The ramifications of myChannel for marketing communicators include:

- Considerably more fragmentation of the target audience of communications campaigns
- Less precise timing of delivery
- Increased opportunity to provide niche information
- Less certainty of how each recipient is receiving the information
- Greater opportunity for innovation in inviting and securing interaction
- The need for new mechanisms for gauging campaign success.

'New' PR is not so difficult to understand, but can be complex to execute. In my January 2008 blog post "[You're in IT](#)"¹⁷, I posit that the communications profession has reached the point of needing information technology to achieve its objectives, in much the same way as many other professions became dependent on IT in previous decades. Part of that IT toolkit is SWA.

Brand

Your stakeholders now collectively define what your brand means, what it stands for, based on their lifelong interactions with your organisation; your services, your products, your people, your partners, your CSR activity, and other stakeholders. You can't *tell* them, you can only make sure your brand values permeate everything you do, continuously, so they end up reaching the conclusion about your brand that you want them to.

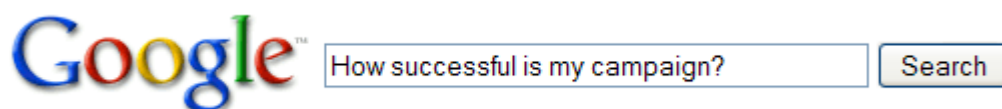
¹⁷ <http://www.marcomprofessional.com/posts/philip.sheldrake/youre-in-it>

And part of the “everything you do” is marketing. Your marketing teams, both in-house and consultancy are converging into a joint influence team, seeking to influence by exercising finely attuned ears and projecting an open, honest and engaging voice. These are the bedrock characteristics of your voice, but you will of course continue to blend in your brand’s particular personality; just so long as you don’t erode that bedrock.

The brand landscape exhibits emergent behaviour (which more or less means it may be unpredictable at times), and I’ve come to call this focus of study Brand Complexity. That’s a subject for a future ebook however, and in this ebook we’re looking at how Social Web Analytics helps to go some way towards serving as that finely attuned ear and acting to inform your voice and, critically, your actions across your whole organisation.

Measurement & evaluation

The evaluation of PR campaign effectiveness is controversial. Forget for a moment the inadequate practitioners that insist all PR must have a benefit so better just get on with it than devote energy to measurement, and you're left with an array of evaluation processes as diverse as the number of agencies.



The idea of return on investment (ROI) is applied casually in marketing, or else politely ignored. For example, when you read the rationale justifying the selection of [the winners of OnMedia’s Best of Broadband Advertising awards 2007](#)¹⁸, only three out of ten make an

¹⁸ <http://alwayson.goingon.com/permalink/post/23421>

attempt to link the campaign to a filip to the client's bottom line. But "Creative" and "ROI" are not synonymous.

Here's a polemic. If our campaigns strive to exert influence, isn't Google search the ultimate measure of the influence achieved, and the change in that influence over time?

A corollary to this is that the ultimate role of a marketing consultancy is content creation and SEO. But before we identify the weaknesses in this claim, here are some supporting points of view:

- With Google spidering most of the offline world (as that content is put online too), all online publications, forums, chat rooms, blogs and social networks, only a search engine can add up the cumulative effect of brand and product mentions and their association with key words, key phrases.
- The search engines strive to deliver the most relevant search results to users. Their algorithmic methods make this measure of relevance more or less equivalent to brand influence and brand momentum.
- The primary objective of PR, of all marketing disciplines, is to inculcate brand loyalty with current customers, influence opinion and behaviour, to establish or correct perceptions, and attract new customers. If the Web is the most important channel to these targets forming an opinion of your brand and products, and if search is the way they get around the Web, then QED search is the ultimate measure of marketing campaign effectiveness.

There are some sweeping assumptions in these claims, and a few flaws in the argument; here are the three primary objections.

The first is that today's search engines ignore sentiment. Northern Rock and UBS came pretty high in search engine results during the

first half of 2008 for some of their associated key words and phrases, but not always for good reasons. Rather, something to do with a credit crunch! Search engines today do not have well developed semantic analysis capability. In other words, they're adept at queries like "Tell me about banks" but less so "Tell me about banks with a good reputation", let alone "Tell me where I should bank".

Secondly, people perform as they are measured. In other words, whilst a particular performance measure may be appropriate in isolation, it should not encourage "gaming" whereby the individual or team concerned becomes persuaded that the objective is simply to score higher whichever way they can rather than achieve a higher score through doing what's 'right'.

Our third objection is connected to my mantra that goes:

The discontented spread their discontent. The neutral say nothing.
The content say nothing. The delighted spread their delight.

Many brands and products spend most of their time in the middle of this spectrum. Consider your own bank for example, or broadband provider, or mobile phone operator. Customers and prospects are mostly either neutral or content and contribute nothing audible and nothing visible for any search engine to stuff into their mathematics. Yet the opinions residing unexpressed in the minds of customers and prospects will exert an influence next time they need to reach a buying decision.

I can't yet envisage a future where the third objection here is shot down, but the potential of the Semantic Web, semantic analysis and interpretation with SWA tools, will have intriguing ramifications for the measurement of marketing campaign effectiveness.

Whilst a controversial measure of marketing and business success, [Net Promoter Score](#)¹⁹ depends today on explicit Q&A with customers; perhaps SWA enables an implicit equivalent approach to the measure. Perhaps, SWA could even help build that illusive ROI formula.

Market research & new product development

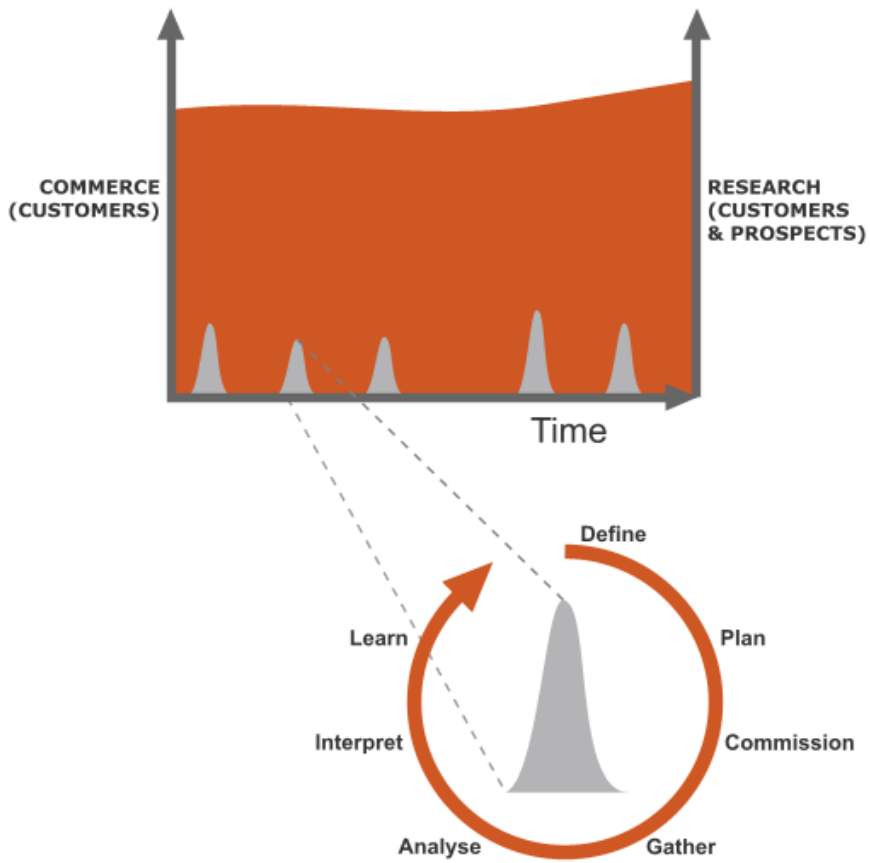
[ESPN pulled the plug on their cell-phone product after investing \\$150m including \\$40m in advertising](#)²⁰. This is precisely the failure market research is intended to prevent. How can business harness customers and prospective customers to improve their hit rate and time to market?

I presented the following simple figures at the market research Insight 2006 show in London to demonstrate the difference between traditional market research and continuous engagement.

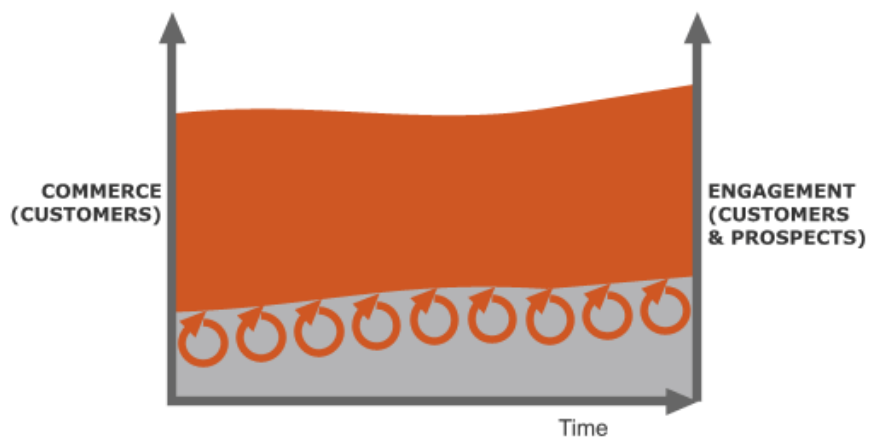
¹⁹ <http://www.netpromoter.com/netpromoter>

²⁰ http://www.businessweek.com/magazine/content/06_44/b4007026.htm

Traditional Market Research



Continuous Engagement Loop



There are dozens of differences between the two approaches. Here's a list of the primary differences:

- Research is ad hoc or regular interval; engagement is continuous
- Research is one-way (and needs the carrot of a prize or payment!); engagement is two-way (mutually rewarding)
- Research is unemotional; engagement is emotional
- Research is independent of loyalty; engagement inculcates brand loyalty
- Research has a tight focus; engagement has a wide focus
- Research deals with sequential parameters; engagement is multi-parametric
- Research is designed to achieve statistical confidence; engagement is designed to detect weak signals.

The disadvantages of traditional market research

Traditional market research is ad hoc or at regular intervals at best. This could mean your last data set is getting on a bit. It could lead you to trying to read between the lines because the last survey didn't ask exactly the question you now need answering. Your market may be speeding up faster than your research frequency. You will probably need to ask new questions, but want to continue trending previous survey data.

Traditional market research is one-way. So what's in it for your respondents? Ever wondered if they're answering your questions conscientiously? Are they likely to benefit or suffer as a consequence of the information they share with you?

Traditional market research is unemotional... so, quite simply, do they care? Ad hoc, one-way, unemotional interaction does not drive brand loyalty.

It has a tight structure, but once you've collected the demographic data, how much time remains to get to the crux of the matter? By what degree can you change the subject? How many times can you change the subject before the respondent's brain starts hurting?

After all that, it's no wonder you need some mathematics to determine statistical confidence.

Research through continuous engagement

Anything and everything is discussed by your customers, prospects and all stakeholders on the Social Web. For each topic, you can choose to interact or just listen.

You can also seed the conversation with topics relevant to your business tomorrow, not just today. Test their reaction. Harvest value-added feedback, qualitative and quantitative.

Traditional research addresses a limited sequence of parameters, whereas you can explore multiple parameters on the Social Web. Your product roadmap may encompass hundreds of parametric permutations, in which case you could choose to present ideas based on "runs" (parametric groupings based on [Taguchi orthogonal arrays](#)²¹) to your most loyal and valued social media participants. You've heard of user-generated-content, well this is consumer-generated-products. (On that note, anyone interested in consumer-created-brands?)

So-called "weak signals", early but faint signs of things to come, are easily overlooked in traditional research as statistically insignificant.

²¹ http://en.wikipedia.org/wiki/Taguchi_methods

But understanding how to identify the most authoritative members in your social media (the *mavens* and *connectors* in the language of Gladwell's "[The Tipping Point](#)"²²), and learning to listen to them, can place you weeks if not months ahead of your competition in timely new product launches.

Moreover, whilst [focus groups](#)²³ aren't a panacea, whether they're held in one room or distributed, it feels right that your chances of coming out with a well-received and successful product are improved. That's, of course, if you subscribe to the conclusions presented by Surowiecki in "[The Wisdom of Crowds](#)"²⁴. Sure, you could ask whether innovations such as the iPod could have been spawned in such a manner, but that misses the point. I'm not advocating this interaction with your key stakeholders courtesy of the Social Web as your sole approach to market research and product development, just that it should be one of your primary approaches.

In conclusion, supplanting or supplementing market research with continuous engagement requires:

- A new strategy
- An implementation framework
- New analyses methods
- Sound corporate performance measurement to close the loop.

²² <http://www.gladwell.com/tippingpoint>

²³ http://en.wikipedia.org/wiki/Focus_group

²⁴ <http://www.randomhouse.com/features/wisdomofcrowds>

For and against

Unsurprisingly, there are advocates and detractors from this point of view. Take [an interview with Bill Neal of SDR Consulting](#)²⁵ for example:

"..But I have some real problems with consumer generated media as a source of credible and reliable information. In many ways it combines the worst elements of non-scientific research – self selection and advocacy – both positive and negative.

"The information they generate may be true, or not true – there is no way to discern which. Therefore, the information generated by those folks is neither credible nor reliable."

However, this perspective could not be more strongly countered by the assertions made in the Cluetrain manifesto:

"A powerful global conversation has begun. Through the Internet, people are discovering and inventing new ways to share relevant knowledge with blinding speed. As a direct result, markets are getting smarter - and getting smarter faster than most companies.

"These markets are conversations. Their members communicate in language that is natural, open, honest, direct, funny and often shocking. Whether explaining or complaining, joking or serious, the human voice is unmistakably genuine. It can't be faked."

I also recommend an article on a related topic, "[Online Polls: How Good Are They?](#)"²⁶, Business Week 16th June 2008, which debates the

²⁵ http://bloombergmarketing.blogs.com/bloomberg_marketing/2006/06/savvy_divas_and.html

²⁶ http://www.businessweek.com/magazine/content/08_24/b4088086641658.htm

pros and cons of interacting with an online community for this specific purpose, and what to look out for if you do pursue this approach.

OK, continuous engagement on the Social Web, enabled by SWA, may not signal the death of traditional market research, but it marks a distinct and influential turning point; a turning point leading companies will adopt to their competitive advantage.



About Social Web Analytics

What is SWA?

I define Social Web Analytics as the application of search, indexing, semantic analysis and business intelligence technologies to the task of identifying, tracking, listening to and participating in the distributed conversations about a particular brand, product or issue, with emphasis on quantifying the trend in each conversation's sentiment and influence.

What are you looking for in a SWA service?

Every organisation is unique, and you will form your own precise requirements spec from the aspects discussed below: indexing, spider capability, semantic analysis, search query structure, APIs and libraries, infrastructure, and commercial, licensing and terms of use.

Indexing

Literally, what proportion of the World Wide Web does this service catalogue? Right now, to my knowledge, no SWA tool has even a fraction of the breadth of Google, but obviously a service that indexes three million URLs may be more useful than one that indexes three hundred thousand.

More critically, however, indexing social networks, such as Myspace and Facebook, is fraught with difficulty, and if one service can index and make sense of these sites more thoroughly than another, this capability may outweigh any consideration of how many millions of sites the spider covers. In other words, quality may trump quantity.

Other factors here include language capability (which languages are critical to you, and which are just nice to have), URL selection (what

sites are included over others), de-duplication (do you want to remove duplicates or keep them in) and critically the ability to cope with dynamic sites (ie, websites that have pages generated on the fly from various data sources rather than static pages). You may also be interested in the rate at which a service is growing these capabilities.

The sources can include all and any kind of website: review sites, forums, chat rooms, social networks, blogs, micro-blogs, wikis, company websites, retail sites with customer feedback, etc. Sources can also include services such as Factiva which provide access to so-called traditional media material.

Actually, such indexing of content from third parties can constitute the entire source of data for a SWA vendor. In other words, such vendors do not author or maintain their own spiders (see below) but rather abdicate that task to others. They contend that this leaves them to focus their efforts on the analysis, interpretation and presentation of that data.

This sounds credible, but you will need to understand how the vendor can then tailor their service to your needs. If, for example, you require that specific sources are added to the indexing, this may be more readily achieved if the vendor controls their own spidering than if they then have to pass on your request to a third party.

Either way, it's worth noting here that SWA is nascent and no two vendors' service data sheets will have the same headings or line items, not least of which is their readiness to compare and contrast their exact indexing and spidering capability. Indeed, some vendors have declined to share this information with me for publishing here, so that alone indicates to me that it's an area you should explore thoroughly.

Spider capability

Spiders are also known as crawlers. A spider is software code charged with exploring the World Wide Web and sending back appropriate data to base.

There are robust, resilient, professional spiders, and there are amateur spiders. The former can interpret and report back on sites more effectively, and they can also cope with the inevitable page serving and Web server responsiveness and network latency problems riddling today's Web. They know where they've been with greater accuracy, and they revisit more frequently. They also comply with the robots exclusion protocol²⁷; essential if you don't want a website owner to get nasty with you and cut you out altogether.

The [Wikipedia page on crawlers](#)²⁸ is useful if you'd like to know more about them.

In my experience, SWA vendors with better than average spidering capability aren't differentiating themselves sufficiently on that basis. This has been explained to me as simply the relative immaturity of the procurement of such services to date; in other words, it hasn't yet featured as highly as I think it should as a point of concern for procurers.

One trend to watch out for: the ability to spider, index and interpret multimedia content (audio and video).

Semantic analysis

This is the most technical aspect of SWA, but it isn't rocket science from the procurer's perspective, thankfully.

²⁷ A voluntary code whereby spiders are programmed to respect the wishes of website owners regarding what they are happy for you to spider, and what they would rather you did not.

²⁸ http://en.wikipedia.org/wiki/Web_crawler

Take a micro-blog statement such as “Apple owners are cooler”. Does this statement relate to the iPod firm, or orchards? Is it to do with street cred or temperature? Could your personal reputation depend on the result if you work with Apple Inc.?

Other challenging company names from a semantic analysis perspective include Creative (innovative?), HP (horse power?), GM (genetically modified), Boots (shoes?), Shell (in the sea? in IT?).

And what about the British propensity to revert to sarcasm: “Wow, my ISP, FatPipes, is so awesome, they know I spend too much time online and take my connection down now and then just to give me a breather.” Could you write software to interpret this (fictitious) statement as negative rather than positive sentiment?

Should you get semantic?

There is no perfect semantic analysis approach today, and there is considerable variation in capability. Moreover, some SWA vendors don’t use semantic analysis at all. The latter breed will claim to distinguish HP, the successful technology company, from Horse Power by constructing more traditional search queries looking for words like “printer”, “PC” and “camera” close to the reference to HP, and, equally, for the absence of references to “car”, “motorbike” or “engine”.

But how they will cope if HP wheels an electric vehicle out of their Silicon Valley garage?

I mentioned above the ability to determine the *sentiment* of references to your brand, also known as the *tonality*. You may find it useful to know, for example, that 60% of references to your product were positive last month, 25% were neutral and 15% were negative. Tracking sentiment over time will help you establish whether things

are going your way or not, and how exactly. But SWA vendors without semantic analysis capability cannot estimate sentiment for you.

Again, this may not be a problem for your organisation depending on your specific needs.

The vendor without semantic analysis capability may allow you to determine the sentiment manually for each item they discover for you, and let you log your conclusion in their system accordingly. This could be fine for a few hundred mentions a month, but if your organisation is likely to generate thousands or tens of thousands of mentions each month, this will soon prove ugly to scale.

Lastly, you will probably want your SWA to ignore spam blogs (blogs created automatically simply to catch clicks and generate affiliate marketing revenue for their owners). Identifying spam relies heavily on semantic analysis.

This [disambiguation page on Wikipedia links to pages on semantic analysis](#)²⁹ should you wish to get down deeper into the science.

Search query structure

Search success may vary considerably on the quality of the search query structure. Exactly how are you communicating what you're looking for to the SWA service?

There are two extremes here, with some vendors having a mix of the two.

At its simplest, the SWA service offers users the same thing the Google homepage does; a search box. Enter search query here. A step up from this is analogous to Google's advanced search where you can be more specific about things such as phrases, exclusions, language and

²⁹ http://en.wikipedia.org/wiki/Semantic_analysis

dates. This is perfect for a quick analysis, for investigating a new business prospect or partner for example, or informing a new product development brainstorm.

The other end of the spectrum is the allocation to you by the SWA vendor of a "search manager". This individual is expert in their company's service, and expert at working with you to construct detailed search queries and, critically, honing them regularly over time from the results that get returned.

The results from the latter are more likely to deliver less false positives (erroneous results such as when "HP" means "Horse Power") and less false negatives (failing to identify true references as such).

As for every parameter here, only you can decide which approach is most appropriate to your needs. It may boil down to aspects such as the degree of ambiguity in your company and brand names for example, and your budget.

APIs and libraries

You'll need to involve your IT team here, although it will be apparent from the next paragraph whether this section is relevant to you.

APIs are application programming interfaces. The availability, capability and documentation of APIs will prove critical to your ability to suck the results out of your SWA vendor's service into another of your IT systems, and wield their service from within other systems, should that be your intention.

In these situations, the service's user interface is less important as your users will be accessing the information via another system altogether.

You probably don't just want to show the numbers however; the story is still best summarised with good looking charts. You'll need, therefore, to have easy and documented access to the provider's Flash libraries, the code that converts the data into Flash images.

Alternatively, a provider may use Silverlight or some other image rendering technology (look out for fast Scalable Vector Graphics as this capability becomes native to modern browsers such as the planned update to Firefox 3.0), and these capabilities have to be understood similarly.

These capabilities are rare at the moment, and surprisingly so I think given the vendors' focus on the Social Web which, in my mind, goes hand-in-hand with other Web 2.0 characteristics such as mashing-up³⁰.

Some vendors are more "Web 2.0" in their thinking, more "mashup" friendly than others. Here's a hint. If you see the vendor referring to *cost per seat* in their pricing sheet then you know they'll welcome your overtures to mashing-up their data as much as the record labels revel in file sharing. *Cost per seat* is, in my opinion, an archaic way to price SWA. Something like *cost per active search query per day* is more appropriate.

Infrastructure

Whilst some Web services can be run from someone's bedroom, given its criticality to your business, you don't want SWA to be one of them!

You will be interested in understanding where the vendor's main tech facility is located, what redundancy such as co-located facilities they have established, and their hours of customer service support.

³⁰ Mashing-up refers to sourcing data from various sources and combining them in previously unachievable yet highly useful ways. A common mashup, for example, is to display various data on a Google Map.

Commercial, licensing and terms of use

What is the pricing structure? Per seat? Per search query? Per language? Annual? Pay-as-you-go?

How does agency licensing work? Many vendors gear their pricing to the single organisation rather than to the agency working across many campaigns for many organisations. Do they offer discounts for multiple campaign tracking? Do they offer a service level agreement?

How does the licensing address the copyright of both the reports generated by the service and of the material it collates? Do they indemnify you for the use of their service for any claims made in relation to patent, copyright or trademark infringement?



The free tools

There's a wealth of free tools and services available to you on the Web. You should make use of these before procuring SWA to give you an idea of what's out there and what you might need to be doing in the longer-term. And, as they're free, there's no reason to stop using them when you've procured a SWA tool; after all, no one tool will give you the full picture.

Here's a selection.

Google, Yahoo!, MSN Live, Ask

<http://www.google.com>

<http://search.yahoo.com>

<http://www.live.com>

<http://www.ask.com>

What better place to start to understand how your organisation and brands are perceived than by conducting a Web search... as this will be a very common way for your stakeholders to look for you and find content about you. And whilst Google currently accounts for just over half of all Web searches, that means a significant proportion are conducted with other search engines. That means you should too.

Try thinking about the search terms each of your stakeholder groups might use; some will include your company name and brand names, and some search terms will not. I'm guessing however that most readers of this ebook will have progressed beyond this sort of recommendation some years back, so I'll move on...

Google Alerts

<http://www.google.com/alerts>

Google Alerts cover "News", "Web", "Blogs", "Groups" and "Video". A Google alert is an email that lets you know if:

- new articles make it into the top ten news results for your Google News search
- new Web pages appear in the top twenty results for your Google Web search
- new blog posts make it into the top ten results for your Google Blog search
- new group posts make it into the top fifty results for your Google Groups search
- new videos make it into the top ten results for your Google Video search.

A sixth type of Google Alert, "Comprehensive", combines the results of the Google Alerts for "News", "Web" and "Blogs".

Setting up Google Alerts means you get updates automatically as they happen without having to run manual searches repeatedly yourself.

Google Trends

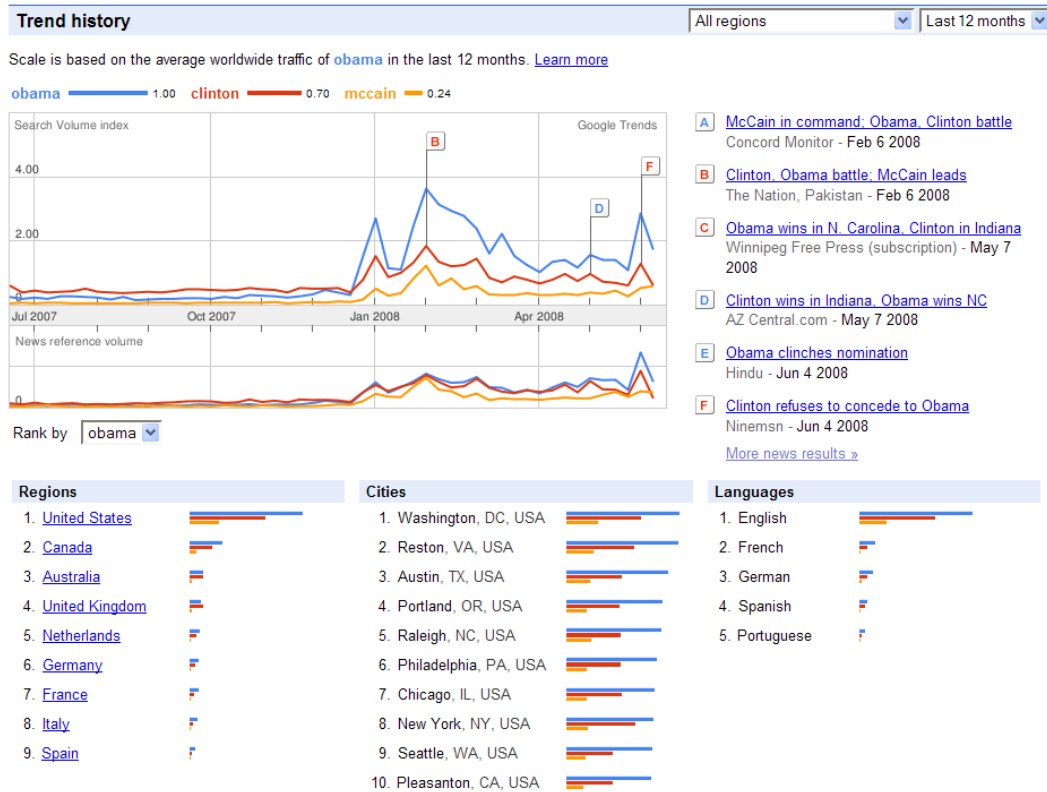
<http://www.google.com/trends>

Google Trends delivers charts that portray how often a particular search term or terms is entered relative to the total search volume across regions of the world, and in various languages.

Popularity is broken down by region, city and language, and it's possible to refine the main graph by region and time period.

You can compare the volume of searches between two or more terms (as per the screenshot here), and the main news explaining spikes in activity is conveniently summarised to the right of the chart.

Screenshot – Google trends



Google Blog Search

<http://blogsearch.google.com>

As per Google's generic Web search, a Google Blog Search can be initiated simply by entering keywords and phrases in the search box, or honed with the advanced search facility available from the Google Blog Search homepage.

Results can be sorted by relevance or by date.

Technorati

<http://www.technorati.com>

The first blog search service, Technorati is still considered by many to be the best. Claiming, as of December 2007, to spider and index over 112 million blogs, the results can be sorted by "authority" and "rank".

Technorati Authority is the number of blogs (not the number of blog posts) linking to a website in the last six months. The higher the number, the more Technorati Authority the blog has.

Technorati Rank is calculated based on how far a blog is from the top. The blog with the highest Technorati Authority is the number 1 ranked blog, so the smaller your Technorati Rank, the closer you are to the top.

Twingly

<http://www.twingly.com>

Just out of private beta as of mid-June 2008, Twingly describes its goal as being "to create the first and best spam-free search engine and to be the most innovative web startup in Europe." They continue "while our eventual goal is nothing short of world dominance, we are currently focusing our efforts on European blogs. Having said this, our index already includes many blogs from all over the world."

Try it, particularly if your geographic focus is Europe. I like it.

IceRocket

<http://www.icerocket.com>

Another well respected blog search engine, IceRocket offers the same basic and advanced search facility as all blog search engines listed here.

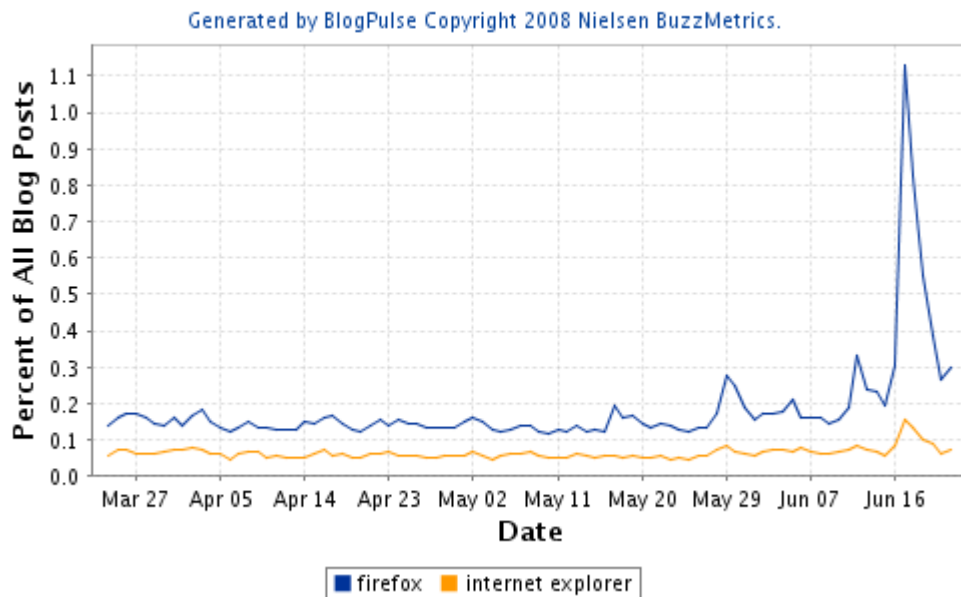
BlogPulse

<http://www.blogpulse.com>

Owned by Nielsen, whose Nielsen BuzzMetrics service features later in this ebook, BlogPulse is “an automated trend discovery system for blogs”.

Screenshot – trend results

Trend Results



Here's their feature list:

- Blog search
- A set of buzz-tracking tools tracking key issues, people, news stories, news sources, bloggers and more
- A fun look at real-world trends as reflected through blogs
- Daily blog stats that measure activity in the world of blogging
- A trend search that allows you to create trend charts comparing buzz in the blogosphere on up to three specific topics

- A Conversation Tracker that follows and captures the discussion, or conversation, that emanates and spreads from individual blogs or individual blog posts
- Blogger Profiles that identify top-ranked blogs and analyze their blog presence, activity and relative influence in the blogging world.

News readers

A news reader is also known as a feed aggregator or feed reader. A news reader is to RSS newsfeeds what Google Mail or Hotmail is to email. In fact, look out for services combining RSS and email in the not too distant future.

Whilst early news readers came as software requiring installation on your PC, the majority of news reader users today prefer Web-based services.

Popular Web-based news readers include [Newsgator](#), [Google Reader](#) and [Bloglines](#).

Alexa

<http://www.alexa.com>

A subsidiary company of Amazon.com that operates a website that provides information on web traffic.

See http://en.wikipedia.org/wiki/Alexa_Internet for more information.

Del.icio.us

<http://del.icio.us>

Del.icio.us is the first social bookmarking website and searching it gives you an idea of what people are bookmarking. This is different to Google's pagerank approach, which looks at how many and what kind

of sites are linking to a webpage. You could say that Joe Public will save a social bookmark more often than he will add a hyperlink to a website, so searching social bookmarking websites will give you a different and equally valuable perspective.

Other popular social bookmarking services include [StumbleUpon](#), [Ma.gnolia](#), [Furl](#), [Simpv](#) and [Faves](#) (previously called Blue Dot),

Digg

<http://www.digg.com>

A social news service, Digg describes itself as follows: "Everything on Digg — from news to videos to images to podcasts — is submitted by our community. Once something is submitted, other people see it and Digg what they like best. If your submission rocks and receives enough Diggs, it is promoted to the front page for the millions of our visitors to see."

Other social news services include [Reddit](#) and [Newsvine](#).

Summize

<http://www.summize.com>

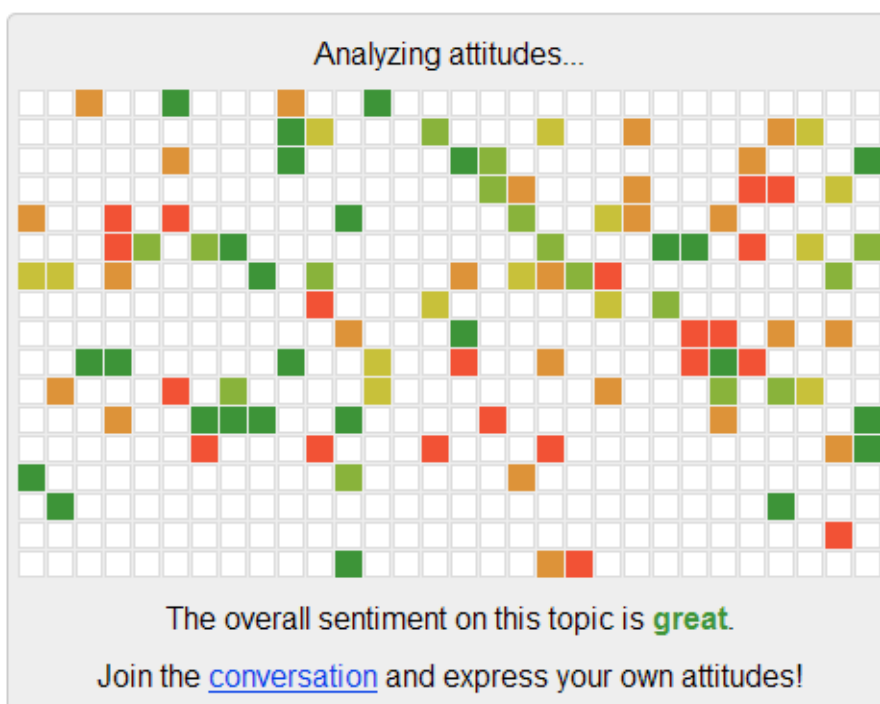
As described on their website: "Summize's mission is to search & discover the topics and attitudes expressed within online conversations. Our home page currently features real time conversations on Twitter. Also check out [Summize Labs](#) for prototypes that harness conversations within blogs and reviews. Additional sources of online conversations are coming soon!"

Screenshot – Summize labs Twitter sentiment

Realtime Twitter Sentiment

Enter a topic in the box below. We use our [search engine](#) to find up-to-the-second tweets about this topic, then automatically analyze the attitudes expressed in those tweets.

Try: obama, BSG, iphone.



The colors indicate the sentiment of words and phrases found in tweets:

- | | | |
|-------------|-------------|----------------|
| great words | so-so words | wretched words |
| swell words | bad words | no sentiment |



The vendors

This is a list of all the vendors approached for inclusion in this ebook and whether they responded with the majority of data I requested, some of it or none.

COMPANY / SERVICE	Most	Some	None
www.Attentio.com	•		
www.Biz360.com		•	
www.Branddimensions.com	•		
www.Buzzlogic.com	•		
www.Cision.com	•		
www.CollectiveIntellect.com	•		
www.CyberAlert.com		•	
www.Cymfony.com		•	
www.DNA13.com	•		
www.DowJonesInsight.com	•		
www.Integrasco.com		•	
www.Kaavacorp.com	•		
www.Magpie.net	•		
www.MillwardBrown.com			•
www.NielsenBuzzmetrics.com		•	
www.Radian6.com		•	
www.RelevantNoise.com			•
www.Umbriacorp.com	31		
www.VisibleTechnologies.com			•
www.Vocus.com		•	

³¹ Umbriacorp apologised for not being able to respond at this time as the company was acquired 1st April 2008 by J.D. Power and Associates, itself part of The McGraw-Hill Companies. They asserted their interest in participating next year.

It's interesting that I'm not a journalist, I'm a chartered engineer, a director of a PR consultancy and a blogger, and in writing this ebook and publishing it for free on the Social Web I'm creating something that is as pertinent to the SWA vendors as the kind of stuff their services unearth for their customers.

So it's no surprise then that, excepting a notable few, the vendors understood the relevance of responding to my emails and phone calls. But none approached me proactively.

I find this a bit odd as I blogged twice about my intentions to write this ebook, well in advance and including relevant keywords, phrases and hyperlinks to most of the companies listed above. The blogs on which I posted attract thousands of unique visitors a month, yet either the posts were not flagged up by the vendors' own tools, or they were and no-one cared at that juncture.

So I guess the SWA vendors aren't yet walking the talk³².

Vendor information and your participation

The information on the following vendor pages is provided by the vendors at my request, and may be supplemented by information taken directly from their current public websites. I have referenced no other sources, nor have I validated the vendor claims.

As you may interpret from the ebook title, I intend to publish an updated ebook each year. I hope that this ebook will enjoy wide circulation and warrant, in true Social Web style, your participation. I would like to include your experiences of wielding these tools in the next edition, and indeed any thoughts you may have on the whole

³² This statement cannot apply to Dow Jones. I can't determine if they would have picked up on the blog posts as they were quickly alerted to my post by a reader and responded accordingly.

subject, so I will be looking out for your comments, blog posts, refbacks, emails, twitters, feeds, ...

As you would expect, I'd also welcome updates from all vendors included here and any I may have omitted or that are yet to launch. If any vendor wishes to volunteer their services to help me do some ongoing social Web analytics on SWA, that would be super!

Attentio

<http://www.attentio.com>



**Simon McDermott,
CEO, Attentio**

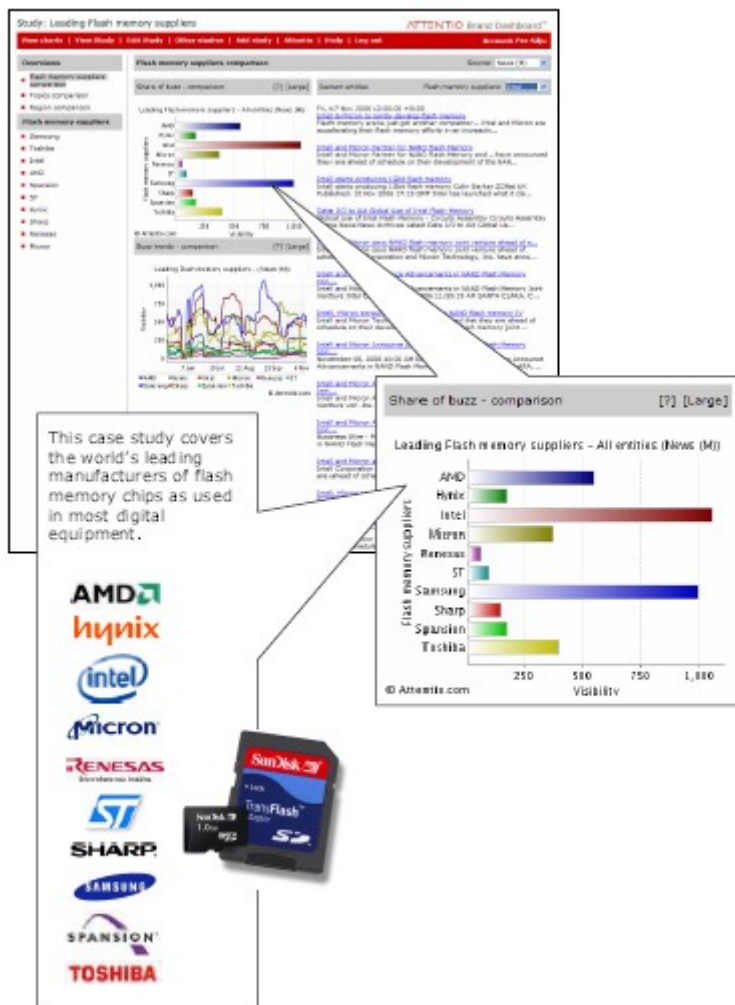
Based in Brussels, Attentio position themselves as Europe's leading provider of real time market intelligence based on social media. Founded in 2004 by serial entrepreneur Per Siljobergsasen (founder of Kelkoo) and Simon McDermott (ex. Intel, Cisco), the service is powered by "a robust search engine, indexing technology and tools to extract information from blogs, forums, news, YouTube". It's also award winning – Red Herring and the European ICT Prize.

The Attentio team of researchers includes Doctorates and Masters in Artificial Intelligence for the production of advanced algorithms in areas such as topic detection, sentiment analysis and influencer identification. Attentio employs advanced semantic analysis (which, for those interested in the science, includes n-grams, patterns, tf*idf and supervised learning).

Attentio indexes data from 600,000 English language and 900,000 other language sources, including all western European languages, three times a day. This number is growing approximately 15% per

quarter. Whilst crawling thousands of European Myspace sites, they do not generally employ spiders but rather subscribe to RSS feeds and buy information from a handful of third parties.

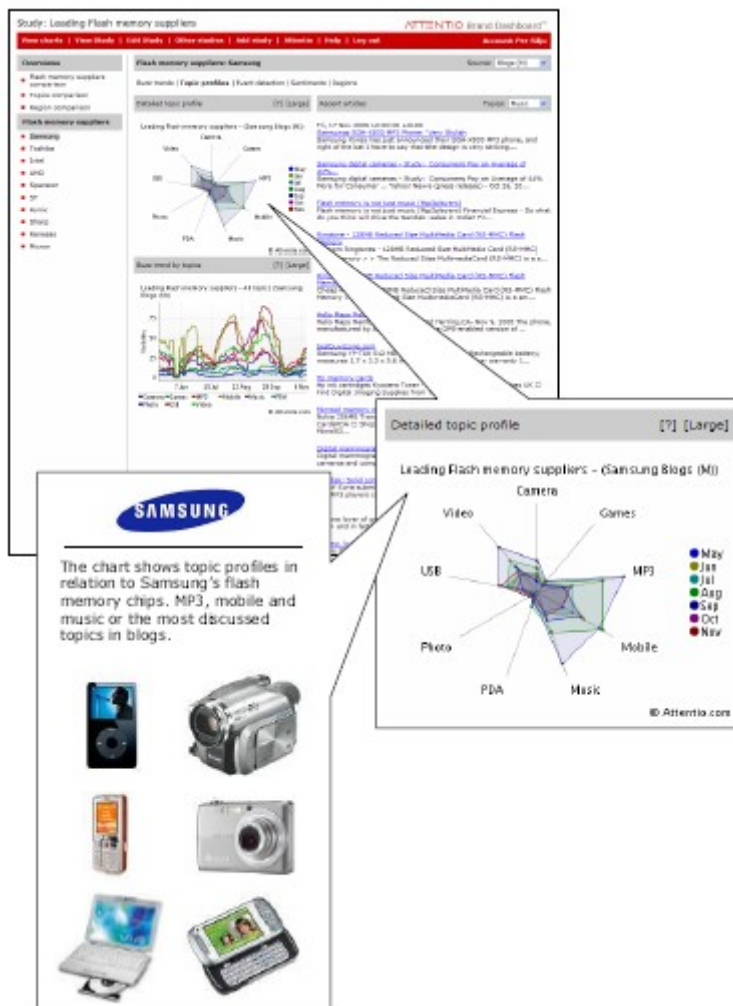
Screenshot – dashboard



The service doesn't allow customers to remove or add sources directly, but they will try to meet requests submitted to them. Customer service follows the central European working day, so Attentio may not be appropriate for US or Asian organisations requiring same day support.

Attentio is in my good books with pricing that is not on a per seat basis. Moreover, they're also geared up to work with agencies with a white label service and an "all you can eat" deal covering all regions and sources for multiple keywords and topics from €15,000 per month.

Screenshot – insight



The client list includes Microsoft, Johnson & Johnson, Samsung, Toyota, Lexus, Intel, Disney Pictures, GfK, Roche Pharmaceuticals, Club Med and "major European PR and digital agencies". If you want a taster of Attentio's output, they have a showcase at

www.trendpedia.com, although this won't exhibit sentiment until later versions.

Attentio's roadmap includes sentiment weighting and blog comment aggregation.

Contact simon@attentio.com.

Biz360

<http://www.biz360.com>



**Brad Brodigan,
CEO, Biz360**

Founded in 2000, Biz360 Inc. is an information-services company focused on transforming information to help businesses make better decisions. They use technology to aggregate, measure and analyse information from traditional news media and social media.

“Biz360 is distinguished by its broad, global network of content sources, superior leading-edge analytical expertise, and outstanding service & support – which enables us to deliver to our clients the most accurate, relevant and insightful information / metrics in the most flexible and consumable formats.”

A privately held company, funded by three Silicon Valley investor firms, they closed their last round of funding in May 2008 for \$10m. They plan to invest the capital in “next-generation products that capitalise on significant, emerging market research opportunities

spurred by the evolution of consumer behavior on the Web.” Clients include AAA, Aetna, AstraZeneca and Sun Microsystems.

Brad Brodigan, CEO, says:

“We are building technology to capitalize on the unprecedented millions of customer product opinions that exist on the Web. We aim to provide decision makers in marketing, research and product development a new and innovative means to harness these opinions in an extremely powerful way.

“Two years ago, people were buying media analysis for defensive purposes – they wanted to know when people were saying bad things. But now companies want to know how to compete better. What’s driving someone to buy one digital camera over another or one car over another? We believe there’s a big shift from defensive to offensive tools.”

But that’s all I can tell you I’m afraid. Biz360’s PR consultancy declined to provide any further detail. Can you believe it? I can’t. They want to help marketing decision makers, just not here!

If you want to know more, then you need to submit a request for proposal and sign a non-disclosure agreement.

Brandimensions

<http://www.brandimensions.com> / <http://www.brandintel.com>



**Bradley Silver,
Co-CEO,
BrandIntel**

Brandimensions offers two services, BrandIntel (the focus here) and BrandProtect.

“BrandIntel translates consumer-generated content into predictive consumer insight through a combination of proprietary technology, iterative human analysis and proven best practices. Using its unique methodology, BrandIntel provides visibility into consumer discussions online and delivers market intelligence that extends beyond traditional research by capturing spontaneous, relevant and emotional content. Through this process, BrandIntel clients get actionable recommendations that can be used to capture market share, boost brand equity and manage market reputation.”

BrandIntel claims customers amongst the Fortune 1000 companies, across sectors including automotive, media and entertainment, financial services and life sciences.

It became immediately obvious that BrandIntel is different from most of the other services in this ebook when, in answer to the request for screenshots, I was informed: "We provide a service rather than a product so screen clips are not applicable."

As such, the company left my questions about indexing, spidering, semantic analysis, APIs etc. blank. The commercial aspect was the only question set answered, and it basically told me that you can hire these guys on a project or ongoing basis. The project entails them digging into the social Web and reporting back on your product / brand, and that of your two closest competitors. You also get a telephone call with their analyst.

Alternatively, you could subscribe to BrandIntel on an ongoing basis. This programme will be adapted based on its prior findings, and entails a weekly call with your assigned analyst.

Without further information about its underlying technology I have no idea how this service might stack up. For all I know, they may subscribe to the service or services of other SWA vendors, which would therefore eliminate them from future editions of this ebook.

BrandIntel is headquartered in Canada, and has offices across North America. No contact details supplied to me; see website for details I guess!

BuzzLogic

<http://www.buzzlogic.com>



**Rob Crumpler,
CEO, BuzzLogic**

Founded in 2004, BuzzLogic is “a technology company and growing ad network enabling marketers and publishers to harness online influence to create and execute effective social media growth strategies.”

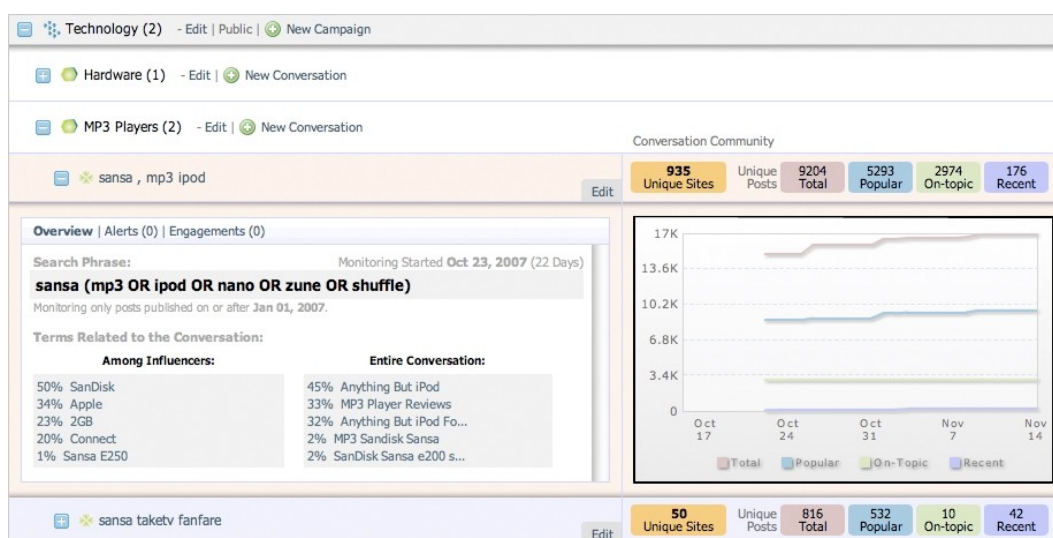
BuzzLogic uses patented algorithms to identify influential online discussions and claims to enable marketers to effectively target their messages to both opinion leaders and the engaged readers who follow them.

Agencies and brands such as Nordstrom, Publicis, PBS, eBay, and McDonald’s use BuzzLogic for a number of initiatives, from blogger relations to direct marketing and Conversation Targeting™ for online ad campaigns.

BuzzLogic describes its service as “built specifically to crawl high quality blog ‘conversations’ – which we define as blogs linking back and forth to each other, ‘talking’ around very specific topics.” They continue to describe their indexing as “based on trust filters, and

[comprising] high quality content deliberately linked to by trusted sources to ensure no spam or 'splogs' touch our customers."

Screenshot – conversation at a glance



Whilst not disclosing the number of sources or indeed how they collate all their data, they do describe their indexing as follows:

"To date, BuzzLogic has indexed hundreds of millions of posts, adding thousands more a day. These include content from personal and corporate blogs, media sharing sites, micro-publishing sites and open social networks. New sites enter into our index every day – be it organically through a blog being linked out to by a trusted source, or by a customer entering a particular URL into the index to understand and track its influence.

"BuzzLogic's Influence Algorithms crawl and analyze content for influence on a daily basis. All sites – regardless of language or origin – get the same treatment. In addition to English, we support 19 other languages. Currently, approximately 80% of the BuzzLogic Conversation Index is comprised of English language sites."

BuzzLogic has appropriate focus on the social networking services. They crawl Myspace as permitted, and are amongst the first vendors here to take notice of Wikipedia and Twitter. If a link is tweeted (the verb referring to making a Twitter post), then this may impact its BuzzLogic ranking.

Screenshot – social map

The screenshot displays the BuzzLogic interface for a workspace named 'iphone'. The top navigation bar includes 'Conversations', 'All Watchlists', 'Ad Targeting BETA', and 'All Reports'. Below this, there are tabs for 'Influencers', 'All Content', 'Watchlist', 'Engaged', and 'Reports'. The main content area is divided into 'Overview' and 'Map'. The 'Overview' section shows a funnel chart with categories: All (475), Influential (13), Popular (325), and On-Topic (309). The 'Map' section shows a network diagram with nodes representing influencers and their connections. Two specific posts are highlighted: 'iPhone review - Engadget' and 'Engadget reviews the iPhone - web2announcer'. The 'iPhone review - Engadget' post has a summary and a publisher URL (http://www.engadget.com/). The 'Engadget reviews the iPhone - web2announcer' post has a summary and a publisher URL (http://www.web2announcer.com/).

Customers may add new sources to a 'watchlist', and equally relegate existing sources to a 'blacklist'.

BuzzLogic has refrained from semantic analysis to date. Rather, they give customers the facility to rate each item for sentiment themselves or outsource this manual task to partner firm [KDPaine & Partners](#)³³.

The service employs “patent-pending algorithms, which take more than a dozen factors into account when calculating influence, including the linking activity around a specific post, the quality, popularity of the surrounding blogs paying attention to a post, and how credible an author is on a specific subject.”

I’ll leave BuzzLogic to describe the output:

“The data is presented via a dashboard that illustrates lists of top influencers, as well as a view that displays individual posts in order of influence. BuzzLogic also offers a Social Map – a 360 degree visual depiction of the linking activity around any given post, complete with which sites are linking in, which sites are being linked out to. Users can drill down to read background information on a blogger, outreach notes from colleagues, and access post content straight from the BuzzLogic dashboard. They can also access reach metrics to gauge the audience size of any blog, as well as discover whether a certain bloggers is influential on any other conversation queries they have running in the system.

“Customers can run reports to pull statistically relevant info to understand the percentage of influencers talking about a specific subject as compared to the rest of the community, as well as reports that map the change in tone and sentiment over time.”

The company shares my enthusiasm for APIs and tells me they have several reporting APIs in development. Of course this also indicates that they have the foresight to avoid per seat pricing.

³³ <http://www.kdpaine.com>

Customers pay a yearly subscription fee, from \$12,000, for complete on-demand access. They are unique here in offering their Conversation Targeting service to advertisers, but that's beyond the current remit of this ebook.

BuzzLogic is geared to agency needs, with appropriate multi-campaign pricing allowing the fee to scale up and down with client wins and losses.

BuzzLogic is based in Northern California's Bay Area, privately held, and secured \$12 million funding in 2007.

Contact Courtney Hughes for sales (chughes@buzzlogic.com) and Valerie Combs for general enquiries (vcombs@buzzlogic.com).

Cision

<http://www.cision.com>

CISION



Niklas Flyborg,
CEO, Cision

Cision AB (www.cision.com) is publicly owned and quoted on the Nordic exchange. The company has around 2,600 employees and offices in the US, UK, Sweden, Canada, Germany, Norway, Finland, Denmark, Portugal, Lithuania, the Netherlands, Hong Kong and China, and partners in another 125 countries.

They have a heritage of monitoring mainstream media for 70 years, which makes them the granddaddy of this group. Cision jumped into Web monitoring “when the Web became a vehicle for communications and marketing”.

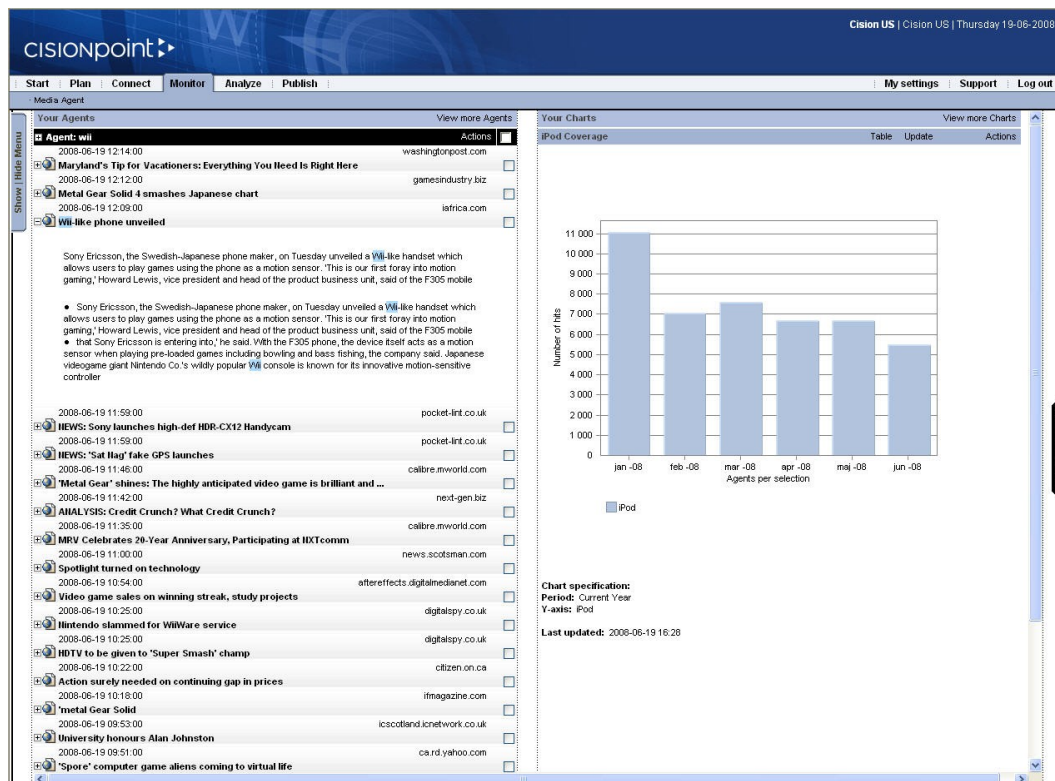
Cision serves “almost all of the world’s largest public relations firms” and a majority of the Fortune 500, as well as NGOs and governmental organizations. Corporate clients include Samsung, Bayer, Hilton, American Express, Sony, Walt Disney, Starbucks, Oracle and Intel.

It would be easy to think of Cision as the old boys trying their best to keep abreast of this modern age, but that would be wrong. Plainly.

Cision 'gets' the Social Web and what it takes to learn most from it. It is also the only vendor to have created a specific website landing page for readers of this ebook, (although they could have made this a discussion page rather than simply an invitation to fill in a form; I'm assured your query will be routed to a contact person in your country).

http://us.cision.com/campaigns/2008_ebook/request.asp

Screenshot – media agent

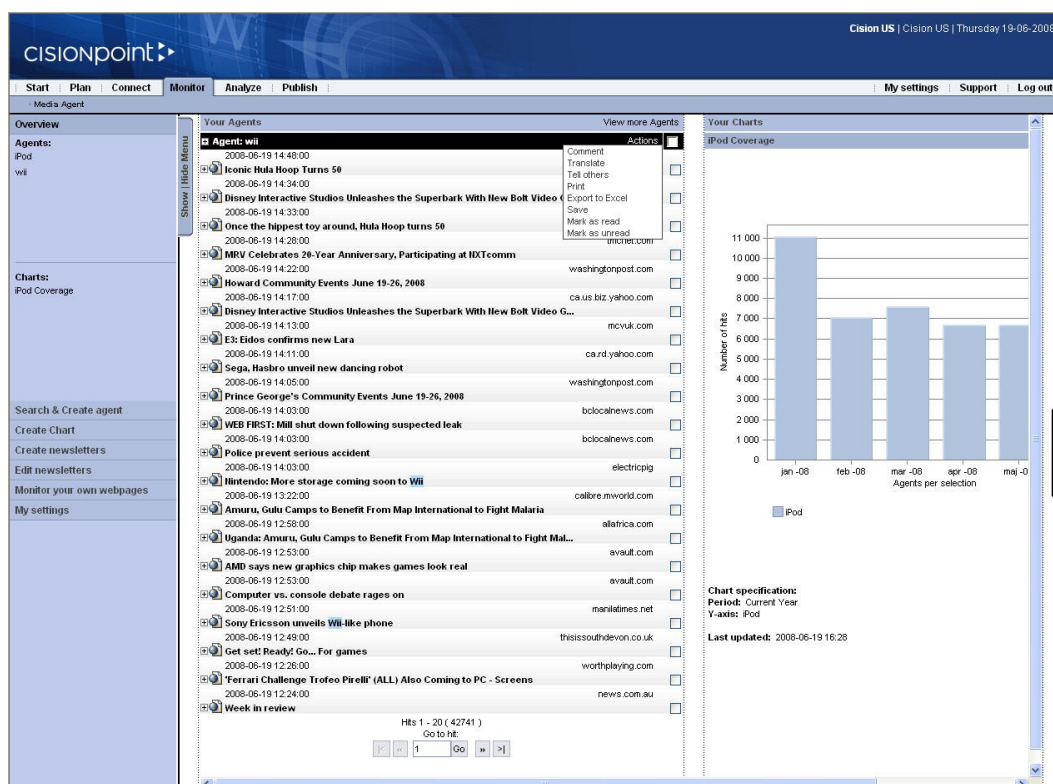


Interestingly, in all my communications with SWA vendors, Cision is the only vendor to have referred to the possibilities of the Semantic Web (so called Web 3.0) that I touched upon in the "Measurement & evaluation" sections.

Let's look at the details. Cision covers 30,000 English language news sites and 25 million English language blogs. They have one of the most extensive language capabilities spanning 70 countries. Everything is

crawled at least once a day, and a team of 200 analysts maintain and grow this database of sources, including direct requests from customers. The depth of spidering required for each source is assessed by an analyst, but it's unclear to me whether the main social networks (Myspace et al) are included.

Screenshot – media agent



Like Cymfony, the Cision service uses natural language processing, a form of semantic analysis based on artificial intelligence and computational linguistics. It has thorough de-duplication and spam elimination capability. Impressively, Cision is one of the few vendors here to declare active R&D into speech-to-text capability for crawling and indexing audio and video content. Such capability will be a significant differentiator in coming years.

Cision claims to have automatic sentiment analysis but also states that this is not something they recommend to their clients. Note to self to understand this position more clearly for the 2009 ebook; is it because they feel no such analysis is sufficiently accurate as yet, or just their own sentiment analysis?

Cision wisely offer both mechanisms for defining search criteria: DIY or in conjunction with one of their Media Analysts. The analytical results are made available as an XML output (a file that can be readily used by another IT system), but there's no current access to charting libraries for the subsequent display of such data in another system.

Lastly, Cision also has the wisdom to price on a per client / campaign / brand basis, but there remains a per seat element.

Contact http://us.cision.com/campaigns/2008_ebook/request.asp.

CollectiveIntellect

<http://www.collectiveintellect.com>



**Don Springer,
CEO, Collective-**

Founded in 2005 and based in Boulder, Colorado, CollectiveIntellect describes their last year as one focused on adding millions of new social media sources and improving data quality with highly automated categorisation and analytic tools. Quite rightly, they point out that analysis is pointless if you don't start with accurate and comprehensive data.

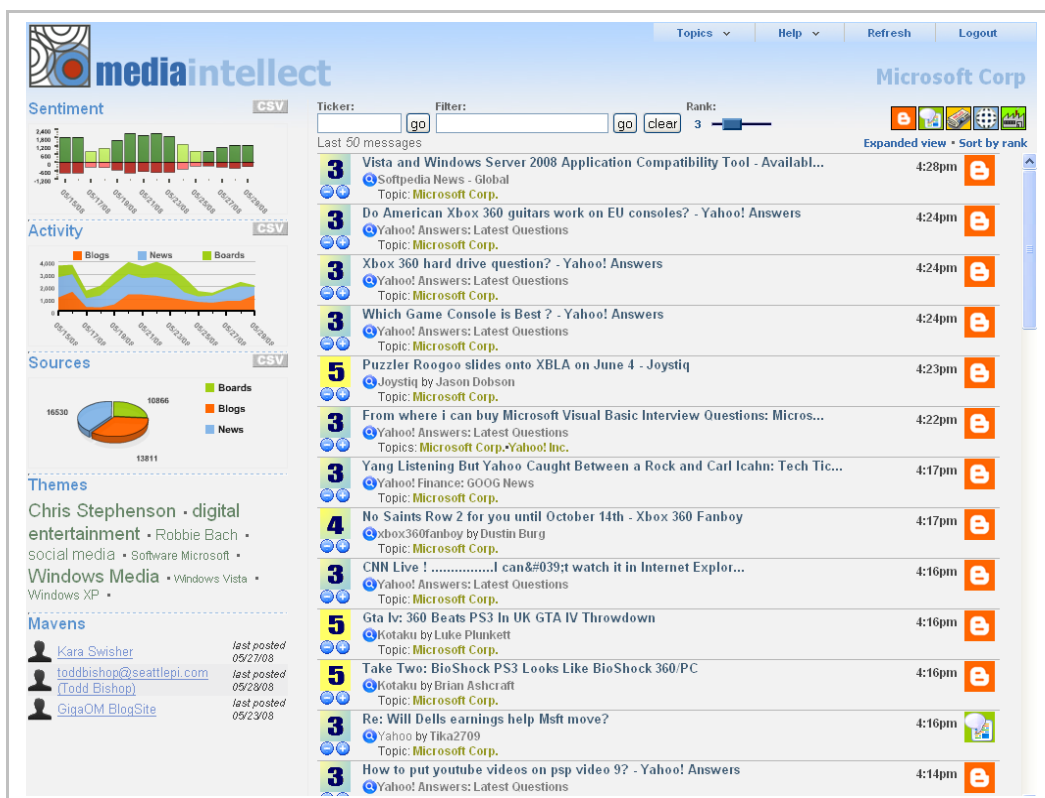
CollectiveIntellect must be doing something right to have attracted custom from the likes of Microsoft, Chrysler, Anheuser-Busch, Pfizer, Dell, Yahoo!, Viacom, Verizon, Levi's and Adobe. They are privately held and have raised \$11m to date.

Alongside the technological service, CollectiveIntellect provides "strategic planning, development of integrated social media plans and landscape analyses, project management and counsel including identifying best practices, assisting with ad hoc questions and co-developing communities." This may or may not explain the absence of

agencies listed in their key customer line up – it appears that CollectiveIntellect may be considered by some agencies as competitive, although I’m sure the company considers themselves more as partners. They have indicated to me their willingness to co-pitch with agencies.

The company recognises the growing demand for quantitative performance measurement of marketing activity, and emphasises its capabilities “to quantitatively [show] how a brand ranks against competitors and to measure the lift of each campaign, social media or not.”

Screenshot – Real-Time Media Intellect Monitor

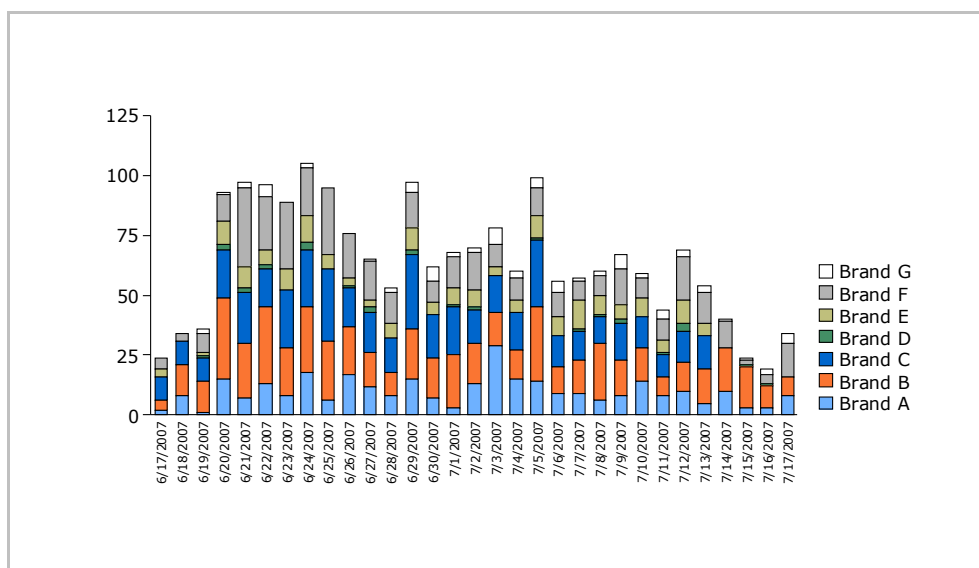


The service indexes over 12 million blogs, 10,000 message boards, 19,000 traditional news sources, and tracks the social networking sites as permitted. All these are English language only at the moment,

although the company plans to extend to Latin based and Asian languages this year, with three million sources already bedded in.

CollectiveIntellect has adopted topic crawling (along with Kaava, the only services to make this distinction). Called TopicNets, this patent-pending spidering technology potentially reveals deeper insight by directing their crawls based on discovering linked topical consistency. This kind of intensity is generally referred to as “deep crawling”.

Screenshot – share of voice



Their obsession with topicality, whilst no doubt an asset of itself, may be a weakness in isolation. Interestingly, CollectiveIntellect dismisses social networking pages as “highly off topic and generally unhelpful”, and instead prefers to focus on the group / community pages within such networks. They also write off micro-blogs such as [Twitter](#) and [Jaiku](#) as “highly irrelevant”.

That’s one perspective I can’t agree with; for example, I’ve found my Twitter community to be a great source of interesting information, recommendations, critique and links (and trivia too I admit). In other words, micro-blogs can be an influential channel. And social

networking pages are as on topic as any chat over a pint might be; random perhaps, influential, possibly.

Saying that, CollectiveIntellect's [Nick Sowden](#)³⁴ is the only staff member of the SWA vendors described here to let me know he's on Twitter and to have subscribed to my Twitter. He's also one of two to have invited me to link up on [LinkedIn](#). To me, 'being' social inspires confidence that they truly 'think' and 'live' social.

Despite my reservations about the company's dismissal of social pages and micro-blogs I recall telling a colleague that CollectiveIntellect has "the magic dust". Here are a few more things, additional to those above, that led me to make such an exclamation:

- Defined failure mechanisms for spiders
- Distinguish between individual authors on a blog
- Advanced de-duping capability, if required
- Easy-to-use manual weighting of sources as needed
- Advanced linguistic statistical analysis helps train the service from example content to "find more content like this"
- Have every analysis capability going – Bayesian networks, language models and latent semantic analysis
- Market leading spam elimination
- Accurate sentiment / tonality assessment that does not require domain specific tuning
- A 17-point assessment of influence
- Availability of documented APIs and libraries, and even widgets, although like me they recognise that the demand is nascent

³⁴ <http://www.twitter.com/collectiveintel>

- Solid technical infrastructure and 24/7 support.

Screenshot – maven list

Influence Rank	Source	Type	Total # Posts	Average Message Rank
1	BIOPACT	BLOG	979	6.7
2	ABC ONLINE	NEWS	837	6.1
3	GUARDIAN UNLIMITED	NEWS	573	6.4
4	TELEGRAPH	NEWS	525	6.2
5	RENEWABLE ENERGY ACCESS	BLOG	325	6.4
6	POWER ENGINEERING	NEWS	189	6.2
7	GREENCAR CONGRESS	BLOG	127	6.4
8	AUTOBLOGGREEN - SEBASTIAN BIANCO	BLOG	126	6.3
9	GREEN CAR CONGRESS	BLOG	102	6.5
10	AUTOBLOGGREEN - SAM ABUELSAMID	BLOG	94	6.4
11	THE HINDU	NEWS	75	6.4
12	ACTIVERAIN BLOGS	BLOG	70	6.2
13	AUTOBLOGGREEN - XAVIER NAVARRO	BLOG	65	6.4
14	RENEWABLE ENERGY ACCESS	NEWS	64	6.3
15	DOMESTICFUEL	BLOG	58	6.4
16	BUSINESS WIRE	NEWS	47	6.3
17	PEAKENERGY	BLOG	43	6.7
18	NATIONAL WIND WATCH: NEWS	BLOG	37	7.1
19	PEAK OIL	BLOG	34	7.4
20	THE AUSTRALIAN	NEWS	34	6.4
21	REPORT ON BUSINESS	NEWS	34	6.2
22	CNN	NEWS	33	6.3
23	AUTOBLOGGREEN - LASCELLES LINTON	BLOG	28	6.6
24	AUTOMOTOPORTAL	BLOG	28	6.0
25	NUCLEAR REGULATORY COMMISSION	NEWS	28	6.2
26	REUTERS	NEWS	27	6.4
27	JALOPNIK - MATT HARDIGREE	BLOG	25	6.4
28	INTERNATIONAL HERALD TRIBUNE	NEWS	23	6.4
29	PLATTS	NEWS	23	6.3
30	WORD NUCLEAR NEWS	NEWS	23	6.2

All of this capability does come at a small price to convenience however. CollectiveIntellect assigns you a “search manager” to help build up your SWA criteria. But as I wrote in the “Search query structure” section, you get back what you put in.

By now, you’ll know I detest per seat pricing, and again CollectiveIntellect comes up trumps. The commercials are based on a per project, per campaign or per topic basis. In fact, the only other

criticism I can muster up is the plainly visible fact that their reports aren't as pretty as the competitions'; not a critical desideratum and easily fixed... come on CollectiveIntellect!

Contact Nick Sowden, Director, Business Development, (telephone +1.303.842.0805, nick@collectiveintellect.com, twitter [@collectiveintel](https://twitter.com/collectiveintel)).

CyberAlert

<http://www.cyberalert.com>



CyberAlert's heritage lies in tracking traditional media sources, indexing 27,000 news sites worldwide, of which just over half are US based and 20,000 of which are in the English language.

The company has added 25 million blogs to its indexing capability, but cannot currently weed out spam blogs. Indeed, despite significant evidence to the contrary, they explain this inability by stating: "There doesn't seem to be any detectible pattern of where they originate or what format they use."

Whilst they do say they will incorporate spam recognition and elimination "once [such software] becomes available", this fact alone betrays the company's relatively immature SWA capabilities.

Their service has no semantic analysis, and they reply on their customers to rate one source as more relevant than another.

CyberAlert provided less information than the majority of vendors here. Why? I don't know, but it means I can't say much more. The only highlight I can report is the fact that pricing is not on a per seat basis.

Cymfony

<http://www.cymfony.com>



**Andrew
Bernstein, CEO,**

There's an interesting fact about Cymfony – U.S. intelligence agencies have been relying on their technology for the best part of a decade. That grabs your attention. It may have grabbed TNS Media's, as they acquired Cymfony in February 2007.

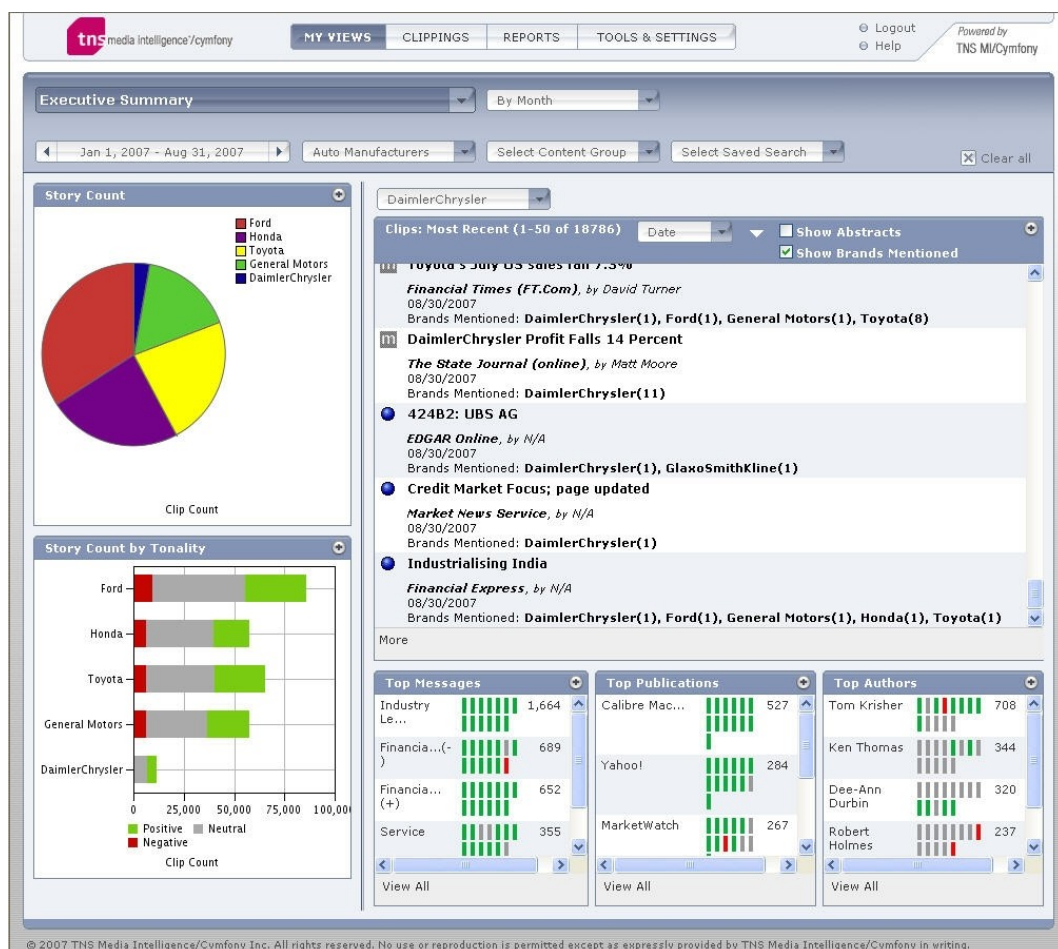
Cymfony monitors 70 million blogs, 30,000 discussion boards and consumer review sites, 40,000 usenet groups and all the major social networking sites and consumer video sites as permitted.

The company claims to monitor over 100,000 traditional media sources. Depending on the third party subscriptions you have or chose to take up, Cymfony will include content from Factiva, Lexis Nexis, Dialog, Critical Mention, Burrelles and other sources to provide a comprehensive view of newspaper, magazine, broadcast, and online editorial sources.

Uniquely, as far as I know, they can also incorporate a customer's own content such as customer emails, call centre records, web-based customer service comments, etc.

Like Cision, the service then applies natural language processing, a form of semantic analysis based on artificial intelligence and computational linguistics, to interpret all the data identified as salient to the search criteria. Importantly, you can set these searches up yourself, or work with your Program Manager and Strategic Analyst to hone the queries, to interpret the data, prepare reports and make recommendations on how to use the findings.

Screenshot – executive summary



Cymfony customers include Mattel, Southwest, DHL, Jiffy Lube, and Microsoft. DHL's Leslie Monreal-Feil, Public Relations Manager, explains why DHL adopted Cymfony:

"We need an effective solution that helps us track and analyze thousands of stories, multiple brands and a wide range of topical issues that affect our business. Cymfony Orchestra makes this possible with its intuitive user interface, customizable reporting and high level of flexibility. It offers easy filtering that allows the user to generate highly tailored reports based on any given need – whether you're tracking a specific issue or analyzing the overall effectiveness of a communications campaign."

Cymfony's Jim Nail, Chief Strategy & Marketing Officer, posted the following observation to the Cymfony blog:

"What should marketers do if the bloom comes off the social media rose? Ignore the anguished cries of VCs and even the doomsday blather that the media may put out. Keep your eye on the people who buy and use your brands. Stay involved with social media because your consumers will be involved with it."

Spot on Jim. Indeed, I'd like to tell you more about the service, but Cymfony was amongst the vendors to stop some way short of answering all my questions. As you know, I can't understand why any vendor of Social Web Analytics services would want to be anything less than fully open and transparent. Perhaps, despite Jim's enthusiasm, they're suspicious of taking part fully in the conversation themselves, so perhaps you'll have better luck getting your questions answered one-on-one.

Of course, my emails may have just slipped between the cracks.

You can email sales@cymfony.com or telephone 001.617.673.6000.

Screenshot – convergence analysis



DNA13

<http://www.dna13.com>



driving great brands



**Chris Johnson,
CEO, DNA13**

Here's dna13's opener:

"Founded in Ottawa, Canada in August 2001, dna13 launched dnaEnterprise – an on-demand Enterprise Communications Management platform with integrated functionality for issue management, media monitoring and analysis."

The private company closed a CDN \$5m investment round in May 2007, and has not been slow in putting the capital to work. In February 2008, dna13 announced Release 7 of their service, adding a number of new features including enhanced item scoring and analytical charting. The company claims more than 100 brands rely on its platform including Manulife Financial, Starbucks Coffee Company, Royal Bank of Canada, Scotiabank, John Hancock Insurance and Fidelity Investments.

The company has chosen to specialise in indexing, analysis and reporting, leaving collection of the data to those that make it their business. Most notably, dna13 and Nielsen BuzzMetrics announced a media content partnership giving dna13 access to the BuzzMetric blog database (and a similar deal has been agreed with Moreover Technologies). It's always interesting to witness this kind of partnership between otherwise competing companies, and more specifically it's interesting to understand what dna13 does with this data that's different to BuzzMetrics.

The service indexes 30 million sources from these third parties, and is striving to increase update frequency from daily to hourly or even more often. The company can process English, Spanish, German, Italian and French for new media, and 75 languages across 200 countries for print.

dna13 does not undertake semantic analysis, so any attempt to quantify sentiment is left to the client to do manually.

The company has two products: dnaEnterprise is their Web-based, "enterprise class, brand and reputation management solution"; and dnaMonitor is their Web-based stand-alone media monitoring solution.

Uniquely for a SWA vendor, as far as I know, dnaEnterprise also includes collaborative workspaces, although I don't believe this system competes head on with leading collaboration software and services from my knowledge of the sector.

"dnaEnterprise promotes topic-specific collaboration among users, agencies and stakeholders, keeping everyone up-to-date and on the same page. Workspaces can be extended to include key spokespeople for issue briefings or global team members' updates on projects in progress. Similarly, PR agencies can extend

campaign workspaces and specific information for client review, comment or approval.”

Speaking of her choice of dnaEnterprise, Beja Rodeck, Senior Manager, Media and Public Relations, RBC Financial Group, says:

“dnaEnterprise has introduced a structured, intelligent and user friendly approach to managing our communications – it has significantly enhanced the overall effectiveness of our media relations department and enable us to proved greater value to the company by integrating the communication functions of our international operations.”

The company did not answer my questions about APIs or libraries, so I’m assuming neither is available.

For information, contact info@dna13.com +1.866.842.1723 (toll free), +1.613.562.0232.

Dow Jones

<http://www.solutions.dowjones.com/insight>

Dow Jones **Insight**



**Clare Hart, EVP,
Dow Jones,
President,
Enterprise
Media Group**

I enjoyed my interaction with Dow Jones. They were more than happy to give me a call and wanted to catch up on how I was doing. They are enthusiastic about the idea of this ebook and possibly the most helpful company here. Delightful.

Factiva was founded in 1999 as a joint venture between Dow Jones & Company and Reuters Group. In 2005 Factiva acquired media-measurement firm 2B Reputation Intelligence Limited and Benchmarking Solutions Limited. Factiva was acquired by Dow Jones in December 2006, and then Dow Jones was acquired by News Corporation in December 2007.

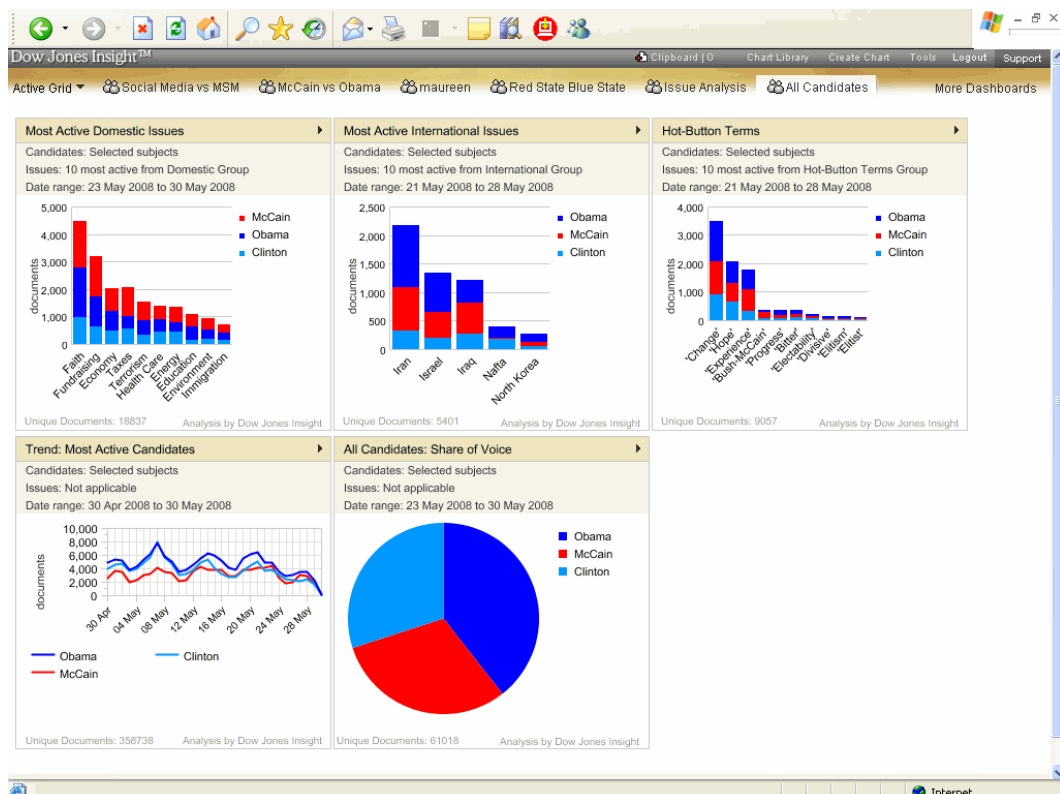
Dow Jones tracks amongst the most languages of any vendor here, including Eastern European and unusually Catalan and Turkish. They index 10,000 of "the most influential" websites (including social

networks as permitted and Wikipedia), millions of blogs and approximately 45,000 discussion boards. Not sure how often they revisit the sources, they simply report “various frequencies”.

The Dow Jones team is currently developing the capability to index the micro-blog services such as Twitter.

The service uses PLSA (Probability Latent Semantic Analysis) to achieve what Dow Jones calls automated favourability analysis; what we call here sentiment or tonality. Their spiders and analysis don’t appear to be quite the standard set by CollectiveIntellect, but appear very thorough nevertheless.

Screenshot – the dashboard



Sentiment and source weighting may be achieved automatically or manually as required, yet interestingly search results will be displayed in (reverse) chronological order only, and not relevance.

Dow Jones is very much the corporate supplier of SWA, if you know what I mean by that. It's a complement and a criticism. They are polished and professional, provide 24/5 customer service and come with all the heritage an organisation could wish for. Clients include Comcast, IMF, Fiserv, Sprint, American Express and Target.

But you won't find much Web 2.0 style thinking here – no APIs or access to libraries, and the dreadful cost per seat approach to pricing.

Contact Peter Smith, Director of Sales, Dow Jones Insight peter.smith@dowjones.com.

Screenshot – the details

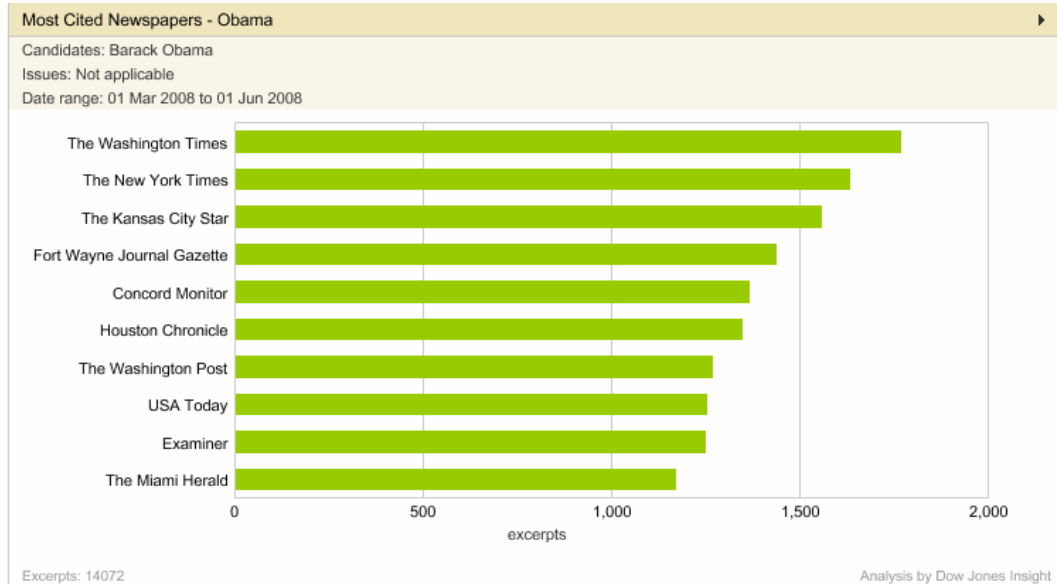
The screenshot shows the Dow Jones Insight interface. At the top, there are navigation tabs for 'Social Media vs MSM', 'McCain vs Obama', 'maureen', 'Red State Blue State', 'Issue Analysis', and 'All Candidates'. The main search details are: Subject: Barack Obama, Issue: 'Change', Date range: May 25 2008 - Jun 01 2008.

On the left, there are two horizontal bar charts. The first, 'Source Types', shows the number of documents from various sources: Web (~650), Press (~450), Blogs (~250), and Boards (~100). The second, 'Source Groups', shows: Newspapers (~550), Blue State Press (~350), Long-Tail Blogs (~250), Magazines (~200), Red State Press (~150), Newswires (~100), Transcripts (~80), Pink States (~50), Gray States (~30), and Government (~20).

Below the charts is a list of sources, including Associated Press Newswires (19), dyn.politico.com (9), The Seattle Times (9), perspectives.com (8), democraticunderground.com (6), Reuters News (7), New West (7), CBS News: 60 Minutes (6), INDS154 (6), and The Politico (6).

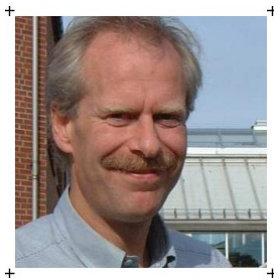
The main content area displays a list of documents. The first document is from 'The Verge' dated Jun 01 2008, with the headline '... even that live moment was low-key. "Y'all ready for change?" he asked. Then he said, "This message has not been endorsed by Barack Obama." It's possible that Jay-Z has flexed only a fraction of his promo muscle because no one is sure what effect his endorsement...'. Other documents include 'Is this goodbye?: He was 'the man from hope'. She was going to be America's first female president. But are the lights finally going out on the Clinton era? Michael Crowley on the future for Hillary and Bill Photograph Cynthia Johnson', 'Obama rewrites the rules for fund-raising', 'Senate rivals, Durbin, take stage in Maine', 'Sunday, June 1: Local firm opposes deregulation', and 'Clinton's tenacity should be applauded'.

Screenshot – most cited



Integrasco

<http://www.integrasco.com>



**Jan Hansen,
CEO, Integrasco**

Some SWA vendors here have breadth, but may lack depth relatively. Integrasco has depth, but does not position itself as having the breadth. As CEO Jan Hansen points out: "... we work high up in the value chain with a few clients that are large leading brands".

It appears that Integrasco goes out and identifies the places the conversation is happening once they have been retained rather than before. They then analyse this dialogue intensively, both technologically and manually. I think I'm seeing a sub-group of such companies forming now, consisting Integrasco, Brandimensions and Kaava.

Jan Hansen is particularly excited at the prospect of distinguishing between content contributed by people in the UK and people in North America. Anyone who has tried to compile a key blogger / commentator list for either of these markets will know how frustrating it is sometimes to determine where this individual lives. Some would argue that it doesn't matter, but it does when your client is paying you

to look after one geography and specifically not another! Sometimes you end up looking for spellings such as "colour" / "color", or "organise" / "organize" to give you a clue.

Screenshot – search

The screenshot displays the Integrasco search interface. At the top left is the logo 'INTEGRASCO WORD OF MOUTH PORTAL'. On the top right, there is a user profile for 'Jan Hansen' with links for 'Profile' and 'Logout', and a 'Market' dropdown menu set to 'Mobile'. Below the navigation bar, there is a search sidebar on the left with options for 'Simple' and 'Advanced' search, a search box containing 'Zn5', and sorting options by 'Date' and 'Score'. The main search results area on the right shows a list of results. The first result is titled 'RE: [MotoZINE Zn5] Un nouveau Motorola' from 'Forumobiles.com', dated 4 days ago. The second result is 'RE: LG launches its third Black Label Series handset: Secret' from 'whatmobile.net', also dated 4 days ago. The third result is 'RE: [vds]nokia 7373 gambatista valli noir' from 'planete-nokia.com', dated one week ago. The fourth result is 'RE: Motorola Announcements' from 'Mobile-Review', dated one week and 3 days ago. Each result includes a brief snippet of the original post and a 'See original post' link.

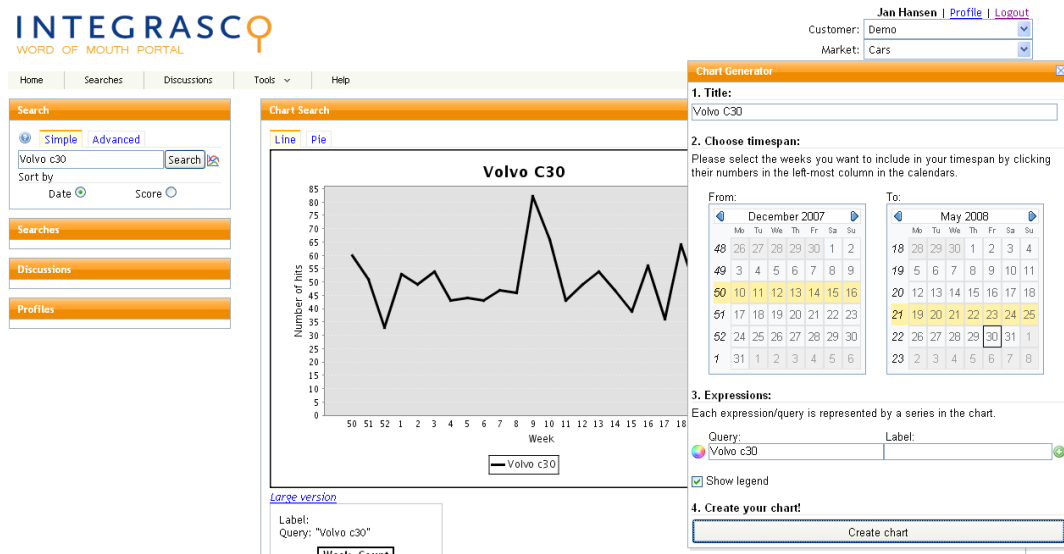
Another development is Integrasco's Automated Product Identifier. Currently in beta, this service learns which products belong to which brand entirely from the analysis of discussion threads, and by matching this to their client's master list can rapidly identify, for example, where information about a new product has leaked.

I'm assuming this requires semantic analysis capability of some variety, but this was not disclosed.

Integrasco focuses on 400 core sources at the moment, all "discussion forums", of which half are English language, and they archive this data for up to two years; far longer than the typical 90 days.

This private Norwegian company does allow export of its data, but does not provide access to charting libraries. Lastly, Jan simply says that pricing is on a per project basis.

Screenshot – stats



Contact Jan Hansen on +47 91 68 40 82 and email Jan.Hansen@Integrasco.no.

Kaava

<http://www.kaavacorp.com>



**Joshua Sinel,
CEO, Kaava**

Kaava has been in business for a long time, at least relative to other vendors here. Starting out in 1996, the company claims to have been the first to offer outsourced online community development, management, and moderation services for companies including AOL, Charles Schwab, CVS, Hallmark, IBM, iVillage, Martha Stewart, MetLife, MTV, and Showtime.

The CEO believes this legacy gives his company the edge, lending it “a unique sensitivity to, and understanding of, the subtle nuances that comprise online consumer dialogue, and the limitations, pit-falls and foundational truths surrounding the use of such organic data for generating consumer observations, findings, insights, and measurements.”

He goes on to list current clients such as American Express, Electronic Arts, Kodak and the Ogilvy family of agencies.

Kaava has “secondary access” (ie, via third parties) to over 60,000 threaded community data sources with topical indexing of over 1,600 such sources; all English language, and do not include the social networks. All data is crawled daily including those sources comprising Kaava’s proprietary topical taxonomy.

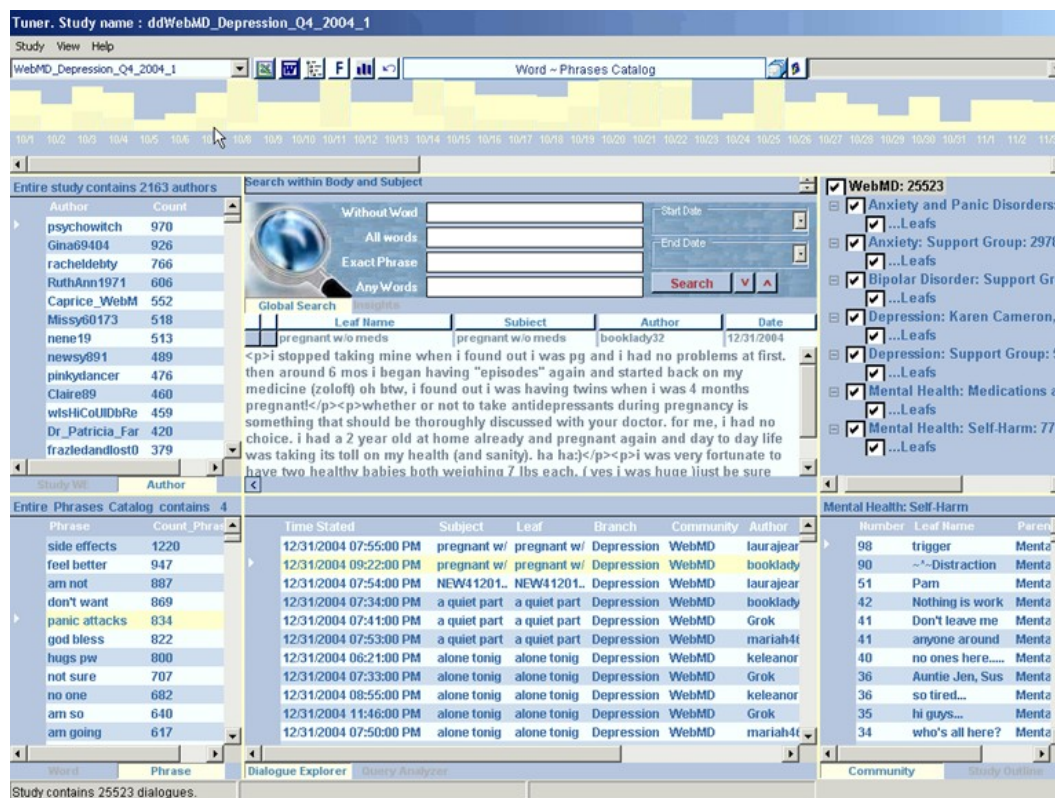
Like CollectiveIntellect, Kaava emphasises topical crawling. Its topical index consists of over 350 top-level topics and their sub-topics, with over 1,600 top-level sources (qualified for topical relevance, activity, and consumer representation) assigned across the taxonomy. These top-level sources are deconstructed and assigned at the conversational thread level to improve quality. Kaava’s upcoming desktop application, the Tuner, is named for its ability to pick out the consumer signals.

Semantic analysis is achieved using a technique Kaava calls Naturalistic Inquiry, which it created based on the collective works of Egon Guba and Yvonna Lincoln.

Based in White Plains, NY, with co-location in Hauppauge, NY, customer service is only available during the East Coast working day. Kaava is currently reviewing its pricing model.

Overall, I remain a little confused as to how Kaava stacks up to the other vendors here. They seem to ‘have something’, yet at the same time I’m not entirely clear from their marketing collateral and responses to my questions what that something is. It may well be that I’m so familiar with the approaches taken by other vendors that I’ve been incapable of getting my head round Kaava’s differential. Naturalistic Inquiry is definitely unique to the company.

Screenshot – the Tuner product



A couple of aspects concern me. Firstly, their marketing includes the statement that "communities are magnets for relevant discussion", perhaps explaining their smaller number of sources. This appears to fly in the face of my assertions in the "Distributed conversations" section somewhat. Secondly, the new Tuner product is a desktop product at a time when most service providers have migrated to or are in the process of moving to browser-based Web services.

I don't believe they should be ignored on this basis however; perhaps you'll have more time to investigate the service in your full time role than I've had during my evenings and weekends. And perhaps their CEO Joshua Sinel will post further explanatory notes.

Joshua wanted me to highlight the following, which I quote verbatim and goes some way to confirming that they may be a different animal:

“Kaava works from project-based custom data sets that are derived from highly targeted subject-matter relevant ongoing discussions, and as such, does not spider and index in the more ‘traditional’ sense that has been defined by the majority of ‘measurement’ companies.

“Kaava’s ongoing data collection and indexing efforts are comprised within its growing topical data library.”

Contact Joshua Sinel, jsinel@kaavacorp.com or 001.914.461.3439.

Magpie

<http://www.brandwatch.net>



**Giles Palmer,
CEO, Magpie**

[Disclosure – Giles is the only CEO here I knew before embarking on writing this ebook, from our mutual participation in www.mashuevent.com.]

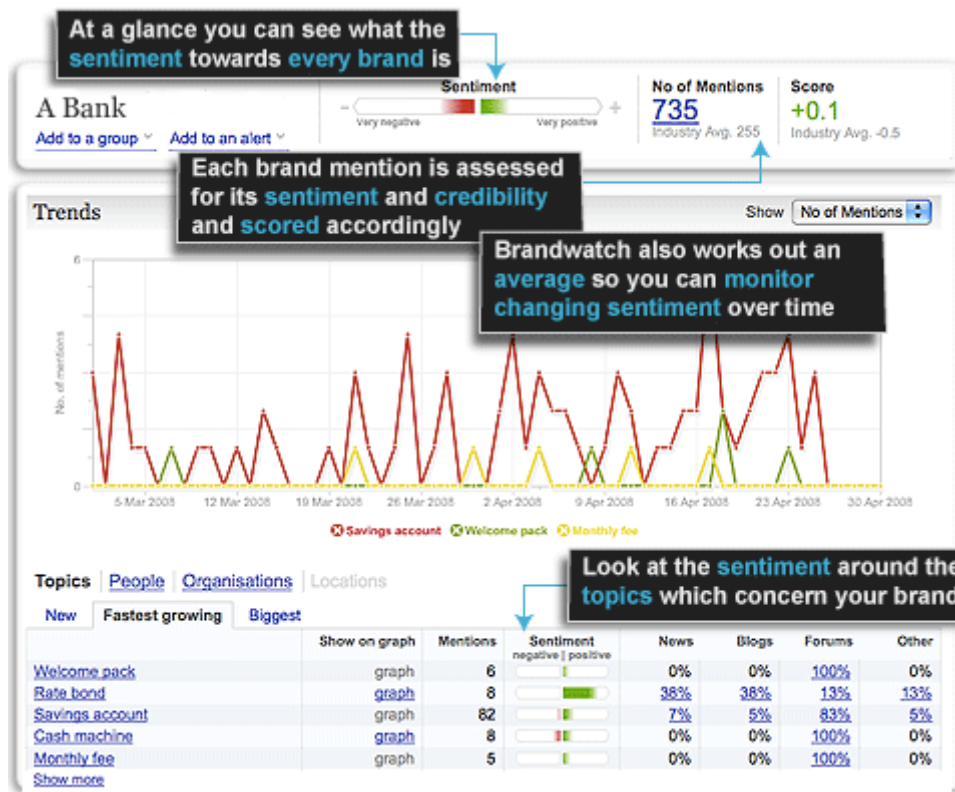
A private company based in London, Magpie's Brandwatch service crawls more than 500,000 English language sites up to 24 times a day. Sites in four other European languages will be added later this year.

The company adds sites in a number of ways: as proposed by customers; following keyword checks across six generic search engines; and identified from popular links in the existing corpus. And they are doing so at a rate exceeding 250,000 per quarter currently.

Brandwatch watches Myspace, Bebo, Facebook et al as permitted. Interestingly, the company is the only SWA vendor here to have pointed out the potential for a brand to invite "friends" so that their social network activity can be monitored. Giles is quick to point out that this must only be on a completely transparent basis, and a

consumer deciding to become a “friend” does not mean that they are then necessarily happy for you to monitor them in such a manner.

Screenshot – the main dashboard explained



Magpie’s spiders are well designed and robust. They are already on to Twitter and Wikipedia, and are acute to spam.

The company prides itself on its semantic analysis capability. It looks like it’s amongst the best of the pack here, so with that knowledge you might want to skip the next quote if this science doesn’t excite you of itself! It’s Giles’ explanation of their approach in this respect:

“Our Semantic Analysis solution is based on a combination of rule-based Natural Language Analysis (for standard NLP functionality) and Machine Learning algorithms (for automation of non-trivial tasks by supervised-learning).

“For standard Natural Language Analysis, we have adopted the GATE framework that has been developed since 1995 and has a proven track record in business use. GATE provides all the low-level NLP functionality (such as tokenisation, sentence detection, and part-of-speech tagging) for our application, as well as has higher-level standard NLP processes (such as Named Entity Recognition). We have refined and customised this framework and fully integrated it into our application.

“For Machine Learning tasks, we use Support Vector Machines and Character-based Probabilistic Language Models as our primary classifier architectures. The classifiers are only as good as the data they were trained on, so we collect and annotate their training data in house, using our own annotation tool and our own data analysts.”

Brandwatch assigns a sentiment score automatically, and this is then weighted by the source’s credibility (from “very low” to “very high”) and the site type (eg, blogs have higher weighting than corporate websites).

Users can then view results by date or relevance, and can filter them by site type, sentiment and source credibility. The user may also set up email alerts with various triggers; should negative sentiment increase uncharacteristically for example.

Data can be exported from Brandwatch in an XML format for ready import into another IT system, and the company is assessing the need for APIs and access to libraries. Giles adds that he’s more than happy to develop this capability on an as-needed basis for larger clients.

Magpie uses London’s Interhouse to offer full redundancy, and Brandwatch comes with a 99.9% uptime SLA. General customer

support is available during UK working hours, but technical support is available 24/7.

Screenshot – brand comparison

A Bank

[Add to a group](#) [Add to an alert](#)

Sentiment

No of Mentions: **735**
Industry Avg. 255

Score: **+0.1**
Industry Avg. -0.5

Groups

View: **Banking** Show: No of Mentions Sort by: **A to Z**

Brand 1 - 20 of 20

Compare your brand to the industry average

Create groups to view select brands side by side and compare by volume or score

Brand (A-Z)	No. of Mentions	Mentions by Site Type					Mentions by Site Credibility					Sentiment	Score
		Blog	News	Forum	Other	V.High	High	Med	Low	V.Low			
Abbey	619	3%	4%	89%	4%	0%	0%	87%	2%	11%		-0.3	
Alliance & Leicester	735	2%	3%	92%	3%	0%	0%						
Bank of Scotland	66	5%	11%	80%	5%	0%	3%						
Barclays	963	2%	3%	93%	3%	0%	0%						
Capital One	752	61%	1%	36%	3%	0%	4%						
Co-operative Bank	315	6%	7%	78%	9%	0%	0%						
First Direct	138	1%	2%	93%	4%	0%	0%						
Nationwide Building Society	62	16%	8%	55%	21%	0%	0%	71%	8%	21%		0.0	
Natwest	520	3%	2%	92%	2%	0%	0%	88%	2%	10%		-1.0	
Northern Rock	195	15%	13%	67%	4%	0%	3%	61%	7%	29%		-0.2	
Royal Bank of Scotland	72	29%	14%	47%	10%	0%	3%	71%	4%	22%		-0.6	
Sainsbury's Bank	26	15%	19%	62%	4%	0%	0%	85%	4%	12%		0.0	
Tesco Personal Finance	1	0%	0%	100%	0%	0%	0%	100%	0%	0%		0.0	
uSwitch (banking)	16	0%	6%	75%	19%	0%	0%	81%	0%	19%		0.0	

Analyse key statistical data across each of your selected brands

Measure where the chatter is emerging, from **Blogs, Forums** or **News** sites and see how influential it is

Edit name of brand group: **Banking** [Delete this Group](#)

Select brands for this group

By Industry | [By Brand \(A-Z\)](#)

- Adhd
- Air-travel
- Au-auto
- Auto
- Banking**
- Beer
- Books
- Business-services
- Car-home-insurance
- Charity
- Child-support

Brands in this group (20 of 50 Max)

- Select All (36)
- Abbey
- Alliance & Leicester
- ASDA Personal Finance
- Bank charges
- Bank of China
- Bank of Scotland
- Barclays
- Britannia
- Cahoot
- Capital One

[Save changes](#) [Cancel](#)

The commercials are per brand per month, but there's also reference to per user in the company's response to me. I can only interpret this as the traditional "per seat" approach I feel is archaic. Magpie offers agencies significant discounts for multiple campaign tracking, and will white label as required. Ben Bose, Content & Media Strategist, iCrossing, is one such white label customer:

"iCrossing has built a reputation as a top quality research and strategy agency within social media, and we need to make sure that we use the highest quality and most innovative tools to support our work. Brandwatch provides an excellent data and analytics source, and it has been developed in line with our, and, more importantly, our clients' needs. Magpie is a dedicated and knowledgeable company, and a pleasure to partner."

Magpie customers include HBOS, Alliance+Leicester, River Island, Sony and Orange.

Contact Dominic Frost, Commercial Director, dominic@brandwatch.net.

Nielsen BuzzMetrics

<http://www.nielsen-online.com>



**David Calhoun,
Chairman & CEO,
The Nielsen
Company**

Formed by companies founded in the late 1990's, Nielsen invested in BuzzMetrics in 2006 and purchased 100% of the company in 2007. Today, the BuzzMetrics service is part of Nielsen Online, itself part of the privately held The Nielsen Company.

The company provides SWA services to "more than 100 top advertisers and agencies, including Microsoft, Toyota and Sony."

The service indexes 40 million blogs, 7,000 message boards and 45,000 groups in three dozen languages. Incredibly, the company claims to be adding 80,000 blogs a day! Customers can add and remove sites from their crawls. The company does not index blog comments however.

The BuzzMetrics spiders appear to be amongst the best here, and whilst they visit the main general social networking sites as permitted, they don't yet go knocking at Wikipedia or the micro-blogs like Twitter.

BuzzMetrics focuses on breadth but not, it seems, depth of analysis – they have disclosed no semantic analysis capability to me. As such, you cannot sort search results in order of relevance or sentiment, but in date order only. Reports are available for message count, most active author, trending, top cited post, top board, top blog, top author and top cited news article or source. Whether the mindset is swinging in your favour however, or your competitor's, is left unanswered (or more precisely "un-estimated").

BuzzMetrics does come with a basic search API, but does not facilitate remote operation from another system or access to charting libraries.

The Tampa Florida company operates a redundant network with a promise of 99% uptime, working hours telephone support, and 24/7 email support. The licensing basis was not disclosed to me.

Contact Brad.little@nielsen.com.

Radian6

<http://www.radian6.com>



**Marcel LeBrun,
CEO, Radian6**

Major Radian6 clients include Weber Shandwick, GolinHarris, DraftFCB Healthcare, Carmichael Lynch, SHIFT Communications, Voce Communications, Social Media Group, Doe Anderson, Moosehead, Aliant, Universal McCann. Interestingly, the customer list leads with agencies.

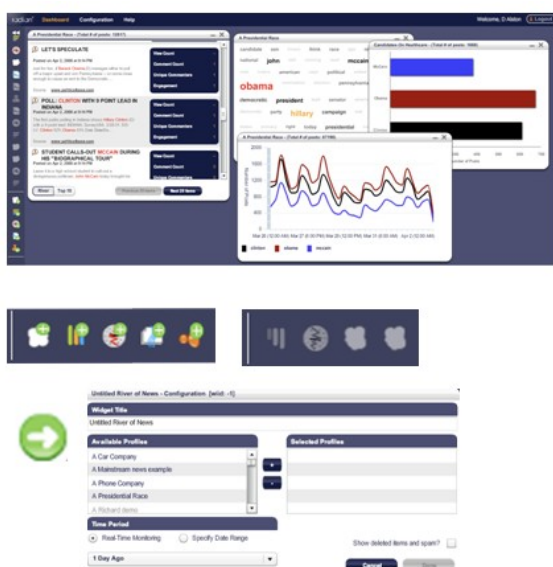
Unfortunately, this privately held Canadian company wasn't particularly forthcoming with information (I don't believe the email attachment with my questions was forwarded internally). So the information here is taken from their company PowerPoint and PDF brochure (including the screenshots if you think them a little hazy).

The source of their data is particularly intriguing. Whilst the company does have a [webpage explaining their crawlers](#)³⁵ to interested webmasters, it appears to focus on RSS updating and comment scraping. They supplement their spidered data by buying source data

³⁵ <http://www.radian6.com/crawler>

in from third parties such as Nielsen BuzzMetrics and [Moreover Technologies](#). You can learn from their corporate PDF that they index over 100 million blogs, top video and image sharing sites (see Blinkx below), over 20,000 mainstream media sources, and “new classes of social media such as Twitter”.

Screenshot – dashboard



I note from their news pages that they have recently partnered with [Blinkx](#)³⁶, the company that bills itself as “the world’s largest and most advanced video search engine”. So the video content Blinkx finds and transcribes should then show up in Radian6’s search results.

Their pricing model is described briefly in their corporate PowerPoint:

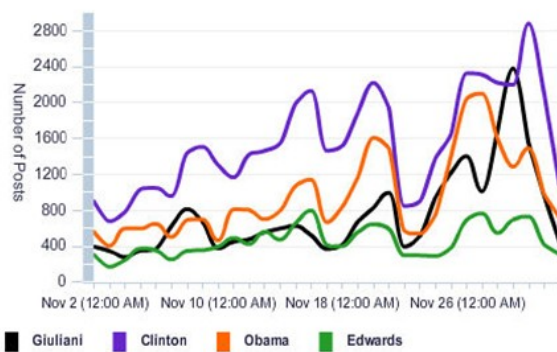
“Unlimited Keywords – Client pays for results not the parameters used. Monthly results volumes can be estimated in advance in order to calculate billing.

³⁶ <http://www.blinkx.com>

“Scalable Pricing – Pricing starts at \$100 / user / month and \$500 / topic profile / month for up to 10,000 items found. Most clients fall into this scenario. Billing commitment is only month to month with discounts for longer term commitments or larger rollouts. 7 day free trial on all new topic profiles.”

Anyone patient enough to have read this whole ebook so far will know by now that I think Radian6 has the billing approach spot on, except for the per user part! Now if only I could find out about their API and library approach.

Screenshot – topic trends



Indeed, if anyone can throw more light on Radian6, please post about your insights. I hope Radian6 can find the resource to post their response to my questionnaire in the coming weeks and participate more fully next year.

Contact inquire@radian6.com.

Vocus

<http://www.vocus.com>



I found Vocus reluctant participants in this ebook, but I have been able to get some information together.

This public company has a well established heritage in traditional media, claiming over 2,600 customers worldwide. They appear to have been caught on the back foot however in respect to SWA. For example, whereas many SWA vendors here index hundreds of thousands of social media sources, if not millions, Vocus claims just under thirteen thousand.

The sources are updated at a frequency aligned to an assessment of their importance, with the most important revisited several times daily, and it appears that the majority of SWA sources are collated via a third party.

Vocus does semantic analysis based on natural language processing (NLP), and they use this to determine sentiment. They also assess what they call "prominence", which appears to be a function of who said what where. A big blogger posting on a big blog equals substantial prominence. The overall "weight" of an item is then a product of both its sentiment and prominence.

Vocus prices on a per seat basis.

###

The conversation continues at www.socialwebanalytics.com.