Urban Operational Experimentation hosted by the National Urban Security Technology Laboratory (NUSTL)

RepKnight Report

December 2015

Authors:
Bhargav Patel, NUSTL

Kris Dooley, NUSTL

NUSTL@hq.dhs.gov

Homeland Security
Science and Technology

Executive Summary

On July 28, 2015 at the New York City Emergency Management's Emergency Operations Center in Brooklyn, New York, the National Urban Security Technology Laboratory (NUSTL) hosted the Department of Homeland Security (DHS) Science and Technology Directorate's (S&T) Urban Operational Experimentation (OpEx), which conducted experiments with RepKnight. This event included first responders from New York, New Jersey, Washington, D.C., and California.

RepKnight is an open-source data platform that performs real-time monitoring and analysis of social media using keywords, trends and sentiment. It is intended to bridge a technology gap in situational awareness by distilling specific knowledge concerning threats, hazards and conditions both passive and active in real-time.

Participating first responders and observers were provided a brief overview of the program and then engaged in an abbreviated training session where representatives from RepKnight explained the software's basic functions and features. This software was designed for the intelligence community, targeting an intelligence analyst as its primary user. After the demonstration and training session, participants created their own projects and searches. Each participant selected search topics; however, the OpEx planning team and RepKnight representatives provided suggestions that might yield more results over the short span of time these searches were set up to run. When a real-time search is initiated, it does not look into archival social media posts. To get useful and relevant results for later, same-day review for this experiment, it was imperative that the search parameters were wide enough and popular enough to yield results within hours. Participants then conducted a feature-based exploration of pre-compiled data from searches created by the experimentation director and RepKnight representatives. For example, in summer 2015, a building collapsed in Brooklyn, New York and that event was used to pre-compile social media results that discussed it.

The evaluators and observers present at the OpEx believed that RepKnight was a useful tool with many unique and interesting features and that it was a partial solution to the type of software tool they desired. Some key features of note were network diagrams that provide visualization into social networks and which people might have more influence amongst them. As participants explored the capabilities of RepKnight, they began to express a desire for a tool that operated more as an autonomous predictive tool. This initiated substantive dialogue that highlighted the constraints of pulling very large amounts of data to process and the financial cost of using large amounts of bandwidth for both a vendor and an organization.

It was apparent to participants that the full scope of functionality and nuances RepKnight provides requires more training and hands-on experience than was allotted for during this experimentation event. RepKnight representatives recommend a two- or three-day course for a trained intelligence analyst; participants from this experiment were only given one hour. Participants were quickly able to share use-case scenarios where a tool like RepKnight would be useful, such as identifying networks of people, and scenarios where RepKnight still has limitations, such as predicting the anomalies with high impact and low frequency or the "unknown unknown." Emergency management participants were focused on a tool whose search parameters can be very broad and places a high focus on studying the rates of change of everything that could be captured (the first unknown) as a means for discovering an impending event (the second unknown). RepKnight is capable of doing this; however, it was not designed with this as a main task for the software to complete. One of the reasons is the end user agency would be faced with very large internet data usage charges for downloading and processing the amount of data that is associated with this type of activity.

Table of Contents

1   Introduction

RepKnight is an open-source data platform that performs real-time monitoring and analysis of social media for keywords, trends and sentiment. The developer is a company by the same name, RepKnight, and the product is marketed in the United States by ADI Technologies, Inc.

RepKnight is a web browser-based system that needs a high speed internet connection[1] to function optimally. Users can create "Projects" that will contain many different customized search parameters known as "tags." These search parameters, unlike most conventional search engines, require Boolean logic-based inputs to properly function and create useful output search results.

On July 28, 2015, responders experimented with RepKnight during the Department of Homeland Security (DHS) Science and Technology Directorate (S&T) Urban Operational Experimentation (OpEx) event, hosted by the National Urban Security Technology Laboratory (NUSTL). Subject matter experts from the New York City Fire Department, New York City Emergency Management, New York City Police Department and Port Authority of New York and New Jersey supported the OpEx by selecting the technologies working with NUSTL scientists to plan the experimentation scenarios and arrange test venues. Responders from these agencies and members of the First Responder Resource Group experimented with the technologies and provided feedback and observations. Table 1 lists all of the technologies included in this event. Technologies assessed during this event were selected with input from local first responder agencies and met capability gaps identified in the Project Responder 4 National Technology Plan for Emergency Response to Catastrophic Incidents (1).

Table 1. Technologies Included in OpEx 2015

Product Name Manufacturer	Description
Situational Head Up Display Avon Protection Systems	Micro liquid crystal display (also known as LCD) with full color widescreen layout built into face shield
Tridion™-9 PerkinElmer	Portable Gas Chromatography/Mass Spectrometry (GC/MS) system that provides identification of volatile and semi-volatile organic hazards in the field in less than three minutes
BioFlash-E Biological Identifier PathSensors	Portable and rapid aerosol sample collection and identification of up to 16 biological threat agents
Fido B2 IBAC FLIR	Networked array of portable biosensors
Internet of Things for First Responders BAE Systems	Networked sensors that use a long-range wireless protocol capable of concrete penetration to send signals through a network aggregator
Knight Robot/HAZPROBE WM Robots	All-terrain robot with a manipulator arm, cameras, and a boring and inspecting device that can drill through walls for bomb tech personnel to inspect suspect abandoned vehicles or objects
RepKnight ADI Technologies	Monitors and analyzes social media with geolocation feature
X-Ray Scanning Rover Smart Imaging Systems	An x-ray scanner integrated into a custom-built robot that is designed to rapidly screen suspicious left-behind bags or parcels on the ground

---

[1] Internet access provided by a network of servers that transfers data via high-speed cable, satellite and wireless connections.

1.1   Purpose

The purpose of the RepKnight operational experimentation was to allow first responders to engage with the new technology and provide feedback directly to vendors and DHS S&T officials. The results were documented in this report and will be shared with the national first responder community. By bringing together law enforcement, first responders and product developers, design changes can be effected early; responders can learn about emerging technologies to enhance mission capabilities; and S&T can gain a better understanding of responder needs and gaps to guide future homeland security investments.

1.2   Objective

This experimentation allowed responders to observe the use of RepKnight in an operational setting, and then offer feedback and suggestions to the developers that can be used to enhance the product capabilities and usability for first responder operations.

1.3   Responder Capability Need

The ability to incorporate information from multiple traditional and nontraditional sources (i.e. crowdsourcing and social media) into incident command operations is one of the situational awareness capability needs consistently identified by first responders participating in the Project Responder (1) studies. RepKnight offers a technology solution that could address this capability need.

1.4   Prototype Description

 RepKnight is an open-source data platform that performs real-time monitoring and analysis of social media for keywords, trends, and sentiment.

1.4.1   Real-Time Stream Analysis

RepKnight is able to aggregate, analyze, and visualize large amounts of data to support decision making. All of the RepKnight Data Analytics Software (RK-DAS) algorithms provide stream-based analysis of the incoming data from diverse open sources. The underlying infrastructure ensures that during extreme bursts in data volumes, RepKnight can continue to provide analysis to the front-line user within ten seconds.

1.4.2   Customizable, Keyword Search Features

RepKnight users have flexibility over keyword or search strings that they may wish to collate. It provides users a tool that will analyze raw data and highlight trends. The advanced search option allows analysts to refine areas of interest with Boolean search strings and enables searches focused on factors such as:

- Geo-location;
- Event;
- Source;
- Date and time;
- Author;
- Sentiment;
- Links; and
- Title.

All RepKnight algorithms are language-agnostic, and data capture supports all languages using the UTF-8 character set.

1.4.3 Sentiment Analysis

A fundamental element of effective open source intelligence is how language is used by the community of interest. RepKnight uses a proprietary, lexicon-based, sentiment engine and the company works with end users to customize the lexicon to the users' areas of interest. Current active lexicons include English, Arabic and Malay. Ongoing research enhances the search engine's ability to deal with mixed sentiment and filter situations that contain a mixture of positive and negative words in order to screen for "false positives."

1.4.4 Trend Analysis

The RepKnight Trend Analyzer algorithm tracks words and phrases to identify those that are "trending," i.e., becoming more popular than usual. Practically, this means that breaking news incidents are highlighted to a decision maker before they become widely known, giving teams time to organize, plan and respond.

1.4.5 Customizable by Audience Segmentation

RepKnight provides audience segmentation by age, gender and geo-location. This allows the end-user of RepKnight to categorize the users who are generating social media posts into demographics that can aid in further analysis.

1.4.6 Customizable by Geographic Segmentation

RepKnight recommends the use of actual location as provided in the open source environment; however, it provides the functionality to search based upon inferred location. This means it can use known data, such as previously known locations of an individual social media poster, to infer where that user is likely submitting posts from.

1.4.7 Qualitative, Data Visual Analytics and Representations (Heat Maps, Charts, Graphs)

RepMaps provides visualization of the analyzed data using word clouds, line graphs, pie charts and heat maps, all of which are interactive and automatically refreshed to incorporate real-time data.

1.4.8 Access to Historical Twitter Data

Searches can be created to capture historical data, within specific date and time ranges (see Figure 1). Additionally, real-time searches can be paused so that data will still appear in visual and predictive analytics.

Figure 1- Twitter Activity for a Set of Search Parameters as a Function of Time

1.4.9 Influencer Identification

Social media accounts/users are categorized into seven types: hotspot, seeker, facilitator, thinker, amplifier, commentator and contributor. Categorization is based on their activity, the attention their comments received, their number of followers and the number they follow, among other factors. These categories are color-coded and are a measure of how much influence the user has.

1.4.10 Standard Web Browser Access with Login Credentials

Administrators can create login credentials for all users through standard web browser access. For enhanced security, three incorrect login attempts will disable the user account, requiring the administrator to re-enable the account.

1.4.11 Compatibility with Internet Explorer 8 and Above

RK-DAS is explicitly compatible with Internet Explorer 8 and above. Additionally, any web browser that supports Javascript language can run RK-DAS, so this includes, but is not limited to:

- Google Chrome;
- Firefox version 4 and above; and
- Safari OSX 8 Snow Leopard and above.

1.4.12 Customizable Functionality to Send Notifications to Users

Alerts are user-configurable. Notifications can be sent to users through in-application notifications or to designated persons by email and text message (see Figure 2).

Figure 2 - Configurable Functionality

1.4.13 Customizable Functionality to Analyze Data Over a Given Period of Time

The RK-DAS filtering system means all analysis can be conducted over a user-configured timeframe. Figure 3, below, shows the RepKnight Dashboard screen, which displays data traffic over an adjustable timeframe.
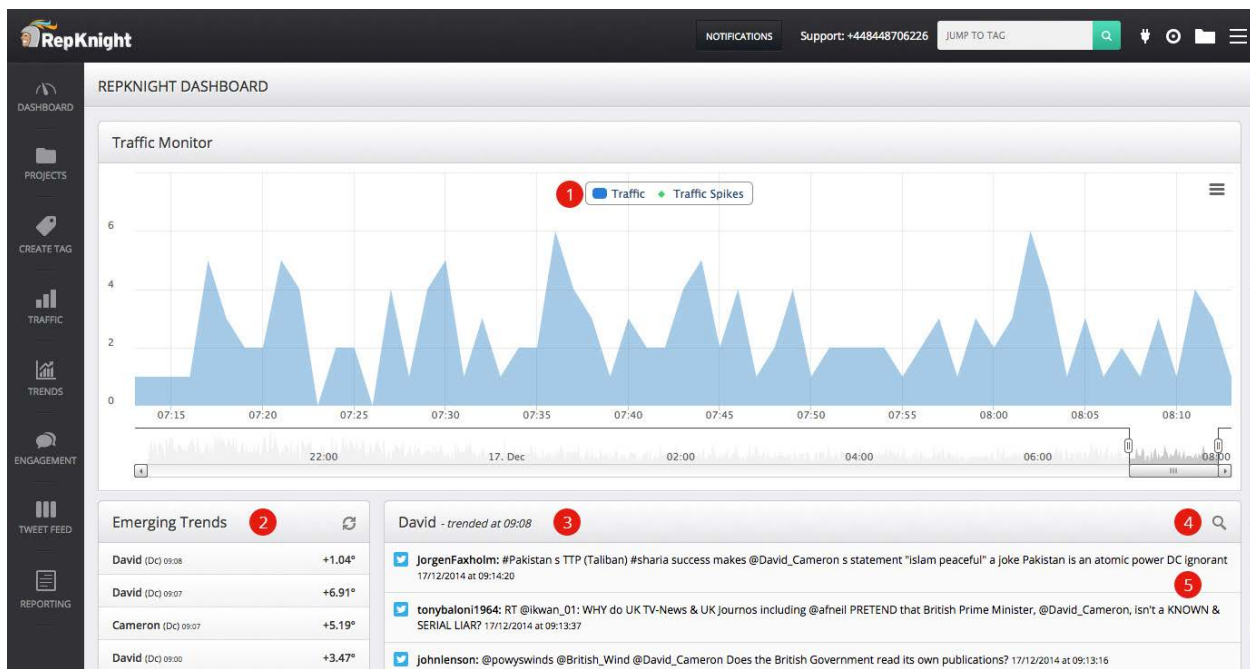
Figure 3 - Sample RepKnight Dashboard displaying traffic monitoring and emerging trends based on search parameters: (1) character displays when spikes have occurred in results; (2) emerging trends; (3) messages containing the word that is trending; (4) all the results returned from the trend and (5) returned results containing the trend.

2 Experimentation Design

A detailed description of the experimentation design can be found in the Experimentation Plan RepKnight (2). The experimentation scenarios were developed with input from first responders, RepKnight product managers and the DHS NUSTL experimentation director for RepKnight to simulate real events and incidents that would demonstrate an expansive level of RepKnight's capability.

2.1 Event Design

This event convened a group of first responders, including law enforcement and intelligence analysts, to experiment with this technology in a simulated operational setting and provide feedback. Participants were asked to input search parameters of interest to them for a real-time media search. It was requested that the search terms should be popular enough that they can populate a relatively large data set of open-source, user generated responses and results. Once these parameters were established, they were set to run, allowing them to collect data over time from various sources. Participants also were able to select which sources they sought this information from (i.e., Twitter and Instagram, only Flickr or All sources). While these custom searches ran, participants were then directed to explore the results of pre-compiled data on several historic events, including the collapse of a four-story building in Brooklyn, New York, and instances of flash flooding.

2.2 Summary of the Operational Experimentation

On July 28, 2015, ten responders with various backgrounds and experiences gathered at the New York City Emergency Management's (NYCEM) Emergency Operations Center (EOC) facility, located in Brooklyn, New York, to participate in the OpEx of RepKnight. This included representatives from the NYCEM, New York City Police Department (NYPD), the Port Authority of New York and New Jersey Police Department, Montgomery County Fire

and Rescue (Maryland), California Governor's Office of Emergency Services and the Boston Fire Department (Massachusetts). The training, demonstration and all experimentation activities were carried out in the EOC.

The NUSTL Experimentation Director provided participants with information on the OpEx program's goals and purpose and an overview of the RepKnight experiment. A RepKnight representative gave an introduction on the software application and some background on how it has been used by other organizations. RepKnight provided a short training session, showing users how to input searches using Boolean logic and multiple search terms, and how to view and filter through results. The timeline of the event constrained this experimentation event to only allotting one hour of training. Responders were then asked to input their own searches parameters. The learning curve for RepKnight is moderately high, meaning the one hour of training was insufficient for inexperienced users or analysts to fully understand all of the functionality and features. Because of this, there were many questions, and one-on-one guidance was required as the participants used the software.

Many participants deviated from the procedures of the experimentation plan; however, they were still exploring software features relevant to them as first responders. Participants conducted archival searches and quickly noticed the difference between an archival search and real-time search. During archival searches, Boolean logic is not required, and searches are based on keyword identification. Archival searches also do not allow the end-user to stipulate the type of content and from which social media sources they are extracted from, as the search is broad in scope.

The experimentation and exploration of RepKnight came to an end a few minutes sooner than anticipated as the data allocation cap for this event was reached. RepKnight pulls data from various sources. Data comes at a cost, which is incurred by RepKnight and charged by an internet service provider. For this event, RepKnight allocated a certain amount of data based on their internal budgets.

Every participant was given an opportunity to input search parameters, explore pre-compiled results and run their searches over a time period between 10 minutes and an hour (this was dependent on the speed the participant worked at). During the event, data collectors shadowed participants to collect observations and feedback on their experience and interaction. After completion of all activities, the experimentation director led the participants in a discussion to provide feedback about RepKnight.

3   Results

All recorded questions and comments are captured in Table A-1 in Appendix A. Feedback and recommendations are discussed below.

3.1   Operational Scenario Comments and Debriefing

First responder participants varied in background, experience and jurisdiction, and this diversity was apparent in their varied feedback.

The NYPD found potential utility in RepKnight's ability to create visual social networks at any given moment. This feature analyzes who is talking to whom on platforms such as Twitter, and incorporates how many followers a person has, how many responses they get and how many people they are following. It then turns all this information into a visual for the analyst to better understand who was communicating and who might be an influential actor in a group. The fact that this could be cross-validated through multiple social media platforms, if the same users are actively posting on Instagram for example, added to its appreciation.

The NYCEM highlighted RepKnight's ability to not only monitor social media, but also respond to individuals or groups of individuals through the same platform as a benefit. RepKnight software could be used to log in to an agency's social media accounts, reply directly to messages, and post on Facebook, Instagram, Twitter or other social media platforms. This feature is particularly useful to help inform the public or respond to disinformation that may

be spreading through social media. RepKnight is able to indicate to its end-user when there is a change in the rate of mention of a certain topic or set of search parameters, providing a user with the potential ability of observing the growth of an incident, whether it is a riot, inclement weather or other issue that will require resources and a response. This feature was of interest to the NYCEM since it is aligned with their objective of discovering trends as early as possible in an attempt to either predict an incident or event, or to be informed faster than conventional methods such as 911 calls during large-scale incidents.

All users provided mixed comments on the user interface and usability. Certain aspects were intuitive and commended, including various visual ways in which results were displayed (i.e., lists, time series charts and social network graphs). Other aspects were difficult to understand or navigate. The Boolean logic required for inputting real-time searches proved to be difficult for some users. This challenge would be mitigated with training. Others noted the ability to rearrange the logic gates by dropping and dragging was very helpful in understanding what they were doing.

There were also several glitches where the website would not render properly (i.e., text overlapping or images being cut off) and freezes (i.e., website was not responsive to actions) that were noted that required users to reset their web browsers and log in again. Some of these issues were browser dependent. One issue noted by first responders occurred when they attempted to change their search parameters. They updated search terms and conditions, which did not appear to update on the screen until several minutes later, indicating there was a delay between what the end user saw and what was actually processed. This caused confusion as to whether the changes were actually completed.

The experimentation event was cut short by a few minutes due to a data limitation being reached. RepKnight allocated a certain amount of data bandwidth for this event and it was reached sooner than anticipated. This initiated substantive dialogue about functionality and the reality of cost, since the allocated data bandwidth for the event was a constraint driven by cost. RepKnight was not designed to seek out the "unknown unknown," a term used by first responders to identify an event that is an anomaly of sorts (i.e., an impending riot, an infrastructure issue, a gas leak, a terrorist attack). However, some of its features and design still allow for first responders to construct a wide aperture search and set trigger thresholds for changes in certain topics, so that they might have a better chance of identifying an event before it has been identified by conventional sources such as 911 or 311, and they could respond more rapidly. Responders noted the ability to use RepKnight in this fashion comes at a high cost of data bandwidth and processing power, which could translate into higher monetary costs.

Overall, the first responder participants noted that RepKnight could serve as a partial solution to their needs of identifying the unknown unknowns. The visual networks, ability to observe the growth of an incident and communication capabilities that spanned the users' social media accounts were highlighted as useful tools by the first responders. There were also areas for improvement identified, which included the user interface, Boolean searches and general usability. RepKnight representatives took note of all suggestions from the first responder participants for consideration during the development of their next version of the product.

4   References

1. **U.S. Department of Homeland Security.** *Project Responder 4 - 2014 National Technology Plan for Emergency Response to Catastrophic Incidents.* s.l. : DHS Science and Technology, July 2014.

2. **NUSTL.** *Urban Operational Experimentation Plan RepKnight.* New York City : Department of Homeland Security Science and Technology Directorate, July 2015. OpEx-T-PL-8.

5   Acronym List

DHS   Department of Homeland Security

FRRG   First Responder Resource Group

NUSTL   National Urban Security Technology Laboratory

NYCEM   New York City Emergency Management

NYPD   New York City Police Department

OpEx   Operational Experimentation

RK-DAS   RepKnight Data Analytics Software

S&T   Science and Technology Directorate

6   Appendix A

Table A. Questions and Comments, Grouped by Topic

Topic	Comment
Account Login	• During the assessment, two evaluators were accidently provided with the same login credentials. This allowed the evaluators to see a security feature introduced by the vendor, which will immediately log-off whoever is logged in to an account once someone else logged in. While the group of evaluators agreed that security and account control is important, they suggested that when a second user logs into a single account to add a prompt or pop-up window that states, "This account is already in use. Would you like to continue to log-in and terminate the other user's session?" Another option is to potentially shift that authority to the initial user, allowing him/her to decide what should become of this new request for login.

Bandwidth	• During the assessment, the evaluators had to stop their searches early due to data uses. The vendor explained they were allotted an allowance of data for the event (due to the costs associated with high data use). The vendor stated there is a pay structure for contracts that include data. Participants expressed the following concerns:
	  o Cost vs. need to keep broad ongoing searches;
	  o Capability for measuring data use in real time;
	  o Remaining within contract allotments; and
	  o Ability to negotiate contracted rates for data.
	• The vendor responded by stating that RepKnight is able to detect large surges in bandwidth use and data pulls and will usually contact the agency before allowing data usage charges to surge. They will often work with the customer agency to understand the nature of the surge, as well as help mitigate false surges (i.e., data surges for topics that lie outside the scope of interest) in the future.

Filtering	• All evaluators liked the feature that multiple search terms can be included in a single search field when separated by a comma while filtering through results. Some of the evaluators noted that other programs required you to build a search one by one; this highlighted RepKnight's efficiency. • The ability to refine results and create additional searches within results that were already obtained was a positive aspect. It could better allow analysts to hone in on events and causes, or unique patterns of interest.

Topic	Comment
	• Would be ideal to be able to perform additional searches within the more visually graphic results screen. **(New York City Police Department)**
Geolocation	• Geolocation can only be utilized if the device the social media post is being submitted from has that functionality turned on; several evaluators noted that geo-inferencing would be ideal. Geo-inferencing is when the software makes an estimated, but informed, guess as to where the message is coming from based on behavior patterns or other influencing factors. **(New York City Police Department, New York City Emergency Management & FRRG)**
Media Results	• Currently, results can be filtered by media type, but there is no visual of the media. In order to see the type of media (image or video), it requires you to click through links and then wait for the media to load. In an emergency situation, this could result in a significant time delay and reduce the efficiency and effectiveness of the social media searches. o After discussing this with the vendor, they indicated that this has already been noted and a request has been submitted to their development team to address this.
Search	• Currently, RepKnight does not allow for a single search input to conduct searches historically and in real-time concurrently. Instead, two separate searches would have to be created and initiated. The evaluators noted that it would be more efficient to be able to run a single search that would use both historical and real-time data simultaneously. **(New York City Emergency Management & New York City Police Department)** o The real-time searches provide a more advanced and Boolean based search method, whereas historical searches are just based on keywords. • A highlight that multiple agencies noted was the usefulness of being able to upload an excel spreadsheet of search terms directly into the software application. **(All)**
Translation	• After running searches, several evaluators received results in languages other than English, which prompted a conversation about the translation capabilities of RepKnight. o RepKnight representatives stated that they currently have the capability of using Google Translate to read results, but the translations are not admissible in court as they cannot guarantee accuracy. RepKnight representatives did note that, if needed, a translation service can be secured and built into a contract for their customers. • It was viewed positively that searches can be conducted in different languages if the search parameters are inputted in those languages. (New York City Emergency Management & New York City Police Department)
Usability	• The evaluators noted it would be useful to be able to search by time increments, i.e., five minutes before a spike, to determine if there was any suspicious activity leading up to the event they are researching. • All evaluators agreed that the program is not intuitive enough for first responders to use "right out of the box." o RepKnight has concurred that this tool is better suited for intelligence analysts with moderate experience; however, they feel that anyone can learn the software if adequate training is provided. They indicated the Boolean search has a learning curve and training typically takes three to four hours.

Topic	Comment

Web-based Functionality	Most evaluators experienced issues with the Internet Explorer browser. Issues included:
	
	  o Freezing search screens and graphs;
	  o The inability to use advanced filters; and
	  o Filters being modified are slow to change and still displays previous tag even if removed (Suggestion to add a status of "pending" next to changes to reduce confusion). **(FRRG)**

Other	Several of the evaluators indicated RepKnight is useful for looking through social media when you have a specific event to search for, but would not be the ideal tool for general monitoring or when attempting to predict an event. • RepKnight has the capability to identify changes in frequency and rate of appearance of identified search terms in social media, thus mounting events can be identified. The data-cost constraint would be the deciding factor in using RepKnight in this use-case.