MIS Quarterly

SPECIAL ISSUE: BUSINESS INTELLIGENCE RESEARCH

BUSINESS INTELLIGENCE AND ANALYTICS: FROM BIG DATA TO BIG IMPACT

Hsinchun Chen
Eller College of Management, University of Arizona,
Tucson, AZ 85721 U.S.A. {hchen@eller.arizona.edu}

Roger H. L. Chiang
Carl H. Lindner College of Business, University of Cincinnati,
Cincinnati, OH 45221-0211 U.S.A. {chianghl@ucmail.uc.edu}

Veda C. Storey
J. Mack Robinson College of Business, Georgia State University,
Atlanta, GA 30302-4015 U.S.A. {vstorey@gsu.edu}

Business intelligence and analytics (BI&A) has emerged as an important area of study for both practitioners and researchers, reflecting the magnitude and impact of data-related problems to be solved in contemporary business organizations. This introduction to the MIS Quarterly Special Issue on Business Intelligence Research first provides a framework that identifies the evolution, applications, and emerging research areas of BI&A. BI&A 1.0, BI&A 2.0, and BI&A 3.0 are defined and described in terms of their key characteristics and capabilities. Current research in BI&A is analyzed and challenges and opportunities associated with BI&A research and education are identified. We also report a bibliometric study of critical BI&A publications, researchers, and research topics based on more than a decade of related academic and industry publications. Finally, the six articles that comprise this special issue are introduced and characterized in terms of the proposed BI&A research framework.

Keywords: Business intelligence and analytics, big data analytics, Web 2.0

Introduction

Business intelligence and analytics (BI&A) and the related field of big data analytics have become increasingly important in both the academic and the business communities over the past two decades. Industry studies have highlighted this significant development. For example, based on a survey of over 4,000 information technology (IT) professionals from 93 countries and 25 industries, the *IBM Tech Trends Report* (2011) identified business analytics as one of the four major technology trends in the 2010s. In a survey of the state of business analytics by Bloomberg Businessweek (2011), 97 percent of companies with revenues exceeding $100 million were found to use some form of business analytics. A report by the McKinsey Global Institute (Manyika et al. 2011) predicted that by 2018, the United States alone will face a shortage of 140,000 to 190,000 people with deep analytical skills, as well as a shortfall of 1.5 million data-savvy managers with the know-how to analyze big data to make effective decisions.

Hal Varian, Chief Economist at Google and emeritus professor at the University of California, Berkeley, commented on the emerging opportunities for IT professionals and students in data analysis as follows:

> So what's getting ubiquitous and cheap? Data. And what is complementary to data? Analysis. So my recommendation is to take lots of courses about how to manipulate and analyze data: databases, machine learning, econometrics, statistics, visualization, and so on.[1]

The opportunities associated with data and analysis in different organizations have helped generate significant interest in BI&A, which is often referred to as the techniques, technologies, systems, practices, methodologies, and applications that analyze critical business data to help an enterprise better understand its business and market and make timely business decisions. In addition to the underlying data processing and analytical technologies, BI&A includes business-centric practices and methodologies that can be applied to various high-impact applications such as e-commerce, market intelligence, e-government, healthcare, and security.

This introduction to the *MIS Quarterly* Special Issue on Business Intelligence Research provides an overview of this exciting and high-impact field, highlighting its many challenges and opportunities. Figure 1 shows the key sections of this paper, including BI&A evolution, applications, and emerging analytics research opportunities. We then report on a bibliometric study of critical BI&A publications, researchers, and research topics based on more than a decade of related BI&A academic and industry publications. Education and program development opportunities in BI&A are presented, followed by a summary of the six articles that appear in this special issue using our research framework. The final section concludes the paper.

BI&A Evolution: Key Characteristics and Capabilities

The term *intelligence* has been used by researchers in artificial intelligence since the 1950s. *Business intelligence* became a popular term in the business and IT communities only in the 1990s. In the late 2000s, *business analytics* was introduced to represent the key analytical component in BI (Davenport 2006). More recently *big data* and *big data analytics* have been used to describe the data sets and analytical techniques in applications that are so large (from terabytes to exabytes) and complex (from sensor to social media data) that they require advanced and unique data storage, management, analysis, and visualization technologies. In this article we use business intelligence and analytics (BI&A) as a unified term and treat big data analytics as a related field that offers new directions for BI&A research.

BI&A 1.0

As a data-centric approach, BI&A has its roots in the long-standing database management field. It relies heavily on various data collection, extraction, and analysis technologies (Chaudhuri et al. 2011; Turban et al. 2008; Watson and Wixom 2007). The BI&A technologies and applications currently adopted in industry can be considered as BI&A 1.0, where data are mostly structured, collected by companies through various legacy systems, and often stored in commercial relational database management systems (RDBMS). The analytical techniques commonly used in these systems, popularized in the 1990s, are grounded mainly in statistical methods developed in the 1970s and data mining techniques developed in the 1980s.

Data management and warehousing is considered the foundation of BI&A 1.0. Design of data marts and tools for extraction, transformation, and load (ETL) are essential for converting and integrating enterprise-specific data. Database query, online analytical processing (OLAP), and reporting tools based on intuitive, but simple, graphics are used to explore important data characteristics. Business performance management (BPM) using scorecards and dashboards help analyze and visualize a variety of performance metrics. In addition to these well-established business reporting functions, statistical analysis and data mining techniques are adopted for association analysis, data segmentation and clustering, classification and regression analysis, anomaly detection, and predictive modeling in various business applications. Most of these data processing and analytical technologies have already been incorporated into the leading commercial BI platforms offered by major IT vendors including Microsoft, IBM, Oracle, and SAP (Sallam et al. 2011).

Among the 13 capabilities considered essential for BI platforms, according to the Gartner report by Sallam et al. (2011), the following eight are considered BI&A 1.0: reporting, dashboards, *ad hoc* query, search-based BI, OLAP, interactive visualization, scorecards, predictive modeling, and data mining. A few BI&A 1.0 areas are still under active development based on the Gartner BI Hype Cycle analysis for emerging BI technologies, which include data mining workbenchs, column-based DBMS, in-memory DBMS, and real-time decision tools (Bitterer 2011). Academic curricula in Information Systems (IS) and Computer Science (CS) often

[1]"Hal Varian Answers Your Questions," February 25, 2008 (http://www.freakonomics.com/2008/02/25/hal-varian-answers-your-questions/).

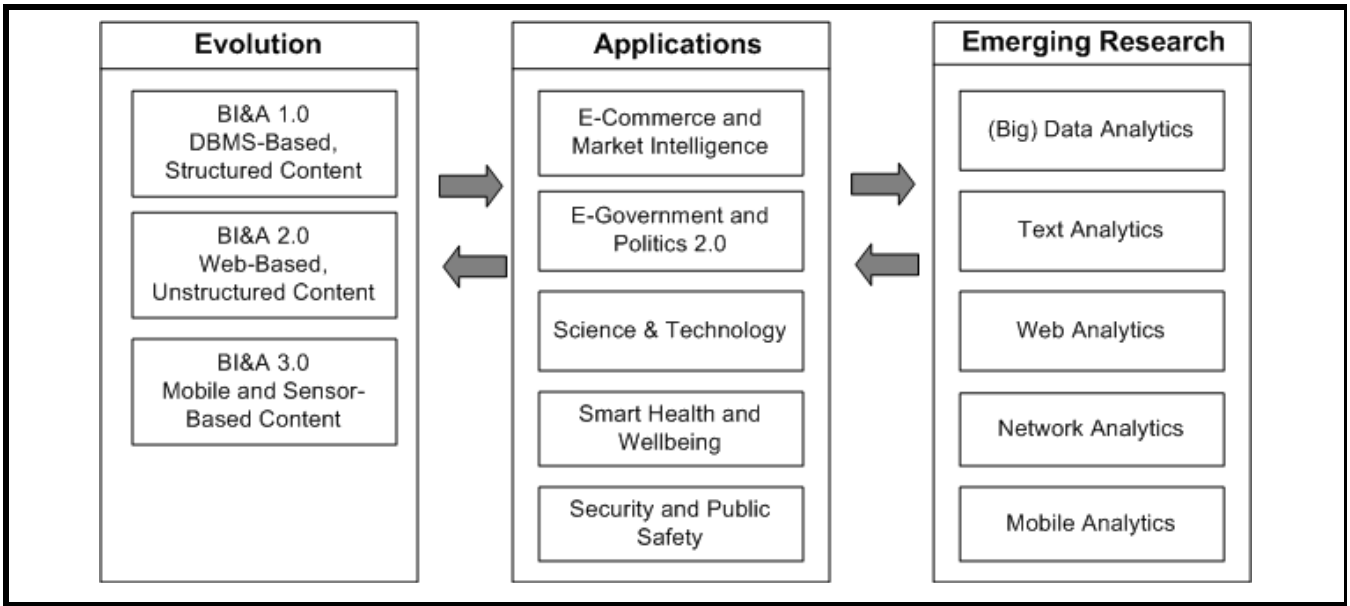

Figure 1. BI&A Overview: Evolution, Applications, and Emerging Research

include well-structured courses such as database management systems, data mining, and multivariate statistics.

BI&A 2.0

Since the early 2000s, the Internet and the Web began to offer unique data collection and analytical research and development opportunities. The HTTP-based Web 1.0 systems, characterized by Web search engines such as Google and Yahoo and e-commerce businesses such as Amazon and eBay, allow organizations to present their businesses online and interact with their customers directly. In addition to porting their traditional RDBMS-based product information and business contents online, detailed and IP-specific user search and interaction logs that are collected seamlessly through cookies and server logs have become a new gold mine for understanding customers' needs and identifying new business opportunities. Web intelligence, web analytics, and the user-generated content collected through Web 2.0-based social and crowd-sourcing systems (Doan et al. 2011; O'Reilly 2005) have ushered in a new and exciting era of BI&A 2.0 research in the 2000s, centered on text and web analytics for unstructured web contents.

An immense amount of company, industry, product, and customer information can be gathered from the web and organized and visualized through various text and web mining techniques. By analyzing customer clickstream data logs, web analytics tools such as Google Analytics can provide a trail of the user's online activities and reveal the user's browsing and purchasing patterns. Web site design, product placement optimization, customer transaction analysis, market structure analysis, and product recommendations can be accomplished through web analytics. The many Web 2.0 applications developed after 2004 have also created an abundance of user-generated content from various online social media such as forums, online groups, web blogs, social networking sites, social multimedia sites (for photos and videos), and even virtual worlds and social games (O'Reilly 2005). In addition to capturing celebrity chatter, references to everyday events, and socio-political sentiments expressed in these media, Web 2.0 applications can efficiently gather a large volume of timely feedback and opinions from a diverse customer population for different types of businesses.

Many marketing researchers believe that social media analytics presents a unique opportunity for businesses to treat the market as a "conversation" between businesses and customers instead of the traditional business-to-customer, one-way "marketing" (Lusch et al. 2010). Unlike BI&A 1.0 technologies that are already integrated into commercial enterprise IT systems, future BI&A 2.0 systems will require the integration of mature and scalable techniques in text mining (e.g., information extraction, topic identification, opinion mining, question-answering), web mining, social network analysis, and spatial-temporal analysis with existing DBMS-based BI&A 1.0 systems.

Except for basic query and search capabilities, no advanced text analytics for unstructured content are currently considered in the 13 capabilities of the Gartner BI platforms. Several, however, are listed in the Gartner BI Hype Cycle, including information semantic services, natural language question answering, and content/text analytics (Bitterer 2011). New IS and CS courses in text mining and web mining have emerged to address needed technical training.

BI&A 3.0

Whereas web-based BI&A 2.0 has attracted active research from academia and industry, a new research opportunity in BI&A 3.0 is emerging. As reported prominently in an October 2011 article in *The Economist* (2011), the number of mobile phones and tablets (about 480 million units) surpassed the number of laptops and PCs (about 380 million units) for the first time in 2011. Although the number of PCs in use surpassed 1 billion in 2008, the same article projected that the number of mobile connected devices would reach 10 billion in 2020. Mobile devices such as the iPad, iPhone, and other smart phones and their complete ecosystems of downloadable applicationss, from travel advisories to multi-player games, are transforming different facets of society, from education to healthcare and from entertainment to governments. Other sensor-based Internet-enabled devices equipped with RFID, barcodes, and radio tags (the "Internet of Things") are opening up exciting new steams of innovative applications. The ability of such mobile and Internet-enabled devices to support highly mobile, location-aware, person-centered, and context-relevant operations and transactions will continue to offer unique research challenges and opportunities throughout the 2010s. Mobile interface, visualization, and HCI (human–computer interaction) design are also promising research areas. Although the coming of the Web 3.0 (mobile and sensor-based) era seems certain, the underlying mobile analytics and location and context-aware techniques for collecting, processing, analyzing and visualizing such large-scale and fluid mobile and sensor data are still unknown.

No integrated, commercial BI&A 3.0 systems are foreseen for the near future. Most of the academic research on mobile BI is still in an embryonic stage. Although not included in the current BI platform core capabilities, mobile BI has been included in the Gartner BI Hype Cycle analysis as one of the new technologies that has the potential to disrupt the BI market significantly (Bitterer 2011). The uncertainty associated with BI&A 3.0 presents another unique research direction for the IS community.

Table 1 summarizes the key characteristics of BI&A 1.0, 2.0, and 3.0 in relation to the Gartner BI platforms core capabilities and hype cycle.

The decade of the 2010s promises to be an exciting one for high-impact BI&A research and development for both industry and academia. The business community and industry have already taken important steps to adopt BI&A for their needs. The IS community faces unique challenges and opportunities in making scientific and societal impacts that are relevant and long-lasting (Chen 2011a). IS research and education programs need to carefully evaluate future directions, curricula, and action plans, from BI&A 1.0 to 3.0.

BI&A Applications: From Big Data to Big Impact

Several global business and IT trends have helped shape past and present BI&A research directions. International travel, high-speed network connections, global supply-chain, and outsourcing have created a tremendous opportunity for IT advancement, as predicted by Thomas Freeman in his seminal book, *The World is Flat* (2005). In addition to ultra-fast global IT connections, the development and deployment of business-related data standards, electronic data interchange (EDI) formats, and business databases and information systems have greatly facilitated business data creation and utilization. The development of the Internet in the 1970s and the subsequent large-scale adoption of the World Wide Web since the 1990s have increased business data generation and collection speeds exponentially. Recently, the Big Data era has quietly descended on many communities, from governments and e-commerce to health organizations. With an overwhelming amount of web-based, mobile, and sensor-generated data arriving at a terabyte and even exabyte scale (*The Economist* 2010a, 2010b), new science, discovery, and insights can be obtained from the highly detailed, contextualized, and rich contents of relevance to any business or organization.

In addition to being data driven, BI&A is highly applied and can leverage opportunities presented by the abundant data and domain-specific analytics needed in many critical and high-impact application areas. Several of these promising and high-impact BI&A applications are presented below, with a discussion of the data and analytics characteristics, potential impacts, and selected illustrative examples or studies: (1) e-commerce and market intelligence, (2) e-government and politics 2.0, (3) science and technology, (4) smart health and

Table 1. BI&A Evolution: Key Characteristics and Capabilities

	Key Characteristics	Gartner BI Platforms Core Capabilities	Gartner Hype Cycle
BI&A 1.0	DBMS-based, structured content • RDBMS & data warehousing • ETL & OLAP • Dashboards & scorecards • Data mining & statistical analysis	• *Ad hoc* query & search-based BI • Reporting, dashboards & scorecards • OLAP • Interactive visualization • Predictive modeling & data mining	• Column-based DBMS • In-memory DBMS • Real-time decision • Data mining workbenches
BI&A 2.0	Web-based, unstructured content • Information retrieval and extraction • Opinion mining • Question answering • Web analytics and web intelligence • Social media analytics • Social network analysis • Spatial-temporal analysis		• Information semantic services • Natural language question answering • Content & text analytics
BI&A 3.0	Mobile and sensor-based content • Location-aware analysis • Person-centered analysis • Context-relevant analysis • Mobile visualization & HCI		• Mobile BI

well-being, and (5) security and public safety. By carefully analyzing the application and data characteristics, researchers and practitioners can then adopt or develop the appropriate analytical techniques to derive the intended impact. In addition to technical system implementation, significant business or domain knowledge as well as effective communication skills are needed for the successful completion of such BI&A projects. IS departments thus face unique opportunities and challenges in developing integrated BI&A research and education programs for the new generation of data/analytics-savvy and business-relevant students and professionals (Chen 2011a).

E-Commerce and Market Intelligence

The excitement surrounding BI&A and Big Data has arguably been generated primarily from the web and e-commerce communities. Significant market transformation has been accomplished by leading e-commerce vendors such Amazon and eBay through their innovative and highly scalable e-commerce platforms and product recommender systems. Major Internet firms such as Google, Amazon, and Facebook continue to lead the development of web analytics, cloud computing, and social media platforms. The emergence of customer-generated Web 2.0 content on various forums, newsgroups, social media platforms, and crowd-sourcing systems offers another opportunity for researchers and practitioners to "listen" to the voice of the market from a vast number of business constituents that includes customers, employees, investors, and the media (Doan et al. 2011; O'Rielly 2005). Unlike traditional transaction records collected from various legacy systems of the 1980s, the data that e-commerce systems collect from the web are less structured and often contain rich customer opinion and behavioral information.

For social media analytics of customer opinions, text analysis and sentiment analysis techniques are frequently adopted (Pang and Lee 2008). Various analytical techniques have also been developed for product recommender systems, such as association rule mining, database segmentation and clustering, anomaly detection, and graph mining (Adomavicius and Tuzhilin 2005). Long-tail marketing accomplished by reaching the millions of niche markets at the shallow end of the product bitstream has become possible via highly targeted searches and personalized recommendations (Anderson 2004).

The Netfix Prize competition[2] for the best collaborative filtering algorithm to predict user movie ratings helped generate significant academic and industry interest in recommender systems development and resulted in awarding the grand prize of $1 million to the Bellkor's Pragmatic Chaos team, which

[2]Netflix Prize (http://www.netflixprize.com//community/viewtopic.php?id=1537; accessed July 9, 2012).

surpassed Netflix's own algorithm for predicting ratings by 10.06 percent. However, the publicity associated with the competition also raised major unintended customer privacy concerns.

Much BI&A-related e-commerce research and development information is appearing in academic IS and CS papers as well as in popular IT magazines.

E-Government and Politics 2.0

The advent of Web 2.0 has generated much excitement for reinventing governments. The 2008 U.S. House, Senate, and presidential elections provided the first signs of success for online campaigning and political participation. Dubbed "politics 2.0," politicians use the highly participatory and multimedia web platforms for successful policy discussions, campaign advertising, voter mobilization, event announcements, and online donations. As government and political processes become more transparent, participatory, online, and multimedia-rich, there is a great opportunity for adopting BI&A research in e-government and politics 2.0 applications. Selected opinion mining, social network analysis, and social media analytics techniques can be used to support online political participation, e-democracy, political blogs and forums analysis, e-government service delivery, and process transparency and accountability (Chen 2009; Chen et al. 2007). For e-government applications, semantic information directory and ontological development (as exemplified below) can also be developed to better serve their target citizens.

Despite the significant transformational potential for BI&A in e-government research, there has been less academic research than, for example, e-commerce-related BI&A research. E-government research often involves researchers from political science and public policy. For example, Karpf (2009) analyzed the growth of the political blogosphere in the United States and found significant innovation of existing political institutions in adopting blogging platforms into their Web offerings. In his research, 2D blogspace mapping with composite rankings helped reveal the partisan makeup of the American political blogsphere. Yang and Callan (2009) demonstrated the value for ontology development for government services through their development of the OntoCop system, which works interactively with a user to organize and summarize online public comments from citizens.

Science and Technology

Many areas of science and technology (S&T) are reaping the benefits of high-throughput sensors and instruments, from astrophysics and oceanography, to genomics and environmental research. To facilitate information sharing and data analytics, the National Science Foundation (NSF) recently mandated that every project is required to provide a data management plan. Cyber-infrastructure, in particular, has become critical for supporting such data-sharing initiatives.

The 2012 NSF BIGDATA[3] program solicitation is an obvious example of the U.S. government funding agency's concerted efforts to promote big data analytics. The program

> aims to advance the core scientific and technological means of managing, analyzing, visualizing, and extracting useful information from large, diverse, distributed and heterogeneous data sets so as to accelerate the progress of scientific discovery and innovation; lead to new fields of inquiry that would not otherwise be possible; encourage the development of new data analytic tools and algorithms; facilitate scalable, accessible, and sustainable data infrastructure; increase understanding of human and social processes and interactions; and promote economic growth and improved health and quality of life.

Several S&T disciplines have already begun their journey toward big data analytics. For example, in biology, the NSF funded iPlant Collaborative[4] is using cyberinfrastructure to support a community of researchers, educators, and students working in plant sciences. iPlant is intended to foster a new generation of biologists equipped to harness rapidly expanding computational techniques and growing data sets to address the grand challenges of plant biology. The iPlant data set is diverse and includes canonical or reference data, experimental data, simulation and model data, observational data, and other derived data. It also offers various open source data processing and analytics tools.

In astronomy, the Sloan Digital Sky Survey (SDSS)[5] shows how computational methods and big data can support and facilitate sense making and decision making at both the macroscopic and the microscopic level in a rapidly growing and globalized research field. The SDSS is one of the most ambitious and influential surveys in the history of astronomy.

---

[3]"Core Techniques and Technologies for Advancing Big Data Science & Engineering (BIGDATA)," Program Solicitation NSF 12-499 (http://www.nsf.gov/pubs/2012/nsf12499/nsf12499.htm; accessed August 2, 2012).

[4]iPlant Collaborative (http://www.iplantcollaborative.org/about; accessed August 2, 2012).

[5]"Sloan Digital Sky Survey: Mapping the Universe" (http://www.sdss.org/; accessed August 2, 2012).

Over its eight years of operation, it has obtained deep, multi-color images covering more than a quarter of the sky and created three-dimensional maps containing more than 930,000 galaxies and over 120,000 quasars. Continuing to gather data at a rate of 200 gigabytes per night, SDSS has amassed more than 140 terabytes of data. The international Large Hadron Collider (LHC) effort for high-energy physics is another example of big data, producing about 13 petabytes of data in a year (Brumfiel 2011).

Smart Health and Wellbeing

Much like the big data opportunities facing the e-commerce and S&T communities, the health community is facing a tsunami of health- and healthcare-related content generated from numerous patient care points of contact, sophisticated medical instruments, and web-based health communities. Two main sources of health big data are genomics-driven big data (genotyping, gene expression, sequencing data) and payer–provider big data (electronic health records, insurance records, pharmacy prescription, patient feedback and responses) (Miller 2012a). The expected raw sequencing data from each person is approximately four terabytes. From the payer–provider side, a data matrix might have hundreds of thousands of patients with many records and parameters (demographics, medications, outcomes) collected over a long period of time. Extracting knowledge from health big data poses significant research and practical challenges, especially considering the HIPAA (Health Insurance Portability and Accountability Act) and IRB (Institutional Review Board) requirements for building a privacy-preserving and trustworthy health infrastructure and conducting ethical health-related research (Gelfand 2011/2012). Health big data analytics, in general, lags behind e-commerce BI&A applications because it has rarely taken advantage of scalable analytical methods or computational platforms (Miller 2012a).

Over the past decade, electronic health records (EHR) have been widely adopted in hospitals and clinics worldwide. Significant clinical knowledge and a deeper understanding of patient disease patterns can be gleanded from such collections (Hanauer et al. 2009; Hanauer et al. 2011; Lin et al. 2011). Hanauer et al. (2011), for example, used large-scale, longitudinal EHR to research associations in medical diagnoses and consider temporal relations between events to better elucidate patterns of disease progression. Lin et al. (2011) used symptom–disease–treatment (SDT) association rule mining on a comprehensive EHR of approximately 2.1 million records from a major hospital. Based on selected International Classification of Diseases (ICD-9) codes, they were able to identify clinically relevant and accurate SDT associations from patient records in seven distinct diseases, ranging from cancers to chronic and infectious diseases.

In addition to EHR, health social media sites such as Daily Strength and PatientsLikeMe provide unique research opportunities in healthcare decision support and patient empowerment (Miller 2012b), especially for chronic diseases such as diabetes, Parkinson's, Alzheimer's, and cancer. Association rule mining and clustering, health social media monitoring and analysis, health text analytics, health ontologies, patient network analysis, and adverse drug side-effect analysis are promising areas of research in health-related BI&A. Due to the importance of HIPAA regulations, privacy-preserving health data mining is also gaining attention (Gelfand 2011/2012).

Partially funded by the National Institutes of Health (NIH), the NSF BIGDATA program solicitation includes common interests in big data across NSF and NIH. Clinical decision making, patient-centered therapy, and knowledge bases for health, disease, genome, and environment are some of the areas in which BI&A techniques can contribute (Chen 2011b; Wactlar et al. 2011). Another recent, major NSF initiative related to health big data analytics is the NSF Smart Health and Wellbeing (SHB)[6] program, which seeks to address fundamental technical and scientific issues that would support a much-needed transformation of healthcare from reactive and hospital-centered to preventive, proactive, evidence-based, person-centered, and focused on wellbeing rather than disease control. The SHB research topics include sensor technology, networking, information and machine learning technology, modeling cognitive processes, system and process modeling, and social and economic issues (Wactlar et al. 2011), most of which are relevant to healthcare BI&A.

Security and Public Safety

Since the tragic events of September 11, 2001, security research has gained much attention, especially given the increasing dependency of business and our global society on digital enablement. Researchers in computational science, information systems, social sciences, engineering, medicine, and many other fields have been called upon to help enhance our ability to fight violence, terrorism, cyber crimes, and other cyber security concerns. Critical mission areas have been identified where information technology can contribute, as suggested in the U.S. Office of Homeland Security's report "National Strategy for Homeland Security," released in 2002, including intelligence and warning, border and transportation

[6]"Smart Health and Wellbeing (SBH)," Program Solicitation NSF 12-512 (http://www.nsf.gov/pubs/2012/nsf12512/nsf12512.htm; accessed August 2, 2012).

security, domestic counter-terrorism, protecting critical infrastructure (including cyberspace), defending against catastrophic terrorism, and emergency preparedness and response. Facing the critical missions of international security and various data and technical challenges, the need to develop the science of "security informatics" was recognized, with its main objective being the

> development of advanced information technologies, systems, algorithms, and databases for security-related applications, through an integrated technological, organizational, and policy-based approach (Chen 2006, p. 7).

BI&A has much to contribute to the emerging field of security informatics.

Security issues are a major concern for most organizations. According to the research firm International Data Corporation, large companies are expected to spend $32.8 billion in computer security in 2012, and small- and medium-size companies will spend more on security than on other IT purchases over the next three years (Perlroth and Rusli 2012). In academia, several security-related disciplines such as computer security, computational criminology, and terrorism informatics are also flourishing (Brantingham 2011; Chen et al. 2008).

Intelligence, security, and public safety agencies are gathering large amounts of data from multiple sources, from criminal records of terrorism incidents, and from cyber security threats to multilingual open-source intelligence. Companies of different sizes are facing the daunting task of defending against cybersecurity threats and protecting their intellectual assets and infrastructure. Processing and analyzing security-related data, however, is increasingly difficult. A significant challenge in security IT research is the information stovepipe and overload resulting from diverse data sources, multiple data formats, and large data volumes. Current research on technologies for cybersecurity, counter-terrorism, and crime-fighting applications lacks a consistent framework for addressing these data challenges. Selected BI&A technologies such as criminal association rule mining and clustering, criminal network analysis, spatial-temporal analysis and visualization, multilingual text analytics, sentiment and affect analysis, and cyber attacks analysis and attribution should be considered for security informatics research.

The University of Arizona's COPLINK and Dark Web research programs offer significant examples of crime data mining and terrorism informatics within the IS community (Chen 2006, 2012). The COPLINK information sharing and crime data mining system, initially developed with funding from NSF and the Department of Justice, is currently in use by more than 4,500 police agencies in the United States and by 25 NATO countries, and was acquired by IBM in 2011. The Dark Web research, funded by NSF and the Department of Defense (DOD), has generated one of the largest known academic terrorism research databases (about 20 terabytes of terrorist web sites and social media content) and generated advanced multilingual social media analytics techniques.

Recognizing the challenges presented by the volume and complexity of defense-related big data, the U.S. Defense Advanced Research Project Agency (DARPA) within DOD initiated the XDATA program in 2012 to help develop computational techniques and software tools for processing and analyzing the vast amount of mission-oriented information for defense activities. XDATA aims to address the need for scalable algorithms for processing and visualization of imperfect and incomplete data. The program engages applied mathematics, computer science, and data visualization communities to develop big data analytics and usability solutions for warfighters.[7] BI&A researchers could contribute significantly in this area.

Table 2 summarizes these promising BI&A applications, data characteristics, analytics techniques, and potential impacts.

BI&A Research Framework: Foundational Technologies and Emerging Research in Analytics

The opportunities with the abovementioned emerging and high-impact applications have generated a great deal of excitement within both the BI&A industry and the research community. Whereas industry focuses on scalable and integrated systems and implementations for applications in different organizations, the academic community needs to continue to advance the key technologies in analytics.

Emerging analytics research opportunities can be classified into five critical technical areas—(big) data analytics, text analytics, web analytics, network analytics, and mobile analytics—all of which can contribute to to BI&A 1.0, 2.0, and 3.0. The classification of these five topic areas is intended

[7]"DARPA Calls for Advances in 'Big Data" to Help the Warfighter," March 29, 2012 (http://www.darpa.mil/NewsEvents/Releases/2012/03/29.aspx; accessed August 5, 2012).

Table 2. BI&A Applications: From Big Data to Big Impact

	E-Commerce and Market Intelligence	E-Government and Politics 2.0	Science & Technology	Smart Health and Wellbeing	Security and Public Safety
Applications	• Recommender systems • Social media monitoring and analysis • Crowd-sourcing systems • Social and virtual games	• Ubiquitous government services • Equal access and public services • Citizen engagement and participation • Political campaign and e-polling	• S&T innovation • Hypothesis testing • Knowledge discovery	• Human and plant genomics • Healthcare decision support • Patient community analysis	• Crime analysis • Computational criminology • Terrorism informatics • Open-source intelligence • Cyber security
Data	• Search and user logs • Customer transaction records • Customer-generated content	• Government information and services • Rules and regulations • Citizen feedback and comments	• S&T instruments and system-generated data • Sensor and network content	• Genomics and sequence data • Electronic health records (EHR) • Health and patient social media	• Criminal records • Crime maps • Criminal networks • News and web contents • Terrorism incident databases • Viruses, cyber attacks, and botnets
	Characteristics: Structured web-based, user-generated content, rich network information, unstructured informal customer opinions	Characteristics: Fragmented information sources and legacy systems, rich textual content, unstructured informal citizen conversations	Characteristics: High-throughput instrument-based data collection, fine-grained multiple-modality and large-scale records, S&T specific data formats	Characteristics: Disparate but highly linked content, person-specific content, HIPAA, IRB and ethics issues	Characteristics: Personal identity information, incomplete and deceptive content, rich group and network information, multilingual content
Analytics	• Association rule mining • Database segmentation and clustering • Anomaly detection • Graph mining • Social network analysis • Text and web analytics • Sentiment and affect analysis	• Information integration • Content and text analytics • Government information semantic services and ontologies • Social media monitoring and analysis • Social network analysis • Sentiment and affect analysis	• S&T based domain-specific mathematical and analytical models	• Genomics and sequence analysis and visualization • EHR association mining and clustering • Health social media monitoring and analysis • Health text analytics • Health ontologies • Patient network analysis • Adverse drug side-effect analysis • Privacy-preserving data mining	• Criminal association rule mining and clustering • Criminal network analysis • Spatial-temporal analysis and visualization • Multilingual text analytics • Sentiment and affect analysis • Cyber attacks analysis and attribution
Impacts	Long-tail marketing, targeted and personalized recommendation, increased sale and customer satisfaction	Transforming governments, empowering citizens, improving transparency, participation, and equality	S&T advances, scientific impact	Improved healthcare quality, improved long-term care, patient empowerment	Improved public safety and security

Table 3. BI&A Research Framework: Foundational Technologies and Emerging Research in Analytics

	(Big) Data Analytics	Text Analytics	Web Analytics	Network Analytics	Mobile Analytics
Foundational Technologies	• RDBMS • data warehousing • ETL • OLAP • BPM • data mining • clustering • regression • classification • association analysis • anomaly detection • neural networks • genetic algorithms • multivariate statistical analysis • optimization • heuristic search	• information retrieval • document representation • query processing • relevance feedback • user models • search engines • enterprise search systems	• information retrieval • computational linguistics • search engines • web crawling • web site ranking • search log analysis • recommender systems • web services • mashups	• bibliometric analysis • citation network • coauthorship network • social network theories • network metrics and topology • mathematical network models • network visualization	• web services • smartphone platforms
Emerging Research	• statistical machine learning • sequential and temporal mining • spatial mining • mining high-speed data streams and sensor data • process mining • privacy-preserving data mining • network mining • web mining • column-based DBMS • in-memory DBMS • parallel DBMS • cloud computing • Hadoop • MapReduce	• statistical NLP • information extraction • topic models • question-answering systems • opinion mining • sentiment/affect analysis • web stylometric analysis • multilingual analysis • text visualization • multimedia IR • mobile IR • Hadoop • MapReduce	• cloud services • cloud computing • social search and mining • reputation systems • social media analytics • web visualization • web-based auctions • internet monetization • social marketing • web privacy/ security	• link mining • community detection • dynamic network modeling • agent-based modeling • social influence and information diffusion models • ERGMs • virtual communities • criminal/dark networks • social/political analysis • trust and reputation	• mobile web services • mobile pervasive apps • mobile sensing apps • mobile social innovation • mobile social networking • mobile visualization/HCI • personalization and behavioral modeling • gamification • mobile advertising and marketing

to highlight the key characteristics of each area; however, a few of these areas may leverage similar underlying technologies. In each analytics area we present the foundational technologies that are mature and well developed and suggest selected emerging research areas (see Table 3).

(Big) Data Analytics

Data analytics refers to the BI&A technologies that are grounded mostly in data mining and statistical analysis. As mentioned previously, most of these techniques rely on the mature commercial technologies of relational DBMS, data warehousing, ETL, OLAP, and BPM (Chaudhuri et al. 2011). Since the late 1980s, various data mining algorithms have been developed by researchers from the artificial intelligence, algorithm, and database communities. In the IEEE 2006 International Conference on Data Mining (ICDM), the 10 most influential data mining algorithms were identified based on expert nominations, citation counts, and a community survey. In ranked order, they are C4.5, k-means, SVM (support vector machine), Apriori, EM (expectation maximization), PageRank, AdaBoost, kNN (k-nearest neighbors), Naïve Bayes, and CART (Wu et al. 2007). These algorithms cover classification, clustering, regression, association analysis, and network analysis. Most of these popular data mining algorithms have been incorporated in commercial and open source data mining systems (Witten et al. 2011). Other

advances such as neural networks for classification/prediction and clustering and genetic algorithms for optimization and machine learning have all contributed to the success of data mining in different applications.

Two other data analytics approaches commonly taught in business school are also critical for BI&A. Grounded in statistical theories and models, multivariate statistical analysis covers analytical techniques such as regression, factor analysis, clustering, and discriminant analysis that have been used successfully in various business applications. Developed in the management science community, optimization techniques and heuristic search are also suitable for selected BI&A problems such as database feature selection and web crawling/ spidering. Most of these techniques can be found in business school curricula.

Due to the success achieved collectively by the data mining and statistical analysis community, data analytics continues to be an active area of research. Statistical machine learning, often based on well-grounded mathematical models and powerful algorithms, techniques such as Bayesian networks, Hidden Markov models, support vector machine, reinforcement learning, and ensemble models, have been applied to data, text, and web analytics applications. Other new data analytics techniques explore and leverage unique data characteristics, from sequential/temporal mining and spatial mining, to data mining for high-speed data streams and sensor data. Increased privacy concerns in various e-commerce, e-government, and healthcare applications have caused privacy-preserving data mining to become an emerging area of research. Many of these methods are data-driven, relying on various anonymization techniques, while others are process-driven, defining how data can be accessed and used (Gelfand 2011/ 2012). Over the past decade, process mining has also emerged as a new research field that focuses on the analysis of processes using event data. Process mining has become possible due to the availability of event logs in various industries (e.g., healthcare, supply chains) and new process discovery and conformance checking techniques (van der Aalst 2012). Furthermore, network data and web content have helped generate exciting research in network analytics and web analytics, which are presented below.

In addition to active academic research on data analytics, industry research and development has also generated much excitement, especially with respect to big data analytics for semi-structured content. Unlike the structured data that can be handled repeatedly through a RDBMS, semi-structured data may call for *ad hoc* and one-time extraction, parsing, processing, indexing, and analytics in a scalable and distributed MapReduce or Hadoop environment. MapReduce has been hailed as a revolutionary new platform for large-scale, massively parallel data access (Patterson 2008). Inspired in part by MapReduce, Hadoop provides a Java-based software framework for distributed processing of data-intensive transformation and analytics. The top three commercial database suppliers—Oracle, IBM, and Microsoft—have all adopted Hadoop, some within a cloud infrastructure. The open source Apache Hadoop has also gained significant traction for business analytics, including Chukwa for data collection, HBase for distributed data storage, Hive for data summarization and *ad hoc* querying, and Mahout for data mining (Henschen 2011). In their perspective paper, Stonebraker et al. (2010) compared MapReduce with the parallel DBMS. The commercial parallel DBMS showed clear advantages in efficient query processing and high-level query language and interface, whereas MapReduce excelled in ETL and analytics for "read only" semi-structured data sets. New Hadoop- and MapReduce-based systems have become another viable option for big data analytics in addition to the commercial systems developed for RDBMS, column-based DBMS, in-memory DBMS, and parallel DBMS (Chaudhuri et al. 2011).

Text Analytics

A significant portion of the unstructured content collected by an organization is in textual format, from e-mail communication and corporate documents to web pages and social media content. Text analytics has its academic roots in information retrieval and computational linguistics. In information retrieval, document representation and query processing are the foundations for developing the vector-space model, Boolean retrieval model, and probabilistic retrieval model, which in turn, became the basis for the modern digital libraries, search engines, and enterprise search systems (Salton 1989). In computational linguistics, statistical natural language processing (NLP) techniques for lexical acquisition, word sense disambiguation, part-of-speech-tagging (POST), and probabilistic context-free grammars have also become important for representing text (Manning and Schütze 1999). In addition to document and query representations, user models and relevance feedback are also important in enhancing search performance.

Since the early 1990s, search engines have evolved into mature commercial systems, consisting of fast, distributed crawling; efficient inverted indexing; inlink-based page ranking; and search logs analytics. Many of these foundational text processing and indexing techniques have been deployed in text-based enterprise search and document management systems in BI&A 1.0.

Leveraging the power of big data (for training) and statistical NLP (for building language models), text analytics techniques have been actively pursued in several emerging areas, including information extraction, topic models, question-answering (Q/A), and opinion mining. Information extraction is an area of research that aims to automatically extract specific kinds of structured information from documents. As a building block of information extraction, NER (named entity recognition, also known as *entity extraction*) is a process that identifies atomic elements in text and classifies them into predefined categories (e.g., names, places, dates). NER techniques have been successfully developed for news analysis and biomedical applications. Topic models are algorithms for discovering the main themes that pervade a large and otherwise unstructured collection of documents. New topic modeling algorithms such as LDA (latent Dirichlet allocation) and other probabilistic models have attracted recent research (Blei 2012). Question answering (Q/A) systems rely on techniques from NLP, information retrieval, and human–computer interaction. Primarily designed to answer factual questions (i.e., who, what, when, and where kinds of questions), Q/A systems involve different techniques for question analysis, source retrieval, answer extraction, and answer presentation (Maybury 2004). The recent successes of IBM's Watson and Apple's Siri have highlighted Q/A research and commercialization opportunities. Many promising Q/A system application areas have been identified, including education, health, and defense. Opinion mining refers to the computational techniques for extracting, classifying, understanding, and assessing the opinions expressed in various online news sources, social media comments, and other user-generated contents. Sentiment analysis is often used in opinion mining to identify sentiment, affect, subjectivity, and other emotional states in online text. Web 2.0 and social media content have created abundant and exciting opportunities for understanding the opinions of the general public and consumers regarding social events, political movements, company strategies, marketing campaigns, and product preferences (Pang and Lee 2008).

In addition to the above research directions, text analytics also offers significant research opportunities and challenges in several more focused areas, including web stylometric analysis for authorship attribution, multilingual analysis for web documents, and large-scale text visualization. Multimedia information retrieval and mobile information retrieval are two other related areas that require support of text analytics techniques, in addition to the core multimedia and mobile technologies. Similar to big data analytics, text analytics using MapReduce, Hadoop, and cloud services will continue to foster active research directions in both academia and industry.

Web Analytics

Over the past decade, web analytics has emerged as an active field of research within BI&A. Building on the data mining and statistical analysis foundations of data analytics and on the information retrieval and NLP models in text analytics, web analytics offers unique analytical challenges and opportunities. HTTP/HTML-based hyperlinked web sites and associated web search engines and directory systems for locating web content have helped develop unique Internet-based technologies for web site crawling/spidering, web page updating, web site ranking, and search log analysis. Web log analysis based on customer transactions has subsequently turned into active research in recommender systems. However, web analytics has become even more exciting with the maturity and popularity of web services and Web 2.0 systems in the mid-2000s (O'Reilly 2005).

Based on XML and Internet protocols (HTTP, SMTP), web services offer a new way of reusing and integrating third party or legacy systems. New types of web services and their associated APIs (application programming interface) allow developers to easily integrate diverse content from different web-enabled system, for example, REST (representational state transfer) for invoking remote services, RSS (really simple syndication) for news "pushing," JSON (JavaScript object notation) for lightweight data-interchange, and AJAX (asynchronous JavaScript + XML) for data interchange and dynamic display. Such lightweight programming models support data syndication and notification and "mashups" of multimedia content (e.g., Flickr, Youtube, Google Maps) from different web sources—a process somewhat similar to ETL (extraction, transformation, and load) in BI&A 1.0. Most of the e-commerce vendors have provided mature APIs for accessing their product and customer content (Schonfeld 2005). For example, through Amazon Web Services, developers can access product catalog, customer reviews, site ranking, historical pricing, and the Amazon Elastic Compute Cloud (EC2) for computing capacity. Similarly, Google web APIs support AJAX search, Map API, GData API (for Calendar, Gmail, etc.), Google Translate, and Google App Engine for cloud computing resources. Web services and APIs continue to provide an exciting stream of new data sources for BI&A 2.0 research.

A major emerging component in web analytics research is the development of cloud computing platforms and services, which include applications, system software, and hardware delivered as services over the Internet. Based on service-oriented architecture (SOA), server virtualization, and utility computing, cloud computing can be offered as software as a

service (SaaS), infrastructure as a service (IaaS), or platform as a service (PaaS). Only a few leading IT vendors are currently positioned to support high-end, high-throughput BI&A applications using cloud computing. For example, Amazon Elastic Compute Cloud (EC2) enables users to rent virtual computers on which to run their own computer applications. Its Simple Storage Service (S3) provides online storage web service. Google App Engine provides a platform for developing and hosting Java or Python-based web applications. Google Bigtable is used for backend data storage. Microsoft's Windows Azure platform provides cloud services such as SQL Azure and SharePoint, and allows .Net framework applications to run on the platform. The industry-led web and cloud services offer unique data collection, processing, and analytics challenges for BI&A researchers.

In academia, current web analytics related research encompasses social search and mining, reputation systems, social media analytics, and web visualization. In addition, web-based auctions, Internet monetization, social marketing, and web privacy/security are some of the promising research directions related to web analytics. Many of these emerging research areas may rely on advances in social network analysis, text analytics, and even economics modeling research.

Network Analytics

Network analytics is a nascent research area that has evolved from the earlier citation-based bibliometric analysis to include new computational models for online community and social network analysis. Grounded in bibliometric analysis, citation networks and coauthorship networks have long been adopted to examine scientific impact and knowledge diffusion. The h-index is a good example of a citation metric that aims to measure the productivity and impact of the published work of a scientist or scholar (Hirsch 2005). Since the early 2000s, network science has begun to advance rapidly with contributions from sociologists, mathematicians, and computer scientists. Various social network theories, network metrics, topology, and mathematical models have been developed that help understand network properties and relationships (e.g., centrality, betweenness, cliques, paths; ties, structural holes, structural balance; random network, small-world network, scale-free network) (Barabási 2003; Watts 2003).

Recent network analytics research has focused on areas such as link mining and community detection. In link mining, one seeks to discover or predict links between nodes of a network. Within a network, nodes may represent customers, end users, products and/or services, and the links between nodes may represent social relationships, collaboration, e-mail exchanges, or product adoptions. One can conduct link mining using only the topology information (Liben-Nowell and Kleinberg 2007). Techniques such as common neighbors, Jaccard's coefficient, Adamic Adar measure, and Katz measure are popular for predicting missing or future links. The link mining accuracy can be further improved when the node and link attributes are considered. Community detection is also an active research area of relevance to BI&A (Fortunato 2010). By representing networks as graphs, one can apply graph partitioning algorithms to find a minimal cut to obtain dense subgraphs representing user communities.

Many techniques have been developed to help study the dynamic nature of social networks. For example, agent-based models (sometimes referred to as multi-agent systems) have been used to study disease contact networks and criminal or terrorist networks (National Research Council 2008). Such models simulate the actions and interactions of autonomous agents (of either individual or collective entities such as organizations or groups) with the intent of assessing their effects on the system as a whole. Social influence and information diffusion models are also viable techniques for studying evolving networks. Some research is particularly relevant to opinion and information dynamics of a society. Such dynamics hold many qualitative similarities to disease infections (Bettencourt et al. 2006). Another network analytics technique that has drawn attention in recent years is the use of exponential random graph models (Frank and Strauss 1986; Robins et al. 2007). ERGMs are a family of statistical models for analyzing data about social and other networks. To support statistical inference on the processes influencing the formation of network structure, ERGMs consider the set of all possible alternative networks weighted on their similarity to an observed network. In addition to studying traditional friendship or disease networks, ERGMs are promising for understanding the underlying network propertities that cause the formation and evolution of customer, citizen, or patient networks for BI&A.

Most of the abovementioned network analytics techniques are not part of the existing commercial BI&A platforms. Significant open-source development efforts are underway from the social network analysis community. Tools such as UCINet (Borgatti et al. 2002) and Pajek (Batagelj and Mrvar 1998) have been developed and are widely used for large-scale network analysis and visualization. New network analytics tools such as ERGM have also been made available to the academic community (Hunter et al. 2008). Online virtual communities, criminal and terrorist networks, social and political networks, and trust and reputation networks are some of the promising new applications for network analytics.

Mobile Analytics

As an effective channel for reaching many users and as a means of increasing the productivity and efficiency of an organization's workforce, mobile computing is viewed by respondents of the recent IBM technology trends survey (IBM 2011) as the second most "in demand" area for software development. Mobile BI was also considered by the Gartner BI Hype Cycle analysis as one of the new technologies that have the potential to drastically disrupt the BI market (Bitterer 2011). According to eMarketer, the market for mobile ads is expected to explode, soaring from an estimated $2.6 billion in 2012 to $10.8 billion in 2016 (Snider 2012).

Mobile computing offers a means for IT professional growth as more and more organizations build applications. With its large and growing global install base, Android has been ranked as the top mobile platform since 2010. This open source platform, based on Java and XML, offers a much shorter learning curve and this contributes to its popularity with IT professionals: 70 percent of the IBM survey respondents planned to use Android as their mobile development platform, while 49 percent planned to use iOS and 35 percent planned to use Windows 7. The ability to collect fine-grained, location-specific, context-aware, highly personalized content through these smart devices has opened new possibilities for advanced and innovative BI&A opportunities. In addition to the hardware and content advantages, the unique apps ecosystem developed through the volunteer community of mobile app developers offers a new avenue for BI&A research. The Apple App Store alone offers more than 500,000 apps in almost any conceivable category as of August 2012;[8] the number of Android apps also reached 500,000 in August 2012.[9] Many different revenue models have begun to emerge for mobile apps, from paid or free but ad-supported apps to mobile gamification, which incentivizes participants (e.g., users or employees) by giving rewards for contributions (Snider 2012). For mobile BI, companies are considering enterprise apps, industry-specific apps, e-commerce apps, and social apps (in ranked order) according to the IBM survey.

The lightweight programming models of the current web services (e.g., HTML, XML, CSS, Ajax, Flash, J2E) and the maturing mobile development platforms such as Android and iOS have contributed to the rapid development of mobile web services (e.g., HTML5, Mobile Ajax, Mobile Flash, J2ME) in various mobile pervasive applications, from disaster management to healthcare support. New mobile analytics research is emerging in different areas (e.g., mobile sensing apps that are location-aware and activity-sensitive; mobile social innovation for m-health and m-learning; mobile social networking and crowd-sourcing; mobile visualization/HCI; and personalization and behavioral modeling for mobile apps). In addition, social, behavioral, and economic models for gamification, mobile advertising, and social marketing are under way and may contribute to the development of future BI&A 3.0 systems.

Mapping the BI&A Knowledge Landscape: A Bibliometric Study of Academic and Industry Publications

In an effort to better understand the current state of BI&A related research and identify future sources of knowledge, we conducted a bibliometric study analyzing relevant literature, major BI&A scholars, disciplines and publications, and key research topics. A collection, transformation, and analytics process was followed in the study, much like a typical BI&A process adopted in other applications.

To discern research trends in BI&A, related literature from the past decade (2000–2011) was collected. Relevant IT publications were identified from several large-scale and reputable digital libraries: Web of Science (Thomson Reuters, covering more than 12,000 of the highest impact journals in sciences, engineering, and humanities), Business Source Complete (EBSCO, covering peer-reviewed business journals as well as non-journal content such as industry/trade magazines), IEEE Xplore (Institute of Electrical and Electronics Engineers, providing access to the IEEE digital library), ScienceDirect (Elsevier, covering over 2,500 journals from the scientific, technical, and medical literature), and Engineering Village (Elsevier, used to retrieve selected ACM conference papers because the ACM Digital Library interface does not support automated downloading). These sources contain high-quality bibliometric metadata, including journal name and date, author name and institution, and article title and abstract.

To ensure data consistency and relevance across our collection, we retrieved only those publications that contained the keywords *business intelligence*, *business analytics*, or *big data* within their title, abstract, or subject indexing (when applicable). The choice of these three keywords was intended to focus our search and analysis on publications of direct relevance to our interest. However, this search procedure may

[8]Apple – iPhone 5 – Learn about apps from the App store (http://www.apple.com/iphone/built-in-apps/app-store.html; accessed August 8, 2012).

[9]AppBrain, Android Statistics (http://www.appbrain.com/stats/number-of-android-apps; accessed August 8, 2012).

Keyword	All Years	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
Business Intelligence	**3,146**	113	104	146	159	229	330	346	394	352	201	334	338
Business Analytics	**213**	0	5	43	4	5	2	9	6	19	16	17	126
Big Data	**243**	0	1	0	0	7	4	3	26	11	41	44	95
Total	**3,602**	**113**	**110**	**149**	**163**	**241**	**336**	**358**	**426**	**382**	**358**	**356**	**560**

Figure 2. BI&A Related Publication Trend from 2000 to 2011

also omit articles that use other BI&A relevant terms (e.g., data warehousing, data mining) but not the three specific keywords in the title or abstract. This kind of limitation is common in bibliometric studies. The collected data was exported as XML records and parsed into a relational database (SQL Server) for analysis. The number of records initially retrieved totaled 6,187 papers. After removing duplicates, the number of unique records totaled 3,602.

Figure 2 shows the statistics and growth trends of publications relating to the three search keywords. Overall, business intelligence had the largest coverage and the longest history. This is consistent with the evolution of BI&A, as the term BI appeared first in the early 1990s. In our collection, business analytics and big data began to appear in the literature in 2001, but only gained much attention after about 2007. The business intelligence related publications numbered 3,146, whereas business analytics and big data publications each numbered only 213 and 243, respectively. While the overall publication trend for business intelligence remains stable, business analytics and big data publications have seen a faster growth pattern in recent years.

Knowledge of the most popular publications, as well as prolific authors, is beneficial for understanding an emerging research discipline. Table 4 summarizes the top 20 journals, conferences, and industry magazines with BI&A publications. (The top 20 academic BI&A authors are identified in Table 6.)

Overall, the largest source of academic business intelligence publications was academic conferences. The Conference on Business Intelligence and Financial Engineering (#1) and Conference on Electronic Commerce and Business Intelligence (#3) are specialized academic conferences devoted to business intelligence. One IS conference ranks #2 in the top-20 list: Hawaii International Conference on Systems Sciences (HICSS), with 370 publications.[10] IEEE holds the majority of conferences on the list through various outlets; several are related to emerging technical areas, such as data mining, Internet computing, and cloud computing. The IEEE International Conference on Data Mining (ICDM) is highly regarded and ranks #5. ACM has two publications in the top-20 list: *Communications of the ACM* and the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Both are well-known in CS. Again, the data mining community has contributed significantly to BI&A. Other technical conferences in CS are also contributing to BI&A in areas such as computational intelligence, web intelligence, evolutionary computation, and natural language processing, all of which are critical for developing future data, text, and web analytics techniques discussed in our research frame-

[10]Two major IS conferences, ICIS (International Conference on Information Systems) and WITS (Workshop on Information Technologies and Systems) may have also published significant BI&A research; however, their collections are not covered in the five major digital libraries to which we have access and thus are not included in this analysis.

Table 4. Top Journals, Conferences, and Industry Magazines with BI&A Publications

Top 20	Academic Publication	Publications	Top 20	Industry Publication	Publications
1	Conf. on Business Intelligence and Financial Engineering	531	1	ComputerWorld	282
2	Hawaii International Conf. on Systems Sciences	370	2	Information Today	258
3	Conf. on Electronic Commerce and Business Intelligence	252	3	InformationWeek	229
4	International Conf. on Web Intelligence and Intelligent Agent Technology Workshops	151	4	Computer Weekly	199
5	IEEE International Conf. on Data Mining	150	5	Microsoft Data Mining	108
6	IEEE International Conf. on e-Technology, e-Commerce, and e-Service	129	6	InfoWorld	86
7	IEEE Intelligent Systems	47	7	CIO	71
8	IEEE Cloud Computing	44	8	KM World	61
9	Decision Support Systems	39	9	CRN (formerly VARBusiness)	59
10	IEEE Congress on Evolutionary Computation	39	10	Stores Magazine	56
11	Journal of Business Ethics	34	11	Forbes	45
12	Communications of the ACM	33	12	CRM Magazine	40
13	European Journal of Marketing	32	13	Network World	39
14	IEEE/ACM International Symposium on Cluster, Cloud, and Grid Computing	31	14	Financial Executive	37
15	International Journal of Technology Management	29	15	Healthcare Financial Management	33
16	ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining	28	16	Chain Store Age	40
17	International Symposium on Natural Language Processing	22	17	Strategic Finance	29
18	IEEE Internet Computing	21	18	Traffic World	28
19	International Conf. on Computational Intelligence and Software Engineering	21	19	Data Strategy	27
20	IEEE Software	20	20	CFO	25

work. Journals are somewhat more limited in their publication volume, although it is notable that the IS journal *Decision Support Systems* made the top 20 list (at #9). A few business school journals also contain related BI&A research in areas such as business ethics, marketing, and technology management. Other major IS publications also published business intelligence related articles, but at a lower rate than the aforementioned sources (see Table 5). Relevant sources from industry tend to be general IT publications, without a specific BI focus (e.g., *ComputerWorld* at #1, *Information Today* at #2, and *InformationWeek* at #3), as shown in Table 4. However, there are some focused sources as well, such as *Microsoft Data Mining* (#5), *KM World* (#8), and *CRM Magazine* (#12), that are more relevant to the BI&A related topics of data mining, knowledge management, and customer relation management. KM and CRM have traditionally been topics of interest to IS scholars.

Table 6 summarizes the top-20 academic authors with BI&A publications. Most of these authors are from IS and CS, with several others from the related fields of marketing, management, communication, and mathematics. Many of these authors are close collaborators, for example, Hsinchun Chen (#1), Jay F. Nunamaker (#18), Michael Chau (#11), and Wingyan Chung (#18) through the University of Arizona connection,[11] and Barabara H. Wixom (#5) and Hugh J. Watson (#5) through the University of Georgia connection. We also report the PageRank score (Brin and Page 1998), a popular metric for data and network analytics, for the BI&A authors based on the coauthorship network within BI&A publications. A higher PageRank score captures an author's propensity to collaborate with other prolific authors. The

[11]Readers are welcome to contact the authors for validation of our data set and results or for additional analysis.

Table 5. Major IS Journals with BI&A Publications

Academic Publication	Publications
Decision Support Systems	41
Communications of the AIS	19
Journal of Management Information Systems	12
Management Science	10
Information Systems Research	9
Journal of the Association for Information Systems	5
INFORMS Journal on Computing	4
Management Information Systems Quarterly	2

Table 6. Top Academic Authors in BI&A

Rank	Name	Affiliation	Discipline	Region	Total	PageRank
1	Hsinchun Chen	University of Arizona, U.S.	IS	North America	19	7.471
2	Shenghong Li	Zhejiang University, China	Math	Asia	16	4.276
3	Yong Shi	University of Nebraska, U.S.	CS	North America	15	3.708
4	Kin Keung Lai	City University of Hong Kong, China	IS	Asia	14	4.780
5	Barbara H. Wixom	University of Virginia, U.S.	IS	North America	8	2.727
5	Hugh J. Watson	University of Georgia, U.S.	IS	North America	8	2.485
5	Elizabeth Chang	Curtin University, Australia	IS	Australia	8	2.381
5	Sheila Wright	De Montfort University, U.K.	Marketing	Europe	8	2.859
5	Matteo Golfarelli	University of Bologna, Italy	CS	Europe	8	1.785
5	Farookh Hussain	University of Technology Sydney, Australia	CS	Australia	8	1.264
11	Michael Chau	Hong Kong University, China	IS	Asia	7	1.788
11	Josef Schiefer	Vienna University of Technology, Austria	CS	Europe	7	2.731
11	Craig S. Fleisher	College of Costal Georgia, U.S.	Management	North America	7	1.042
14	Lingling Zhang	Towson University, U.S.	Communication	North America	6	2.328
14	Olivera Marjanovic	University of Sydney, Australia	IS	Australia	6	2.464
16	Xiaofeng Zhang	Changsha University of Science and Technology, China	IS	Asia	5	2.393
16	Stefano Rizzi	University of Bologna, Italy	CS	Europe	5	1.683
18	Jay F. Nunamaker	University of Arizona, U.S.	IS	North America	4	2.792
18	Wingyan Chung	Santa Clara University, U.S.	IS	North America	4	1.761
18	Zahir Urabu	Brunel University, U.K.	Management	Europe	4	2.241

analysis reveals broad and even contribution of authors from North America, Asia, Europe, and Australia, reflecting the diversity and international interest in the field of BI&A.

The last set of analyses investigated the content of BI&A publications from 2000–2011. Mallet (McCallum 2002), a Java-based open-source NLP text analytics tool, was used to extract the top bigrams (two-word phrases) for each year. A few bi-grams were combined to form more meaningful BI-related trigrams such as "customer relation management" and "enterprise resource planning." These keywords were then ranked based on their frequency, and the top 30 keywords displayed using the tagcloud visualization. More important keywords are highlighted with larger fonts as shown in Figure 3. For example, competitive advantage, big data, data warehousing, and decision support emerged as the top four topics in the BI&A literature. Other BI&A related topics such as customer relation management, data mining, competitive

Figure 3. Tagcloud Visualization of Major Topics in the BI&A Literature

intelligence, enterprise resource planning, and knowledge management were also highly ranked. Overall, the topics extracted were highly relevant to BI&A, especially for its managerial and application values, although most of the detailed technical terms, as described in the previous research framework sections, were not present. This could be attributed to the tendency of authors to use broad terminologies in article titles and abstracts.

BI&A Education and Program Development

BI&A provides opportunities not only for the research community, but also for education and program development. In July 2012, Columbia University and New York City announced plans to invest over $80 million dollars in a new Center for Data Science, which is expected to generate thousands of jobs and millions of dollars in tax revenues from 100 startup companies over the next 10 years (Associated Press 2012). BI&A is data science in business. Job postings seeking data scientists and business analytics specialists abound these days. There is a clear shortage of professionals with the "deep" knowledge required to manage the three V's of big data: volume, velocity, and variety (Russom 2011). There is also an increasing demand for individuals with the deep knowledge needed to manage the three "perspectives" of business decision making: descriptive, predictive, and prescriptive analytics. In this section, we describe BI&A education in business schools, present the challenges facing IS departments, and discuss BI&A program development opportunities. We also provide some suggestions for the IS discipline in addressing these challenges (Chiang et al. 2012).

Education Challenges

BI&A focuses on understanding, interpretation, strategizing, and taking action to further organizational interests. Several academic disciplines have contributed to BI&A, including IS, CS, Statistics, Management, and Marketing, as shown in our bibliometric study. IS programs, in particular, are uniquely positioned to train a new generation of scholars and students due to their emphasis on key data management and information technologies, business-oriented statistical analysis and management science techniques, and broad business discipline exposure (e.g., Finance, Accounting, Marketing, and Economics).

Since its inception approximately 45 years ago, IS as an academic discipline has primarily focused on business needs in an era when the major challenges involved the management of internal business and transaction data. In the age of big data, these problems remain, but the emphasis in industry has shifted to data analysis and rapid business decision making based on huge volumes of information. Such time-critical decision making largely takes place outside of the IS function (i.e., in business units such as marketing, finance, and logistics). Can IS programs serve the needs of these business decision makers? Can we provide courses in data mining, text mining, opinion mining, social media/network analytics, web mining, and predictive analytics that are required for marketing and finance majors? We should also ask ourselves about the skill sets needed by students. Should we recruit students with strong math and statistical skills, for example? We contend that a new vision for IS, or at least for some IS programs, should address these questions.

BI&A presents a unique opportunity for IS units in business schools to position themselves as a viable option for edu-

cating professionals with the necessary depth and academic rigor to tackle the increased complexity of BI&A issues. IS programs housed within business schools have access to a variety of business courses, as well as courses intended to improve communication and presentation skills. It is also common for business schools to house management science and statistics faculty in the same IS unit.

BI&A Knowledge and Skills

BI&A education should be interdisciplinary and cover critical analytical and IT skills, business and domain knowledge, and communication skills required in a complex data-centric business environment.

Analytical and IT skills include a variety of evolving topics. They are drawn from disciplines such as statistics and computer science for managing and analyzing both structured data and unstructured text. Coverage of these topics ranges from BI&A 1.0 to BI&A 3.0. The academic programs intended to produce BI&A professionals should consider these analytical and IT skills as suggested in Table 3 of our research framework.

To provide useful insights and decision-making support, the BI&A professionals must be capable of understanding the business issues and framing the appropriate analytical solutions. The necessary business knowledge for BI&A professionals ranges from general familiarity with the areas of Accounting, Finance, Management, Marketing, Logistics, and Operation Management, to the domain knowledge required in specific BI&A applications, some of which are discussed earlier and summarized in Table 2.

The importance of an organization-wide culture for informed fact-based decision making for business analytics is emphasized by Davenport (2006). To support such a culture, BI&A professionals need to know not only how to turn raw data and information (through analytics) into meaningful and actionable knowledge for an organization, but also how to properly interact with and communicate this knowledge to the business and domain experts of the organization.

Program Development

BI&A provides a unique opportunity for IS units in business schools to develop new courses, certificate programs, and degree programs charged with preparing the next generation of analytical thinkers. There are many options for delivering BI&A education. Because of the depth of knowledge required, graduate programs are the obvious choice. Viable program development options in delivering BI&A education include

- creating a Master of Science (MS) degree in BI&A
- creating a BI&A concentration in existing MS IS programs
- offering a graduate BI&A certificate program

The first option requires the effort of developing a new program. A few universities have embarked on this endeavor. A nonexhaustive list includes North Carolina State University, Saint Joseph's University, Northwestern University, the University of Denver, Stevens Institute of Technology, and Fordham University. New York University will launch its new program in May 2013. New MS degree programs can be designed explicitly to attract analytically strong students with undergraduate degrees in areas such as mathematics, science, and computer science, and to prepare these students for careers, not only in the IS or IT groups in industry, but also in functional areas such as research and development, marketing, media, logistics, and finance.

The second option leverages existing MS IS programs with a BI&A concentration that would supplement the already existing curriculum in IT, data management, and business and communication courses with additional analytics coverage. This option has been adopted by a number of schools including the IS departments at Carnegie Mellon University and the University of Arizona. This option provides BI&A knowledge and skills for students who will primarily find careers in IS groups in industry.

For working IT professionals who wish to expand into BI&A, a part-time MS or certificate program (the third option) offer practical and valid alternatives. These certificate programs can be delivered online or on-site and need to provide the skills that will complement the current IT or business experience of IT professionals, and/or provide technical and analytical skills to business professionals in non-IT areas. Online programs that are currently available include Northwestern University's MS in Predictive Analytics and Stanford University's Graduate Certificate on Mining Big Data. In addition, IS programs can help design a BI&A concentration in MBA programs to help train a new generation of data- and analytics-savvy managers.

A key to success for a BI&A program is to integrate the concept of "learning by doing" in the BI&A curriculum via hands-on projects, internships, and industry-guided practicum. Big data analytics requires trial-and-error and experimentation. Strong relationships and partnerships between academic programs and industry partners are critical to foster the experiential learning aspect of the BI&A curriculum.

Papers in this Special Issue

The idea for this special issue began in May 2009, when Detmar Straub, the Editor-in-Chief of *MIS Quarterly*, solicited suggestions for special issues from the editorial board members. We submitted the special issue proposal on Business Intelligence Research in August 2009, with the call-for-papers approved and distributed at the 30th Annual International Conference on Information Systems (ICIS) in December of that year. Submissions to this special issue needed to relate to *MIS Quarterly's* mission with strong managerial, organizational, and societal relevance and impact. In addition to the Design Science approach (Hevner et al. 2004; March & Storey 2008), rigorous and relevant BI-related research using management science (modeling, optimization), information economics, and organizational and behavioral methodologies (case studies, surveys) was also welcomed. A total of 62 manuscripts was received by October 2010. In the following 20 months, six of the manuscripts went through three or four review rounds and were then accepted for this issue.

The six papers address various aspects of the BI&A research framework presented in this introduction paper (see Table 7). All six papers are based on BI&A 1.0, with three also based on BI&A 2.0. The first three papers by Chau and Xu, Park et al., and Lau et al. focus on BI&A 2.0 with applications on e-commerce and market intelligence using text, web, and network analytics. In the next two papers, both Hu et al. and Abbasi et al. work in the category of BI&A 1.0 with a focus on security, but Hu et al. use network analytics whereas Abbasi et al. emphasize security analysis and data analytics. Finally, Sahoo et al. also work in BI&A 1.0, with direct application to e-commerce and market intelligence using web and data analytics.

In "Business Intelligence in Blogs: Understanding Consumer Interactions and Communities," Michael Chau and Jennifer Xu recognized the potential "gold mine" of blog content for business intelligence and developed a framework for gathering business intelligence by automatically collecting and analyzing blog content and bloggers' interaction networks. A system developed using this framework was applied to two case studies, which revealed novel patterns in blogger interactions and communities.

Sung-Hyuk Park, Soon-Young Huh, Wonseok Oh, and Sang Pil Han in their paper, "A Social Network-Based Inference Model for Validating Customer Profile Data," argue that business intelligence systems are of limited value when they deal with inaccurate and unreliable data. The authors proposed a social network-driven inference framework to determine the accuracy and reliability of self-reported customer profiles. The framework utilizes the individuals' social circles and communication patterns within their circles. To construct the specific inference and validation model, a combination of methods was used, including query processing, statistical inference, social network analysis, and user profiling. The authors analyzed over 20 million actual mobile call transactions and their proposed social network-based inference model consistently outperformed the alternative approaches.

In "Web 2.0 Environmental Scanning and Adaptive Decision Support for Business Mergers and Acquisitions," Raymond Lau, Stephen Liao, K. F. Wong, and Dickson Chiu analyzed company mergers and acquisitions (M&A). Online environmental scanning with Web 2.0 provides top executives with opportunities to tap into collective web intelligence to develop better insights about the socio-cultural and political-economic factors that cross-border M&As face. Grounded on Porter's five forces model, this research designed a due diligence scorecard model that leverages collective web intelligence to enhance M&A decision making. The authors also developed an adaptive business intelligence (BI) 2.0 system, which they applied to Chinese companies' cross-border M&A activities.

In their paper, "Network-Based Modeling and Analysis of Systemic Risk in Banking Systems," Daning Hu, J. Leon Zhao, Zhimin Hua, and Michael Wong analyzed systemic risk in banking systems by treating banks as a network linked with financial relationships, leading to a network approach to risk management (NARM). NARM analyzes systemic risk attributed to each individual bank via simulation based on real-world data from the Federal Deposit Insurance Corporation. NARM offers a new means for predicting contagious bank failures and determining capital injection priorities at the individual bank level in the wake of financial crisis. A simulation study showed that, under significant market shocks, the interbank payment links became more influential than the correlated bank portfolio links in determining an individual bank's survival.

In their paper, "MetaFraud: A Meta-Learning Framework for Detecting Financial Fraud," Ahmed Abbasi, Conan Albrecht, Anthony Vance, and James Hansen employed a design science approach to develop MetaFraud, a meta-learning framework for enhanced financial fraud detection. A series of experiments was conducted on thousands of legitimate and fraudulent firms to demonstrate the effectiveness of the framework over existing benchmark methods. The research results have implications for compliance officers, investors, audit firms, and regulators.

The paper by Nachiketa Sahoo, Param Vir Singh, and Tridas Mukhopadhyay, "A Hidden Markov Model for Collaborative Filtering," reports on the analysis of making personalized recommendations when user preferences are changing. The

Table 7. Summary of Special Issue Papers Within the BI&A Research Framework

Authors and Titles	Evolutions	Applications	Data	Analytics/ Research	Impacts
Chau and Xu, "Business Intelligence in Blogs: Understanding Consumer Interactions and Communities"	BI&A 2.0 on social media & network analytics	Market intelligence on consumers and communities	User-generated content extracted from blogs	• Text and network analytics • Community detection • Network visualization	Increased sales and customer satisfaction
Park et al., "A Social Network-Based Inference Model for Validating Customer Profile Data"	BI&A 1.0 & 2.0 on social network analysis and statistical analysis	Market intelligence in predicting customers' profiles	Self-reported user profiles and mobile call records	• Network analytics • Anomaly detection • Predictive analytics	Personalized recommendation and increased customer satisfaction
Lau et al., "Web 2.0 Environmental Scanning and Adaptive Decision Support for Business Mergers and Acquisitions"	BI&A 1.0 and 2.0 on scorecards and web analytics	Market intelligence on environmental scanning	Business information extracted from Internet and proprietary financial information	• Text and web analytics • Sentiment and affect analysis • Relation mining	Strategic decision making in mergers and acquisitions
Hu et al., "Network-Based Modeling and Analysis of Systemic Risk in Banking Systems"	BI&A 1.0 on statistical analysis	Systemic risk analysis and management in banking systems	U.S. banking information extracted from FDIC and Federal Reserve Wire Network	• Network and data analytics • Descriptive and predictive modeling • Discrete event simulation	Monitoring and mitigating of contagious bank failures
Abbasi et al., "MetaFraud: A Meta-Learning Framework for Detecting Financial Fraud"	BI&A 1.0 on data mining and meta-learning	Fraud detection	Financial ratios, and organizational and industrial-level context features	• Data analytics • Classification & generalization • Adaptive learning	Financial fraud detection
Sahoo et al., "A Hidden Markov Model for Collaborative Filtering"	BI&A 1.0 on statistical analysis	Recommender systems with changing user preferences	Blog reading data, Netflix prize data set, and Last.fm data	• Data and web analytics • Statistical dynamic model • Collaborative filtering	Personalized recommendations

authors proposed a hidden Markov model based on collaborative filtering to predict user preferences and make the most appropriate personalized recommendations for the predicted preference. The authors employed real world data sets and simulations to show that, when user preferences are changing, there is an advantage to using the proposed algorithm over existing ones.

Summary and Conclusions

Through BI&A 1.0 initiatives, businesses and organizations from all sectors began to gain critical insights from the structured data collected through various enterprise systems and analyzed by commercial relational database management systems. Over the past several years, web intelligence, web analytics, web 2.0, and the ability to mine unstructured user-generated contents have ushered in a new and exciting era of BI&A 2.0 research, leading to unprecedented intelligence on consumer opinion, customer needs, and recognizing new business opportunities. Now, in this era of Big Data, even while BI&A 2.0 is still maturing, we find ourselves poised at the brink of BI&A 3.0, with all the attendant uncertainty that new and potentially revolutionary technologies bring.

This *MIS Quarterly* Special Issue on Business Intelligence Research is intended to serve, in part, as a platform and conversation guide for examining how the IS discipline can better serve the needs of business decision makers in light of maturing and emerging BI&A technologies, ubiquitous Big Data, and the predicted shortages of data-savvy managers and of business professionals with deep analytical skills. How can academic IS programs continue to meet the needs of their traditional students, while also reaching the working IT professional in need of new analytical skills? A new vision for IS may be needed to address this and other questions.

By highlighting several applications such as e-commerce, market intelligence, e-government, healthcare, and security, and by mapping important facets of the current BI&A knowledge landscape, we hope to contribute to future sources of knowledge and to augment current discussions on the importance of (relevant) academic research.

Finally, the six papers chosen for this special issue are themselves a microcosm of the current state of BI&A research. These "best of the best" research papers showcase how high-quality academic research can address real-world problems and contribute solutions that are relevant and long lasting—exactly the challenge that our discipline continues to face.

Acknowledgments

We wish to thank the Editor-in-Chief of *MIS Quarterly*, Detmar Straub, for his strong support for this special issue from its inception. He shared the belief that business intelligence and analytics is an emerging and critical IS research area. We appreciate the continued support from the incoming Editor-in-Chief, Paulo Goes, and his feedback on an earlier version of this paper. We also thank Janice DeGross and Jennifer Syverson from the *MIS Quarterly* office for their professional editorial support and Cathy Larson for her support and assistance in managing the manuscripts and coordinating the review process.

We are grateful to our excellent group of 35 associate editors (listed below) and the reviewers (too numerous to name) who carried out the review process in a timely manner while still meeting *MIS Quarterly's* high expectations of scholarly quality. We thank the authors of these 62 submissions who chose to submit their research to our special issue. We are especially indebted to the associate editors who handled the six accepted papers of the special issue. They and the reviewers they invited offered valuable critiques and suggestions throughout the review process. This special issue would not have been possible without their efforts.

The research reported in this article was partially supported by the following sources: National Science Foundation (NSF CMMI-1057624, CMMI-0926270, CNS-0709338), Defense Threat Reduction Agency (DTRA HDTRA-09-0058), J. Mack Robinson College of Business at the Georgia State University, Carl H. Lindner College of Business at the University of Cincinnati, and the Eller College of Management at the University of Arizona. We also thank the following colleagues for their assistance or comments: Ee-Peng Lim, Ted Stohr, Barbara Wixom, Yukai Lin, and Victor Benjamin.

Special Issue Associate Editors

Gediminas Adomavicius, University of Minnesota
Sue Brown, University of Arizona
Michael Chau, University of Hong Kong
Cecil Chua, University of Auckland
Wendy Currie, Audencia, Ecole de Management
Andrew Gemino, Simon Fraser University
Paulo Goes, University of Arizona
Alok Gupta, University of Minnesota
Paul Jen-Hwa Hu, University of Utah
Hemant Jain, University of Wisconsin – Milwaukee
Robert Kauffman, Singapore Management University
Vijay Khatri, Indiana University
Gondy Leroy, Claremont Graduate University
Ting-Peng Liang, National Chengchi University
Ee-Peng Lim, Singapore Management University
Vijay Mookerjee, University of Texas at Dallas
Sridhar Narasimhan, Georgia Institute of Technology
Jeffrey Parsons, Memorial University of Newfoundland
H. Raghav Rao, The State University of New York at Buffalo
Raghu T. Santanam, Arizona State University
Balasubramaniam Ramesh, Georgia State University
Ramesh Sharda, Oklahoma State University
Matti Rossi, Aalto University School of Economics
Michael Jeng-Ping Shaw, University of Illinois, Urbana-Champaign
Olivia Sheng, University of Utah
Keng Siau, Missouri University of Science and Technology
Atish Sinha, University of Wisconsin – Milwaukee
Alexander Tuzhilin, New York University
Vijay Vaishnavi, Georgia State University
Doug Vogel, City University of Hong Kong
Chih-Ping Wei, National Taiwan University
Barbara Wixom, University of Virginia
Carson Woo, University of British Columbia
Daniel Zeng, University of Arizona
J. Leon Zhao, City University of Hong Kong

References

Adomavicius, G., and Tuzhilin, A. 2005. "Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions," *IEEE Transactions on Knowledge and Data Engineering* (17:6), pp. 734-749.

Anderson, C. 2004. "The Long Tail," *WIRED Magazine* (12:10) (http://www.wired.com/wired/archive/12.10/tail.html).

Associated Press. 2012. "Columbia U Plans New Institute for Data Sciences," July 30 (http://www.cbsnews.com/8301-505245_162-57482466/columbia-u-plans-new-institute-for-data-sciences/, accessed August 3, 2012).

Barabási, A. 2003. *Linked: How Everything Is Connected to Everything Else and What it Means for Business, Science, and Everyday Life*, New York: Plume.

Batagelj, V., and Mrvar, A. 1998. "Pajek: A Program for Large Network Analysis," *Connections* (21), pp. 47-57.

Bettencourt, L. M.A., Cintrón-Arias, A., Kaiser, D. I., and Castillo-Chávez, C. 2006. "The Power of a Good Idea: Quantitative Modeling of the Spread of Ideas from Epidemiological Models," *Physica A* (364), pp. 513-536.

Bitterer, A. 2011. "Hype Cycle for Business Intelligence," Gartner, Inc., Stamford, CT.

Blei, D. M. 2012. "Probabilistic Topic Models," *Communications of the ACM* (55:4), pp. 77-84.

Bloomberg Businessweek. 2011. "The Current State of Business Analytics: Where Do We Go from Here?," Bloomberg Business-

week Research Services (http://www.sas.com/resources/asset/busanalyticsstudy_wp_08232011.pdf).

Borgatti, S. P., Everett, M. G., and Freeman, L. C. 2002. *UCInet for Windows: Software for Social Network Analysis*, Harvard, MA: Analytic Technologies.

Brantingham, P. L. 2011. "Computational Criminology," Keynote Address to the European Intelligence and Security Informatics Conference, Athens, Greece, September 12-14.

Brin, S.,and Page, L. 1998. "The Anatomy of a Large-Scale Hypertextual Web Search Engine," *Computer Network and ISDN Systems* (30), pp. 107-117.

Brumfiel, G. 2911, "High-Energy Physics: Down the Petabyte Highway," *Nature* (469), pp. 282-283.

Chaudhuri, S., Dayal, U., and Narasayya, V. 2011. "An Overview of Business Intelligence Technology," *Communications of the ACM* (54:8), pp. 88-98.

Chen, H. 2006. *Intelligence and Security Informatics for International Security: Information Sharing and Data Mining*, New York: Springer.

Chen, H. 2009. "AI, E-Government, and Politics 2.0," *IEEE Intelligent Systems* (24:5), pp. 64-67.

Chen, H. 2011a. "Design Science, Grand Challenges, and Societal Impacts," *ACM Transactions on Management Information Systems* (2:1), pp. 1:1-1:10.

Chen, H. 2011b. "Smart Health and Wellbeing," *IEEE Intelligent Systems* (26:5), pp. 78-79.

Chen, H. 2012. *Dark Web*: *Exploring and Mining the Dark Side of the Web*, New york: Springer.

Chen, H., Brandt, L., Gregg, V., Traunmuller, R., McIntosh, A., Dawes, S., Hovy, E., and Larson, C. A. (eds.). 2007. *Digital Government: E-Government Research, Case Studies, and Implementation*, New York: Springer.

Chen, H., Reid, E., Sinai, J., Silke, A., and Ganor, B. (eds.). 2008. *Terrorism Informatics: Knowledge Management and Data Mining for Homeland Security*, New York: Springer.

Chiang, R. H. L., Goes, P., and Stohr, E. A. 2012. "Business Intelligence and Analytics Education and Program Development: A Unique Opportunity for the Information Systems Discipline," *ACM Transactions on Management Information Systems* (3:3), forthcoming.

Davenport, T. H. 2006. "Competing on Analytics," *Harvard Business Review* (84:1), p. 98-107.

Doan, A., Ramakrishnan, R., and Halevy, A. Y. 2011. "Crowdsourcing Systems on the World-Wide Web," *Communications of the ACM* (54:4), pp. 86-96.

Fortunato, S. 2010. "Community Detection in Graphs," *Physics Reports* (486:3-5), pp. 75-174.

Frank, O., and Strauss, D. 1986. "Markov Graphs," *Journal of the American Statistical Association* (81:395), pp. 832-842.

Freeman, T. 2005. *The World is Flat: A Brief History of the Twenty-First Century*, New York: Farrar, Straus, and Giroux.

Gelfand, A. 2011/2012. "Privacy and Biomedical Research: Building a Trust Infrastructure—An Exploration of Data-Driven and Process-Driven Approaches to Data Privacy," *Biomedical Computation Review*, Winter, pp. 23-28 (available at http://biomedicalcomputationreview.org/content/privacy-and-biomedical-research-building-trust-infrastructure, accessed August 2, 2012).

Hanauer, D. A. , Rhodes, D. R., and Chinnaiyan, A. M. 2009. "Exploring Clinical Associations Using '-Omics' Based Enrichment Analyses," *PLoS ONE* (4:4): e5203.

Hanauer, D. A., Zheng, K., Ramakrishnan, N., and Keller, B. J. 2011. "Opportunities and Challenges in Association and Episode Discovery from Electronic Health Records," *IEEE Intelligent Systems* (26:5), pp. 83-87.

Henschen, D. 2011. "Why All the Hadoopla?" *Information Week*, November 14, pp. 19-26.

Hevner, A., March, S. T., Park, J., and Ram. S. 2004. "Design Science Research in Information Systems," *MIS Quarterly* (28:1), pp. 75-105.

Hirsch, J. E. 2005. "An Index to Quantify an Individual's Scientific Research Output," *Proceedings of the National Academy of Sciences of the United States of America* (102:46), pp. 16569-16572.

Hunter, D. R., Handcock, M. S., Butts, C. T., Goodreau, S. M., and Morris, M. 2008. "ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Network," *Journal of Statistical Software* (24:3) (http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2743438/).

IBM. 2011. "The 2011 IBM Tech Trends Report: The Clouds are Rolling In...Is Your Business Ready?," November 15 (http://www.ibm.com/developerworks/techntrendsreport; accessed August 4, 2012)..

Karpf, D. 2009. "Blogsphere Research: A Mixed-Methods Approach to Rapidly Changing Systems," *IEEE Intelligent Systems* (24:5), pp. 67-70.

Liben-Nowell, D., and Kleinberg, J. 20007. "The Link-Prediction Problem for Social Networks," *Journal of American Society for Information Science and Technology* (58:7), pp. 1019-1031.

Lin, Y., Brown, R. A., Yang, H. J., Li, S., Lu, H., and Chen, H. 2011. "Data Mining Large-Scale Electronic Health Records for Clinical Support," *IEEE Intelligent Systems* (26:5), pp. 87-90.

Lusch, R. F., Liu, Y., and Chen, Y. 2010. "The Phase Transition of Markets and Organizations: The New Intelligence and Entrepreneurial Frontier," *IEEE Intelligent Systems* (25:1), pp. 71-75.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Byers, A. H. 2011. "Big Data: The Next Frontier for Innovation, Competition, and Productivity," McKinsey Global Institute (http://www.mckinsey.com/insights/mgi/research/technology_and_innovation/big_data_the_next_frontier_for_innovation; accessed August 4, 2012).

Manning, C. D., and Schütze, H. 1999. *Foundations of Statistical Natural Language Processing*, Cambridge, MA: The MIT Press.

March, S. T., and Storey, V. C. 2008. "Design Science in the Information Systems Discipline," *MIS Quarterly* (32:4), pp. 725-730.

Maybury, M. T. (ed.). 2004. *New Directions in Question Answering*, Cambridge, MA: The MIT Press.

McCallum, A. 2002. "Mallet: A Machine Learning for Language Toolkit," University of Massachusetts, Amherst (http://mallet.cs.umass.edu/).

Miller, K. 2012a. "Big Data Analytics in Biomedical Research," *Biomedical Computation Review* (available at http://biomedicalcomputationreview.org/content/big-data-analytics-biomedical-research; accessed August 2, 2012).

Miller, K. 2012b. "Leveraging Social Media for Biomedical Research: How Social Media Sites Are Rapidly Doing Unique Research on Large Cohorts," *Biomedical Computation Review* (available at http://biomedicalcomputationreview.org/content/leveraging-social-media-biomedical-research; accessed August 2, 2012).

National Research Council. 2008. *Behavioral Modeling and Simulation: From Individuals to Societies*, Committee on Organizational Modeling: From Individuals to Societies, G. L. Zacharias, J. MacMillan and S. B. Van Hemel (eds.), Board on Behavioral, Cognitive, and Sensory Sciences, Division of Behavioral and Social Sciences and Education, Washington, DC: The National Academies Press.

O'Reilly, T. 2005. "What Is Web 2.0? Design Patterns and Business Models for the Next Generation of Software," September 30, (http://www.oreillynet.com/pub/a/oreilly/tim/news/2005/09/30/what-is-web-20.html).

Pang, B., and Lee, L. 2008. "Opinion Mining and Sentiment Analysis," *Foundations and Trends in Information Retrieval* (2:1-2), pp. 1-135.

Patterson, D. A. 2008. "Technical Perspective: The Data Center Is the Computer," *Communications of the ACM* (51:1), p. 105.

Perlroth, N., and Rusli, E. M. 2012. "Security Start-Ups Catch Fancy of Investors," *New York Times*, Technology Section, August 5.

Robins, G., Pattison, P., Kalish, Y., and Lusher, D. 2007. "An Introduction to Exponential Random Graph (p*) Models for Social Networks," *Social Networks* (29:2), pp. 173-191.

Russom, P. 2011. "Big Data Analytics," TDWI Best Practices Report, Fourth Quarter.

Sallam, R. L., Richardson, J., Hagerty, J., and Hostmann, B. 2011. "Magic Quadrant for Business Intelligence Platforms," Gartner Group, Stamford, CT.

Salton, G. 1989. *Automatic Text Processing*, Reading, MA: Addison Wesley.

Schonfeld, E. 2005. "The Great Giveaway," *Business 2.0* (6:3), pp. 80-86.

Snider, M. 2012. "More Businesses Getting Their Game On," *USA Today*, July 30.

Stonebraker, M., Abadi, D., DeWitt, D. J., Madden, S., Pavlo, A., and Rasin, A. 2012. "MapReduce and Parallel DBMSs: Friends or Foes," *Communications of the ACM* (53:1), pp. 64-71.

The Economist. 2010a. "The Data Deluge," Special Report on Managing Information, Technology Section, February 25 (http://www.economist.com/node/15579717).

The Economist. 2010b. "All Too Much," Special Report on Managing Information, Technology Section, February 25 (http://www.economist.com/node/15557421).

The Economist. 2011. "Beyond the PC," Special Report on Personal Technology, October 8 (http://www.economist.com/node/21531109).

Turban, E., Sharda, R., Aronson, J. E., and King, D. 2008. *Business Intelligence: A Managerial Approach*, Boston: Pearson Prentice Hall,

U.S. Office of Homeland Security. 2002. *National Strategy for Homeland Security*, Washington, DC: Office of Homeland Security.

van der Aalst, W. 2012. "Process Mining: Overview and Opportunities," *ACM Transactions on Management Information Systems* (3:2), pp. 7:1-7:17.

Wactlar, H., Pavel, M., and Barkis, W. 2011. "Can Computer Science Save Healthcare?" *IEEE Intelligent Systems* (26:5), pp. 79-83.

Watson, H. J., and Wixom, B. H. 2007. "The Current State of Business Intelligence," *IEEE Computer* (40:9), pp. 96-99.

Watts, D. 2003. *Six Degrees: The Science of a Connected Age*, New York: W. W. Norton.

Witten, I. H., Frank, E., and Hall, M. 2011. *Data Mining: Practical Machine Learning Tools and Techniques* (3rd ed.), San Francisco: Morgan Kaufmann.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., Yu, P.S., Zhou, Z.-H., Steinbach, M., Hand, D. J., and Steinberg, D. 2007. "Top 10 Algorithms in Data Mining," *Knowledge and Information Systems* (14:1), pp. 1-37.

Yang, H., and Callan, J. 2009. "OntoCop: Constructing Ontologies for Public Comments," *IEEE Intelligent Systems* (24:5), pp. 70-75.