SENTIMENT ANALYSIS AND SOCIAL MEDIA IN HSCB

By Dr. John Boiney

The HSCB Modeling Program is supporting a number of projects conducting research on social media, drawing on publicly available data from non-U.S. sources. One of the main foci for this research is on assessing the value of social media data for understanding and forecasting sociocultural behavior primarily through the use of sentiment analysis techniques. Although many other approaches can be used to analyze social media, such as Least Squares's triangulation (discussed below), sentiment analysis is well grounded in both theory and practice, and as such is a very effective way to analyze these data. HSCB modeling does not focus only on sentiment analysis, however; the program is also addressing technical challenges associated with processing very large volumes of open source data—including data from social media—and structuring them sufficiently so that they can be used in computational models or formal analysis. Better understanding and use of social media data are critical contributors to realizing the vision of a "Social Radar.[1]"

Part of the reason for the increasing attention paid to social media is that it can be a source of data on how populations are feeling—their attitudes, emotions, sentiments, and opinions. **Strategic Analysis Enterprises** hypothesizes that automated sentiment analysis can serve as a cost-effective substitute for formal polling and surveys when those techniques are not feasible due to operational or strategic constraints. The project is building and validating text analysis software using natural language processing techniques, and has been testing its hypothesis using blog and other text data. **Lockheed Martin** is leading an advanced prototyping effort called iSENT that will provide daily automated collection and storage of open-source text and allow users to generate temporal and frequency-driven visualizations, including analytic trending, sentiment volumetrics, geographic mapping, and network analysis and visualizations.

Another project supported by the HSCB Modeling Program is researching techniques for rapidly detecting meaningful change in mood as expressed in microblogs. This work, led by **MITRE**, builds on "Linguistic Inquiry and Word Count 2007" (LIWC) software that can be used to represent the relative levels of various emotion categories in social media and

1 Maybury, Mark (2010). "Social Radar for Smart Power." The MITRE Corporation.

Continued on page 8

ISSUE Nº 13 SUMMER 2012

WELCOME
Introduction and Welcome to the HSCB Program
Page 2

FEATURES
HSCB Participation in CCDM
Page 2

HSCB's Unique and Critical Role
Page 3

New HSCB Projects
Page 5

W-ICEWS Transition to the ISPAN GAP CIE Program
Page 5

Social Media Confernce Summary
Page 6

Social Media Specialists Meeting Summary
Page 8

SBIR Program
Page 11

Social Media Technology Research and Commercialization
Page 13

SPOTLIGHT
DataCards—Collaborative Data Discovery and Information Sharing for Sociocultural Data
Page 9

WELCOME

Welcome to the thirteenth issue of the Human Social Culture Behavior (HSCB) Modeling Program newsletter. In this issue we examine the wide variety of projects that the program has funded, and how the HSCB Program is adapting to new challenges. We also look at HSCB participation in the Cross-Cultural Decision Making (CCDM) conference, part of the Applied Human Factors and Ergonomics (AHFE) conference that will take place in San Francisco in July. Lieutenant David Combs (USN) and John Bornmann discuss our affiliation with AHFE and provide an overview of how our HSCB projects fit into the CCDM conference.

Several articles present specific programs in greater detail. In his article on page 9, Brian Efird discusses the DataCards project hosted at National Defense University, which began as a database for identifying the variety of sources for sociocultural and other data in Afghanistan. DataCards has since expanded to support the U.S. Northern Command and the Departments of Homeland Security and Justice, and will soon include a number of tools that will make it more effective for the user.

David Foster's article on page 5 focuses on new projects that the HSCB Program has begun to fund this year, including both new research teams and expansions of existing projects that are moving closer to transition. Starting on page 5, Philippe Loustaunau highlights the completion of the HSCB process with a look at our successful Worldwide Integrated Crisis Early Warning System (W-ICEWS) program and its transition to a Program of Record through the Integrated Strategic Planning and Analysis Network (ISPAN) Global Adaptive Planning Collaborative Information Environment (GAP CIE). This issue also discusses some of our ongoing projects, with Ted Stump's description of selected Small Business Innovative Research (SBIR)-related projects and John Boiney's article detailing sentiment analysis and social media analysis tools in our portfolio. As a supplement to this article, Jennifer Mathieu outlines research and commercialization in social media technology, noting many of the different projects in the private and government sectors that focus on this new and dynamic field.

Social media is likely to play a large part in sociocultural research in the future, and the HSCB Program is prepared to engage with this new form of data. Les Servi's article on page 7 and Elizabeth Lyon's article on page 6 discuss two recent social media events in which our HSCB-funded researchers participated. To extend our engagement with the academic and government researchers working in this field, our research teams will attend more such events in the future, including participating in sessions at the CCDM conference.

I hope you enjoy this issue of the HSCB newsletter, and look forward to seeing you at the July CCDM conference.

CAPT Dylan Schmorrow, MSC, USN, PhD

Deputy Director, Human Performance, Training and BioSystems Research Directorate
Program Manager, Human Social Culture Behavior Modeling Program
Office of the Assistant Secretary of Defense (Research and Engineering)

HSCB Participation in the Cross-Cultural Decision Making Conference

By Dr. John Bornmann

This year the Human Social Culture Behavior (HSCB) Modeling Program will participate in the 4th International Applied Human Factors and Ergonomics (AHFE) conference, primarily at the 2nd International Cross-Cultural Decision Making (CCDM) conference to be held on 21–25 July 2012. The CCDM conferences provide a venue for communication between HSCB-funded research teams and teams from other areas of government, academia, and industry. Two years ago, HSCB-funded research teams participated in the 1st International CCDM conference; this year we hope that we will continue developing the relationships built at that event.

Both the AHFE and CCDM conferences encompass broad intellectual areas with reasonable overlap between the two. The primary focus of the CCDM conference is on the intersections between psychosocial theory informed by the social sciences and methods of computational modeling informed by computer science and mathematics. There are many research challenges that will be addressed at this year's CCDM conference, including but not limited to unification and standardization of data, validation

Continued on page 4

PUBLISHED BY:

Strategic Analysis, Inc.
4075 Wilson Blvd.
Suite 200
Arlington, VA 22203

Layout and Design
Amy Cauffman

Technical Editor
Becky Afergan

HSCB's Unique and Critical Role in Cross-Cultural Decision Making

By LT David Combs

This year's 2nd annual International Conference on Cross-Cultural Decision Making (CCDM), to be held 21–25 July in San Francisco, California, focuses on "improved decision making across a variety of cultural constructs, including geographical, historical, sociological, organizational, team, and technology interactions." The Assistant Secretary of Defense for Research and Engineering's (ASD(R&E)) Human Social Culture Behavior (HSCB) Modeling Program, led by CAPT Dylan Schmorrow, is uniquely positioned to contribute to the CCDM field. In many ways, the mission of the HSCB Program is a near-perfect fit with the CCDM conference. The HSCB mission is to study the influence of cultural, social, and cognitive factors on human behavior, with a focus on using research results to build models and tools for analysts, warfighters, and decision makers. HSCB achieves its mission through its Understand, Detect, Forecast, and Mitigate research framework.

Figure 1. Understand, Detect, Forecast, and Mitigate Framework

Understand

Improved cross-cultural decision making requires enhanced understanding of countries, cultures, and people. HSCB-funded researchers conduct critical, on-the-ground research designed to increase our knowledge of the sociocultural features and dynamics of different countries and regions. Dr. Mansoor Moaddel (Eastern Michigan University) recently completed a survey of eight Middle Eastern countries that examines the drivers of Islamic fundamentalism. Other researchers, including Dr. Ron Fricker and Lieutenant Commander W. Kulzy of the Naval Postgraduate School (NPS), are probing the nature of trust across cultures. NPS is partnering with the Department of State and U.S. Africa Command to examine variables that affect African citizens' trust in government.

Detect

Detecting operationally relevant sociocultural signatures is critical in cross-cultural decision making and is a natural outcome of enhanced cultural understanding. The MITRE Corporation leads a social radar prototyping initiative that shows great promise for detecting operationally relevant signals in global-scale open source data. The HSCB Program funds a number of other detection-related projects, such as 361 Interactive's CultureGear effort, which integrates natural decision making theory and cognitive task analysis into a computer-based training system designed to enhance general cross-cultural competence. U.S. ground force trainees who use the CultureGear trainer will be prepared to detect abnormalities in the environments and cross-cultural situations within which they operate.

Forecast

Cultural understanding paired with advanced detection tools places users in a position to forecast potential cross-cultural events of interest. Forecasting, in turn, is key to nuanced cross-cultural decision making. HSCB funds the Worldwide Integrated Crisis Early Warning System (W-ICEWS). A joint effort of Lockheed Martin Advanced Technology Laboratories, Strategic Analysis Enterprises, Lustick Consulting, BBN Technologies, and Duke University, W-ICEWS uses open source data and theory-grounded models to forecast nation state instability across 167 countries.

Mitigate

Building on the platform of enhanced understanding, detection, and forecasting, mitigation represents the manifestation of CCDM in the development and implementation of informed courses of action. The HSCB Program funds a team from the University of California San Diego and the University of California Los Angeles to examine methods of analyzing conflict and stability based upon geographical data and other causal factors such as religious and economic differences within a population. The mitigation component of this four-step process will not only allow decision makers to develop and implement various courses of action, but will also provide a robust ability to forecast the effects of their courses of action and measure the effectiveness of their hypothesized solutions.

HSCB and CCDM

The HSCB Program supports one of the world's most advanced interdisciplinary research communities with an interest in enhancing cross-cultural competence and awareness among leaders and analysts through improved cultural understanding, detection methods, forecasting approaches, and mitigation techniques. This year's international conference on CCDM will host more than 70 papers generated by HSCB-funded researchers, just a few of whom are mentioned in this or the previous article. We are happy to participate in the CCDM conference being hosted by AHFE. The HSCB community looks forward to opportunities to learn from and collaborate with other CCDM participants at this year's conference.

HSCB Participation in the Cross-Cultural Decision Making Conference

Continued from page 2

and verification of that data, and assuring that these data remain solidly grounded in psychosocial theory but maintain utility.

While approaches to data handling and verification have been performed in many other fields, human behavior is complex and highly nuanced, making it impossible to develop a simple and universal technique. Therefore, one of the major challenges for CCDM researchers is to define scalable models that possess the correct balance between general and specific approaches. While general or universal approaches can be potentially more useful, they often lack the granularity and robustness which more specific approaches possess. More information about AHFE and the CCDM conference in particular can be found at http://www.ahfe.org .

Participation in the 2nd International CCDM Conference highlights the ongoing importance of the HSCB Program and its role in the wider academic community. Over 70 presenters at this year's CCDM conference will discuss their HSCB-funded projects, both current and past activities. Eight of this year's sessions are primarily dedicated to presenting research conducted by HSCB-funded projects, although almost all of these sessions also include participants from academia and private industry who will present their own work. Fifteen sessions featuring primarily outside researchers and other presenters also incorporate HSCB-funded projects. All of these sessions indicate how the HSCB Program works concurrently and collaboratively with other researchers to solve problems of common interest to the wider research community.

Measuring Performance & Effectiveness of HSCB Methods for Improving Decision Making is a joint session with presenters from both the HSCB Program and the RAND Corporation discussing techniques to validate and verify many of the behavioral models in use today. Several other modeling sessions at this year's CCDM conference include a wide variety of HSCB-funded researchers, academics, government employees, and researchers from other areas of private industry. For example, the **Applications of Human, Social, Culture Behavioral Modeling Technology, Architecture for Socio-Cultural Modeling; Multifarious Modeling, and Verification and Validation** sessions all focus on improving the current state of the art in modeling, from construction of new models to delivery of completed ones.

To improve the effectiveness of our service members in the field, they must be effectively prepared for their operational environments. For this reason, improving training represents a major focus of the HSCB Program, and many HSCB-funded research teams will discuss their projects at the **Tactical Culture Training: Narrative, Personality, & Decision Making and Perceptual Training for Cross-Cultural Decision Making** sessions. These sessions will involve presenters from HSCB and the Office of Naval Research (ONR) describing ways to improve the training and effectiveness of our warfighters, especially their ability to work within foreign cultures and to perceive and understand non-verbal and other culturally specific cues among local populations.

Once service members have achieved competency in these areas, trainers and leaders must also address how to understand and use those competencies most effectively. The CCDM conference will feature two sessions on **Strategic & Tactical Considerations for Cross-Cultural Competency**. The first session focuses on modeling and assessing cross-cultural competency; while the majority of presenters are HSCB funded the session also includes presentations from Minerva- and Office of Naval Research (ONR)-funded researchers. The second session includes three HSCB-funded presenters, as well as four other research teams describing different elements of cross-cultural competence, its definition, and measurements.

The **Understanding Diverse Populations** session includes HSCB-funded researchers from Milcord and the Naval Postgraduate School discussing the various ways in which operational information can be automatically ingested, as well as other measures for understanding populations that can be used to improve the effectiveness of our own deployed forces in a variety of missions. These two research teams will be joined by presenters from Global Cognition and Charles River Analytics, also HSCB-funded researchers, to supplement their findings.

The **Operational Use & Requirements session will include many HSCB performers, as well as pr**esenters from the Air Force Research Lab. They will describe current and past research and how that research is fulfilling the needs of warfighters today, on the ground and in operational environments. One entire session at the CCDM conference is dedicated to discussion of the HSCB-funded Worldwide Integrated Crisis Early Warning System (W-ICEWS) project and other early warning systems. Philippe Loustaunau's article in this newsletter describes the W-ICEWS project in greater detail.

The **When Social Networks Form Through Social Media: Impact & Implications Across Cultures** and the **Social Media & Culture** sessions both include HSCB-funded performers as well as academic, government, and private industry research teams. The sessions will cover ways to utilize new and social media to enhance our ability to understand other cultures and communities.

These are simply a few snapshots of the sessions and presentations that the HSCB Program is supporting at this year's CCDM conference. We are very proud to be working with and alongside so many other research teams in a wide variety of fields. We hope that you plan to attend the conference. Be sure to mark the dates and times of the sessions that sound especially interesting so that you can participate in the discussions.

New HSCB Projects Getting Underway

By Dr. David Foster

2012 promises to be an exciting year for the HSCB Modeling Program as a number of innovative new projects get underway. These projects run the gamut from leveraging remote sensing capabilities to assessing crowd-sourced information and analyzing tactical courses of action.

Can signatures in overhead imagery be used to provide surrogate indicators of well-being and governance? Draper Labs believes they can, and its Remote Sensing and Indicators of Well-Being and Governance project will explore, analyze, and characterize the HSCB-relevant information that can be measured or inferred from commercial satellite imagery. Draper Labs aims to measure the reliability of these indicators and also to assess and validate the indicators' applicability across different geographic regions, with the potential to vastly increase the ability to remotely measure this information.

To assist commanders and staffs in planning and allocating their resources, Strategic Analysis Enterprise's (SAE) Subregional Instability Modeling at the Provincial Level project will provide stability/instability forecasts for country sub-regions (i.e., provinces), further improving our understanding of the conditions that could indicate upcoming violence and instability. SAE's efforts will center on increasing the breadth, depth, and quality of data for provincial-level analyses and on developing new and maturing old methods to better geo-locate events and sentiments given incomplete information.

Just how accurate is crowd-sourced information? Lockheed Martin's Establishing Trust in Crowds project will investigate and develop general-purpose, trust-based techniques and algorithms for assessing crowd-sourced information by leveraging dynamic social networks and analyst feedback. Lockheed Martin will evaluate the Trust in Crowds algorithms using realistic and real-world data within a prototype decision support tool.

Need help evaluating potential courses of action? Perceptronics' Enhanced Course of Action Analysis by Integration of Decision and Social Influence Modeling with Multi-Agent System Technology (CADSIM) is designed to deliver tactical and strategic course of action analysis capabilities to predict and assess how friendly actions result in behavioral outcomes by the adversary and the general population. It also looks at how those behavioral outcomes impact the adversary's and population's decisions and future actions. CADSIM combines Social Influence Theory, Influence Diagrams and Multi-Agent Systems technology. This provides commanders and staffs with a novel capability to express the complex relationships among populations in their area of operations and the goals and strategies for each population. In addition, it projects the likely second- and third-order effects of friendly courses of action.

Identifying pertinent groups within large social media datasets can be challenging, but Least Squares is developing a methodology for identifying coordinated groups of individuals through triangulation analysis across different social data sets, including new social media, more traditional communications media, and expert knowledge. This project, called Triangulating Agency Networks in Social Media for Social Group Identification and Monitoring, will use an interdisciplinary team to extend technologies and methodologies developed under previous government-funded efforts. The team will apply these methods of leveraging massively parallel computers to address existing computational challenges in exploiting very large data sets.

The HSCB Modeling Program eagerly awaits initial findings from these promising new projects.

W-ICEWS Transition to the ISPAN GAP CIE Program

By Dr. Philippe Loustaunau

iTRACE, a capability of the Worldwide Integrated Crisis Early Warning System (W-ICEWS), has been deployed on SIPRNet via U.S. Strategic Command's (USSTRATCOM) Integrated Strategic Planning and Analysis Network (ISPAN) Global Adaptive Planning Collaborative Information Environment (GAP CIE). iTRACE delivers automated event extraction, coding, analysis, and visualization to planners and analysts to examine trends in events ranging from cooperation to violent attacks. This represents the first transition of an HSCB Program capability to a program of record.

ISPAN GAP CIE is an ACAT 1AM acquisition program executed by the Air Force's Electronic Systems Center, with the Office of the Assistant Secretary of Defense for Acquisition, Technology, and Logistics (AT&L) serving as the milestone decision authority. The program provides a capability to support Joint Operations Planning through a common framework that meets the requirements of all Combatant Commands.

W-ICEWS is the next iteration of the software developed under the Defense Advanced Research Programs Agency (DARPA) that coupled open source data analysis with forecasting models to detect political instability. Under the HSCB Modeling Program, managed through the Office of Naval Research, W-ICEWS has expanded from its U.S. Pacific Command (PACOM) and U.S. Southern Command (SOUTHCOM) focus to a worldwide capability that provides automated event extraction from news reports for 175 countries, computational models that forecast instability events (iCAST), and sentiment analysis tools (iSENT). W-ICEWS will transition the iCAST and iSENT capabilities to ISPAN GAP CIE over

Continued on page 10

Social Media Conference Summary

By Elizabeth Lyon and Becky Afergan

In December 2011, the Office of Naval Research (ONR) held a two-day meeting that brought together experts from both government and industry to discuss the topic of current and emerging technology and applications in the social media domain. The purpose of the meeting was primarily to expose government program managers to commercial industry's perspective on where the social media industry is heading and to provide insight to commercial industry with regard to federal research thrust areas. Over the course of the two-day event, participants heard from representatives from commercial industry (social media products and services), the government research community, and individuals with direct experience and firsthand knowledge of the use of social media for applications relevant to national security. Held in San Jose, California, to take advantage of industry expertise, participants provided insight into innovative technologies and emerging trends in the social media industry.

Current and emerging applications, technologies, and research in the social media domain are currently helping the government gain a better understanding of the content and structure of social media, as well as its effectiveness as a communications mechanism. One of the general observations from the workshop was that both commercial industry and the Department of Defense seek to use social media data in order to develop knowledge on the perceptions, attitudes, and beliefs of populations; forecast behaviors, and better understand the direct and indirect effects of potential actions; and formulate and deliver timely and culturally attuned messages.

The meeting was broken up into various sessions to better allow for group discussion. The first was on real-world case studies in social media and included a panel discussion on incidents with national security implications where social media analysis has been used to improve awareness and understanding. These case studies included the Arab Revolution, humanitarian relief, and disease tracking. This session made a strong case for why social media is important to the Department of Defense and national security. The next session focused on federal government research and included briefings by government program managers representing a variety of different research and development organizations. The purpose of this session was to provide an overview to commercial industry representatives of the U.S. Government of research investments in social media related technologies and to outline the driving requirements for that research. Participating organizations included: Army Research Office (ARO), Air Force Research Laboratory (AFRL), Office of Naval Research (ONR), the Defense Advanced Research Projects Agency (DARPA), and the Intelligence Advanced Research Projects Activity (IARPA). The next session held was on future trends and technologies of social media platforms and consisted of a panel of commercial firms that have developed and deployed social media platforms. Platforms were defined as any Internet applications that allow users to enter and post content (in the form of text, images, video, audio, etc.) that can be seen, downloaded, and responded to by others who have access. Social media platforms participating in the workshop included Facebook, Google, and Yahoo!.

The second day of the meeting was primarily devoted to commercial companies that have developed social media analytics applications. In general, these applications process data collected and stored on social media databases (platforms) and transform the user data into useful information about behaviors and trends. The first session focused on those firms that developed underlying technologies for analysis which were source independent (data could be from social media or from other documents). These technologies addressed sentiment analysis and related text analytic methods. Representatives from SAS, Lexalytics, and inTTENSITY each presented their current products and provided insight into where they thought the industry was headed. The next two sessions of the day included commercial firms that offer social media analytic products or services. The workshop closed with a presentation by Lauren Gelman, a professor at Stanford University, who provided an overview of the current issues regarding data privacy and legal trends in social media, focusing

Figure 1. Site visit to Facebook Headquarters

Social Media Conference Summary

Continued from previous page

on the collection, dissemination, and processing of personal data via social media.

Government representatives also had the opportunity following the last day of the conference to tour Topsy and Facebook headquarters. The group met with representatives from the respective companies to discuss social media.

Key challenges in social media analytics articulated during the meeting included:

- Methods for ingest, categorization, and analysis on large, unstructured and diverse data sets.
- Temporal correlation of social media content to future events. Development of multivariate time series models robust to non-stationary, noisy data to reveal patterns that precede events. Use of novel techniques to estimate causality in time series.
- Validity of social media data: misinformation, deception, and characterization of automated content generation - bots in social media.
- Can online social media be used to effectively measure sentiment? Veracity of social media data to ground truth beliefs and behaviors. Accuracy of sentiment analysis given context and dissimilarity of language - culture.
- Validity / bias of open source data and detection of bias / deception.
- Text analytics: limited amount of text; emoticons-shorthand-jargon; multiple and different languages; sarcasm.

Social Media Specialists Meeting Summary

By Dr. Les Servi

In April 2012, participants from the North Atlantic Treaty Organization (NATO) and other organizations met in Tallinn, Estonia, to discuss international research and technical developments in the field of social media. The Social Media Specialists meeting adopted a comprehensive approach, providing a forum to increase understanding of ways to use social media in military applications. It gave participants the opportunity to engage with social media platforms and applications to enhance awareness and understanding of these different forms of social media and how best to engage with social media analysis.

The conference was organized into six sessions, guiding participants through a definition of the problem of working with social media through grounded analysis techniques. The program included thirty-four technical presentations. In addition, assessments of risks and opportunities and recommendations for analysts and other personnel were provided at the conclusion of the meeting.

The first session was devoted to defining the problem and providing a framework for further discussions. Presentations focused on how social media reframes understanding of knowledge and connectedness, as described by Dr. John Verdon, and Dr. Ivy Estabrooke's discussion of the current state of the art in using social media as a proxy for opinion data. Potential future applications were also highlighted by Mr. Barry Costa, who described the development of a "social radar" that can assist today's diplomats, developers, and defenders with strategic- to operational-level situation awareness, as well as with course of action analysis.

The second session of the conference offered a multi-national perspective on examples and implementation of social media. Dr. Jörg Jacobs and Capt. Valeria Diefenbach discussed a German Bundeswehr public affairs campaign that utilized social media platforms, and Dr. Jacquelyn Crebolder presented the results of using social media to increase team effectiveness among Canadian armed forces. Other presentations addressed demographics of Estonian social media users, implementation in the Comprehensive Approach, and integration of social media into command and control systems. The analysis of Estonian social media users highlighted an especially important issue:

Figure 1. Social Radar

the assumption that social media is ubiquitous and familiar to an entire population can often be false, and can lead to incorrect conclusions.

Military applications and case studies were the subject of the third session, especially the use of social media by military personnel and the potential challenges that such use can create for the military community. One of the main conclusions that came out of this session was that senior leaders, especially those unfamiliar with social media, need to develop and maintain proper guidance for themselves and their units. Other presentations in this session focused on specific examples of social media analysis, covering topics such as the use of Twitter during crises, team building through virtual social networking, and monitoring extremist groups through social media content.

The final sessions addressed social media tools and analytical techniques. A wide variety of tools and technologies were described in the fourth session, with discussions covering robust taxonomies, the use of social media by red as well as blue forces, data collection and analysis via smart phones, spatial agent-based modeling incorporating geo-tagged social media data, and deception and psychological warfare through social media. Speakers noted that analytical techniques can be used on social media data to help users gain rapid

Continued on next page

SENTIMENT ANALYSIS AND SOCIAL MEDIA IN HSCB

Continued from cover

other texts.[2] This HSCB project is using LIWC on large volumes of microblogs to generate a series of ratios across time. The research uses a mathematical algorithm to detect "breakpoints" at which emotions shift substantially. The selection of breakpoints is unbiased by global events or by any person's subjective views of global trends. Work currently underway focuses on assisting with forecasting the future direction of emotions.

Sentiment or emotion analysis tends to focus on seeing patterns across highly aggregated data on feelings and sentiments expressed by individuals. Considerable ongoing research centers on what social media data can tell us about how groups form and function. Traditional social science approaches to group identification rely on techniques such as fieldwork and surveys. **Least Squares** is leading a project that looks at how the emerging science of social network analysis, using massive corpora of social communication data, can complement traditional approaches and better support group identification. The research team is conducting three studies of social groups, using a methodology that combines traditional methods (e.g., fieldwork, surveys) triangulated against quantitative group discovery (e.g., cliques, k-shells) on social media data.

Trust is a key element in the effective functioning of groups. **Lockheed Martin** hopes to use crowd sourcing techniques and algorithms applied to open source data to assess the veracity of that data. Specifically, Lockheed Martin is researching techniques for automated assessment of the trustworthiness of open source data, as an improvement on manual information veracity assessment, which is slow, costly, and sometimes not possible. The team is applying its technologies to blogs and news event reports.

One major concern with use of social media to support operations is the potential for bias or deception to introduce distortion. **Raytheon/BBN** has a strong history of applied natural language processing, and is leveraging that work to train a statistical model that takes the next step in sentiment analysis. Rather than simply evaluating binary sentiment as traditional techniques do, their research aims to capture both the direction and magnitude of any given sentiment. As shown in the illustration below, this project will use rich syntactic and semantic context features to capture indications of bias in both English and Arabic. Semi-supervised training and clustering over retrieved passages will maximize accuracy with limited training data. This work is intended to provide users with a retrieval and bias-ranking capability for data collection and dissemination systems. It will support analysts' efforts to track and measure perception on a topic across time and within regions.

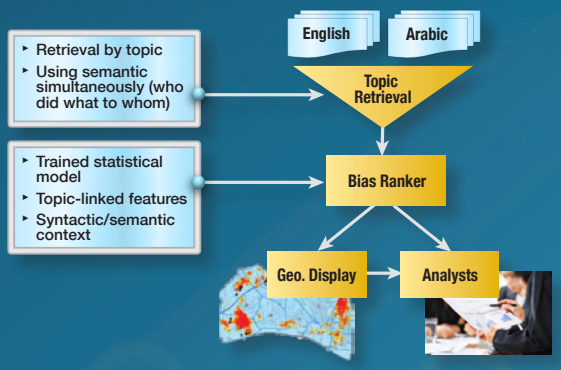

Figure 1. Sentiment Analysis and Social Media in HSCB

Work funded through the HSCB Modeling Program draws from a wide variety of data sources, including survey data, databases produced and published by government, non-government organizations, global news, and more. At present, social media is a relatively small but critical area of focus for the HSCB Program. Social media appear to offer a rich source of publicly available data, but further research is needed to fully demonstrate its operational value for the Department of Defense and its mission partners.

2 See Pennebaker, Booth, & Francis, (2007), Linguistic Inquiry and Word Count: LIWC 2007 – Operator's manual; and Pennebaker, Chung, Ireland, Gonzales, & Booth (2007), The development and psychometric properties of LIWC 2007. Both at LIWC.net.

Social Media Specialists Meeting Summary

Continued from page 7

situational awareness, discover emergent ideas, and forecast breakpoints in the moods and attitudes of social media users. These sessions highlighted a wide variety of subject matter, as presenters drew on data collected from Estonia, Nigeria, and virtual worlds such as Second Life.

As the conference presentations clearly demonstrated, analysis of social media is not a simple task. Decision makers must remain aware of the risks inherent in such endeavors, including deception and the possible damage and destruction of information and data. However, given the environments where most of military forces operate today, in which perception and opinion are as important as kinetic operations and military strength, social media analysis provides great opportunities for understanding, awareness, and agility in response, often at a reduced cost in both time and resources.

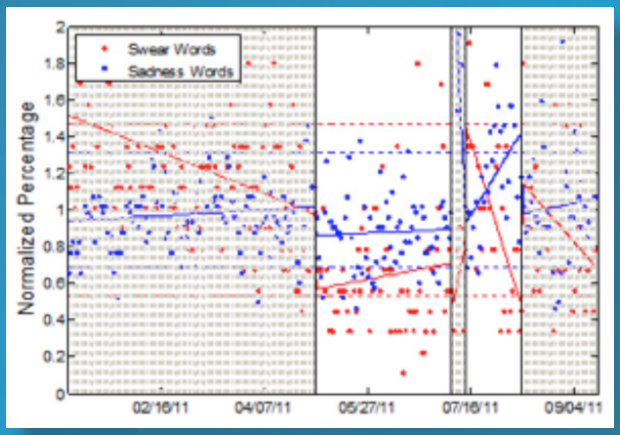

Figure 2. LICW: Breakpoint analysis of normalized percentages of UWCs swears and sadness words in daily aggregates of Tweets, using a linear model

DataCards — Collaborative Data Discovery and Information Sharing for Sociocultural Data

By Dr. Brian Efird

Senior Research Fellow at the Center for Technology and National Security Policy

National Defense University
www.datacards.org
Email: brian.efird@ndu.edu
Tel: (202) 685-2397

1. The Problem

The lack of a central repository, authority, or even reason for collecting data that can be used to evaluate and analyze sociocultural issues confounds the ability of the Department of Defense (DoD) to address metrics for assessment of sociocultural analysis. In the best of circumstances, the types of data sources that could be used are stove-piped among a number of different organizations within the military channels of the United States, not to mention the interagency entities in a given theater, Non-Governmental Organizations (NGOs) and other private organizations, academia, and other foreign governments. Since the area of sociocultural data and information is not well established or well understood, each of these organizations tends to collect different data for its own purposes in its own way and store the data (if it is saved at all) without publicizing its existence. Furthermore, what is collected, how it is defined, where it is stored, and how its existence is communicated is highly dependent on the personality of leadership and staffs.

Generally, the most effective approach to finding data has been to use the "buddy network." That is, users access an informal social network of colleagues and friends through a series of phone calls, emails, and conversations to determine what data are available and where they may be located. If a user is lucky, a trusted agent might even be able to provide an assessment of how "good" the data are once located, though this is haphazard at best. The DataCards tool was created with this problem in mind, so that anyone seeking data could rapidly check the possible existence of specific sources of sociocultural data that might be relevant, and obtain an initial assessment as to the suitability of the data for their purposes.

2. DataCards: A Modest Step in the "Right" Direction

DataCards is a structured wiki that uses "cards" (similar to card catalog cards or baseball cards) to index and describe key details (tagged as metadata) regarding sources of sociocultural and related data. The intent is to maintain a simple, yet powerful and broadly used capability to support the whole of government and related communities, and to make it self-sustaining. For the effort to succeed, a community of interested parties, data owners, analysts, and anyone with knowledge of data would have to share their information broadly to help overcome the identified problem.

The specific objectives of DataCards include:

- Make sources of data discoverable
- Reduce search costs for data
- Provide a conduit to discover and share data sources between and among non-traditional, academic, defense, and intelligence communities

DataCards is designed to provide all users with an easy-to-use capability to contribute new information on sources of data, modify and update knowledge regarding known data sources, and share knowledge about other sources that may be of interest. Data sources identified include a range of data "types," including quantitative sources, imagery, multimedia repositories, qualitative reporting, as well as other formats. So far DataCards has focused on unclassified data sources, often those from non-traditional channels or contained in stove-piped organizations. However, versions of DataCards are also available on the NIPRNet and SIPRNet.

DataCards uses Semantic MediaWiki (much like Wikipedia and Intellipedia) to create a collaborative platform for the user. Access is controlled by a National Security Agency (NSA)-approved firewall and by vetting users, who are provided with usernames and passwords for access.

3. Way Ahead

The DataCards capability is continually being refined and updated - both in functionality and content. DataCards originally focused on Afghanistan in support of the International Security Assistance Force (ISAF) and CENTCOM, but has since expanded worldwide. The capability has been deployed as part of the efforts to counter transnational criminal organizations, specifically to help in identification of sources across the whole of US government and its partners. DataCards is directly supporting NORTHCOM, as well as the Departments of Homeland Security and Justice.

Under development are specific capabilities to make the DataCards platform become more useful for users. These include:

- Link Checker to check for dead links on existing cards
- Data Harvester to harvest data from already identified sources, and then improve search
- Data Card Crawler to identify new data sources and harvest potential data
- Hard Drive Crawler to identify and characterize potential data of interest on a hard drive "brick"
- DataCards Explorer for faster and more accurate identification of useful data cards within the current framework
- Evaluation of Socio-Cultural Data, based on a set of transparent and reasonably objective criteria aggregated into an Amazon-like "5 star system" for evaluation of data sources

W-ICEWS Transition to the ISPAN GAP CIE Program

continued from page 5

the next two years. Currently, W-ICEWS is being used across the J-Codes at USSOUTHCOM, at the J8 at USPACOM, at the J9 at USSTRATCOM, and at the USSTRATCOM Center for Combatting Weapons of Mass Destruction (SCC WMD). ISPAN GAP CIE provides a capability to support Joint Operations Planning through a common framework that meets the requirements of all Combatant Commands. The Office of the Assistant Secretary of Defense for AT&L serves as the milestone decision authority. This is the first transition of an HSCB Program capability to a program of record.

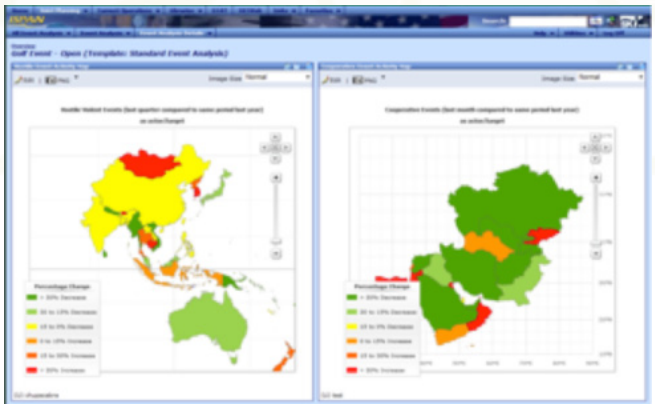

Figure 1. ISPAN

Figure 2. ISPAN

SBIR Program

By Ted Stump

The Small Business Innovative Research (SBIR) program provides the Office of the Secretary of Defense (OSD) an alternate acquisition avenue for innovative research and development by small businesses. SBIR projects in the human, social, culture, behavior (HSCB) domain have been primarily focused on specific topic areas where technology gaps in OSD's HSCB Program have been identified. These topic areas are ideally suited for small, innovative companies that may or may not have any experience in dealing with the Department of Defense (DoD). Recent HSCB SBIR investments include work in the "Development of Social, Cultural and Communication Skills (Learning the Human Terrain)," "Network Construction and Analysis Tools," "Visualization Methods and Tools for Human, Social, Cultural, and Behavioral Models," and "Analytical Tools for Local Economic Analysis" areas.

The "Development of Social, Cultural and Communication Skills (Learning the Human Terrain)" area seeks to develop a serious game that efficiently teaches the social, cultural, and communication skills necessary for Marines and soldiers to function in the irregular warfare and stability and reconstruction environments. The game is intended to provide capabilities for real-time cognitive state assessment, automatically tailoring the challenges of the game to the individual's current knowledge and skill level. Aptima, Inc. has been awarded SBIR Phase II funding to develop and transition its Cultural Awareness for Military Operations (CAMO) project, which is envisioned as a way to provide computer-based training in cross-cultural skills to warfighters prior to deployment. It is based upon a constructivist learning strategy that emphasizes the discovery of fundamental aspects of the target domain in training images that invite comparison/contrast. This approach is highly conducive to the training of cross-cultural skills by teaching "how to think" rather than forcing users to memorize rules and procedures. The Marine Corps Center for Advanced Operational Culture Learning (CAOCL) will serve as the operational sponsor

Continued on next page

SBIR Program

continued from previous page

for CAMO. CAOCL has expressed an interest in using CAMO to meet its requirement for pre-deployment training in "Operational Culture" as defined in Operational Culture for the Warfighter: Principles and Applications.

Figure 1. CAMO Module for Predeployment Training in "Operational Culture"

"Automated Network Construction" seeks to develop tools that automate/speed the construction of dynamic meta-data on social networks to aid analysis of the interdependencies of complex social networks. These tools should help the analyst determine network relationships through automation and "wizard" help/dialog systems. Additionally, the reduction of construction time should allow for more detailed actionable intelligence and/or faster delivery to supported units. CarleyTech will soon begin SBIR Phase II pursuing a human + machine approach that involves text mining, visual data entry, and cleaning aids to improve multisource data collection, fusion and cleaning. CarleyTech's approach is aimed at enabling the rapid construction of network data while staying attuned to the level and source of uncertainty and to the mission focus. CarleyTech will utilize text mining, machine learning, standard thesauri, and import wizards to rapidly extract from raw texts details on networks connecting who, what, where, how, and why attributes of the nodes and links. The project will also apply statistical techniques for inferring missing data, and will develop network assessment capabilities to assess the impact of coding choices on resultant network assessments. CarleyTech will employ mission-driven data model refinement and extraction, uncertainty tracking, general, Diplomatic, Informational, Military, and Economic/Political, Military, Economic, Social, Infrastructure and Information (DIME/PMESII) and adversarial ontologies, topic modeling for thesaurus construction, multi-source data entry, network fusion, and linkages to novel collection technology such as crowd sourcing and cell phones. Core benefits of the proposed technology are customizable rapid data entry from diverse individuals, uncertainty tracking, novel data collection methods, and the ability to make use of network analytics even with small hand-held devices. Key features include automated guidance for ontology improvements and intelligence collection efforts.

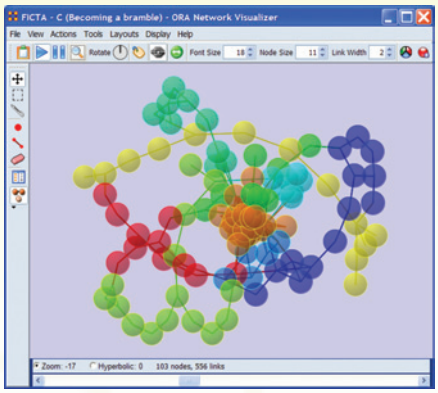

Figure 2. ORA Network Visualizer

"Dynamic Meta-Network Analysis" seeks to develop a series of tools, techniques, and models for collecting data on and reasoning about covert networks in the face of incomplete or uncertain information. Dynamic meta-network analysis is an emergent scientific field that extends the power of thinking about networks to the realm of large-scale, dynamic systems with multiple co-evolving networks involving cognitively realistic agents. This activity includes the creation of methods and tools to visualize dynamic meta-data on social networks and to assist in the analysis of the interdependencies

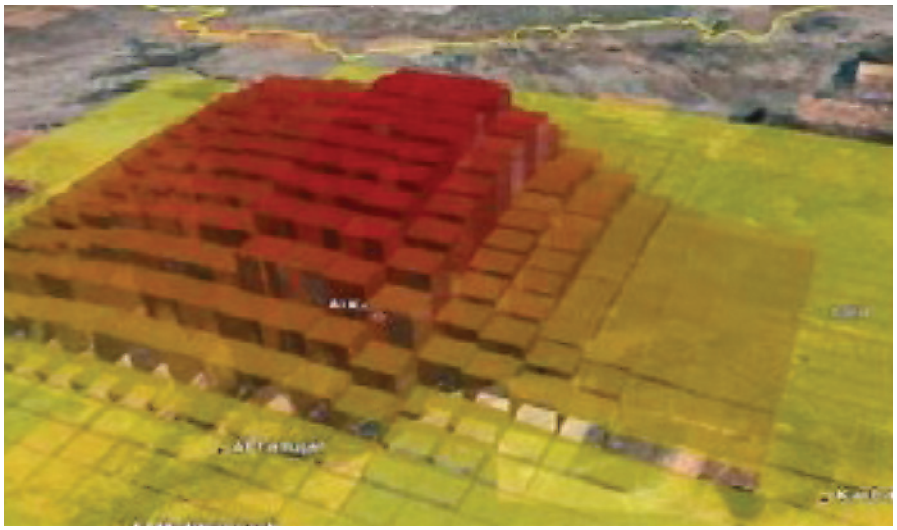

Figure 3. DAVINCI Heat Map Overlay

Continued on next page

SBIR Program

continued from previous page

of complex social networks. With a meta-network perspective, a set of networks connecting various entities such as people, groups, knowledge, resources, events, or tasks is combined to describe and predict system behavior.

Decisive Analytics Corporation has been awarded a SBIR Phase II award in this topic area for the Structured Probabilistic Attributes, Relationships and Knowledge (SPARK) project. The SPARK method of generating social networks automatically builds social network models from unstructured text and captures fine details about relationships expressed in the source text. The models can then be exported to a variety of common formats. The SPARK Phase II system uses dynamic structure discovery algorithms to expose the fundamental patterns and trends that govern network structure and behavior. SPARK's intuitive and flexible user interface will be designed, evaluated, and refined in the operational setting of the U.S. Army Intelligence and Security Command (INSCOM) Futures Lab.

Analysis of HSCB model output introduces significant challenges in data visualization and requires new approaches for representing complex, dynamic data sets. The "Visualization Methods and Tools for Human, Social, Cultural, and Behavioral Models" topic seeks to develop visualization and related tools for translating HSCB model outputs to decision-support products. CHI Systems has been awarded SBIR Phase II funds to develop the Data Analysis Visualization and INtegrated framework for Cultural Intelligence (DAVINCI). In this effort, CHI Systems will create a framework that will display militarily relevant entities, events, and courses of action (COAs) in a cultural context, allowing users to gain familiarity with socio cultural conditions in a population. The bulk of the DAVINCI effort will entail the construction of new visualization capabilities to enhance users' understanding of HSCB model output, and new controls to support operational users as well as analysts. It will also provide a set of mappings between a sub-set of DIME behaviors and the information elements needed to support those behaviors.

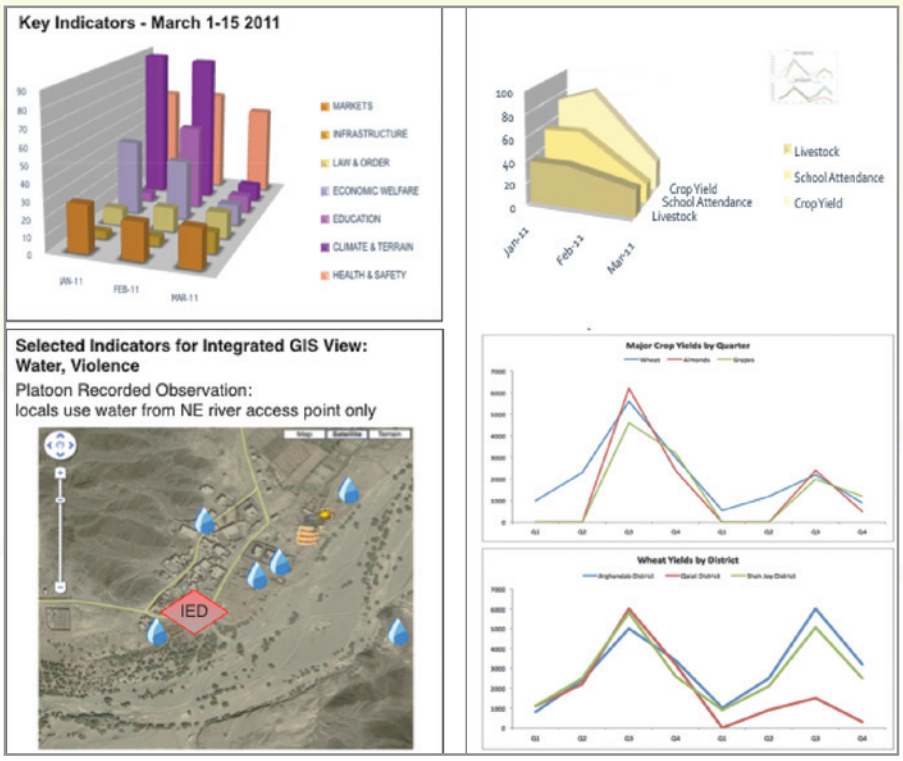

Figure 4. Key Economic Indicator Display

The "Analytical Tools for Local Economic Analysis" area seeks to provide an analytical tool that will enable Marines/soldiers at the company or lower levels to assess local economic conditions and develop COAs to improve the economy. To understand and influence the changing local economic situation, Marines and soldiers need tools that allow them to collect and analyze local economic data from both formal and informal (black) markets in order to inform their COAs as they relate to supporting local economic development. Altusys has been selected for a SBIR Phase II award to develop a tool that facilitates informed engagement within local environments in conflict areas, enabling the collection of high-quality local information, the extraction of valuable insights, and enhanced decision making in counter-terrorism and counter-insurgency situations. These techniques provide methods and tools for turning routine local observations and interactions into insights and action, even in areas where information is not readily accessible. To create the tool, Altusys will lead a multidisciplinary team of software engineers, user researchers, experts in economic analysis, modeling, and development, and experts in human data collection in dynamic areas.

Social Media Technology Research and Commercialization

By Dr. Jennifer Mathieu

The landscape of social media sites is constantly shifting. New web sites and application services appear, others disappear, and the popularity of sites and services can change very quickly. To remain up to date with the ever-changing environment of social media sites, services, and discussions, enterprise-level systems must be kept current with the latest data sources and services that allow for monitoring and analysis of social media. Modular, service-oriented architecture environments help analysts, developers, and technologists to understand the social landscape and to apply the information that can be gleaned from it in their decision making. Research and commercial environments provide valuable experience and design maturity before programs transition tools to use in the field.

	Gnip	Topsy	PeopleBrowser	MorningSide Analytics
Archived Data		X	X	
Streaming Data	X			
Links to Data				X
Limited Analytic Services		X	X	X
Advanced Analytic Services		X	X	X
API Access	X	X	X	X
Dashboard Access		X		
Report Generation		X	X	X

To support non-kinetic engagement with local populations, the Department of Defense and other organizations must understand the conversations taking place in social media. Social media can be organized into six types or modalities as illustrated in Figure 1, which indicates examples of social media applications (inspired by existing representations).[1,2] The Human Social Culture Behavior (HSCB) Modeling Program is supporting a number of projects that conduct research on the value of social media, such as sentiment analysis of blogs and microblogs (Twitter) for understanding sociocultural behavior (discussed in this newsletter in the article titled "Sentiment Analysis and Social Media in HSCB"). The other modalities that may be of interest to the Department of Defense in the near term are content communities (YouTube, flickr, instagram), social networking sites (Facebook, Google+, LinkedIn, foursquare), and news commenting sites (reddit, Mashable engadget, digg). Longer term interest may include collaborative projects (Wikipedia) and virtual worlds (World of Warcraft, Second Life, Star Wars Old Republic). Social media technology research and commercial applications must ultimately be able to harvest, store, process, and aggregate data at scale for data types that include blogs (LiveJournal, Wordpress), microblogs, picture-sharing, video blogging (vlogs), wall-postings, email, instant messaging, crowdsourcing, etc.[3]

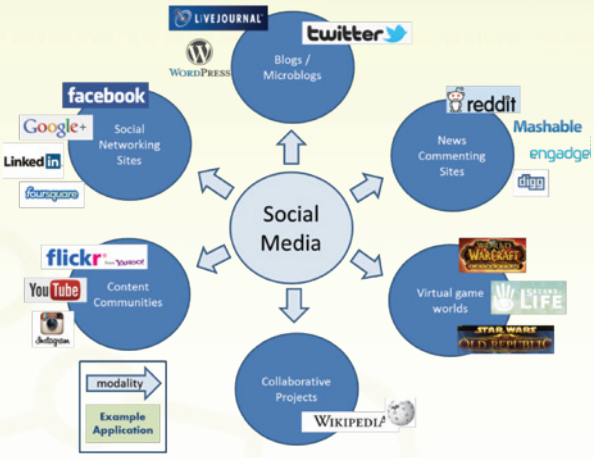

Figure 1. Social media organized into six modalities

Social networking is a subset of social media. The term refers to "creating, maintaining and interacting with one's social network using Internet technologies."[4] Academic publications in the social networking arena appear in a relatively small number of journals. The top five journals are First Monday (20 articles), Journal of Computer-Mediated Communication (18 articles), Cyber Psychology & Behavior (13 articles), Lecture Notes in Computer Science (9 articles), and Computers in Human Behaviour (8 articles)4. A more recent literature review on academic work related to influence in Twitter yielded over 10 conference papers (Association for Computing Machinery, ACM; Association for the Advancement of Artificial Intelligence, AAAI; and Institute of Electronics and Electrical Engineering, IEEE) with a distinct focus on marketing. For example, significant research efforts have centered on attempts to craft viral messages or to predict the likelihood that an individual message will be retweeted (forwarded as a tweet). Other projects attempt to quantify the influence of a user rather than of a single message, often by correlating user attributes (such as number of followers, number of replies) with message propagation.

Various commercial providers, aggregators, and analysts focus on social media data, and the HSCB Program is watching industry developments closely. Companies focused on access to Twitter, for example, include Gnip, Topsy, PeopleBrowsr, DataSift, SAS, and Infochimps. Other commercial companies such as Klout, PeerIndex, Twitalyzer, and MorningSide Analytics rate the influence of users in Twitter and/or blogs. These companies offer various services including archived data, streaming data, and/or links to data (full or partial feeds), limited analytic services (metadata, word counts), advanced analytic services (influence, sentiment, cluster analysis), Application Programming Interface (API) data access, dashboard data access, and report generation. The HSCB Program is investigating ways of transitioning enterprise-level environments such as those produced by these companies to process social media data at scale and integrate it with other data sources.

1 Kaplan, A.M. and Haelein, M. 2010. Business Horizons, 53: 59-68.
2 Cavazza, F. Social Media Landscape. URL: http://www.fredcavazza.net/2008/06/09/social-media-landscape/
3 Wikipedia. Social Media. URL: http://en.wikipedia.org/wiki/Social_media
4 Richter, D., Riemer, K., vom Brocke, J. 2011. Business & Information Systems Engineering, 2:89-101.