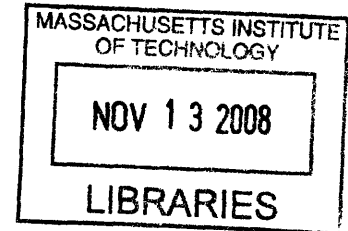


Social Media Analytics in Business Intelligence Applications

by

Bobby Lo



Submitted to the Department of Electrical Engineering and Computer Science
in Partial Fulfillment of the Requirements for the Degree of

Master of Engineering in Electrical Engineering and Computer Science
at the
Massachusetts Institute of Technology

May 2008

©2008 Bobby Lo. All rights reserved.

The author hereby grants to M.I.T. permission to reproduce and to distribute publicly paper and electronic
copies of this thesis document in whole and in part in any medium now known or hereafter created.

Author [Signature]
Department of Electrical Engineering and Computer Science
May 8, 2008

Certified by [Signature]
Michael A. Cusumano
Thesis Supervisor
Sloan Management Review Distinguished Professor of Management & Engineering Systems

Accepted by [Signature]
Arthur C. Smith
Professor of Electrical Engineering
Chairman, Department Committee on Graduate Theses

ARCHIVES

Social Media Analytics in Business Intelligence Applications

by
Bobby Lo

Submitted to the Department of Electrical Engineering and Computer Science
on May 8, 2008 in Partial Fulfillment of the
Requirements for the Degree of Master of Engineering
in Electrical Engineering and Computer Science

Abstract

Social media is becoming increasingly important in society and culture, empowering consumers to group together on common interests and share opinions through the Internet. The social web shifts the originators of content from companies to users. Differences caused by this dynamic result in existing web analytic techniques being inadequate.

Because people reveal their thoughts and preferences in social media, there are significant opportunities in business intelligence by analyzing social media. These opportunities include brand monitoring, trend recognition, and targeted advertising. The market for social media analytics in business intelligence is further validated by its direct application in the consumer research market.

Challenges lie ahead for development and adoption of social media analytics. Technology used in these analytics, such as natural language processing and social network analysis, need to mature to improve accuracy, performance, and scalability. Nevertheless, social media continues to grow at a rapid pace, and organizations should form strategies to incorporate social media analytics into their business intelligence frameworks.

Thesis Supervisor: Michael A. Cusumano

Title: Sloan Management Review Distinguished Professor of Management & Engineering Systems

Table of Contents

1	Introduction.....	9
2	Research Question and Methodology.....	11
3	Traditional Business Intelligence.....	13
3.1	Definition and History.....	13
3.2	Applications	15
3.3	Business Intelligence Components.....	15
3.3.1	Extract, Transform, and Load	16
3.3.2	Data Warehousing	16
3.3.3	Analysis.....	17
3.3.4	Reporting.....	17
3.4	Industry Growth.....	18
4	Analytics for the Information Web.....	19
4.1	History	19
4.2	Measurement Techniques.....	19
4.2.1	Web Server Log Analysis	19
4.2.2	Embedded JavaScript Tagging.....	20
4.2.3	Custom Integrated Solution	21
4.2.4	Remote Sources / Panels.....	22
4.3	Market Size	22
4.4	Major Players.....	23
4.4.1	Alexa (Toolbar)	23
4.4.2	comScore (Panel)	24
4.4.3	Compete (Multiple)	25
4.4.4	Google Analytics (JavaScript)	26
4.4.5	Omniiture.....	27
4.4.6	Nielsen (Panel).....	27
4.4.7	WebTrends	28

5	Evolution of the Internet – The Social Web.....	31
5.1	Social Media Landscape – Types of Social Media	32
5.1.1	Blogs	32
5.1.2	Forums.....	34
5.1.3	News Aggregation	37
5.1.4	Social Bookmarking	39
5.1.5	Social Networking.....	40
5.1.6	Wikis	44
5.2	Social Media Trends.....	45
6	Social Media Analytics	47
6.1	Differences from Web Analytics	47
6.2	Brand Monitoring	47
6.2.1	Brand Integrity.....	48
6.2.2	Brand Sentiment	49
6.3	Trend Recognition	54
6.4	Targeted Advertising	55
6.4.1	Advertising Market Trends.....	55
6.4.2	Advertising in Social Media.....	56
6.4.3	Business Intelligence Applications	58
6.5	Consumer Applications	59
6.5.1	Consumer Trends	59
6.5.2	Product Recommendations	59
6.5.3	Network Aggregation	60
6.5.4	Reputation Monitoring.....	60
7	Relation to Consumer Research Services	61
7.1	Traditional Consumer Intelligence	61
7.2	Impact of Online Analytics.....	62
8	Technological Challenges.....	67

8.1	System Functions	67
8.1.1	Data Acquisition	67
8.1.2	Data Storage	68
8.1.3	Data Analysis	69
8.1.4	User Interface / Visualization	69
8.2	Analysis Modules.....	70
8.2.1	Natural Language Processing (Content Analysis).....	70
8.2.2	Social Network Analysis	72
8.2.3	Other Modules.....	73
9	Companies in Social Media Analytics.....	75
9.1	Biz360.....	75
9.2	BuzzLogic.....	76
9.3	Nielsen BuzzMetrics	78
9.4	Radian6.....	80
9.5	TNS Media Intelligence/Cymfony	82
9.6	Visible Technologies.....	83
10	Summary and Conclusion	85
11	References	89

Figures

Figure 1 - Alexa Daily Reach Comparison Chart.....	24
Figure 2 - Compete Monthly Visits Comparison Chart	25
Figure 3 - Google Analytics Dashboard	26
Figure 4 - Omniture Dashboard.....	27
Figure 5 - WebTrends Dashboard.....	29
Figure 6 - Blog Example.....	33
Figure 7 - Forum Topic Listing.....	35

Figure 8 - Digg Front Page.....	37
Figure 9 - Digg Top Stories Sidebar	38
Figure 10 - del.icio.us.....	40
Figure 11 - MySpace Home Page	42
Figure 12 - Wikipedia article on social media.....	45
Figure 13 - Biz360 BlogView Brand and Reputation Dashboard.....	76
Figure 14 - BuzzLogic Social Map Visualization.....	77
Figure 15 - BuzzLogic Ad Targeting Interface	77
Figure 16 - BuzzMetrics Brand Association Map Example	79
Figure 17 - BuzzMetrics BlogPulse.....	80
Figure 18 - Radian6 Dashboard	81
Figure 19 - Cymfony Dashboard.....	83
Figure 20 - Screenshot of TruCast's Response Manager	84

1 Introduction

Decision support systems are widely used for business intelligence, providing technology-based, data-driven analytics. The information produced from these systems guide business decisions with information such as market analysis, performance metrics, and consumer trends. As an example, in customer analytics, data on customer behavior help companies learn more about specific markets. This information can then be applied to product development and marketing.

Prior to widespread use of the Internet, analytics companies aggregated data for business intelligence from offline data sources such as company sales data, print media, and surveys. Obtaining some of this analytics information might involve paying research firms a considerable sum of money and receiving results over a few months. Specialized marketing research firms monitored specific industries and sold specialized reports to companies in those industries. One example is the Nielsen Ratings, which conducted regular surveys of households to learn about the audiences of television programming.

The Internet revolutionized society and culture by making an unprecedented scale of information available to people through a widespread, connected network. Even in its early stages, the Internet proved to be a disruptive technology with email and instant messaging allowing people to exchange information quickly and efficiently compared to mainstream forms of communication at the time. As the Internet has evolved rapidly, society and culture have also evolved to accommodate the new technologies. As society changes with technology, the analytics industry must also evolve to serve companies with data needed to make key business decisions.

The Internet brought in a wealth of new information and evolved the analytics industry with the introduction of web analytics. Now, individual companies could track behavior from users browsing their websites. Companies could then analyze the captured data using web analytics software to determine areas of interest on their website and a range of customer

demographics. The web analytics industry also introduced tools to allow comparisons on a global level. Alexa, characterizing itself as a “web information” company, provides estimates of page visits for the top websites, creating a popularity rank.

With the explosion of online social media, commonly known as “Web 2.0,” people are interacting directly with each other and sidestepping traditional silos of information on the Internet. Social media is very powerful because they bring together people with things in common and encourage the exchange of ideas. Now in addition to tracking consumer behavior on company websites, companies can now see what users are saying on the Internet and receive feedback almost in real time. History in the software industry has shown that the best places to look for new software product and service opportunities are new platforms and niche applications (Cusumano, 2004).

The focus of this paper is to explore and identify opportunities for business intelligence applications in the area of social media analytics. This paper will examine the market opportunities in social media analytics, the current approaches to analytics, and the key technological drivers which enable such an approach to analytics.

The paper begins with an introduction on business intelligence, decision support systems, and web analytics by looking at the history of the industry, major players, and service offerings. The following section explores the transformation of the Internet as the social web emerges, looking at the different types of social media. Once the context of social media has been established, the paper looks at applications for social media analytics and later on profiles several companies in this space.

Because social media analytics are strongly technology dependent, one section of the paper explores the technological challenges. Many specialized technologies are key to the success of such a service, such as network analysis algorithms for processing opinions found online.

2 Research Question and Methodology

The subject of this paper is opportunities for business intelligence in the area of social media analytics. Social media is relatively new and past analytics methods appear to be ineffective in capturing the available information. The key research question is to determine whether opportunities exist in social media analytics and then explore the technical challenges of performing such analytics.

Therefore, the research methodology is structured as follows:

1) Understand the current business intelligence industry and structure

Information on the current market will provide context of the analytics industry. This section includes a history of business intelligence and decision support systems and how the web analytics industry formed. Profiles of major companies in the web analytics industry provide insight into service offerings.

2) Define social media and evaluate the social media landscape

In order to determine if social media will have a strong impact on companies, one must first define social media, see how social media varies from traditional media, and then look at the implications caused by the growth of social media. Major companies for each type of social media serve as examples. Trends and market research will help in understanding how much of an impact social media will have now and in the near future.

3) Explore opportunities in social media analytics

If social media will have a significant impact on society, then the next step is to see what opportunities are created by social media that are different from traditional web analytics. This section will explore potential analytics applications that can be made possible using the content found in social media.

4) Evaluate technical requirements in analyzing social media

Since the content and format of social media are very different from other media, new approaches must be taken to effectively analyze the media and extract meaningful results. With technology being such an important factor to the success of social media analytics, it is important to understand what technologies are required and what challenges lie ahead.

3 Traditional Business Intelligence

3.1 Definition and History

In 1958, IBM researcher H. P. Luhn realized that information was being generated at an increasing rate such that automation would be necessary to manage information. In his paper, *A Business Intelligence System*, Luhn introduces an automated system that would automatically abstract information from documents, encode the extracted information in meaningful ways, and then disseminate the information to relevant parties (Luhn, 1958). Luhn defines the term *action point* as “specific activities carried out by individuals, groups, departments, divisions, or even larger units.” Therefore, the goal of a business intelligence system is to supply relevant information to support these action points in a quick and efficient manner through the use of technology.

Luhn’s system became part of the much larger research field of decision support systems, where in the 1960s and 70s, researchers were experimenting with computerized quantitative models for decision making. In 1962, Jay Forrester at MIT’s Sloan School of Management led the System Dynamics Group and was instrumental in the development of one of the first data-driven decision support systems, which turned out to be an air defense system (Power, 2007). By the 1970s, the cost of mainframes had fallen enough for large companies to invest in decision making systems. As a result, the field of decision support systems took off, with the term *decision support systems* appearing for the first time in a 1971 Sloan Management Review article. A formal definition of decision support system is “a class of information system that draws on transaction processing systems and interacts with the other parts of the overall information system to support the decision-making activities of managers and other knowledge workers in organizations” (Sprague & Carlson, 1982).

In the following years, data-driven DSS known as executive information systems and executive support systems became increasingly popular. Database companies like Teradyne

allowed the construction of data warehousing systems, where analysis could be performed on large collections of historical data. New analysis techniques and technologies, such as Online Analytical Processing (OLAP) allowed companies to perform analysis at a much deeper level. By 1985, Procter & Gamble had a DSS that linked sales information and retail scanner data (Power, 2007). These data-driven DSS led to the widespread use of the term *business intelligence*. Business intelligence is defined as “the process of turning data into information and then into knowledge” (Golfarelli, Rizzi, & Cella, 2004).

As technology advanced rapidly through the 90s, so did business intelligence. New database technologies from companies like Oracle allowed ever-increasing data warehouses. Wal-Mart already had over 5TB worth of data warehouse storage in 1995 to mine customer sales data. Around that same time, the Internet was also affecting business intelligence.

The Internet transformed the business intelligence market by bringing more applications online. Widespread use of the Internet and the thin-client characteristic of web browsers allowed application service providers to host the DSS infrastructure remotely and create centralized online portals for enterprise customers, which offered services such as knowledge management, business intelligence, and collaboration services.

Today, business intelligence plays a major role in the operation of businesses. Because BI is a very diverse field with an innumerable number of applications, the market continues to grow and new solutions are being developed to analyze different data. Companies such as Business Objects, SAS, and Oracle offer large BI platforms and continue to be major players in this industry. For this paper, we look at how analytics for social networks can be applied to business intelligence.

3.2 Applications

Because business intelligence has the broad definition of turning data into knowledge, there are many possible applications for BI. Business Objects classifies its BI applications into three broad categories (Business Objects, 2008):

- **Enterprise performance management**

EPM applications deliver performance data and metrics to help workers monitor the performance of an organization and make strategic decisions. Company goals and best practices can be monitored and shared among members of the organization.

- **Information discovery and delivery**

This class of reporting applications help workers find and use information available within the organization. These applications provide querying and analysis tools to dive into data, with reporting capabilities to deliver the end result.

- **Enterprise information management**

EIM applications integrate information from sources across the enterprise and make the relevant knowledge available to decision makers. These applications may be used for real-time data processing and integration to provide up-to-date information.

3.3 Business Intelligence Components

In order to produce meaningful results from large collections of data, business intelligence usually involves a multi-step process as the initial data is obtained, processed, analyzed, and then output to the end user. Hence, BI systems involve a set of components which are tightly integrated.

3.3.1 Extract, Transform, and Load

In BI systems, Extract, Transform, and Load (ETL) technology is responsible for obtaining data from internal and external sources and converting the data into a format that can be stored in the system by the data warehouse. During the extraction phase, ETL may pull relevant data from internal sources such as Customer Relationship Management (CRM) systems, Supply Chain Management (SCM) systems, Enterprise Resource Planning (ERP) systems, and from various external sources of data, depending on the task (Bitpipe, Inc., 2008). Because the data obtained from extraction may be in different formats, ETL technology is responsible for transforming the data to fit the format defined by the target dataset. In some cases, some aggregation of data may occur based upon the rules placed on the data. Finally, ETL loads the transformed data into the data warehouse for analysis (Karel, 2007).

Because ETL is responsible for managing the massive amounts data and preparing them for analysis, it is estimated that ETL planning may take up to 70% of the time and effort in a data warehousing strategy (Kimball, 2004).

3.3.2 Data Warehousing

A data warehouse can be defined as a “subject-oriented, integrated, time-varying, non-volatile collection of data that is used primarily in organizational decision making” (Inmon, 1992). In other words, these data warehouses store the large volumes of data processed by ETL. Typically data warehouses consist of a clustered network of storage systems running relational database systems with high availability (Aertsen, 2007). Since data stored in the warehouses are not expected to change after being written, the requirements differ between operational databases and the databases used for warehousing. For example, on-line transaction processing involves a varied mix of reads and writes to the database while warehousing supports more of a write-once model. Due to these functional differences, the

data warehouse is typically a separate component from an organization's operational databases (Chaudhuri & Dayal, 1997).

3.3.3 Analysis

Data in warehouses are typically modeled in a hierarchical multidimensional format allowing for analysis on many different criteria. On-line analytical processing (OLAP) presents data in the form of "cubes," allowing aggregation along one or more dimensions (Chaudhuri & Dayal, 1997). As a result, OLAP of data stored in warehouses can provide a wealth of information that can be easily summarized. For example, a retailer might use OLAP processing on sales data to quickly roll up individual transaction records into a summary of sales volume over time. While the concept of OLAP cubes may sound simple, each roll up operation involves an underlying SQL query into the database. As aggregation or drill-down criteria increase in dimensions, the SQL database query increases significantly in complexity. As a result, highly complex analysis may end up taking a significant amount of time to process.

Another type of analysis that has been developing in recent years is predictive analysis based upon data mining techniques. Whereas OLAP analyses look at historical data, data mining is forward-looking, allowing for discoveries of interesting correlations in the data. One example of data mining use is to predict what other products complement a specific product when people are shopping in a retail store.

3.3.4 Reporting

The reporting component in a BI system is where the end-user interacts with the entire system. A user interface allows the end-user to review current reports that have been generated by the system. Reports are defined by specific criteria and may be recurring. For example, a sales organization might have a sales report that runs every week to roll up current sales data. In some instances the end user may wish to run a one-time custom report. Although the underlying data used in reports may be complex, the reporting interface should

allow for the easy design of reports. The results of these reports can then be exported for further analysis.

Another key type of reporting in BI systems is a dashboard report. These reports typically provide overall visual indicators of company health based upon defined key performance indicators (KPIs). This control panel-type view allows users to quickly identify underperforming areas in an organization.

3.4 Industry Growth

In 2006, the worldwide business analytics market was \$19.3 billion, growing 11.2% from the previous year (Vesset, 2007). Analysts at IDC and Gartner predict the market to continue to grow over the next five years at a CAGR of 8-10%. The top six firms in order of revenue were: Oracle, SAS, SAP, IBM, Microsoft, and Business Objects. Each had 2006 revenues of over \$1 billion, with the top six representing a combined total of 47.4% of the global market.

Analysts at IDC currently see two key growth areas over the next decade. First, more data will be collected by organizations and analyzed through business analytics tools. As organizations upgrade technologies to collect additional data, analytics tools must be optimized to process the combinations of structured and unstructured data, with more rich media content included over time. Second, business intelligence tools are being made available to a wider group of people within organizations. Previously, these decision support systems were available only to the highest levels of management. Today, organizations are finding increased value in offering analytic tools to more staff.

4 Analytics for the Information Web

4.1 History

During the beginning of the Internet, only basic website usage reporting tools were available. These tools would analyze web server logs and then calculate basic page visit statistics. Using these tools, site operators would have an approximation of the number of visitors to the website, visiting patterns, and an idea of popular pages on the website.

As businesses started using their online channel strategically to drive sales, marketing, and product distribution, new web analytics solutions emerged that would help measure the online performance of a company and business impact for various departments within the company. For website designers, web analytics would provide feedback on the usability of the site and allow further engineering to enhance the user experience online. For the marketing department, marketers could measure the success of campaigns by gauging interest on relevant areas of the website. Business development could observe what online affiliates were providing the most value to the company. These are some examples of business intelligence that web analytics can provide.

4.2 Measurement Techniques

4.2.1 Web Server Log Analysis

There are three major web analytic techniques. The most basic method is web server log file analysis. When a customer visits a website, the web server makes a record of each data transaction that occurs between the client and server.

These logs generally contain the following data fields:

- IP address of client – identifies the internet service provider of the user, which can yield additional geographical information.

- Time of access – allows the analyzer to piece together a visit history of the user and aggregate hit data across all visitors to determine peak times
- Page or object requested – tracks which pages the user visits
- Referrer – provides information on how the user found the website; for referrals from search engines, the search term is commonly included
- Internet browser information – shows what internet browser the visitor is using

Here is an example of a typical entry for the Apache web server:

```
18.95.5.102 - - [25/Feb/2008:03:31:23 -0500] "GET /index.html HTTP/1.1" 200 1507
"http://www.google.com/search?q=MIT" "Mozilla/4.0 (compatible; MSIE 7.0; Windows NT 5.1)"
```

In this example, the visitor's IP address is 18.95.5.102, an address that can be traced to Cambridge, Massachusetts, USA. The visitor, using Microsoft Internet Explorer 7.0 on Windows, landed on the main page of the server's web site (index.html) after clicking on a link returned by Google for the search term 'MIT'.

4.2.2 Embedded JavaScript Tagging

JavaScript tracking involves embedding JavaScript code within one or more pages of websites. Most JavaScript tracking solutions embed the code within all pages of a website. When done this way, this method of tracking can offer a superset of the information captured in web server log analysis.

Use of JavaScript offers three major advantages. First, the JavaScript tracker does not need to reside on the same web server hosting the website, allowing analytics to be offloaded to a separate server or a third-party service. Second, the tracker allows greater discretion on what information to capture and on which pages. Lastly, use of programmatic JavaScript allows

more information about the user to be captured, which includes an option of setting a cookie to identify the visitor in the future.

In addition to the information that can be obtained from web server log analysis, JavaScript tracking can record information about the visitor's screen resolution. Custom events can be programmed into a web page to determine whether or not a user interacted with a specific part of a website. Finally, a remotely hosted JavaScript tracker has the potential of recording browsing information for a particular user across multiple websites through the use of cookies.

The main disadvantage of using JavaScript tracking is not receiving any tracking data if the user has disabled JavaScript.

4.2.3 Custom Integrated Solution

Using a custom designed solution offers the greatest flexibility in defining measurements but may introduce a significant level of complexity in implementation. In a custom solution, a web developer can add the tracking code directly in the server-side webpage code, such that tracking data is recorded while the web server is generating the output to the client. A custom solution may include JavaScript tracking code on the client side to obtain data about interactions on a specific web page.

A custom solution has the capability of tying web browsing information with information from other sources. For example, when a user visits an e-commerce website, the tracking solution can record information about the products that the user has viewed. When the user creates an account, that browsing history can be linked to the user's personally identifiable information. A tracking cookie stored on the user's computer helps the website identify future visits by the user. With the web browsing data and the purchasing data combined, the company can then mine the data using business intelligence tools to identify patterns and optimize recommendations.

4.2.4 Remote Sources / Panels

In the field of web analytics, competitive data can provide insight into the position of a company's online presence relative to other websites. However, companies generally will not freely expose their page visit data to others. Therefore, alternative sources of information must be used to estimate this data.

One popular method of obtaining data is to install a tracking device on computers, recording the websites that users visit. These programs commonly come in the form of Internet browser toolbars which offer users some unique functionality in exchange for the user's browsing history and some demographic data. Other opt-in panels compensate users in exchange for more intrusive tracking software.

Another source of data is the logs of Internet Service Providers (ISP). Since all public internet traffic must go through an ISP and traffic is usually unencrypted, some ISPs have chosen to install hardware which extract the addresses of the websites that customers request and store statistical data on traffic to those sites. ISPs then sell the aggregated data to web analytics providers.

4.3 Market Size

The web analytics market peaked during the dot-com era and collapsed during the burst, but the market has steadily grown since 2003, when companies started shifting from tactical to strategic use of their web channel (Gassman, 2006). Companies are in the process of integrating the online channel with other channels, resulting in web analytics with cross-channel attributes. For example, a retailer's website may drive sales in the retail channel, and salespeople might prompt users online for their telephone numbers to make an up-sell.

There are many vendors in the web analytics industry, but the overall market size is only about half a billion dollars, which indicates that the market is currently undervalued and in a growth stage. Analytics strategies and delivery models vary by vendor, ranging from

specialized, vertical integration companies which incorporate web analytics on-site, and broader analytics offerings using a hosted, software as a service model.

Key issues that are currently hindering growth of web analytics include poor user experience, difficulty of integration, and failure of management to properly define key performance indicators for tracking. A March 2007 survey of organizations using web analytics organization showed that only 15% of organizations had fully integrated web analytics data into the decision-making process (Peterson & Bayriamova, 2007). The survey also determined the top five concerns about web analytics to be: ability to use data for decision-making purposes (65%), accuracy of the data (55%), depth of understanding of the data throughout the organization (48%), quality of analysis (43%), and ease of use for reporting solution (41%). The overall conclusions of the survey found that web analytics still faces tremendous obstacles for adoption in organizations, with customers finding web analytics to be difficult to use with questionable accuracy and unclear return on investment.

4.4 Major Players

This section represents major players in the web analytics market with the companies chosen to represent the various technologies and service offerings available in the industry. Of the companies chosen, Alexa, comScore, Compete, and Nielsen offer services that allow comparison across multiple websites because their data sources are external to target websites. Google Analytics, Omniture, and WebTrends offer detailed web analytics for site owners, meaning that the information is based on data gained internal to the website, through embedded JavaScript tags or web server logs.

4.4.1 Alexa (Toolbar)

Now owned by Amazon.com, Alexa Internet introduced one of the first website recommendation engines in 1996, powered through a toolbar that was installed in major Internet browsers. Alexa's toolbar tracks the pages that users visit and aggregates the data to

produce its traffic rankings on websites. In return, toolbar users receive suggestions on websites related to the current site being browsed. Alexa does not release the number of toolbar users, though the company claims that the number is in the millions (Alexa, 2008).

On its website, Alexa provides several statistics on websites:

- Reach – defined as the percent of global Internet users that visit the website
- Page Views – average number of pages viewed per user per day
- Traffic Rank – a ranking calculated from Reach and Page Views to determine a website's overall rank relative to other websites

Tools also allow traffic and trend comparisons among multiple websites.

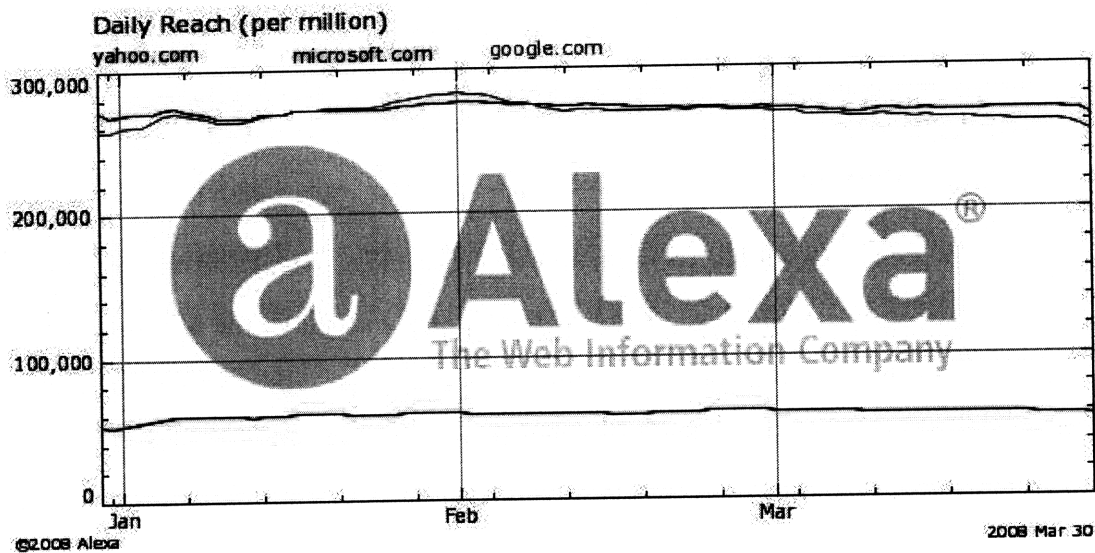


Figure 1 - Alexa Daily Reach Comparison Chart

4.4.2 comScore (Panel)

comScore describes itself as a "global Internet information provider" providing insight into consumer behavior that can be applied to marketing, sales and trading strategies (comScore, Inc., 2008). The company obtains its data on consumer behavior through an opt-in panel of

over 2 million Internet users. These users install special monitoring software on their computers that give comScore access to web browsing and purchasing behavior.

comScore Media Metrix is comScore's measurement product, which provides statistical data such as site traffic, audience demographics, internet trends, and competitive intelligence. comScore also offers a full-service marketing product which can tap into its panel for information such as brand analysis, industry benchmarks, and consumer segmentation. These reports would be compiled based upon demographic data provided by panelists, web usage data, and custom surveys given to panelists.

4.4.3 Compete (Multiple)

Now owned by TNS (a large BI and market research service), Compete offers web analytics similar to Alexa. In addition to a base of opt-in users via its toolbar that tracks browsing data, Compete also licenses consumer data directly from Internet Service Providers, which provides less biased data. Measurements provided by Compete are similar to Alexa, but Compete also offers "Search Analytics" which are top search keywords for specific websites or categories (Compete, 2008).

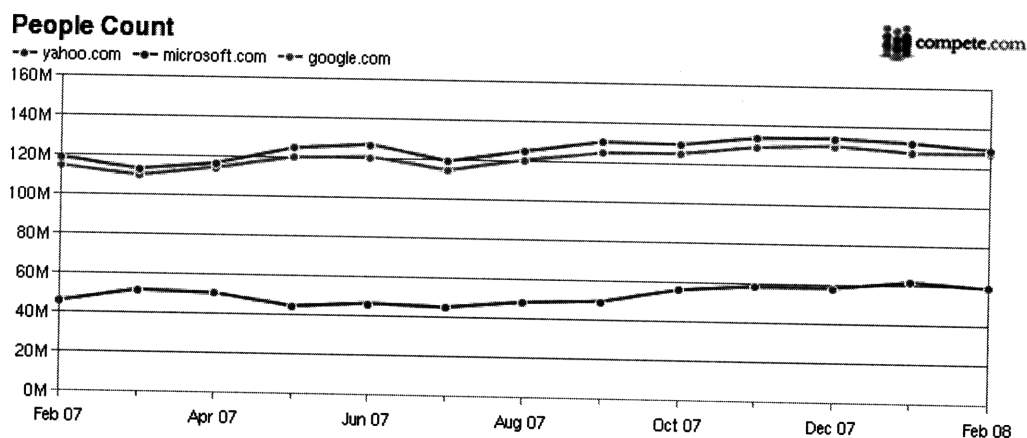


Figure 2 - Compete Monthly Visits Comparison Chart

4.4.4 Google Analytics (JavaScript)

Formerly Urchin Web Analytics, Google Analytics is a traffic measurement solution for individual websites. Website owners integrate Google Analytics JavaScript code into their web pages to allow tracking of user behavior on the website (Google, Inc., 2008).

Google Analytics provides a dashboard showing trends on the website being monitored. Statistics include unique visits per day, geographic information, search keywords, traffic sources, and ecommerce tracking.



Figure 3 - Google Analytics Dashboard

4.4.5 Omniture

Omniture provides one of the most sophisticated web analytics solutions available on the market for owned web sites. The SiteCatalyst product integrates with target websites and allows detailed tracking and reporting. Omniture's products include BI technologies such as the ability to define key performance indicators (KPIs) and create alerts based upon these defined metrics (Omniture, 2008).

In addition to standard web metrics (page views, referring site, etc), Omniture offers multi-site analysis for user tracking across multiple web properties, visitor segmentation based upon how visitors behave on web sites, and an advanced querying interface for mining traffic data. Omniture has also introduced some basic social media integration into its analytics platform which allows identification of top consumer generated media and blogs based upon site contribution.

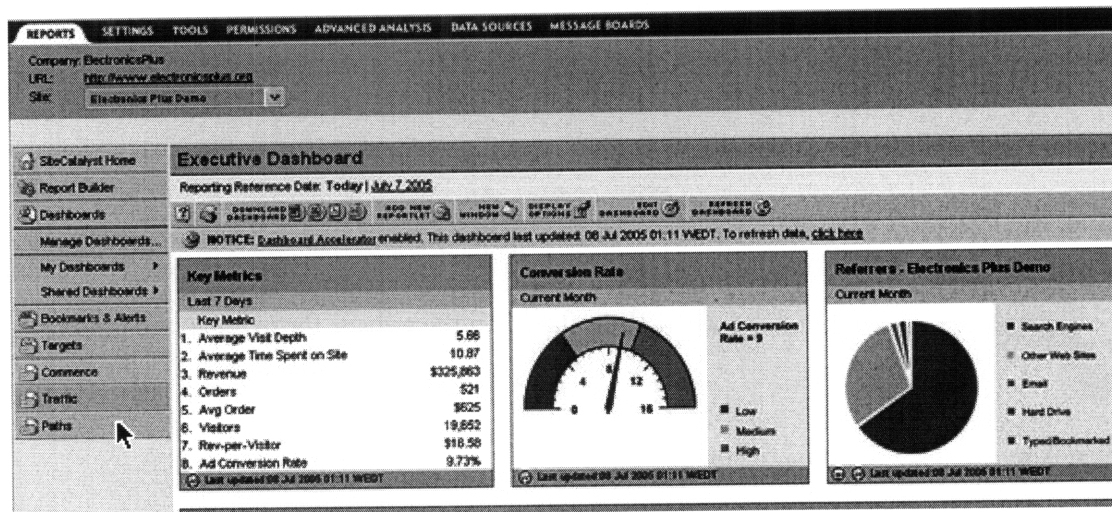


Figure 4 - Omniture Dashboard

4.4.6 Nielsen (Panel)

The Nielsen Company is one of the largest market research companies specializing in audience measurement and consumer behavior. Nielsen NetRatings is one of the two online

analytics services offered by Nielsen Online (the other, Nielsen BuzzMetrics, will be covered in a later section).

Nielsen NetRatings is similar to comScore in that panel members are recruited through random telephone number dialing, and a software program on panelists' computers monitor web browsing activity. NetRatings also provides detailed consumer behavior reports through custom surveys (The Nielsen Company, 2008).

4.4.7 WebTrends

WebTrends offers a suite of web analytics tools similar to Omniture. WebTrends Analytics provides sophisticated visitor tracking, with performance dashboards and detailed reports on traffic data. Special features include usability analysis and site navigation analysis. In addition, WebTrends can integrate offline data, such as sales transaction data and user demographics data into its analysis (WebTrends Inc., 2008).

WebTrends also introduces the concept of visitor engagement as a measurement, trying to quantify the value of each website visitor and creating actionable items. For each action that a visitor performs on a website receives a score. When certain combinations of actions are met, the score may reach a threshold that might trigger actions, such as a targeted marketing email or a customized online promotion displayed on the website for that particular user.

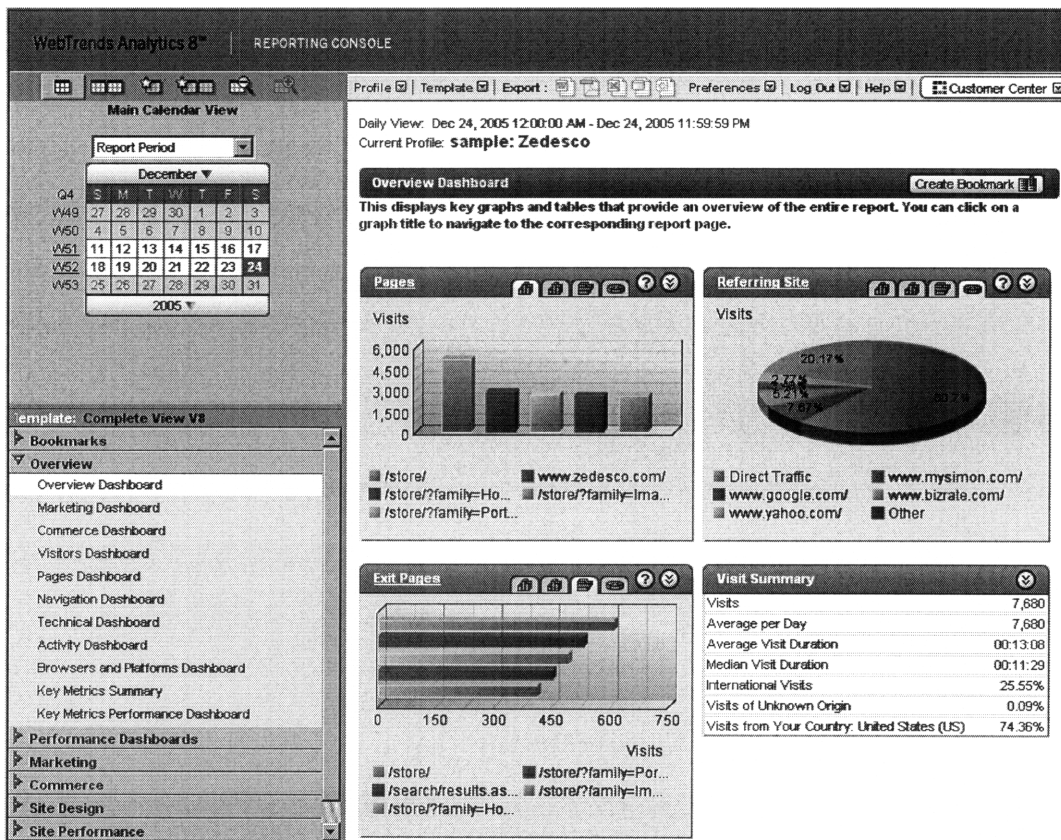


Figure 5 - WebTrends Dashboard

5 Evolution of the Internet – The Social Web

The social web has a very broad definition, encompassing many different forms of communication on the internet. Social media reflects the content in the social web. Because this paper focuses on analytical opportunities, the definition of social media used in this paper includes all types of online media which support or focus on user-generated content. In this regard, social media includes blogs, wikis, forums, and social networks.

The social web took many years before jumping to the exponential growth of users and explosion of social networking websites today. It was not until 1994 when the Internet experienced a 350,000% growth rate that the first large social networking website, Classmates.com, was born (Scholz, 2007). In the same year, Ward Cunningham started developing WikiWikiWeb and publically introduced it in the following year, introducing the concept of a collaborative workspace. In 1997, Slashdot became the first weblog and allowed users to comment on stories, building up a large tech community. Two years later, Blogger.com and Xanga allow users to have their own blogs online.

Social networking picked up steam in the early 2000s with Friendster, a site that grew quickly in 2002 but ultimately failed in its execution, losing its users to MySpace which came along the following year (Boyd & Ellison, 2007). Around that same time, specialized social networks such as LinkedIn for professionals and Facebook for college students emerged and gained popularity quickly. Social bookmarking also started to take off around this time and built communities around common interests, with Delicious introduced in 2003 and Digg soon after (Scholz, 2007).

With many different social media types having been introduced and barriers to participation on the Internet disappearing, the social web expanded rapidly after 2004, with new social platforms being introduced every month. This explosion reflected the shift in Internet culture as more and more people were getting actively involved in online communities.

Today, the social web has been tightly integrated with society and culture. People are joining communities sharing common interests, actively voicing opinions, and making original contributions to content on the Internet. These online communities are having a significant impact on how businesses operate. Instead of businesses driving sales with heavy advertising, consumers are driving decisions using suggestions from their peers. Surveys show that 90 percent of consumers trust word-of-mouth suggestions and factor them into their own decisions (Schneider, 2006).

5.1 Social Media Landscape – Types of Social Media

“Social media” is a broad term that encompasses many different types of media. This section serves as an overview of some of the major classes of social media:

- Blogs (e.g. TechCrunch)
- Forums / Bulletin Boards
- News Aggregation (e.g. Digg, Reddit)
- Social Bookmarking (e.g. del.icio.us, StumbleUpon)
- Social Networking (e.g. MySpace, Facebook)
- Wikis (e.g. Wikipedia)

5.1.1 Blogs

Also known as a web log, a blog is a website that focuses on a topic of interest run by one or more authors which write entries, “posts,” to the site. These entries may be of any length and are displayed on the site in reverse chronological order so that the newest postings occur at the top of the page. Blog readers can subscribe to the blog to be notified of new postings and can write comments on blog posts.

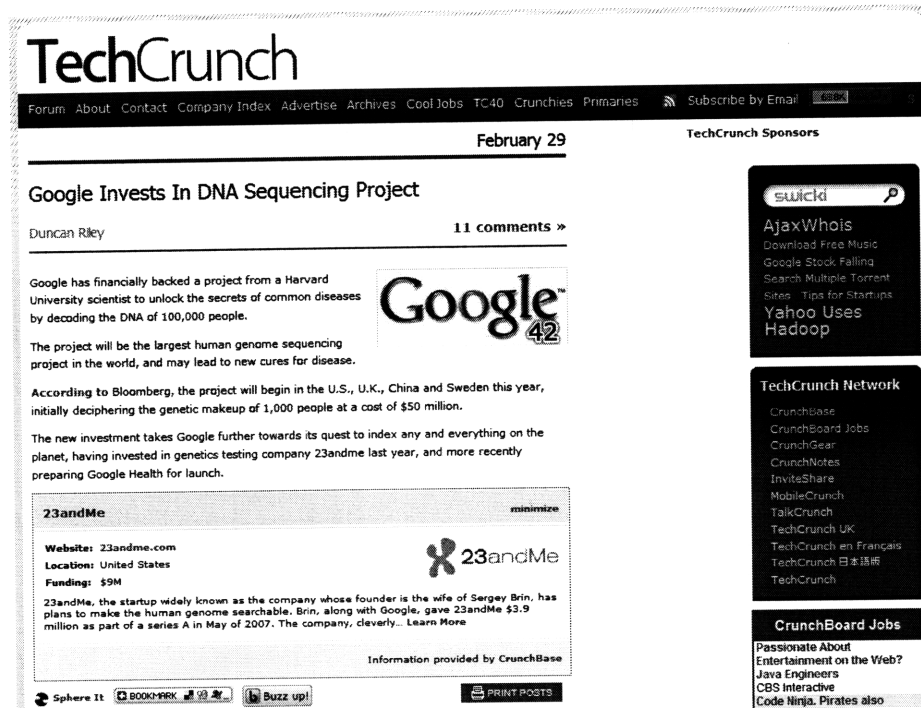


Figure 6 - Blog Example (Techcrunch technology news blog)

Blogs originated in the mid-1990s as personal online journals. At that time, journal writers had to manually update their websites with new journal entries. With the introduction of several websites offering blog hosting services in 1999, including Blogger and LiveJournal, the number of blogs exploded. People realized that blogs were a powerful platform for news and opinions, with information arriving much faster than mainstream media. Blogs were an effective way to allow people to voice their opinions. The social aspect of blogs was amplified with commenting systems which allowed blog readers to share their opinions about blog postings with the author and each other. In addition, “blogrolls” and “trackbacks” functioned as a network for blogs and acted as an indicator of the blog’s influence.

Today, blogs are an established part of society. There are over 100 million blogs in the blogosphere, with over 175,000 new blogs started per day and 1.6 million posts per day (Technorati, 2008). The top 100 blogs cover a variety of different subjects, including

electronics, social networking, political corruption, funny cat pictures, celebrity gossip, and consumer advocacy.

Blog services such as Blogger, Blogspot, LiveJournal, and Xanga offer blog creation and hosting services, making it easy for anyone to start a blog. Professional blogs use an advertising model to generate revenue. Google AdSense contextual advertisements are common, but some of the larger blog networks run their own advertising platform.

Top 15 Blogs

Ranked by authority, Source: Technorati (2/29/08)

1.	Electronics	Engadget (engadget.com)
2.	Electronics	Gizmodo, the Gadget Guide (gizmodo.com)
3.	Technology	Techcrunch (techcrunch.com)
4.	News	Breaking News and Opinion on The Huffington Post (huffingtonpost.com)
5.	Culture	Boing Boing (boingboing.com) Boing Boing is a weblog of cultural curiosities and interesting technologies.
6.	Technology	Lifehacker, tips and downloads for getting things done (lifehacker.com)
7.	Technology	Ars Technica (arstechnica.com) Ars Technica the Art of Technology. News, analysis, and in depth coverage of technology.
8.	Internet	Mashable! The Social Networking Blog (mashable.com) The world's #1 social networking blog
9.	Politics	Blog di Beppe Grillo (beppegrillo.it) Grillo's gags have tackled financial scandals and political corruption
10.	Humor	icanhascheezburger.com (icanhascheezburger.com) Funny lolcats, cat pictures, cat macros, funny captions (and other animals)!
11.	Politics	Daily Kos: State of the Nation (dailykos.com)
12.	Entertainment	TMZ.com – Entertainment News, Celebrity Gossip and Hollywood Rumors (tmz.com)
13.	Technology	ReadWriteWeb (readwriteweb.com) ReadWriteWeb is a popular weblog that provides Web Technology news, reviews and analysis.
14.	Internet	ProBlogger (probblogger.net) Helping Bloggers Earn Money. Blog Tips and News for the Professional and Hobby Blogger.
15.	Entertainment	Celebrity gossip juicy celebrity rumors Hollywood gossip blog from Perez Hilton (perezhilton.com) Perez Hilton celebrity gossip blog is the source for Hollywood gossip, juicy celebrity rumors, celeb scandals, and the latest celebrity gossip

5.1.2 Forums

Internet forums, also known as discussion boards, are places on the internet where people can start and participate in discussions. A forum generally has multiple categories to select from. Within a category, the participant is presented with a list of current topics, known as

threads. Once the participant clicks on the thread, the individual posts that constitute the discussion are listed in chronological order. If the participant wishes to post a reply to the thread, the reply is added to the end of the list of posts. The participant can also start new discussion topics.

Topic Listing
Gaia Forums » Entertainment » Entertainment Discussion [View Forum Moderators]

Entertainment Discussion
Subforums:

- MTV**
Chat about MTV and Gaia's Virtual Hollywood.
- Books**
Discuss your favorite books.
- Independent/Obscure Movies & Television**
Discuss independently produced or obscure movies and television shows here!
- Toys & Collectibles**
From oldies to the latest gadgets, talk about your favorites here.
- Celebrities**
Where you creepy fanboys/girls can go to rant about how sezzzy your favorite celeb is.
- Lord of the Rings**
Discuss the movies, the books and the world of J.R.R. Tolkien.
- Harry Potter**
Discuss the movie and the books! Don't worry, Muggles are allowed.

Explore Guilds in General

Forum Pulse	Replies	Author
Disney Movies	116	A0F
Star Wars: The Official Thread	22791	Aldaron_Tauron
What should I watch next?	New	karama
Old Animated Movies From Forever Ago?! where have they gone?	5	The Lady Magistrate
LOST	16708	lemonhobbit
Food Network	15	Aces9306
I Just Thought of a Great New Contender for the AvP Series	6	Yeaaa-boiii
Maximum Ride ROX Who agrees?	44	tessss1
The Official Kyle XY Thread	2337	Wicked_Vampeera

Figure 7 - Forum Topic Listing (Gaia Online – Entertainment Subforum)

Prior to the introduction of web-based forums, discussions could be facilitated on dial-in or locally networked bulletin board systems (BBS) and then later on newsgroups (Usenet). Therefore, web-based forums are an extension of these discussion technologies into a more widely reachable format.

Internet forums are one of the best mediums for discussions, with forums being available for nearly every special interest. The sizes of forums vary widely depending upon audience and topic. For example, a general discussion forum that branches out into many different categories would appeal to a wider base than forum that only focused on one specific interest.

Because forums function as online communities, some forums have imposed membership requirements for participation. Many forums use an advertising model as a means of revenue generation, but some forums impose membership fees. More exclusive forums can require prospective members to have received an invitation from a current member in order to join, or a new member might be given limited access to a forum until certain requirements are met for full membership. However, the vast majority of online communities is free to join and supported by advertising.

Top 15 Forums

Ranked by post count, Source: Big Boards (2/29/08)

1.	Gaia Online (gaiaonline.com)	1287	11.71
	Anime roleplaying community	(m posts)	(m users)
2.	IGN boards (boards.ign.com)	175	1.26
	General video gaming		
3.	Nexopia (nexopia.com)	153	1.29
	Picture rating, sex and relationships		
4.	FaceTheJury (forums.facethejury.com)	142	0.55
	Real life picture rating		
5.	Vault Network Boards (vnboards.ign.com)	117	0.63
	Role playing games		
6.	Offtopic.com	95	0.19
	General		
7.	Doctissimo.fr	94	1.44
	French medical help and health		
8.	d2jsp (forums.d2jsp.org)	87	0.32
	Diablo 2 video game		
9.	Jogos (forums.jogos.uol.com.br)	72	0.16
	Games (Brazilian)		
10.	BuddyPic (buddypic.com)	66	1.59
	Picture rating		
11.	Something Awful (forums.somethingawful.com)	57	0.11
	General		
12.	Fok.nl	57	0.21
	General (Dutch)		
13.	Christian Forums (christianforums.com)	45	0.22
	Christian-based forum and online community		
14.	VW vortex forums (forums.vwvortex.com)	43	0.38
	Volkswagen and related car brands		
15.	PBNation (pbnation.com)	42	0.39
	Paintball related		

5.1.3 News Aggregation

News aggregation websites allow users to submit web pages containing any type of content. These submissions contain a title, a short description or excerpt, a category, and a link to the source material. The submissions get pooled on the website, where users can vote on the article. Popular articles end up on the front page of the site for about a day, keeping things up to date. Finally, users can make comments on submissions and discuss the topic. The commenting system also uses a voting system, where highly rated comments are displayed and low rated comments are hidden.

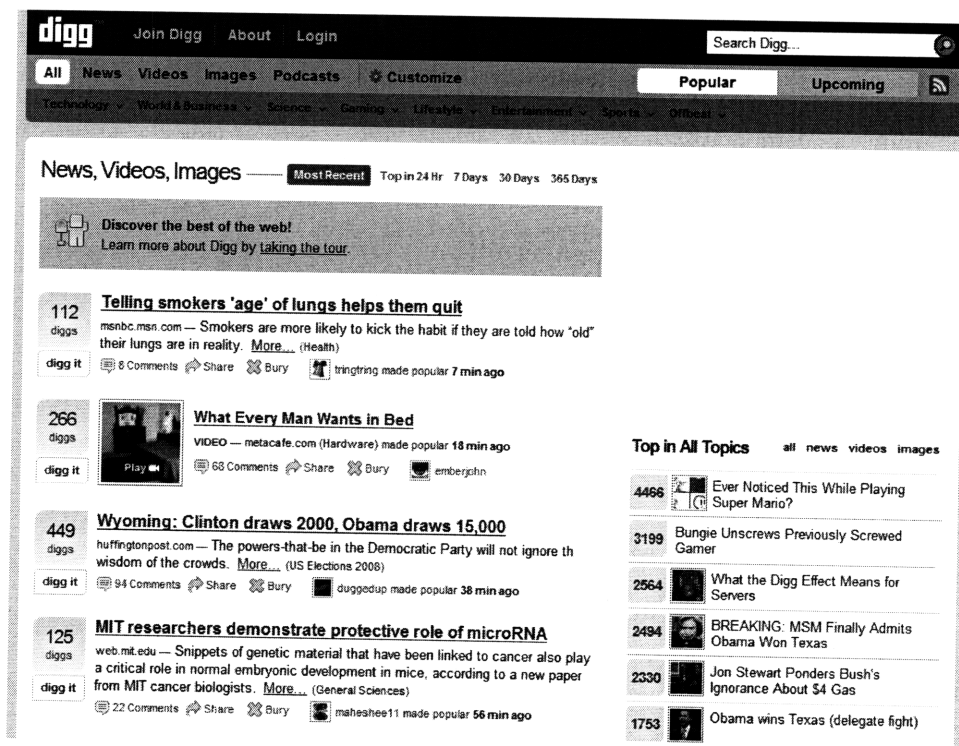


Figure 8 - Digg Front Page

In the news aggregation space, there are two main websites: Digg and Reddit. Although the two sites look very different from each other, with Digg having high contrast graphics and Reddit being mostly text-based, the two sites operate with the same fundamental principles. For Digg, users “Digg” an article if they like it, and articles reaching a threshold of a certain number of “digg” appear on the front page. Reddit is similar, except users can also vote

negatively, which would decrease an article's rating. In addition, Reddit offers a personalization layer on top of news aggregation, so that articles a certain user expressed likes or dislikes about are factored in a recommendation system which filters out future articles by relevance to the user (a social bookmarking concept).

These social news sites have been very successful and influential, with Digg estimated to have about 20 million visitors per month (Compete, Inc., 2008). Sites that appear on the front page of these aggregator sites end up receiving an enormous amount of traffic. Both Digg and Reddit earn revenue through advertisements featured on the sites. Like Google, advertisers are not able to buy their way into being featured with a high ranking.

Despite using a democratized system of promoting content, both Digg and Reddit have faced many criticisms. First, the site is influenced significantly by Internet culture. Many items that end up on the front page of the sites are not newsworthy, as defined in the traditional sense. Instead, content linked from the front page can be anything that is interesting to online viewers at the time, including "cool" videos from YouTube, a funny comic, or pictures of cute puppies. As a result, these news sites appeal mostly to the younger tech generation and articles are heavily biased toward the

demographics' interests. For example, on March 8, 2008, a breakdown of the top 10 stores could be classified as follows: 2 video game stories, 6 pro-Obama political stories, an anti-Bush political story, and a video from YouTube.

Social news sites have also been criticized for being easily manipulated. A study of Digg's "Top Diggers" list in 2006 revealed that the top 100 submitters controlled 56% of Digg's

Top in All Topics		all news videos images
6050		Ever Noticed This While Playing Super Mario?
2652		Obama: I'll discard Bush's unconstitutional executive orders
2060		Obama wins Texas (delegate fight)
2009		Bush vetoes bill banning waterboarding
1616		Pedophile goes nuts at reporter on Australian news
1499		Chicago Tribune: Hillary Has No Foreign Policy Experience
1487		Obama: "You Won't See Me as a VP Candidate"
1343		What Every Man Wants in Bed
1247		Wyoming: Clinton draws 2000, Obama draws 15,000
1182		SEGA updates Dreamcast.com for DC's 10th anniversary !

Figure 9 - Digg Top Stories Sidebar

front page content (Fishkin, 2006). Similar numbers were found on Reddit and other sites. Since then, Digg has removed the Top Diggers list from public view and has announced changes to the ranking algorithm.

5.1.4 Social Bookmarking

Social bookmarking is similar to social news aggregation, in that all of the articles submitted are done by the users. Social bookmarking sites allow users to bookmark sites and have the bookmarks stored centrally, enabling the bookmarks to be accessible from any computer rather than being stored locally in a web browser. The social aspect of bookmarking is that users can see what their friends have bookmarked. In addition, the sites function as recommendation engines, pairing up a user's interests with other users that have bookmarked similar sites.

The two major social bookmarking sites on the Internet are del.icio.us (owned by Yahoo) and StumbleUpon (owned by eBay). While del.icio.us promotes itself as a pure social bookmarking site where all bookmarks can be stored in one place and users can see their friends' bookmarks, StumbleUpon promotes itself as a website discovery tool. StumbleUpon users install a toolbar on their Internet browser. As users browse the Internet, they click a thumbs up or down button to indicate their interest in the site. Then, when the user is interested in finding a new site to visit, the user clicks the "Stumble" button, and the toolbar will redirect the web browser to a site chosen based upon its recommendation system.

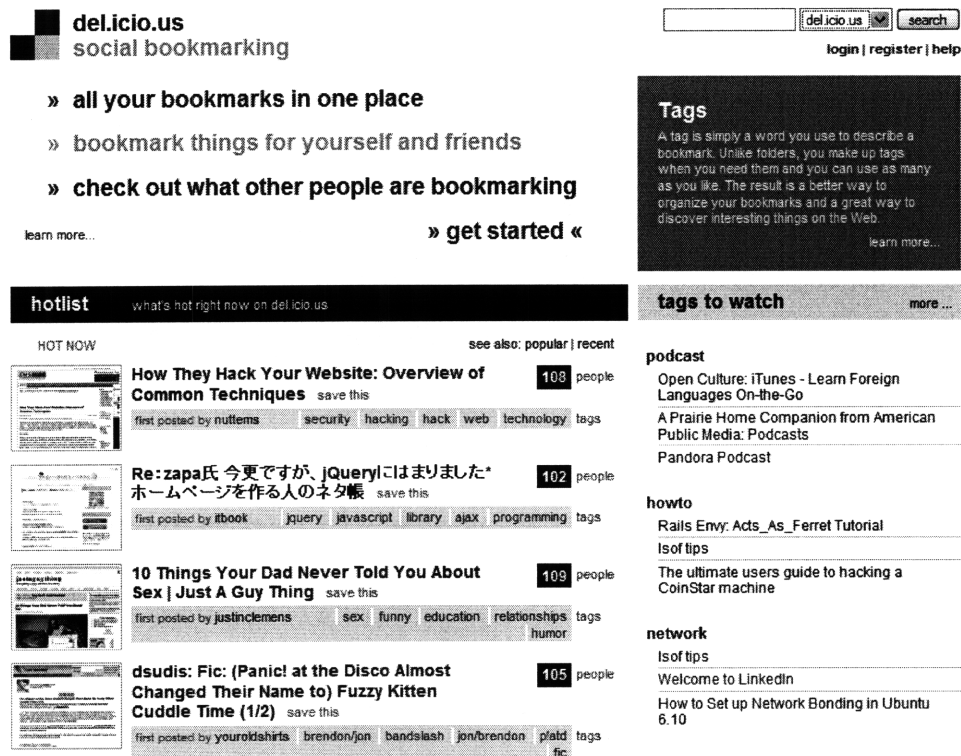


Figure 10 - del.icio.us

del.icio.us is owned by Yahoo!, however, the revenue model for the site is unclear (at the time of writing, del.icio.us had no advertising). StumbleUpon allows advertisers to insert their sites into the recommendation pool, so that when users match certain interest criteria, they can be directed to an advertiser's site. In addition, advertisers can receive analytical data about their websites, such as visit statistics. Note that this advertising model potentially undermines the integrity of the social recommendations.

5.1.5 Social Networking

Social networking services allow people to find people and connect based upon a shared interest. These networks typically have users create an online profile when joining and encourage users to keep their profiles up to date. Then, the networks act as a directory, where users can search and add other users to their list of connections. Typically, adding a

connection requires approval of both parties. Once the connection has been made, the connected users can see each others' list of connections.

It should be noted that while the “networking” part of social networking sites implies helping people make new connections, social networking sites are primarily used for communication between existing contacts (Boyd & Ellison, 2007). Many social networks have features that go beyond the basic functionality of network connections. These features include messaging services, event planning, message boards, and blogging. The landscape of social networking sites is very diverse, meaning that there is not one social network that meets every need. The social network sites generally focus on one vertical application, so active Internet users end up with multiple social network memberships. The proliferation of social networking sites in the past few years has also spawned a new breed of web applications that aggregate information from multiple social networks, so that users do not have to individually log into each social network to get updates on connections.

Online social networking was first introduced in 1997 with SixDegrees.com, which had a user base that swelled to the millions before closing in 2000 due to a failed revenue model (Boyd & Ellison, 2007). The next significant social networking service launched in 2002 as Friendster, which introduced the friends of friends concept. Friendster gained rapid popularity over the years until scalability issues resulted in a mass exodus of users to newer sites. Meanwhile, more specialized social networking sites started to emerge in 2003, with LinkedIn, a business oriented networking site, becoming one of the most successful. Social networking was also expanding internationally – Google introduced its own social networking site Orkut which failed to gain traction in the U.S. but has since secured substantial market share in other countries.

At present, MySpace and Facebook are the two largest and most popular social networking sites. MySpace was founded in 2003 as an alternate from Friendster, and achieved a large following that began from music bands choosing to make MySpace their home. What set

MySpace apart from other sites at the time was the freedom that the site gave to its members in customizing profile pages. MySpace was unusual in that it allowed direct editing of the web page code used in profile pages so that users could freely express themselves on their public profile pages. In addition, MySpace's community features were highly effective in attracting users. Another factor contributing to MySpace's success was that the site allowed minors to join, which resulted in significant network effects at a time when other sites rejected all minor applicants due to privacy issues. MySpace's other features include instant messaging, video sharing, mobile phone access, and classifieds.



Figure 11 - MySpace Home Page

Facebook was founded in 2004 as a college social networking site. The site quickly became popular as students could look up the profiles of other students and get status updates. In its early stage, there were many college-specific features, such as support for looking up other people living in the same dormitory or sharing the same classes (those features have since been removed as the site expanded its audience). One major innovation was that recent

actions made by friends would be listed on their profiles. Then later on, the “feed” feature was introduced on the home page, where updates from all friends would be aggregated into a chronological feed so that a user could get up-to-date quickly. Facebook also introduced a “wall” feature allowing people to leave messages that would be visible to other profile viewers.

Facebook is now open to everyone, with various types of joinable networks based upon university, high school, business, or geography. Facebook’s feature set has also expanded dramatically after introducing Facebook Platform, which allows third-party developers to write applications that integrate with the social network. Users can add applications tailored to their interests, such as a travel map application where the user can plot all the places that the user has been to. In some cases, applications act like a network within a network. For example, a stock market application would allow users to create fantasy stock portfolios and compete with other users that have installed the same application.

Both social networks operate around an advertising revenue model. Ads are presented to users as they browse the social networks. Companies may also sponsor their own profiles or events and add them to the social network.

Top 20 U.S. Social Networks
Ranked by monthly visits, Source: Compete (2/29/08)

1.	MySpace	General	65.7 (m monthly visitors)	955.1 (m monthly visits)
2.	Facebook	General	28.6	326.4
3.	Classmates.com	Reunions	12.0	22.5
4.	myYearbook	General	3.0	20.0
5.	Bebo	General	3.5	19.3
6.	BlackPlanet.com	African American	2.1	13.7
7.	Hi5	General	2.4	11.8
8.	LinkedIn	Professional	3.8	11.1
9.	Tagged	General	2.3	10.6
10.	Reunion.com	Reunions	6.7	10.1
11.	Friendster	General	1.9	8.6
12.	Orkut	General	0.5	7.0
13.	Flixster	Films	3.3	6.7
14.	fubar	Dating / Virtual Bar	1.7	6.6
15.	Tickle	Online Quizzes	2.1	6.3
16.	Cafemom	Mothers	1.6	6.0

17.	Xanga	General	1.9	5.9
18.	Yuku	Hosted Communities	0.9	5.4
19.	Twitter	General	0.6	4.2
20.	Ning	Hosted Communities	1.2	3.8

Top Worldwide Social Networks

Ranked by monthly visitors, Source: ComScore (1/08)

1.	MySpace	109.3 (m monthly visitors)
2.	Facebook	100.7
3.	Hi5	38.3
4.	Friendster	34.1
5.	Orkut	26.9
6.	Bebo	22.4

5.1.6 Wikis

Wikis are collaborative websites where the content can be easily edited by a group of people, usually the public. The theory behind wikis is that if the content is allowed to be openly changed, then the quality and quantity of information will improve as more users read and contribute.

The most widely used wiki is Wikipedia, an online encyclopedia. Any user is free to edit existing articles or add new articles, with a group of volunteers which privileged permissions to enforce editing policies and resolve disputes. Many other wikis exist on the Internet for niche subject matter. For example, video game enthusiasts have created a wiki that goes in-depth on specific video games. Organizations have also found wikis useful for internal knowledge management.

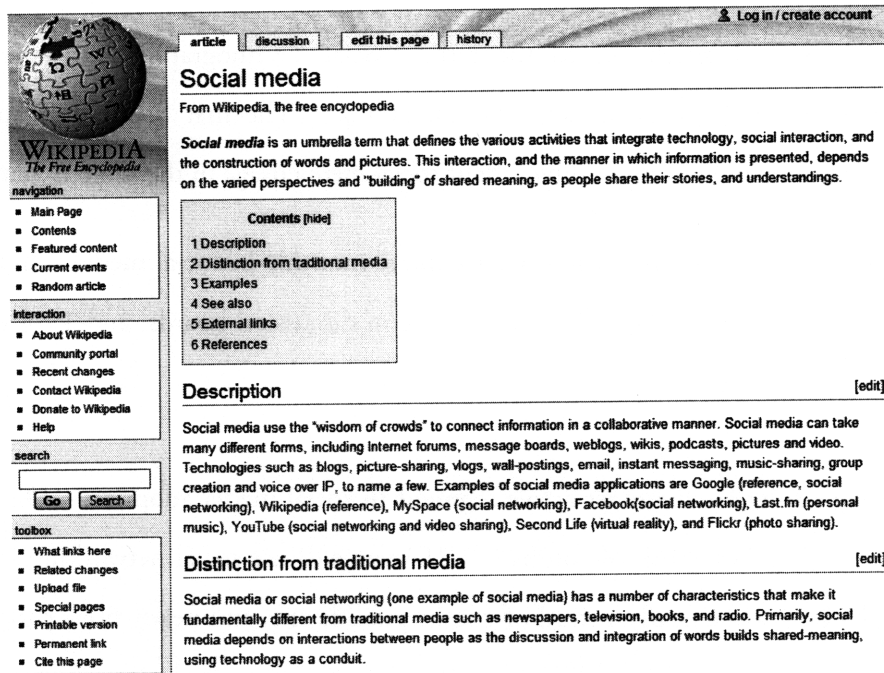


Figure 12 - Wikipedia article on social media

Because most online wiki sites allow any user to edit content, many people have questioned the accuracy of the content. This is because inaccurate information posted on the wiki will remain on the site until someone else with sufficient knowledge on the topic can make corrections. Furthermore, disagreements on facts or bias on sensitive topics may be difficult to resolve. Nevertheless, a 2005 study of 42 science articles by the journal *Nature* found that Wikipedia had a comparable error rate to Britannica (Giles, 2005).

5.2 Social Media Trends

Growth of social media has been fairly rapid and continues at a brisk pace. Many factors contribute to this growth, including: the continued spread of Internet access globally, increased spread and availability of social media service providers, growing network effects from existing social networks, and the continued integration of social media into society and culture.

In 2007, Forrester Research coined the term “Social Technographics” to refer to the demographics of social media participants (Li, 2007). The first Social Technographics report revealed some interesting findings about social media usage in the United States:

- **Overall, 48% of adults currently participate monthly in social media** – In addition, 22% of adults read blogs, 21% participate in discussion boards, and 19% use social networking sites.
- **Young adults are the most active segment in social media adoption** – On a weekly basis, nearly two-thirds of youths (12-21 years old) use social networking sites, 40% read a blog while nearly a third of youths publish one, and nearly 30% participate in discussion boards. On a daily basis, 41% of youths visit social networking sites.
- **Penetration of social media will grow** – The average age of the 52% of non-participating adults is 50, implying that penetration by age groups will increase over time. It should also be noted that even though this segment does not directly participate in social media, people in this group are still affected indirectly through friends or traditional media coverage.

These factors suggest that social media will continue to increase in importance as it continues to penetrate and spread in society. Hence, organizations should begin to incorporate social media into their overall business intelligence strategy, resulting in significant demand for effective social media analytics.

6 Social Media Analytics

6.1 Differences from Web Analytics

The content on social media websites is fundamentally different from traditional web sites. In particular, the source of content has shifted from website publishers to web users. As part of this shift, new classes of websites have emerged as part of the Web 2.0 environment. While certain metrics provided by web analytics are still valid, the nature of social media presents new analytics possibilities that cannot be captured with current web analytic techniques.

This is not to say that web analytics are not useful in the context of social media. Web analytics can be extended to monitoring social media sites and will yield useful information. For example, web analytic providers will be able to create lists of the most popular websites within each social media category, and deep website analytics on individual social media sites might yield information on the top pieces of media or activity patterns. Where web analytics stops and social media analytics starts is when the analysis takes into account relationships between users, the unique organizational structure of social media, or the actual content within the media. For example, social media analytics can examine the structure and relationship of users within a social network and identify key contributors.

This section serves to explore these new possibilities enabled by analytics in social media.

6.2 Brand Monitoring

Brands are a critical asset for many companies. While it may be difficult to put an exact monetary value on a brand, some experts believe that for some companies, the value of the brand constitutes over half of the company's net worth (Deveau, 2007). Therefore, the use of company brands on the Internet is an area of concern for companies, and it would be in a company's best interest to keep tabs on uses of the brand on the Internet.

In the online space, brand monitoring takes on two distinct definitions. The first definition of brand monitoring involves monitoring websites for inappropriate usage of company brands. This was the service associated primarily with brand monitoring during the early 2000s when social media was still in its infancy. This type of service will be referred to as brand integrity monitoring. The newer definition of brand monitoring involves social media analytics, with applications for marketing.

6.2.1 Brand Integrity

The concept of brand integrity monitoring has been around years before the popularization of social media, with businesses in the space promoting effective ways for brand owners to enforce proper usage of brands. As the Internet became a more important part in shaping company brands, unscrupulous companies and individuals would try to profit off of well known brands through activities which misrepresented these brands.

According to BD-BrandProtect, one of the key players in the brand integrity monitoring market, there are many various ways brands can be misrepresented (Joy, 2006), including:

- Phishing – attempts by scammers to trick victims into providing sensitive information (such as bank account passwords) to a website or email address pretending to be the legitimate company
- Unauthorized Associations – websites using company brands or trademarks in unlicensed manners, which may mislead the public into making an association between the company brand and the unauthorized activities or organizations
- Cybersquatting/Typosquatting – registration of top level Internet domains with names or slight misspellings of names belonging to brands; these names are usually used maliciously, with the unauthorized registrants serving advertisements, offering competitor goods, or attempting to sell the name to the brand owner

The economic costs of brand misuse to both brand owners and consumers can be very high, with companies spending millions of dollars to build and maintain their brands. Gartner Group estimated that phishing scams alone cost consumers \$2.8 billion in 2006 (Klein, 2007), with lost revenue to companies and erosion of consumer trust in companies.

As a result, there are many companies that offer monitoring services for brand integrity. Key players include BD-BrandProtect, NameProtect, MarkMonitor, and Cyveillance (Klein, 2007). These companies use a combination of automated web crawlers to find instances of name usage and employees for use assessment to monitor brands and trademarks on the Internet. Once improper usage is verified, these companies typically have effective legal departments and connections into law enforcement agencies to ensure swift enforcement action.

In terms of social media, these services do not monitor what users are saying about a brand online or the overall perception of the brand. It would be difficult, if not impossible, for such a service to have negative comments on a brand be removed from the Internet, due to free speech. Therefore, these services are limited to enforcement of proper brand use, so while negative comments by users on bulletin boards cannot be censored, there are still opportunities for these services to expand into social media. One application would be copyright enforcement of videos submitted to video sharing sites, such as YouTube, or removal of copyrighted material posted on bulletin boards, blogs, and wikis.

6.2.2 Brand Sentiment

Something new with brand monitoring in social media is being able to monitor brand sentiment. This involves understanding what consumers are thinking and saying about company brands and products on the Internet, which is primarily through social media. In this context, Forrester Research defines brand monitoring as “the ongoing analysis of mainstream media and consumer-generated content to identify trends relevant to a company’s marketing activities and competitive landscape” (Kim, 2006).

In the days before the Internet became mainstream, brand monitoring would be done by specialized firms, such as clipping services, field agents, focus groups, and ad-hoc research. This was successful because feedback about brands remained highly centralized in mainstream media channels, such that consumer sentiment was largely shaped by just a few sources.

However with social media, the shift away to online communities and the fragmentation of media means that companies now have a significantly greater amount of media to cover, and the speed at which new information propagates on the Internet has called for automated monitoring services (Kim, 2006). No longer can companies get an accurate measurement of brand sentiment just by looking at mainstream news reports and surveys. Consumers are increasingly empowered to post their own opinions about a company online, and these comments can make dramatic effects on a company's perception. This is because user opinions can be seen by anyone with access to the Internet, so opinions of products are no longer isolated to local social networks. As a result, companies must be able to track consumer sentiment as it develops on the Internet and be able to react quickly.

6.2.2.1 Product Sentiment

Due to the Internet's ability to bring together people with common interests and act as a platform to share ideas and opinions, consumers are starting to turn more to word of mouth recommendations from other consumers online, rather than professional review sites, which are only able to review a limited selection of products and may be biased by company marketing departments. In a 2006 survey of Internet usage, over 35% of respondents reported using the Internet to read customer ratings and reviews at least once a month (Li, 2007).

Epinions.com was the pioneer in user-submitted online reviews, having been founded in 1999 as a site solely modeled on reviews written by consumers. Epinions would keep track of products in the marketplace, categorize them, and then invite consumers to submit

reviews. Each review consisted of a somewhat lengthy write-up by a consumer and then a star rating that could be aggregated to form an overall rating for the product. These reviews allowed consumers to see which products were best in class and also allowed consumers to discover alternative products. To encourage review submissions, Epinions paid their reviewers through a revenue sharing plan based upon the popularity of reviews.

Within a few years, many websites were soliciting customer reviews. Today, many large online retailers, such as Amazon.com, Buy.com, and Staples all allow customers to write reviews that appear directly on the product page. Shopping comparison site PriceGrabber.com also solicits reviews from consumers.

Reviews are not just limited to products that can be purchased in stores. For instance, dedicated websites exist for travel-related reviews, such as TripAdvisor, Expedia, and Yahoo! Travel. Epinions also allows reviews of airlines and hotels. Yelp.com is a local review site that supports reviews of local businesses and is a very popular source for restaurant ratings.

While review sites are great sources for consumers to find a large amount of reviews, it is important to note that these sites are not the only sources to measure consumer sentiment. Social media has given consumers a wide platform to express satisfaction or discontent with products and services. People can share their opinions on discussion boards, blogs, and social network sites. Because social media has a community, network effect around it, opinions in social media may be more influential than opinions on general review websites, because the participants have a connection with each other within their network or community. In addition, social media sites allow interactions such that products can be discussed and consumers can engage experts for opinions, whereas review sites are typically one-way showing only written reviews and offering no room for interaction.

Business Intelligence Applications

Using social media analytics, all of the different opinions expressed on the Internet can be mined, aggregated, and categorized. Once placed within the business intelligence framework, the data can be queried and analyzed.

Given that consumer feedback is a very important part of the product development cycle for companies, the most direct application of analyzing consumer opinion is in product research and development. Applications for product R&D include:

- Receiving early feedback on new products – qualitative measures of consumer sentiment can serve as indicators of a new product's performance in the market
- Monitoring product quality – changes in consumer sentiment over time may be attributed to quality issues
- Determining product issues – a filter for extreme levels of individual positive or negative sentiment opinions can provide insight into reasons why consumers are strongly liking or disliking a product
- Planning new product features or product revisions – planners can mine opinion across multiple products to determine the importance of individual features and key customer needs

There are also applications outside of R&D. Other users of product sentiment monitoring include:

- Sales representatives / buyers – check overall sentiment for a company's product line to determine which products are worth selling; retailers can look at sentiment as another measure to help estimate demand for products
- Marketing – product sentiment can provide marketers with feedback on marketing efforts and competitive intelligence

- Management – use sentiment measures as a key performance indicator in their BI dashboards to monitor the performance of different product divisions

6.2.2.2 Company Image

The Internet and social media are effective platforms for consumers to share their opinions. The previous section discussed the proliferation of customer reviews across many different types of websites, ranging from dedicated review sites to special interest communities. Customer opinion on the Internet is not limited to reviews of products and services. Customers also discuss companies themselves and talk about their positive or negative experiences.

When a community member has positive experiences with a company, that person may be inclined to recommend the company when the opportunity arises, such as another member asking for recommendations on where to purchase a particular product. Likewise, the opposite may apply to people receiving negative experiences.

What really contrasts the power of social media to the offline world is that consumer experiences with companies may become magnified online, where thousands of people may learn about the experience. Particularly negative experiences can be especially damaging to company image, since the outrage of consumers may translate into a story that is viral in nature.

Business Intelligence Applications

Mining sentiment on a company name can provide information on how a company is faring overall. After all, a company's brand name may represent a significant portion of that company's equity. Therefore, such a metric on company brand sentiment might be integrated into a BI dashboard.

Customer sentiment mining also has novel uses in the area of customer relations. In the past, customers would take their complaints up with customer service. However, because the

Internet and social media have made customer grievances public, companies can now be more proactive in having tough customer issues resolved. In nightmare situations where a customer complaint may be a public relations disaster, analytics can be used to set up an early warning system, allowing customer relations to resolve issues before they spiral out of control on the Internet, ending up in significant damage to the brand. Such a system would flag posts with extremely negative sentiment or posts receiving usually high levels of interest on the Internet.

For management, opinion monitoring on the company's brand could also serve as a means of quality control. By analyzing opinions on the company, executives can understand the key strengths and weaknesses of the company as perceived by the general public and observe how the company's image changes over time in response to company initiatives.

6.3 Trend Recognition

The social web can help companies see consumer trends and apply the knowledge to product planning and marketing. Because consumers can reveal their preferences online through social media, these preferences can be aggregated to show likes and dislikes of consumers. These preferences can then be broken down by different criteria such as demographics, based upon the sources of information used from the social web.

There are various applications of trend recognition. In particular, trend recognition is a key component in market research. Social media analytics can help identify trends such as growing segments for business expansion, shifts in consumer demand across different products and categories, changes in consumer attitudes, and emerging or "hot" topics of interest.

Using the social web as a source for researching trends yields several benefits. First, the cost of research will be significantly lower than the traditional methods of monitoring consumer trends, in which special market research firms would conduct surveys, run panels, and

aggregate sources of data such as retail sales information. Another major benefit is the speed at which data can be collected and analyzed. Market research firms publish new reports periodically, with custom research requests taking weeks or months to fulfill. Using automated analysis, new data appearing on social media can be used to update models at speeds near real-time, allowing early identification of new trends. This can offer a significant competitive advantage. Finally, the diversity of demographics represented in social media may allow for deeper and more precise research, which may find trends that would otherwise remain undetected in surveys and panels.

6.4 Targeted Advertising

6.4.1 Advertising Market Trends

Online advertising spend continues to grow as marketers shift their advertising dollars from traditional media such as direct mail and newspapers to online interactive media. A 2007 survey by North American Technographics found that while consumers typically spent 29 percent of their media time on the Internet, advertising on the Internet only constituted 8 percent of advertising spending (VanBoskirk, 2007). This is in sharp contrast to television spending – consumers spent an average of 37 percent of their time watching television, which is on par with the 32 percent of overall advertising spending. These results show that much of the online advertising market has yet to be tapped.

Internet advertising began with simple display ads on web pages, either in the form of text linking to the advertiser's website or a colorful banner advertisement. Pricing models were generally based upon a fixed price per number of impressions or a price for each click-through. The problem with this type of online marketing was that the ads were not very targeted. Although advertisers could narrow down the audience by filtering websites based upon their demographics, advertisers still had very little control over the audience and quality of visitors.

Contextual advertising soon emerged in the form of search engine marketing. Now search engines would serve advertisements that were relevant to what was being searched. Advertisers could then bid on individual keywords that were relevant to the product or service being offered. Google refined contextual advertising with its AdSense program, in which relevant ads were served on third-party websites based upon the content of the specific page being viewed.

Having been around the longest, online display ads and search engine ads have been the primary forms of online advertising. The 2007 US Interactive Marketing Forecast by Forrester Research estimated the size of the online advertising market in 2008 to be \$24.4 billion, constituting just 10 percent of the overall advertising market in the United States. Of the online advertising market, it is estimated that 43 percent of spending is on search engine marketing and 31 percent on online display marketing.

By the end of 2012, Forrester forecasts online advertising to grow to \$61.3 billion representing 18 percent of total advertising. Search engine and online display combined marketing spend will double from 2008 levels, but the strongest areas of growth will be in social media and online video marketing – with levels five times 2008 spend.

6.4.2 Advertising in Social Media

The social web has spawned communities that bring users together on areas of interest, allowing companies to fine tune their advertising so that the advertising is seen by the target demographic. This increases the success rate of advertising campaigns and potentially lowers overall advertising costs.

The traditional forms of online advertising are still applicable in the category of social media: online display ads and contextual search engine ads on community sites can be highly relevant to the community. Because people tend to spend much more time on social media

sites than other online media, marketers should theoretically have more success in reaching their target demographic.

However, the effectiveness of these ads on social media sites is debatable. Unofficial studies have shown that text ads on social media sites receive significantly fewer clicks than on traditional websites. A quick online search on Facebook's ad effectiveness reveals a whole variety of blog posts and articles on the ineffectiveness of Facebook ads. One explanation for this shortcoming is that when users are browsing social media sites, the users are simply not in the right mindset to look at advertising. Another explanation is that while social media can provide very targeted ads based on demographics, it is still difficult to contextualize the advertisement.

Therefore, marketers are still exploring how to effectively advertise in social media and have started to make use of features within the social media platform. The following are examples of interactive online advertising:

- Advertising on MySpace and Facebook through specialized profile pages on the social network sites, allowing customers to interact with the profiles. Adidas, a well-known footwear brand, created a page on MySpace that allowed users to interact with the brand. The result was a page that influenced people at a fifth of the cost as compared to television advertising (VanBoskirk, 2007).
- Paying consumers to post positive opinions on products and services on social media, blurring the line between opinions and advertisements. Startup companies such as PayPerPost.com provide a platform to link up companies with bloggers and similar companies exist targeting social networking sites and discussion boards.
- Creating brand experiences that span across multiple types of interactive media, such as online discussion boards and polls to create a community and receive feedback and casual games featuring products to engage customers (known as advergames).

6.4.3 Business Intelligence Applications

There are two main areas in which social media analytics can apply to targeted advertising:

The first area is in the identification of relevant customer segments and active communities. Social media analytics can help determine the right audiences to market by searching through online discussions for key characteristics and features of products being marketed. By looking at the profile information for people engaged in relevant discussions, analytics can potentially assemble demographic profiles.

Once the audience has been determined, analytics can be helpful in determining the most valuable sites to advertise on. To determine this, analytics would have to take multiple factors into account, such as the amount of activity on individual social media websites and the relationships between each website to other websites. Such a list of most valuable sites would help marketers prioritize ad spending, ensure proper ad exposure to target audiences, and prevent waste on irrelevant or minor websites.

The second area for social media analytics is in deep network analysis on individual social media sites. Deep social network analysis can analyze the relationships between people in a community or network and identify centrality points, or people who are most influential in a particular area. For social networking sites, this type of analysis might identify individuals with the most connections to other people or the most varied connections. Similarly for online discussion boards, factors in the analysis might include the quantity of posts made by an individual, the average number of replies in threads started by the individual, and possibly reputation for forums that incorporate some type of reputation system.

By identifying opinion leaders in a network, companies can have special, custom tailored advertising campaigns to those influential individuals. Companies may send those individuals free products or services for review or even try to recruit those individuals as evangelists for the company, effectively converting them into marketing tools. These approaches can have a huge return on investment for companies, since the endorsement of

opinion leaders in a network can influence a large number of people while the cost of these targeted campaigns is minimal.

6.5 Consumer Applications

Social media analytics may be highly useful to consumers because they can assist individuals in finding information effectively given all the activity happening on social media. As a result, a form of social media analytics might be integrated as a part of a company website to enhance the end user experience or a business model might emerge based upon a consumer analytics offering.

Furthermore, a company offering social media analytics for business intelligence might have a consumer offering to complement its business offering, since the offerings could potentially share the same data set. Consumer-facing analytics applications may also benefit by engaging consumers to feed more data into the system.

This section looks at possible consumer applications of social media analytics.

6.5.1 Consumer Trends

Similar to businesses using analytics to see trends on the social web, consumers can also find out what are the hottest things occurring on the social web. Social bookmarking sites such as Digg and Reddit allow users to submit interesting media on the Internet and vote on the hottest topics, but an analytics approach could construct a similar site which lists active topics on the Internet and points to discussions.

6.5.2 Product Recommendations

Online news sites and blogs generally feature product reviews written by professional writers. Online shopping sites such as Amazon let customers write reviews for the products that they purchase, and discussion boards generally include lots of product discussion. Social media analytics can aggregate commentary from the Internet and help consumers make informed

decisions when purchasing products. Furthermore, content analysis could also help consumers make comparisons with similar products and provide insight into particular brands.

6.5.3 Network Aggregation

With social networks appearing for many vertical segments, such as Facebook for college students and LinkedIn for professionals, people are signing up for multiple networks and managing multiple profiles. One application of social media analytics might be to aggregate network linkages for a particular user across all of the user's network memberships and provide a centralized view of profiles and updates to the user's networks.

6.5.4 Reputation Monitoring

Because anyone is free to post anything on the Internet, undesirable information about individuals may appear on the internet and damage reputations. As a result, people occasionally "Google" themselves on the Internet to see what is being said about them. For celebrities, business executives, and other image-conscious people, social media analytics can offer insight into how individuals are being perceived on the web.

7 Relation to Consumer Research Services

Because social media relates to consumer attitudes, perhaps the best case for social media analytics can be made by framing the opportunities in terms of the consumer research market. For decades, consumer research firms like J.D. Power & Associates and Nielsen have been telling companies how they are perceived by consumers based upon traditional media such as television and print media. These firms collect data through multiple offline sources, such as consumer panels and surveys.

Now with consumers turning to the Internet, web and social media analytics can provide a digital age equivalent of traditional consumer intelligence. This section outlines the case of online media analytics as a complement to the services currently offered by the traditional consumer research firms.

7.1 Traditional Consumer Intelligence

J.D. Power & Associates and Nielsen are the two major players in the traditional consumer intelligence industry. Together, these firms offer a wide array of services, ranging from audience measurement to product quality assessment.

J.D. Power & Associates was founded in 1968 and describes itself as “a global marketing information firm that conducts independent and unbiased surveys of customer satisfaction, product quality, and buyer behavior” (J.D. Power and Associates, 2008). Based in Los Angeles County along with multiple international offices, the company’s 750 analysts conduct a variety of consumer studies. J.D. Power is best known for work in the automotive industry, where the company publishes rankings of the top automobiles and gives awards for high customer satisfaction. The company also serves a variety of other industries, such as healthcare, insurance, sports, telecom, and hotel. Reports made by J.D. Power might include tracking of customer satisfaction over time, key customer satisfaction issues, shopping experience, and product quality.

The Nielsen Company has a long history in consumer intelligence and is the result of many mergers and acquisitions. The combination of many different companies has allowed Nielsen to brand itself as “the world’s leading provider of marketing information, audience measurement, and business media products and services” (The Nielsen Company, 2008). The company is best known for its research in the traditional media industry. ACNielsen, which was founded in 1923 in Chicago and now under the larger Nielsen brand, offers the Nielsen Ratings which does audience measurements.

While J.D. Power focuses exclusively on analysis drawn from consumer surveys, Nielsen draws from a combination of sources. In addition to surveys, Nielsen uses special tracking services for detailed monitoring of consumer behavior. These “in-home scanners” can reveal information such as television viewing habits for media tracking and purchasing habits for retail data. Nielsen also has access to other supporting data, such as retail store sales data.

7.2 Impact of Online Analytics

Web analytics and social media analytics combined can potentially offer new pathways to performing the analysis done through traditional methods. In many ways, online analytics could potentially displace offline and manual data collection methods. In general, the aim of the consumer research companies is to survey consumers on different criteria, such as actions, experiences, habits, preferences, and satisfaction, depending on the type of study. With the Internet and social media, consumers have natural outlets to voice information related to these criteria. Therefore, it would seem that these new online outlets can be used for data collection and analysis.

Consider how online analytics could apply to the following research services:

- Product quality and overall satisfaction measurement

Quality and satisfaction surveys are generally based upon multiple criteria, such as product and brand reputation, customer service, and store experience. These criteria

are then transformed into numerical scores, and the individual scores from each criterion are weighted and combined to form an overall score. Similarly, social media analytics can potentially provide the same insight. Blogs and discussion forums can reveal consumer sentiment, and filtering can help identify specific customer issues. In addition, review sites where consumers submit numerical ratings can be harvested and factored into the measurements.

- Audience measurement

Traditional media, such as television, print, and CD/DVDs, require audience measurement services to gauge levels of consumer interest. These measurement services are best known for producing the top ten lists for the different media types. An online method for measurement could involve the use of both web analytics and social media analytics. To a certain extent, audience size is correlated to online interest, so web analytics could provide insight using page views as one metric. Larger audience sizes should also correlate to higher activity in social media, with a greater level of consumer-generated content.

- Advertising effectiveness

Similar to audience measurement, advertisers need to gauge the levels of consumer interest that result from advertising campaigns. Aside from increases in sales, audience measurement has traditionally been measured by conducting phone or mail surveys. However, a Berkeley study on web visitation behavior found that offline advertising increases online page visits as a result of increased consumer awareness (Ilfeld & Winer, 2001). Therefore for television ads, web analytics could be used to measure the degree of spikes in online traffic after each television ad appearance. For ads on other media, web analytics can provide a smaller degree of insight on effectiveness. Social media can also be a valuable channel for measuring effectiveness.

As consumer awareness increases, there should be more discussion and activity around the product or service advertised. In many cases, very successful ads generate significant levels of online buzz, with consumers discussing the ad itself. As a result, it is not uncommon to see good television ads featured by YouTube.

These examples show how a combination of web and social media analytics can bring consumer research into the digital age. As consumer usage of the Internet increases, companies will have an alternate source of data aside from research services that mainly rely on offline surveying.

As an added advantage, harvesting the research data from the Internet allows faster results with near-instantaneous feedback in some cases. This contrasts to some types of consumer research where specialized firms have to be commissioned for reports that can take weeks or months to finish. In addition, this data-driven analytics approach can be integrated into business intelligence frameworks, enabling companies to combine this consumer research data with other data sources, such as offline sales data or CRM databases.

Using online analytics cannot fully replace the services offered by consumer research firms, as there are several disadvantages and limitations:

First, there are significant limitations on the type of data found on the Internet. While there may be an abundance of social media destinations on the topic of automobiles, people are less likely to discuss more specialized topics such as healthcare and utilities (both of which are industries served by J.D. Power). Even in areas where there is much discussion online, there are criteria where there may be insufficient information to draw conclusions using only online data. For example, research on car dealer satisfaction may require feedback on customer experiences at local dealerships. Such granularity should not be expected online.

Next, the data on the Internet can be “noisy,” meaning that in some cases it may be difficult to obtain an accurate representation. Online rating sites can be easily manipulated with

people submitting a massive amount of ratings. Paid individuals can submit biased reviews of products or services. Furthermore, there is inherent noise in the process of capturing and analyzing unstructured data on the Internet. In applications such as audience measurement and advertising effectiveness, it is difficult to rely only on web analytics as a metric, given other variables that lead people to visit websites.

Finally, consumer research firms offer analytic consulting services and have access to certain private data sources that may be valuable. Nielsen, for example, has the ability to make complete profiles of households, which provides a different perspective on consumer habits and needs. The company also has access to specialized data like retail sales data, store inventory, and performance data.

Online analytics of the web and social media bring consumer research into the digital age. With an abundance of consumer data available, companies can gain readily harvest this data using a computerized approach and incorporate the data into a business intelligence framework. However, this powerful combination cannot replace consumer research firms like J.D. Power and Nielsen. Rather, the use of web and social media analytics complements these service offerings, offering a new source for data. In fact, consumer research firms are in the process of adopting social media analytics to boost their offerings, as evidenced in Nielsen's acquisition of BuzzMetrics.

8 Technological Challenges

Harvesting and processing data from the social web imposes strong technological requirements for effective results. The structure nature of the social web, with its many linkages from place to place on the Internet is very different from traditional media. The entire process from data mining to output can be broken into several stages, known here as the processing pipeline.

The technological requirements for social media analytics will vary depending on the feature set that is offered. Instead of narrowly defining a fixed feature set and describing the requirements for executing that set, this section will take a more general approach by first proposing a general purpose analytics platform and then exploring potential modules that can plug into the core architecture for data analysis.

The first section, system functions, details the main functions that are part of a social media analytics system. In many ways, the requirements parallel the infrastructure of traditional business intelligence systems. The second section, analysis modules, looks at some types of analyses specific to social media.

8.1 System Functions

8.1.1 Data Acquisition

Harvesting data from online sources can arrive as structured or unstructured data. When data is structured, that means that all data arrives using a standardized format. This might be used in situations where the analytics platform publishes an API which allows data sources to feed the raw data into the system. For example, if a social network agreed to let the analytics platform data mine the network, the social network could export data from its network and feed it into the analytics platform.

However, unstructured data is more common, where there is no standard format. These are typical online websites, where each site has its own presentation and data format. Extracting data from these websites is a two step process: first, the website must be crawled in a manner similar to how web search engines discover websites. Then, data must be extracted out of the pages, using algorithms to determine where the content is and how the content is arranged on the page. Additional algorithms might then categorize the data and put the data into a structured data format.

For a general data analytics platform, the number and variety of data sources will have a noticeable effect on the quality of results. For a more complete picture of what users are doing and saying on the internet, unstructured data will be the predominant source of data.

8.1.2 Data Storage

Social network analytics works by processing large volumes of data. The content on the Internet is expansive and growing at an exponential rate as more users contribute to content online. As a result, a high performance storage system must be capable of accepting large volumes of data at a rapid pace and running complex data queries on behalf of the data analysis modules. Since large volumes of data will need to be stored across multiple disks and most likely multiple machines, the system must be optimized to run over a distributed network and build in a level of redundancy.

A database is also required to store the processed results from the raw data for later retrieval. Each type of analysis performed on the data may output different sets of results, so the amount of processed data may be larger than the raw data.

Finally, the database should store operation information for the system, such as settings, user accounts, and queued tasks.

8.1.3 Data Analysis

The data analysis portions of the platform determine how documents are classified and aggregated. When the data acquisition module feeds data for analysis, the data analysis module will need to decompose the data and extract meaningful information from the data.

Information the data analysis module might provide include:

- General context of information
- Summaries of information
- Sentiment
- Relationships of information with other sets of information
- Metrics
- Actionable items

As a result, various technologies will need to be explored, including artificial intelligence classifiers, natural language processing, graph algorithms, and ranking algorithms.

8.1.4 User Interface / Visualization

Because a data analysis system processes a large volume of data, a well-designed user interface is necessary to provide an effective way for users to interpret the results. The user interface should allow users to configure the parameters of the analytics platform to analyze target information. Then, the user interface should allow the user to drill down into the data to extract relevant information. This might include charts and graphs to show statistics and trends and other visual elements for more complex data, such as relationship graphs.

8.2 Analysis Modules

8.2.1 Natural Language Processing (Content Analysis)

8.2.1.1 Background

Natural language processing (NLP) is a part of artificial intelligence trying to tackle the problem of understanding human language. NLP makes it possible for computers to algorithmically determine what the subject of a piece of writing is about and more importantly for social media analytics, determine the sentiment of the writing.

Languages are very difficult and complex structures, with an infinite number of variations in expressing an idea. As a result, there are numerous challenges to solve: Sentence structure is inherently complicated with phrases and clauses that can be nested. Words can have multiple meanings, making it important for machines to determine context. Verbs have many different conjugations. Each language has unique characteristics, making it difficult to create a general algorithm that works out of the box for other languages. Many more challenges exist, which is why NLP is a very active field of research in artificial intelligence.

In order to parse sentences, machines need to know the structural rules of sentence composition and have a word bank containing words tagged with part of speech (Berwick, 2008). Then, a parsing algorithm creates a tree structure based upon the part of speech of words and the location of the words. The branches in the tree allow machines to determine the role each word plays in the sentence and what ideas are formed.

Today, NLP research continues to advance with increased computational power and better machine learning algorithms. Research today involves probabilistic algorithms with machine learning to determine how to interpret ambiguous sentences.

8.2.1.2 Applications

A primary application of NLP in social media analytics is to determine the topic and context of user-generated content, such as blog or forum posts. Given a body of text, NLP can

determine the subject matter in the text and determine the relevance of that body of text to different topics. This is important for social media analytics because in subject-specific applications such as brand monitoring, it is important to determine what content is most relevant to the current subject.

Another key application of NLP is in sentiment analysis. Once NLP algorithms parse the content and determine the relationships in the content, sentiment can be calculated by looking at positive or negative words associated with the subject. For example, an analysis of a sentence containing “*Product X performs very poorly*” would associate the noun *X* with the modifier *poorly* and therefore assign a negative sentiment score to *X* for that particular sentence.

Sentiment analysis remains a very difficult challenge because of language complexity. Sentiment scores can be skewed in many ways, such as highly complex sentences (such as ones including quotations), discussions that have multiple subjects (such as comparisons), and human sarcasm.

As an example, the following texts might be misinterpreted by NLP algorithms to have the wrong sentiment:

To say that Product X is the worst product is unfounded.

Magazine Y found Product X to be best in class. I think their assessment is wrong.

I heard Product X sucks, but I wanted to get your opinion on it.

Because of the language complexities, accuracy in sentiment analysis is still relatively low which can result in a lot of noise in aggregated sentiment calculations.

8.2.2 Social Network Analysis

8.2.2.1 Background

Social networks, in which people build relationships with others through some common interest, can be visualized as a large graph with people as nodes and connections as links between the nodes. Social network analysis analyzes the structure of the graph and extracts meaningful organizational data out of the graph. Formally defined, social network analysis “is the mapping and measuring of relationships and flows between people, groups, organizations, computers, web sites, and other information/knowledge processing entities” (Krebs, 2008).

There are multiple different analyses that can be performed on social networks. Some of these measurements include:

- Degree centrality – determines the number of direct connections to a node, which allows one to determine which nodes are most active and which function as network hubs
- Betweenness Centrality – determines which nodes are critical to information flows in a network; in other words, this determines which nodes function as bridges between different constituencies
- Closeness Centrality – determines which nodes have the best access to all other nodes in the network; these nodes can efficiently monitor other nodes in the network
- Network Reach – determines what kind of reach a node has given one, two, or three degrees away

Social network analysis algorithms currently face two major challenges. The first challenge is in creating scalable algorithms. As more nodes are added to a network, the running time of the algorithm increases exponentially, making it very difficult to analyze large networks. The second challenge is in creating effective visualizations for the networks. Current methods of

drawing dots and lines end up in a large and messy mass. Researchers are still working on better visualizations so people can drill into the network graph effectively.

8.2.2.2 Applications

Applications for social network analysis include:

- Determining the most influential people in a social network or Internet forum – algorithms such as betweenness centrality can generate a list of people with the most influence in the network. This information can then be used by marketers to understand how these individuals affect the overall network, and targeted advertising campaigns may single out those individuals.
- Determining the relative importance of different websites – Google and other search engines use a form of network analysis involving the links between different websites to determine the rank of websites. This type of ranking is important in determining which websites are most authoritative and influential for a particular subject and is useful for determining which websites to focus more for monitoring and targeted advertising.

8.2.3 Other Modules

Natural language processing and social network analysis are two major components in social media analytics. However, there are other supporting modules, one of which is text mining.

Text mining is a broad field of data analysis. Text mining includes natural language processing and sentiment analysis, but there are also other text mining applications. Keyword clustering and concept mining allows algorithmic determination of keywords and key phrases in content. With text mining, content crawled on social media can be automatically tagged and categorized. As practical examples, this technology can automatically determine the competitors for a company or a company's product line. This approach of automated discovery can also yield interesting and unexpected results, such as

identifying new entrants into the market. This information can also be visualized on a radar-type chart that plots keywords and associations.

9 Companies in Social Media Analytics

This section profiles a select group of companies active in the social media analytics space. The companies vary significantly in size and maturity, showing that startups see an opportunity in the social media analytics space while existing intelligence companies are adapting with their own social media offerings.

9.1 Biz360

Founded in San Mateo, California in 2000, Biz360 started off as a market intelligence company focusing on tracking print, online, and broadcast media. The company has been very successful, with its last round of financing, a \$10.5M Series C, taking place in 2003. By then, the company had already built an impressive client list, including 3Com, Bank of America, Oracle, and Sun Microsystems (Biz360 Inc., 2003). Social media monitoring did not begin until the end of 2005.

The Market360 platform aggregates content across multiple types of media in order to give customers a complete picture of media activity. This coverage includes both online and offline media, such as newspapers, magazines, broadcast, web, and social media (Biz360 Inc., 2008). An online dashboard provides access to analytics information, including metrics such as reach, coverage, and sentiment. Within social media, Biz360 primarily monitors blogs and to a lesser extent, forums. Biz360's sentiment calculation is advertised as Point-of-View Sentiment, which requires users initially to manually assign sentiment rankings in order for the system to learn the user's point of view.

Digital Music - Brand and Reputation Dashboard

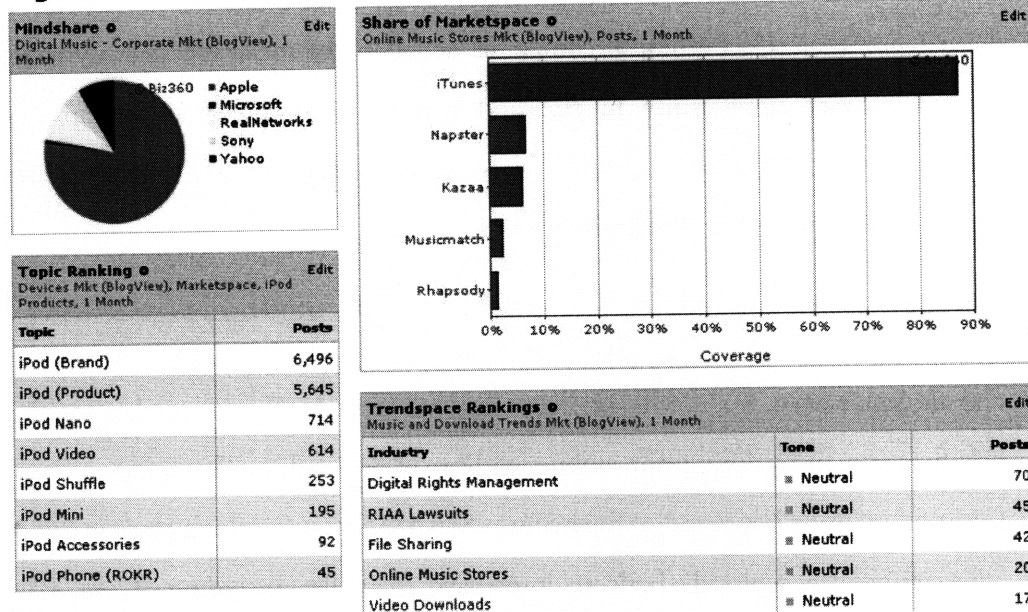


Figure 13 - Biz360 BlogView Brand and Reputation Dashboard

9.2 BuzzLogic

Based in San Francisco, BuzzLogic was founded in 2004. In 2007, BuzzLogic raised \$12.1M in Series A funding. The company formally launched its enterprise product in April 2007 to track online influence.

The software-as-a-service offering provides users with a dashboard, on which keywords to be monitored can be defined. BuzzLogic crawls “millions of blogs, social networks, and other social media, as well as corporate sites and approximately 8,000 mainstream media sites” (BuzzLogic, 2008). The corporate website does not specifically state that the service monitors forums. Once keywords have been defined, BuzzLogic produces a list of influential bloggers and posts, with analytics information such as a social network map. Unlike most social media analytics services, BuzzLogic does not use computer-based sentiment calculation, reasoning that computer algorithms were too inaccurate. Instead, users must manually rate the sentiment of each post in order to see change over time.

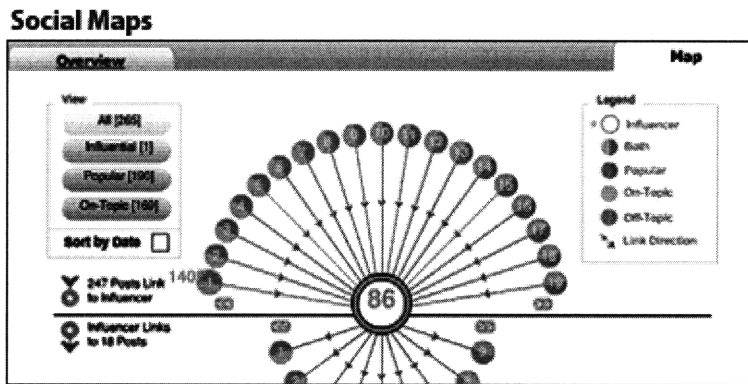


Figure 14 - BuzzLogic Social Map Visualization

BuzzLogic introduces the notion of Conversation Targeting, which is a form of ad targeting across multiple websites. The service claims to allow advertisers to place ads across conversations as they spread from blog to blog. However, only Google AdWords are supported as the means of placing advertisements.

Ad Targeting

Conversations Sites	Daily Imps	Selected Sites	Daily Imps
John's iPhone Project (6)		Influencers & Traffic (3)	
Influencers (1)		www.engadgetmobile.com/	500k+
2 www.cellphonebeat.com/	0-10k+	www.gizmodo.com/	500k+
		www.tumblr.com	10k-100k
Instinks (5)		Additional Sites (1)	
1 into www.engadgetmobile.com/ (2)		www.mashable.com	500k+
www.cellphonebeat.com/	0-10k+		
www.lifehacker.com/	10k-100k		
2 into www.cellphonebeat.com/ (2)			
3 into www.gizmodo.com/ (1)			

Figure 15 - BuzzLogic Ad Targeting Interface

BuzzLogic targets advertiser, marketing, and PR professionals as the primary customers of its service. On its website, BuzzLogic claims its clients include eBay, Nordstrom, and Southwest Airlines. Pricing for the service is sold on a per-conversation query basis, with entry level pricing at \$12,000 per year for 10 conversation queries.

On April 22, 2008, BuzzLogic announced the acquisition of ActiveWeave, which develops BlogRovR, a free browser plug-in for consumers that recommends related stories when browsing blogs. This acquisition allows BuzzLogic to better monitor the interactions between blogs, while offering consumers a service in return. This is one of the first social

media analytics companies to sponsor a consumer offering that plugs into its analytics framework.

9.3 Nielsen BuzzMetrics

Headquartered in New York with seven regional offices, BuzzMetrics has a long history in helping companies understand word-of-mouth behavior and is now one of the dominant players in the social media analytics space. BuzzMetrics evolved from the combination of multiple companies working on different technologies (Dowdell, 2008):

- BuzzMetrics, a company specializing in word-of-mouth consumer interactions was originally founded in New York City in 1999.
- Trendum was founded in Israel in 2000, specializing in natural language processing.
- Intelliseek was founded in Cincinnati in 1997 originally as an enterprise search company. In 2001, the company merged with PlanetFeedback, a website which allowed consumers to voice their feedback on businesses and allow businesses to respond. Combined, Intelliseek evolved into a market intelligence company, specializing in consumer insight and data mining.
- Trendum partnered with The Nielsen Company (then known as VNU) in 2005 to acquire BuzzMetrics, with Trendum using the BuzzMetrics name. The following year, BuzzMetrics acquired Intelliseek.
- In 2007, Nielsen decided to acquire all remaining shares of BuzzMetrics, integrating the company into its Internet division as Nielsen BuzzMetrics.

BuzzMetric's main product offering is BrandPulse, which monitors blogs, forums, newsgroups, and consumer review sites (BuzzMetrics, 2008). BrandPulse uses a hosted ASP delivery model that has a web-based reporting interface. BrandPulse engagements typically last for at least a year, with 2006 pricing of \$80,000 to \$500,000 depending on the number of users and number of brands being monitored. BrandPulse also offers one-off consulting

reports for specific questions that need to be answered with BuzzMetric's experts, with individual reports starting at \$20,000 (Schneider, 2006). In addition to the general service offering, BuzzMetrics has developed five industry practices for special knowledge and experience: automotive, consumer electronics, nutrition, pharma-health, and television.

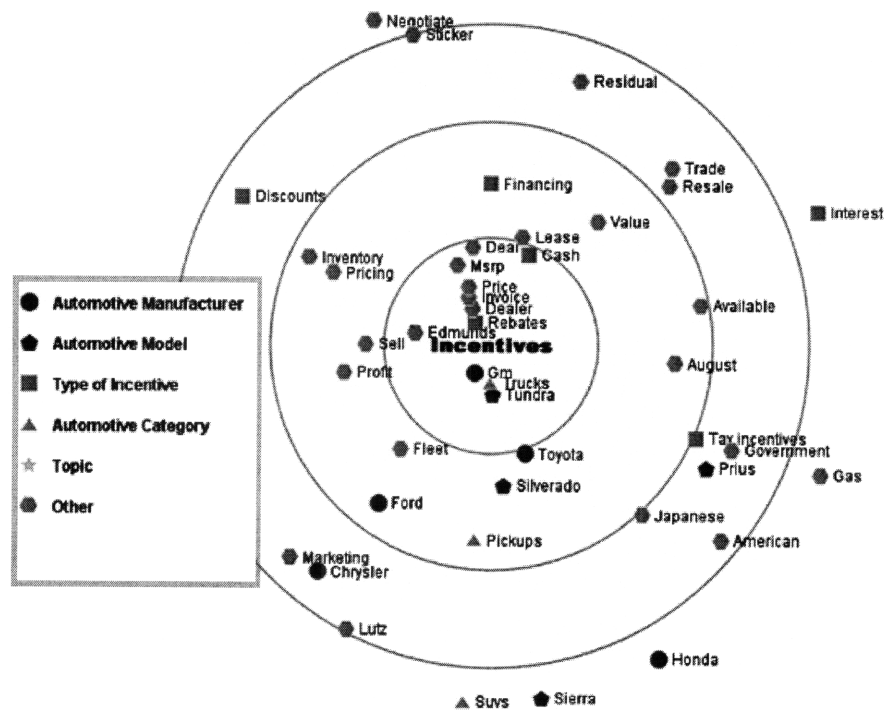


Figure 16 - BuzzMetrics Brand Association Map Example

The client list for Nielsen BuzzMetrics contains major corporations, such as Comcast, General Motors, Microsoft, Procter & Gamble, and Sony. Since BuzzMetrics is owned by Nielsen, a multi-billion dollar market intelligence company, BuzzMetrics is well positioned in the market and can access resources available across Nielsen's many divisions.

BuzzMetrics also offers BlogPulse, a free online service for tracking blogs which originally began as an Intelliseek service. In many ways, the service can be described as Alexa for blogs. The website crawls blogs and determines basic statistics such as top blog posts, videos, and news stories being discussed. The website also allows users to input topic keywords and compare trends, using the frequency of the term occurring in blog posts as the trend metric.

As of April 2008, BlogPulse was testing a new feature which would allow users to request blog profiles. These profiles contain basic analytic information, such as a graph of the blog's trend, the blog's rank, a listing of recent posts and blogs linking to the site, and links to similar blogs (BlogPulse, 2008).

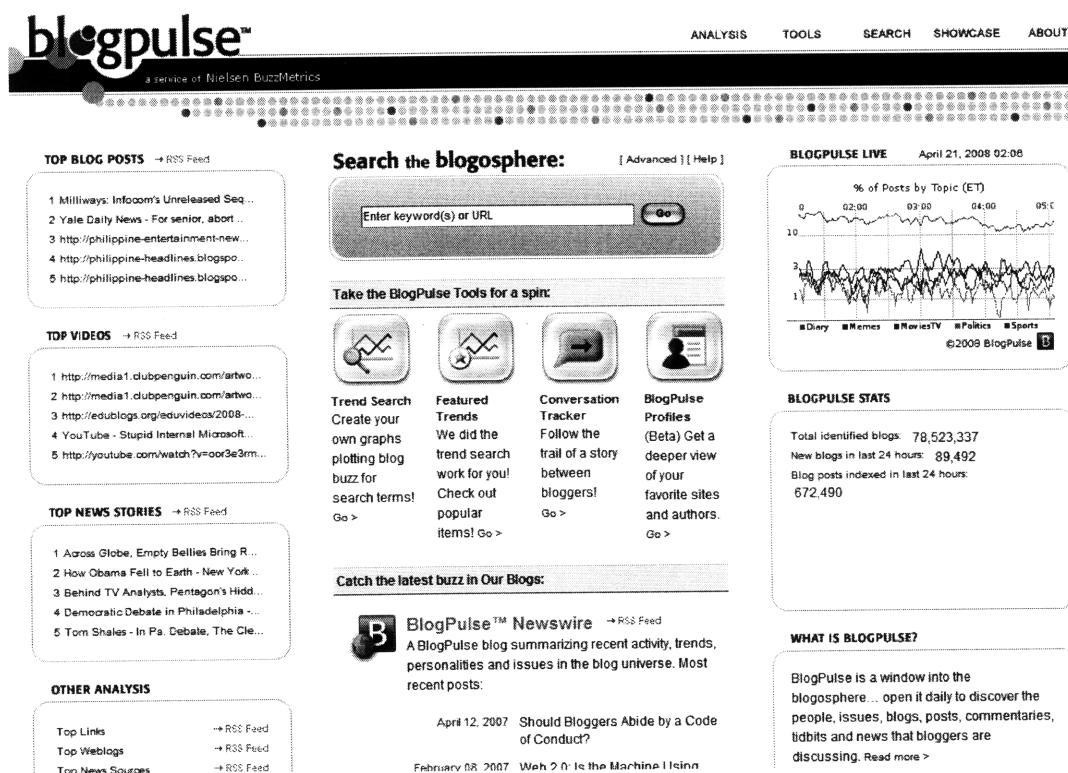


Figure 17 - BuzzMetrics BlogPulse (www.blogpulse.com)

9.4 Radian6

Founded in 2006, Radian6 is a small Canadian company located in Fredericton, New Brunswick. In social media monitoring, the company targets primarily advertising agencies and PR professionals. The company received an initial \$4M round of funding in April 2007 with BCE Capital Inc. as the lead investor (Radian6, 2008). As of November 2007, the company employed 23 people, with plans to double in size over the following year.

Radian6 monitors “all forms of social media including blogs, top video sharing sites, image sharing sites, microblogging sites, online mainstream media and other sites as they become available.” Radian6 uses a software-as-a-service model, with an online Flash-based interface.

Based upon the descriptions and screenshots provided on the company website, Radian6’s interface appears more suited for data collection. Users must set up the system by inputting keywords for monitoring, such as brand names and product names. Then, social media topics matching the keywords are reflected on the dashboard. Users can then drill down into individual pieces of content to see more detail and analytics information, such as comments and viewership. Radian6’s interface uses a widget model, where each piece of information appears in a separate window.

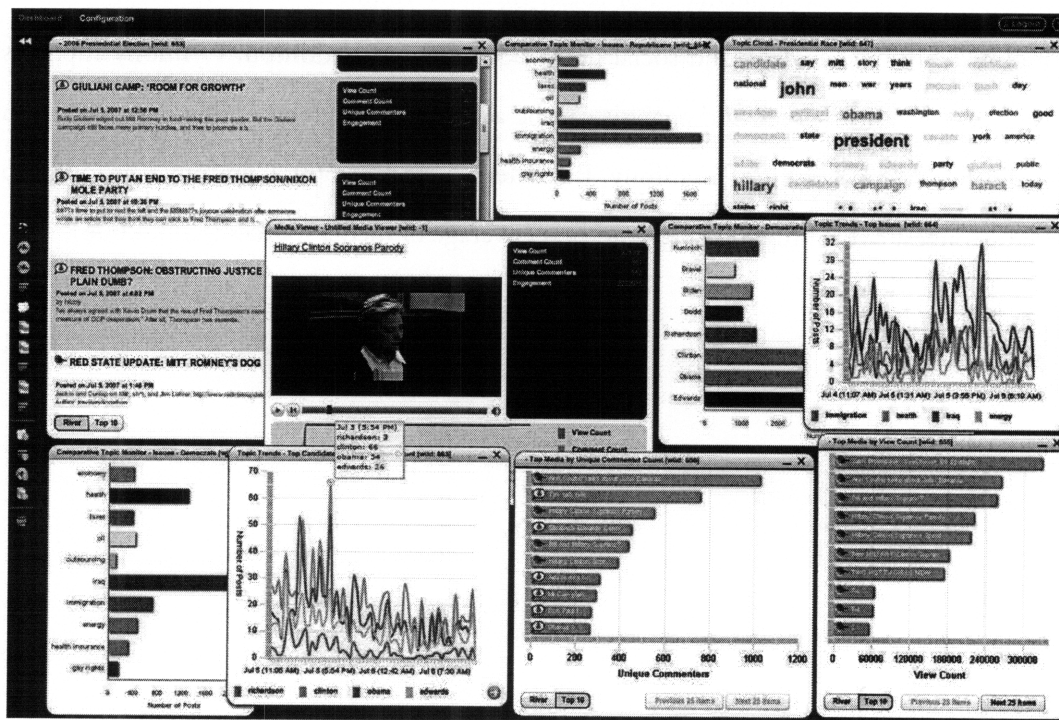


Figure 18 - Radian6 Dashboard

9.5 TNS Media Intelligence/Cymfony

Founded in Newton, Massachusetts in 1996, Cymfony is another well-established player in the general media intelligence market. Similar to Biz360, the company offers real-time media tracking and analysis and added social media monitoring in January 2005. The company's last round of financing was a \$7M Series D in 2005 before being acquired in 2007 by TNS Media Intelligence, one of the largest U.S. providers of competitive intelligence. Today, the company is located in nearby Watertown with a second office in Bangalore, India (Cymfony, 2008).

Cymfony's primary offering is its Orchestra platform, which aggregates content from both traditional media and social media. For social media, Cymphony monitors blogs, forums, newsgroups, consumer review, and social networking sites. In addition, Cymfony supports 13 languages. Real-time data and analysis can be accessed through an online dashboard. In a 2006 Forrester comparison of Cymfony, BuzzMetrics, and Biz360, Cymfony was found to have the strongest overall functionality and best coverage of media sources (Kim, 2006).

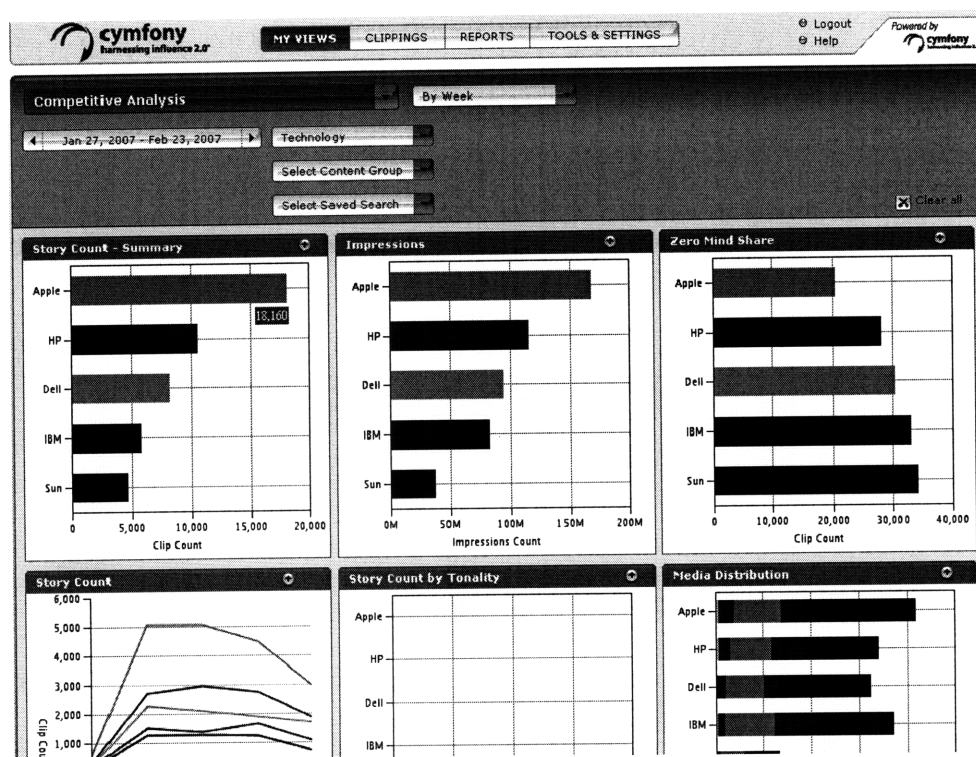


Figure 19 - Cymfony Dashboard

Pricing information for Cymfony was not available. The company also pairs its Orchestra offering with consulting services, which are weekly or monthly engagements where dedicated experts deliver custom reports. Cymfony clients include ConocoPhillips, Doubleclick, LG, Southwest Airlines, and Verisign.

9.6 Visible Technologies

Visible Technologies was founded in 2003 with headquarters in Seattle and offices in New York and Boston. In 2006, the company received \$3.5M in Series A funding from WPP, a multinational communications services company. In September of 2007, Visible Technologies raised a \$12M Series B round led by Ignition Partners and WPP (Visible Technologies, 2008).

The company offers two products: TruCast, which analyzes social media, and TruView, which monitors reputation on search engines. TruCast monitors consumer-written content

on social media, including blogs, social networks, and forums. An online dashboard allows users to analyze topics by subject and sentiment and view analytical data such as trends and influential users. TruCast differentiates itself from other services by allowing users to converse with consumers and write responses directly on social media websites, through a central control panel.

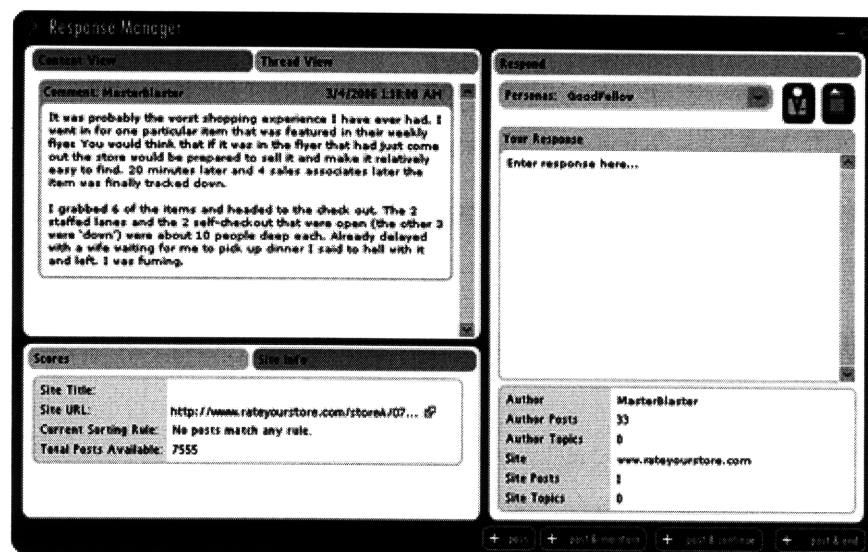


Figure 20 - Screenshot of TruCast's Response Manager (Source: CenterNetworks)

Visible Technologies is focused on the enterprise market. The company has about 20 enterprise level clients as of October 2007, including a division from Microsoft, a Fortune 25 computer manufacturer, and one of the Big 3 automakers (Stern, 2007). Pricing is in the range of \$15,000 to \$25,000 per month.

10 Summary and Conclusion

Business intelligence and decision support systems have become critical components in the everyday operations of organizations. Since its introduction about half a century ago, business intelligence has evolved as new technology became available. Faster processors meant more analytics possibilities, and data warehousing advances enabled the rapid collection of vast amounts of data for processing. Meanwhile, business intelligence vendors continued to refine product offerings to allow decision makers to access the information they need effectively.

With the arrival of the Internet, business intelligence had to adapt to new models and new sources of data. This spawned the web analytics industry, which provided companies with new sources of information and new types of analyses. The survey of web analytics companies in section 4 offered a look into the many offerings now available ranging from basic statistics packages on webpage visits to highly sophisticated solutions that might tie in online experiences with other offline sources for a full lifecycle analysis.

The Internet and web analytics also continued the trend of increasing information arriving at a faster pace. This increased access to information help companies make more informed decisions quicker. As use of business intelligence systems increase, the speed at which companies can react to changing conditions will be necessary for companies to stay competitive in the market.

The next disruptive force is social media. Section 5 showed that the main forms of social media, such as blogs, forums, and social networking sites, are growing at a rapid pace. Social media allows people to extend their networks by helping them stay in contact with friends and making new acquaintances through special interest communities. Consumer habits are changing, with a majority of Internet users participating in a form of social media, which are having an effect in shaping consumer attitudes and demands.

Social media presents new opportunities and challenges for companies. With consumers becoming more and more active in creating content on the Internet, companies have a new and rich source of data to draw from. The real-time nature of the Internet means that companies can draw upon the consumer-generated content found on social media for the latest intelligence. The spread of social media also means that information is more readily available for consumers, so negative customer sentiment may spread quickly and impact company performance.

Given its rapid growth and affect on society and culture, social media will be an important new area of business intelligence analytics. Section 6 offered some key applications of social media analytics. Social media analytics will be critical in brand and product monitoring applications, which can enable early feedback and action.

Similarly, social media analytics can add a lot of value in areas of market research and advertising. The scale and speed at which consumer input can be gathered and analyzed using social media is unparalleled, displacing traditional methods of obtaining consumer input, such as surveys and panels. Section 7 presented a case study on the impact of online analytics in the consumer research market and validated its use. In particular, online analytics can complement traditional consumer research, with the computerized analytics of online content becoming a new source of data for business intelligence frameworks. Hence, social media is a force that cannot be ignored in the world of business intelligence, and companies should be forming strategies for social media.

With the strong business cases presented for the mining of social media and a growing market, there will be many challenges ahead:

1) Companies need to see the importance of social media analytics

Many companies have no strategy for including social media in business intelligence.

Since the area is relatively new, companies are unaware of the potential applications

of social media analytics. Success of the industry will rely on the ability of early adopters to gain insight from social media.

2) Social media analytics companies have to fine-tune their offerings and feature sets

The survey of companies in section 9 presented a snapshot of current offerings. These offerings are relatively new given a young market, so it is expected that current products and services are immature. Service providers are still working to determine the most effective ways of presenting data from social media and the features that are valuable to decision makers. Current offerings may overwhelm users with too many charts, graphs, and other findings. In addition, social media analytics have yet to be effectively integrated into enterprise business intelligence solutions, so systems such as ones from Oracle and SAP can draw upon social media.

3) Technology must improve in natural language processing and other algorithms

Natural language processing algorithms have made great progress, but there is still much room for improvement in terms of accuracy. Manual sentiment ranking by humans used by some companies will not be viable in the long-term. Other algorithms such as ones used social network mapping also have areas for improvement.

Social media analytics in business intelligence holds a lot of potential. It will be interesting to see how the market develops as social media usage increases and the technology to analyze social media improves. There are certainly many business intelligence applications for social media, and opportunity awaits companies willing to embrace social media as part of their strategy.

11 References

- Aertsen, J. (2007, November 16). *Technical BI Architecture*. Retrieved 16 2008, February, from kJube BI: <http://www.kjube.be/tnenopxe/index.php?section=63>
- Alexa. (2008). *Alexa*. Retrieved March 15, 2008, from Corporate Website: <http://www.alexacom>
- Berwick, R. C. (2008, February). *Lecture Notes*. Retrieved April 25, 2008, from MIT 6.863J/9.611J Natural Language Processing: <http://web.mit.edu/6.863/www>
- Bitpipe, Inc. (2008). *Business Intelligence (BI) Overview*. Retrieved January 2008, from Bitpipe: <http://businessintelligence.bitpipe.com/bi>
- Biz360 Inc. (2003, May 19). *Leading Market Intelligence Software Company Biz360 Secures \$10.5 Million Series C Financing, Achieves 184% Growth, Announces New CEO*. Retrieved April 22, 2008, from Granite Ventures, LLC: <http://www.granitevc.com/html/news/2003Q2/052103.html>
- Biz360 Inc. (2008). *Biz360*. Retrieved April 23, 2008, from Corporate Website: <http://www.biz360.com>
- BlogPulse. (2008). *Nielsen BuzzMetrics' BlogPulse*. Retrieved April 21, 2008, from Corporate Website: <http://www.blogpulse.com>
- Boyd, D. M., & Ellison, N. B. (2007). Social Network Sites: Definition, History, and Scholarship. *Journal of Computer-Mediated Communication*, 13 (1), 11.
- Business Objects. (2008). *Business Intelligence*. Retrieved January 2008, from Business Objects: <http://www.businessobjects.com/businessintelligence/>
- BuzzLogic. (2008). *BuzzLogic*. Retrieved April 22, 2008, from Corporate Website: <http://www.buzzlogic.com>

BuzzMetrics. (2008). *Corporate Website*. Retrieved April 20, 2008, from Nielsen BuzzMetrics: <http://www.nielsenbuzzmetrics.com>

Chaudhuri, S., & Dayal, U. (1997). An Overview of Data Warehousing and OLAP Technology. *SIGMOD Rec.*, 26 (1 (Mar. 1997)), 65-74.

Compete. (2008). *Compete*. Retrieved March 15, 2008, from Corporate Website: <http://www.compete.com>

comScore, Inc. (2008). *comScore*. Retrieved March 15, 2008, from Corporate Website: <http://www.comscore.com>

Cusumano, M. A. (2004). *The Business of Software: What Every Manager, Programmer, and Entrepreneur Must Know to Thrive and Survive in Good Times and Bad*. New York: Free Press.

Cymfony. (2008). *TNS Media Intelligence/Cymfony*. Retrieved April 23, 2008, from Corporate Website: <http://www.cymfony.com>

Deveau, D. (2007, November 21). *It's 11 p.m. Do you know where your logo is?* Retrieved April 11, 2008, from [globeandmail.com: http://www.theglobeandmail.com/servlet/story/RTGAM.20071121.wtq-1107BrandProtect/BNStory/?pageRequested=all](http://www.theglobeandmail.com/servlet/story/RTGAM.20071121.wtq-1107BrandProtect/BNStory/?pageRequested=all)

Dowdell, J. (2008). *Nielsen Buzz Metrics Market Research Marketing Company Profile*. Retrieved April 20, 2008, from Marketing Shift (Online Marketing Blog): <http://www.marketingshift.com/companies/marketing/market-research/nielsen-buzzmetrics.cfm>

Fishkin, R. (2006, July 20). *Top 100 Digg Users Control 56% of Digg's HomePage Content*. Retrieved March 8, 2008, from SEOMoz: <http://www.seomoz.org/blog/top-100-digg-users-control-56-of-diggs-homepage-content>

Gassman, B. (2006, March 31). MarketScope for Web Analytics, 2Q06. Stamford, CT: Gartner, Inc.

Giles, J. (2005). Internet encyclopaedias go head to head. *Nature* (438), 900-901.

Golfarelli, M., Rizzi, S., & Cella, I. (2004). Beyond Data Warehousing: What's Next in Business Intelligence? *Proceedings of the 7th ACM international workshop on Data warehousing and OLAP* (pp. 1-6). Washington DC: ACM.

Google, Inc. (2008). *Google Analytics*. Retrieved March 15, 2008, from Corporate Website: <http://www.google.com/analytics>

Ilfeld, J. S., & Winer, R. S. (2001). *Generating Web Site Traffic: An Empirical Analysis of Web Site Visitation Behavior*. Berkeley: University of California at Berkeley.

Inmon, W. (1992). *Building the Data Warehouse*. John Wiley.

J.D. Power and Associates. (2008). *Corporate Website*. Retrieved April 30, 2008, from J.D. Power and Associates: <http://www.jdpower.com>

Joy, K. (2006, December 19). *BD-BrandProtect Perspective: Protecting the Integrity of Your Brand from Online Risk Exposure*. Retrieved April 12, 2008, from BD-BrandProtect: <http://www.brandprotect.com>

Karel, R. (2007). *The Forrester Wave: Enterprise ETL, Q2 2007*. Cambridge: Forrester.

Kim, P. (2006, September 13). The Forrester Wave: Brand Monitoring, Q3 2006. Cambridge: Forrester.

Kimball, R. (2004, December 4). *The 38 Subsystems of ETL*. Retrieved February 16, 2008, from Intelligent Enterprise: Better Insight for Business Decisions: <http://www.intelligententerprise.com/showArticle.jhtml?articleID=54200319>

Klein, M. (2007, January 13). *Protecting your business against online brand abuse*. Retrieved April 9, 2008, from WTN News: <http://wistechnology.com/articles/3623/>

Krebs, V. (2008). *Social Network Analysis*. Retrieved 26 2008, April, from Orgnet: <http://www.orgnet.com/sna.html>

Li, C. (2007). *Social Technographics*. Cambridge: Forrester Research, Inc.

Luhn, H. (1958). A Business Intelligence System. *IBM Journal of Research and Development* , 2 (4), 314-319.

Olszak, C., & Ziemba, E. (2006). Business Intelligence as a Key to Management of an Enterprise. *InSITE* (pp. 855-863). Informing Science.

Omniiture. (2008). *Omniiture*. Retrieved March 15, 2008, from Corporate Website: <http://www.omniiture.com>

Peterson, E. T., & Bayriamova, Z. (2007). *March 2007 Global Survey - Global Use and Attitudes about Web Analytics Technology and Processes*. Web Analytics Demystified.

Power, D. (2007, March 10). *A Brief History of Decision Support Systems*. Retrieved November 29, 2007, from DSSResources.COM: <http://dssresources.com/history/dsshhistory.html>

Radian6. (2008). *Radian6*. Retrieved April 22, 2008, from Corporate Website: <http://www.radian6.com>

Schneider, K. (2006, December 17). Brands for the Chattering Masses. *New York Times* .

Scholz, T. (2007, September 26). *A History of the Social Web (draft)*. Retrieved December 1, 2007, from Collectivate.net: <http://www.collectivate.net/journalisms/2007/9/26/a-history-of-the-social-web.html>

Sprague, R. J., & Carlson, E. (1982). *Building Effective Decision Support Systems*. Englewood Cliffs: Prentice-Hall, Inc.

Stern, A. (2007, October 5). *Conversation with Visible Technologies VP Marketing Blake Cahill*. Retrieved April 22, 2008, from CenterNetworks: <http://www.centernetworks.com/visible-technologies-blake-cahill>

Technorati. (2008). *About Us: Technorati*. Retrieved February 29, 2008, from Technorati: <http://technorati.com/about/>

The Nielsen Company. (2008). *Nielsen//NetRatings*. Retrieved March 15, 2008, from Corporate Website: <http://www.nielsen-netratings.com/>

VanBoskirk, S. (2007). *US Interactive Marketing Forecast, 2007 to 2012*. Cambridge: Forrester.

Veset, D. e. (2007). *Worldwide Business Analytics Software 2007-2011 Forecast Update and 2006 Vendor Shares*. Framingham: IDC.

Visible Technologies. (2008). *Visible Technologies*. Retrieved April 22, 2008, from Corporate Website: <http://www.visibletechnologies.com>

WebTrends Inc. (2008). *WebTrends*. Retrieved March 15, 2008, from Corporate Website: <http://www.webtrends.com>