BlogBrain Ops: Proposal for a Semi-automatic Social Web Mining and Cyberinfluence Decision-support Tool for Info Ops Teams

Maged N. Kamel Boulos
Faculty of Health, University of Plymouth, Plymouth, Devon, UK
Email: m nkamelboulos@plymouth.ac.uk

Abstract—In this article, the author is proposing to develop BlogBrain Ops, a novel semi-automatic, Social Web cyberinfluence decision-making tool for 21st century Info Ops teams in the Social Internet age. Using visual analytics, computational social network analysis, opinion mining and sentiment analysis, captology and other methods and techniques, the tool will provide its operators with the kind of ‘Social Web Situation Awareness’ intelligence that is needed to inform and guide the design and deployment of suitable online Info Ops responses (cyberinfluence interventions) that achieve maximum impact.

Index Terms—Social Web Mining, Info Ops, Cyberinfluence, Topic-sentiment Analysis for Mass Opinion Measurement, ‘Viral’ Social Marketing, Technosocial Predictive Analytics, Visual Analytics, Captology, Blogs, Twitter, Facebook, War on Terror, Bioterrorism, Public Health Surveillance

I. INTRODUCTION

Analysing Social Web (e.g., blogs, Twitter, etc.) post aggregates in real or near-real-time can give us a good indication of the prevailing public opinion(s) of corresponding communities regarding different matters of interest [1]. It can tell us about the general public mood and where ‘the wisdom of the (online) crowds’ is pointing, acting like a measure of the psychosocial “pulse” and beat of online societies. As more and more people are getting connected these days to the Internet all over the world, online societies are rapidly becoming a good mirror of offline, conventional societies, and the Social Web is quickly becoming a strategic place of choice to reach out to them and influence them on a large scale in ways that were never possible a few decades ago, e.g., by using ‘viral’ (rapidly spreading) videos and other forms of ‘viral’ social marketing techniques [2].

Moreover, by tracking the change in Social Web post aggregates over time after some Social Web opinion-shaping intervention has been made, one can monitor and assess the effectiveness of such an intervention and tweak or retarget it as necessary.

The author is proposing to develop a semi-automatic, Social Web/networked social media (NSM) cyberinfluence decision-making tool for Info Ops purposes (BlogBrain Ops). Info Ops or ‘Information Operations’ is a military concept involving, among other things, the use and management of information technology, e.g., for social network analysis, in pursuit of a competitive advantage over an opponent. The proposed interactive tool will provide its operators with the kind of ‘Social Web Situation Awareness’ intelligence that is needed to inform and guide the design and deployment of suitable online Info Ops responses (cyberinfluence interventions) that achieve maximum impact.

The Web-based tool will use Social Web post aggregates (English-language blog posts initially, but this can be later extended to also cover micro-blog (e.g., Twitter and identi.ca) and Facebook posts and non-English-language posts) to inform Info Ops teams’ cyberinfluence intelligence gathering and tactics/manipulation in real or near-real-time. Figure 1 shows a high-level concept diagram of the proposed tool, its possible different components or main modules, and the ways these modules might interconnect with each other. An ‘analytics dashboard’, with appropriate user-configurable search wizards and trends visualisation options, will offer ‘at a glance’ near-real-time Social Web (blogosphere/NSM) trends based on suitable monitoring indicators, and will allow the tool operator to interactively identify Social Web points of interest/key players/hot foci (e.g., strong opinion-shaping hubs in the blogosphere and potential foci of online radicalisation), as well as points for potential Info Ops interventions in the blogosphere and other related NSM that are linked to it such as YouTube (e.g., strategic places where adequate ‘viral’ video response “injections” can be made by Info Ops teams for maximised impact).

The tool will draw on principles, methods and current research from social psychology, linguistic analysis (including opinion mining and sentiment analysis [3]), social informatics, computational social network analysis [4], and visual analytics [5]. Computational social network analysis involves the study of relationships among social entities such as communications among members of an online group. It is an interdisciplinary field that brings together researchers from the social and behavioural sciences, statistics, graph theory, machine learning and data mining.
Previous analyses of linkages between blogs have shown that community forming in the blogosphere is not a random process but is the result of common interests binding like-minded bloggers together [6]. Identifying and classifying blogger clusters by exploiting bloggers’ profiles and social networks will be a main feature of the proposed tool.

BlogBrain Ops will also use concepts borrowed from social media marketing such as those pioneered by Ivy Worldwide, Inc. (http://www.ivyworldwide.com/) and others, as well as Stanford’s science of captology (http://captology.stanford.edu/) for determining the most effective intervention “injection” points to achieve its goals. (Pioneered by Stanford University Persuasive Technology Lab, captology involves the design of interactive computing products such as Web-based material for the purpose of changing people’s opinions, attitudes or behaviours.)

The tool can target both the battle-front and the home-front, also informing the design of Info Ops interventions aimed at supporting the psychosocial well-being and coherence/stability of the latter during times of conflict or mass stress.

Previous work done by others has produced many interesting analyses (e.g., [7,8,9]), but none of these can be run as an ‘on-demand tool’ for Info Ops Social Web intelligence gathering and intervention in real or near-real-time. Some online tools also exist (e.g., Radian6 - http://www.radian6.com/), but they are focused mainly on marketing of brands and products and are not suitable for proper online opinion and sentiment mining and shaping in areas relevant to Info Ops.

The tool’s ‘interactive analytics dashboard’ shown in Figure 1 will feature user interfaces that rely heavily on visual analytics techniques. Visual analytics is an emerging area of research and practice aimed at leveraging the excellent capabilities of humans in terms of visual information exploration.

The main advantages of visual data exploration over automatic data mining techniques from statistics or machine learning are [10,11]:

- Visual analytics can easily deal with highly heterogeneous and noisy data; it is intuitive and requires no understanding of complex mathematical or statistical algorithms or parameters by its operators and it is invaluable when little is known about the data and the exploration goals are vague; and
- Visual analytics can tackle hard analytical problems where neither the machine nor the human alone can efficiently and effectively find a solution.

II. PROPOSED SYSTEM: BLOGBRAIN OPS—A COMPLETE AND FLEXIBLE APPLICATION TO ASSIST WITH THE DISCOVERY, CONTINUOUS MONITORING, ANALYSIS AND ALERT MANAGEMENT OF BLOGS OF SPECIFIC INTEREST

Rather than starting from scratch and to limit initial prototyping costs, a proposed first version of BlogBrain Ops (Figure 2) might utilise some of the already built components from previous tools such as BlogHarvest [6] (http://sourceforge.net/projects/blogharvest/), an Open Source blog crawler/aggregator, mining and search framework, in addition to introducing some newly developed components that would enhance the visual and textual analytical abilities of the proposed system. These latter components are expected to evolve over successive versions of the system, building on and further expanding existing blog mining and discourse parsing methods described in [1, 12-18] and implemented in some experimental systems such as OPTIMISM (Opinion Text Mining from Social Meshe - http://xldb.di.fc.ul.pt/wiki/Optimism) and Reaction (Retrieval, Extraction and Aggregation Computing Technology for Integrating and Organizing News - http://xldb.di.fc.ul.pt/wiki/Reaction).

BlogBrain Ops will be built with password-only access and will be designed to run on Info Ops’ secure intranets. It will be developed using common programming
languages and tools such as JAVA (http://www.oracle.com/technetwork/java/index.html) and/or PHP (Hypertext Preprocessor, a Web scripting language - http://php.net/). Running on an Apache Web server system (http://apache.org/) and using MySQL (a database platform supporting the Structured Query Language - http://mysql.com/) for its database, the application will be accessible via any standard HTML—HyperText Markup Language browser interface (e.g., Mozilla Firefox - http://www.mozilla.com/en-US/firefox/).

BlogBrain Ops will be initially tested on blogs written in the English language, but there will be the provision for the processing of other languages in future versions of the application, with language categorisation for the handling of multiple languages. The proposed first prototype can also be used in some initial testing of the application on blogs in non-English languages, in particular those written in Arabic. A full trial of the application on other language blogs, to identify and resolve any language handling issues, could be considered as a potential future project. Socio-cultural and linguistic-analysis differences among non-English-language users of the Social Web/NSM are very unique. Studies on Western online populations will not always be directly or fully applicable to Middle Eastern (e.g., Arabic [7])/Asian societies and battle-fronts, hence the need to further explore this research direction.

A. End User Involvement

Like any other Information and Communication Technology—ICT system/service, formal user involvement in the design, prototyping and testing phases of BlogBrain Ops will be crucial in order to ensure system success and end user acceptance of it. BlogBrain Ops should be developed in very close collaboration with Info Ops test users, who would, as a minimum, meet with system developers to discuss the project objectives once at the beginning of the development, as well as, say, one third the way through the development term to assess progress and trial an early pilot, thereby ensuring any feedback might be used to shape the design and ensure maximum usability of a more functional, feature-complete first version of the prototype application. Project developments should therefore take place in an iterative fashion, over at least two development cycles (cycle 1: early pilot; cycle 2: first version prototype); a total duration of 8-12 months should be adequate. The final output will be a fully working and tested BlogBrain Ops application (first version), plus code and database for intranet installation. Documentation should also be provided to explain the code and installation process, in addition to an end user guide. The application is expected to be intuitive and self-explanatory to use, with contextual help provided throughout; therefore it will not require a lengthy or complicated manual or user training, but simply an outline of the functionalities and instructions on how to get started.

B. Main BlogBrain Ops Components’ Description

Blog Finder: A super user would enter a small number of initial ‘seed’ blog URLs (Uniform Resource Locators—Web addresses). Those blogs would then be automatically crawled to find titles, descriptive metadata, RSS (Really Simple Syndication) feeds and blog rolls (other blogs recommended by the blog author). Blogs listed in blog rolls would then also be automatically crawled by the Blog Finder for the same purposes. A log would be kept of which blogs link to which, to be used to mark blogs with a popularity rating, and also for use in the Friend Finder component listed below.

Feed Assistant: This is a tool to assist a super user with the manual verification of the blogs that have been automatically found by the Blog Finder, before they are added to the BlogBrain Ops index for continuous monitoring. The tool would provide an administrative dashboard for a super user to scan through newly identified feeds, check them for quality before accepting or dismissing them, and then to categorise them and tag them as wished. Some manual intervention at this initial stage will be essential to stop spam and irrelevant blogs entering the system and to maximise the ability for intelligent analysis, but this tool will ensure that manual intervention is able to be completed with minimal repetition and with as much speed as possible. The Feed Assistant will also include the ability to identify feeds that are no longer working, to replace them with new feed URLs when necessary, to remove feeds, or turn off their continuous monitoring, when they are decided to be of no further use.

Feed Parser: This is an internal component that parses the indexed feeds at regular intervals as defined in the configuration. New feed items found are stored in the database. Any feeds not parsing correctly are logged and error messages are available for access via the Feed Assistant as detailed above.
creating new blogs and linking/positioning them strategically in the videos, posting well-crafted online comments at strategic places where messages and video responses using same tags as original posts and interventions can be "injected") include specially-tailored viral video actual interventions but would suggest suitable points where these influencing people and shaping public opinion, attitudes and responses.

Possible Info Ops interventions (BlogBrain Ops will not deliver the specified Topic Tracker. Once a Topic Tracker has been identified and tagged with the tracker as they arise, must be met for a blog item to be marked with that particular Topic Tracker. Once a Topic Tracker has been set up, all new blog items meeting the criteria will be identified and tagged with the tracker as they arise, available for viewing and further analysis, including frequency trends. It will be possible to set up alerts monitors (including e-mail notification) when specified, infrequently occurring Topic Trackers are triggered, or for when there is a sudden increase in activity within a specified Topic Tracker.

Feed Browser: The Feed Browser provides the ability to browse the latest updates to the blogs indexed, via different categories and tags, either by those defined at the feed level or automatically assigned at blog item level.

Search Engine: This consists of standard search engine functionality, providing the ability for advanced queries on all blog items indexed by BlogBrain Ops.

Tag Clouds and Tag Trends: Blog items are frequently tagged by the blog authors with keywords and phrases that they consider to be of particular relevance. These tags will be identified and made available via a variety of visual displays known as ‘tag clouds’ which will help the user to quickly find topics of particular importance, and to drill into those topics to find other related blogs items. In addition, a log will be kept of the frequency of use of each tag over time, so that the user will be able to view charted trends of any tag, and also to see those tags that have risen in popularity over a period of time.

Topic Trackers and Alerts: Super users will be able to set up key topics for identification and tracking within the blog items. Each Topic Tracker will be based on a query, at its simplest a key word or phrase, but it may be an advanced query that may be a combination of criteria that must be met for a blog item to be marked with that particular Topic Tracker. Once a Topic Tracker has been set up, all new blog items meeting the criteria will be identified and tagged with the tracker as they arise, available for viewing and further analysis, including frequency trends. It will be possible to set up alerts monitors (including e-mail notification) when specified, infrequently occurring Topic Trackers are triggered, or for when there is a sudden increase in activity within a specified Topic Tracker.

Friend Finder: It will be possible to run a friend analysis on any single blog, to see a visual map of who that blog is recommended or quoted by, and also which other blogs that blog author recommends or quotes. It will be possible to navigate across the map by clicking from one blog to another. Blogs analysed will be given a score of popularity and these scores will be logged for future trend tracking over time. This will provide the basis of a tool that could be further developed in a future project to identify trends in blogs rising or falling in popularity over time.

Thought Dissemination and Mood/Opinion Trackers: This is an experimental, but important, component that will attempt to identify those blogs that initiate discussion on particular topics, and those blogs that are influential in the dissemination of ideas (and thus could be candidate points for Social Web intervention “injections”1). It will also attempt to mark discussions as positive or negative through the identification of key language phrases — such functionality would be tested only on the English language, but if found to be effective, a similar approach could then be explored in other relevant languages.

III. ADDITIONAL VALUE OF PROPOSED TOOL OUTSIDE TIMES OF CONVENTIONAL CONFLICTS / AT TIMES OF PEACE

These days most people, especially youths and teenagers, are connected 24/7 via their Internet-enabled mobile phones, which are also frequently used to access and post to various Social Web applications such as blogs, Twitter and Facebook. The proposed tool can help not only Info Ops personnel, but also health communicators at times of peace, gain proper ‘Social Web Situational Awareness’ and understanding to help them tailor, target and monitor suitable Social Web interventions aimed at ensuring the psychosocial well-being, stability and support of the home-front at times of mass stress. Such measures facilitated by the proposed tool can be easily extended/applied outside times of conventional armed conflicts to help shape public opinion, attitudes and responses, e.g., at times of natural disasters, massive disease outbreaks/pandemics, and even during the on-going war on terror (man-made disasters, including bio-terrorist events such as the US anthrax outbreak of 2001). Communicating with the masses and gaining their support (cyberinfluence), with the aim of mitigating such mass stress events, is an art, and informatics is all about the proper and effective communication of information that people can readily put into practice. Using the Social Web and ‘viral’ communication techniques to reach out to people and send them ‘stay calm’ and ‘don’t panic’ messages, as well as telling them the appropriate things to do and how to be watchful, etc., all fall under the umbrella of this ‘psychosocial well-being’ and ‘Social Web Situational Awareness’ that BlogBrain Ops is targeting.

IV. FINAL NOTES

A more advanced prototype of BlogBrain Ops can be further developed (when/if resources become available) to realise the full vision shown in Figure 1. At the time of writing (March 2011), funding has yet to be secured to progress with the current proposal, start developing the initial prototype, and report actual screenshots of the tool interfaces and real-world results of its use. However, because of the importance of the underpinning concepts, the author thought it would be a much desirable service to the research community to disseminate the proposal at its current stage and open up the author’s vision for debate and readers’ suggestions in this brief communication article. The current proposal forms part of the author’s Social Web/NSM research agenda that he proposed back in 2010 [19].

1 Possible Info Ops interventions (BlogBrain Ops will not deliver the actual interventions but would suggest suitable points where these interventions can be “injected”) include specially-tailored viral video messages and video responses using same tags as original posts and videos, posting well-crafted online comments at strategic places where allowed, suitable banner ads/AdWords campaigns on target blogs, creating new blogs and linking/positioning them strategically in the corresponding part of the blogosphere, etc. all of which build on captology and social marketing strategies and techniques, and aim at influencing people and shaping public opinion, attitudes and responses.
Note: All URLs in this manuscript have been last accessed on June 7, 2011.


Maged N. Kamel Boulos is Associate Professor in Health Informatics at the University of Plymouth, Devon, UK. He previously worked as Lecturer in Health Informatics at the University of Bath, UK, where he was instrumental in developing the online MSc programme in Healthcare Informatics, and worked before that at City University, London, UK. As well as his medical degree and Master in Dermatology, he holds a Master in Medical Informatics (2000) from King’s College, University of London, and a PhD in Measurement and Information in Medicine (2002) from City University, London, both in UK. Maged teaches and has over 90 publications on a specialist range of medical and public health informatics topics, including telehealthcare and Semantic Web/knowledge management applications. He is L3 Member of UKCHIP, the UK Council for Health Informatics Professions, and Senior Member of IEEE (Institute of Electrical and Electronics Engineers) and IEEE EMBS (Engineering in Medicine and Biology Society). He is the founder and Editor-in-Chief of the Open Access, MEDLINE-indexed International Journal of Health Geographics (http://ij-healthgeographics.com). He has undertaken consultancies/commissioned reviews and has been Principal Investigator on various activities funded by national and international bodies, including London School of Economics; UK TSB (BERR/DTI) KTP Programme; UK Arts & Humanities Research Council; KFSHRC (King Faisal Specialist Hospital and Research Centre), Jeddah; Qatar National Research Fund; Public Health Agency of Canada (funding one of his PhD students); World Health Organization EMRO and EURO; and European Commission (CAALYX and eCAALYX EU-funded eHealth projects).
Aims and Scope

Journal of Emerging Technologies in Web Intelligence (JETWI, ISSN 1798-0461) is a peer reviewed and indexed international journal, aims at gathering the latest advances of various topics in web intelligence and reporting how organizations can gain competitive advantages by applying the different emergent techniques in the real-world scenarios. Papers and studies which couple the intelligence techniques and theories with specific web technology problems are mainly targeted. Survey and tutorial articles that emphasize the research and application of web intelligence in a particular domain are also welcomed. These areas include, but are not limited to, the following:

- Web 3.0
- Enterprise Mashup
- Ambient Intelligence (AmI)
- Situational Applications
- Emerging Web-based Systems
- Ambient Awareness
- Ambient and Ubiquitous Learning
- Ambient Assisted Living
- Telepresence
- Lifelong Integrated Learning
- Smart Environments
- Web 2.0 and Social intelligence
- Context Aware Ubiquitous Computing
- Intelligent Brokers and Mediators
- Web Mining and Farming
- Wisdom Web
- Web Security
- Web Information Filtering and Access Control Models
- Web Services and Semantic Web
- Human-Web Interaction
- Web Technologies and Protocols
- Web Agents and Agent-based Systems
- Agent Self-organization, Learning, and Adaptation
- Agent-based Knowledge Discovery
- Agent-mediated Markets
- Knowledge Grid and Grid intelligence
- Knowledge Management, Networks, and Communities
- Agent Infrastructure and Architecture
- Agent-mediated Markets
- Cooperative Problem Solving
- Distributed Intelligence and Emergent Behavior
- Information Ecology
- Mediators and Middlewares
- Granular Computing for the Web
- Ontology Engineering
- Personalization Techniques
- Semantic Web
- Web based Support Systems
- Web based Information Retrieval Support Systems
- Web Services, Services Discovery & Composition
- Ubiquitous Imaging and Multimedia
- Wearable, Wireless and Mobile e-interfacing
- E-Applications
- Cloud Computing
- Web-Oriented Architectures

Special Issue Guidelines

Special issues feature specifically aimed and targeted topics of interest contributed by authors responding to a particular Call for Papers or by invitation, edited by guest editor(s). We encourage you to submit proposals for creating special issues in areas that are of interest to the Journal. Preference will be given to proposals that cover some unique aspect of the technology and ones that include subjects that are timely and useful to the readers of the Journal. A Special Issue is typically made of 10 to 15 papers, with each paper 8 to 12 pages of length.

The following information should be included as part of the proposal:
- Proposed title for the Special Issue
- Description of the topic area to be focused upon and justification
- Review process for the selection and rejection of papers.
- Name, contact, position, affiliation, and biography of the Guest Editor(s)
- List of potential reviewers
- Potential authors to the issue
- Tentative time-table for the call for papers and reviews

If a proposal is accepted, the guest editor will be responsible for:
- Preparing the “Call for Papers” to be included on the Journal’s Web site.
- Distribution of the Call for Papers broadly to various mailing lists and sites.
- Getting submissions, arranging review process, making decisions, and carrying out all correspondence with the authors. Authors should be informed the Instructions for Authors.
- Providing us the completed and approved final versions of the papers formatted in the Journal’s style, together with all authors’ contact information.
- Writing a one- or two-page introductory editorial to be published in the Special Issue.

Special Issue for a Conference/Workshop

A special issue for a Conference/Workshop is usually released in association with the committee members of the Conference/Workshop like general chairs and/or program chairs who are appointed as the Guest Editors of the Special Issue. Special Issue for a Conference/Workshop is typically made of 10 to 15 papers, with each paper 8 to 12 pages of length.

Guest Editors are involved in the following steps in guest-editing a Special Issue based on a Conference/Workshop:
- Selecting a Title for the Special Issue, e.g. “Special Issue: Selected Best Papers of XYZ Conference”.
- Sending us a formal “Letter of Intent” for the Special Issue.
- Creating a “Call for Papers” for the Special Issue, posting it on the conference web site, and publicizing it to the conference attendees.
- Information about the Journal and Academy Publisher can be included in the Call for Papers.
- Establishing criteria for paper selection/rejections. The papers can be nominated based on multiple criteria, e.g. rank in review process plus the evaluation from the Session Chairs and the feedback from the Conference attendees.
- Selecting and inviting submissions, arranging review process, making decisions, and carrying out all correspondence with the authors. Authors should be informed the Author Instructions. Usually, the Proceedings manuscripts should be expanded and enhanced.
- Providing us the completed and approved final versions of the papers formatted in the Journal’s style, together with all authors’ contact information.
- Writing a one- or two-page introductory editorial to be published in the Special Issue.

More information is available on the web site at http://www.academypublisher.com/jetwi/.