Specialised Media Monitoring Tool to Observe Situational Awareness

Heloise Pieterse¹, Carien Van 't Wout², Zubeida Khan¹ and Chris Serfontein¹
¹Council for Scientific and Industrial Research, Pretoria, South Africa
²Capgemini, Amsterdam, Netherlands

hpieterse@csir.co.za carien.wout@gmail.com zdawood@csir.co.za cserfontein@csir.co.za

Abstract: The 21st century witnessed the rise of the digital age via the evolution of the Internet. A direct result of the established digital era is the growth of digital media. Although traditional media, such as print and broadcasting, remains relevant, the low accessibility and limited reach impact the effectiveness of such media. In comparison, digital media, such as social media, provides access to vast collections of open-source information. The open-source information available on digital media platforms is current and easily accessible. However, conversion of open-source information into intelligence becomes a time-consuming and quite ineffective process due to the large quantities of information available. Open-source intelligence (OSINT) tools attempt to address such limitations by offering technological solutions that provide access to up to date, as well as easily searchable digital information. Thus, OSINT tools, such as Shodan and Maltego, streamline the conversion of open-source information into intelligence. This paper introduces the design of a new but specialised media monitoring tool to assist with the monitoring of open-source information towards enabling situational awareness. The key design feature of this new media monitoring tool is to offer versatility by guiding the collection of open-source information. Controlling the amount of information gathered in turn improves the efficiency of the processes used to extract intelligence. The obtained intelligence, thus, enables the observation of situational awareness of chosen aspects within a secure environment. Also discussed in this paper is the motivation behind the design of the specialised media monitoring tool, as well as various design considerations, are taken to ensure effective usage of the tool. The paper concludes by discussing the various benefits offered by the specialised media monitoring tool concerning observing situational awareness.

Keywords: Open-Source Information, Digital Media, Social Media, Monitoring, Intelligence, Situational Awareness

1. Introduction

The digital age ushered in a new era characterised by a sudden change from traditional industrial-based processes to an economy primarily driven by information and communication technology (ICT) (Forsyth, 1999). The success of the digital age is due to the evolution of the Internet, which permits the retrieval, sharing and transmission of information via digital media (Haris, 2016). Digital media, such as online newspapers, discussion boards and social media, provides access to vast collections of open-source information. Open-source information is best defined as "unclassified data available to the public" (Williams & Blum, 2018) that anyone can lawfully obtain by request or observation. Lately, open-source information has become increasingly valued (Noble, 2004) due to the proliferation of such information on the Internet, which is current and easily accessible. However, due to the sheer volume of readily available open-source information, conversion of the information into intelligence can be laborious and ineffective.

Technological solutions, such as Open-Source Intelligence (OSINT) tools, attempt to address the challenge of obtaining vast quantities of open source information for intelligence. The purpose of OSINT tools is to assist with the automated retrieval, collection, filtering and distribution of open-source information. Media monitoring systems and real-time search engines, which are examples of OSINT tools, provide analysts with a vast influx of open-source information from multiple data media sources across the world (Best, 2011). However, the present challenge is no longer accessing timely information but processing and analysing that information to obtain intelligence. Thus, to avoid information overload (Best, 2011) and enable the collection of highly accurate information (disregarding fake information), it is necessary to guide the collection of open-source information. Although various OSINT tools currently exist, such tools either limit the collection of open-source information to a single platform or collect from multiple digital sources, which impacts the effective processing and analysis of information. Therefore, sophisticated intelligence tools are required that provide analysts with access to rapid but relevant information, as well as techniques to produce intelligence insight and attain insight.

This paper introduces the design of a new specialised media monitoring tool that allows for directed, as well as geographically positioned searches to be conducted. The purpose of the specialised media monitoring tool is to assist analysts to observe situational awareness, which is defined as having awareness about what is happening in an environment and using the information to make informed decisions (Gheisari, Irizarry, & Horn, 2010).

To overcome known limitations of existing OSINT tools, the specialized media monitoring tool incorporates four distinct features: (i) monitoring pre-selected digital media platforms, (ii) geographically guide searches for information, (iii) direct the extraction of information, and (iv) structuring of the collected information. By guiding the collection of open-source information, the amount of information gathered is controlled, which in turn improves the efficiency of the deployed data mining and pre-analysis techniques to extract intelligence. The intelligence obtained from the specialised media monitoring tool enables the observation of situational awareness and the monitoring of changes of chosen aspects within a secure environment.

The remainder of this paper is structured as follows. Section 2 discusses the motivation and purpose of the newly designed specialised media monitoring tool, as well as various design considerations taken to develop the tool. In Section 3, the various components of the tool are identified and described. The benefits and uses of the specialised media monitoring tool are deliberated in Section 4. Section 5 concludes the paper.

2. Background

Open-source information is not what it used to be and the revolution in ICT has fundamentally changed the distribution and accessibility of such information. As of today, open-source information dominates the universe of intelligence analysts and is unlikely to change in the foreseeable future (Gannon, 2014). Using open-source information, analysts can produce intelligence products, which plays an increasingly important role in decision-making processes (OSINT and Intelligence, 2013). To assist with and ensure the timeous delivery of intelligence products, analysts often rely on free or commercial off-the-shelf OSINT tools to improve the efficiency of information retrieval and extraction. Although various OSINT tools exist, the envisioned aim and purpose of such tools may not always effectively address the needs of intelligence analysts. The need for specialised and sophisticated OSINT tools still remains.

The remainder of this section discusses the motivation for the design and development of a new OSINT tool, called the specialised media monitoring tool. Furthermore, the section also highlights the intended purpose of the tool, as well as design considerations, are taken to guide the development of the tool.

2.1 Motivation

The World Health Organization (WHO) on March 11, 2020, declared the novel coronavirus (COVID-19) outbreak a global pandemic (WHO Press Conference on COVID-19, 2020). With the arrival of the COVID-19 pandemic came social changes and nation-wide lockdowns worldwide. Such sudden and radical changes can have a significant impact on human behaviour and societies, as witnessed by the recent Black Lives Matter global movement (Maqbool, 2020). Using the specialised media monitoring tool, digital media can be continuously monitored for public opinion, emerging sensational events, as well as perceiving of national security risks.

Apart from monitoring news from a reactive or sensation standpoint, the monitoring of events and phenomena for trend analysis and other pattern detection, may provide valuable and early detection of changes in the security environment that can be addressed timeously. This may include the detection of fracture points within a specific geographical society on both national and smaller scale, as well as societal concerns that can be identified and monitored for escalation.

Zooming into a practical use case, in Mumbai, law enforcement has set up a social media lab to monitor social media posts for intelligence gathering of the public. Trained policemen took turns in shifts to monitor the social media feeds of the public together with specialised software that flags content that could be potentially troublesome. This reinforces the motivation for the need for an automated media monitoring tool (Trending now: Cops to keep a keen eye on social media, 2013).

Therefore, the specialised media monitoring tool enables continuous observation of situational awareness regarding chosen aspects within a security environment

2.2 Purpose

The immediate challenge for security and business intelligence fraternities is the expanding volume of opensource information. Technological solutions have proven to be part of the answer to the open-source information challenge (Gannon, 2014). There is, however, a need for the intelligence community to invest more in technology that offers powerful processing and analytical capabilities to assess and exploit the vast amounts of open-source information. Currently, various media monitoring and OSINT tools exist to assist analysts with the transition of open-source information into actionable intelligence. Though, as emphasised earlier in the paper, these tools are often limited, or the sheer volume of information collected can influence effective processing. For example, Maltego, developed by Paterva, is a reconnaissance tool for open-source information and allows for the visualisation of the collected information in a graph format. The true power of Maltego lies with the built-in or custom-created transforms, which allows for the analysis of real-world relationships between people, webpages, domains and networks (Maltego, 2020). However, for intelligence analysts, a learning curve might be required to optimally use Maltego, which can in turn limit the adequate use of the tool to obtain intelligence. Another example is Hootsuite, a social media management platform that permits the scheduling, publishing and distribution of content from 35 social media platforms (Hootsuite, 2020). Although Hootsuite consolidates open-source information from multiple social media platforms, the vast amounts of information can overwhelm intelligence analysts.

While the discussions above merely provided a brief critical overview of two existing OSINT tools, it becomes clear that these tools are somewhat rigid in their design. Such inflexibility impacts the efficiency of the OSINT tools to transition open-source information into intelligence. The need for OSINT tools that support adaptability, as well as flexibility concerning open-source information collection and analysis, are clear. The specialised media monitoring tool attempts to fill this void left behind by current OSINT tools. The purpose of the specialised media monitoring tool is to offer versatility using the following strategic features:

- Monitoring pre-selected digital media platforms: that offers current, relevant and up to date opensource information that can easily be obtained and analyzed.
- **Perform geographically guide searches for information**: by limiting the amount of open-source information collected to prevent information overload and focus on areas of interest.
- **Direct the extraction of information**: to include elements beyond the obvious, such as text, that can offer insight into situational awareness.
- Structuring of the collected information: to enable and allow for quick and easily attainable intelligence.

The various design considerations taken to enable the specialised media monitoring tool to support versatility are captured in the following section.

2.3 Design Considerations

Versatility is a fundamental aspect of the specialised media monitoring tool. Enabling versatility requires the following design considerations to drive the construction of the specialised media monitoring tool (Albin, 2003):

- **Extensibility**: allow for new capabilities to be added to the software without major changes to the underlying architecture.
- Modularity: provide well-defined, independent components which leads to better maintainability of the software.
- Usability: well-constructed user interface (UI) that is efficient and satisfying to use, as well as easy to learn
- Scalability: permits the software to easily adapt to increasing data, additional functions or more users.
- Modifiability: allow the software to grow and change over time.

By guiding the design of the specialised media monitoring using the above-mentioned design consideration, the software tool will be able to grow, adapt and meet the ever-changing requirements of users. Therefore, versatility will be supported, preventing the specialised media monitoring tool from becoming obsolete in the near future. Taking note of these critical design considerations, the following section discusses the various architectural components of the specialised media monitoring tool.

3. Components of the Specialised Media Monitoring Tool

The architectural design of the specialised media monitoring tool consists of the following essential components: (i) data source, (ii) collection methodology, (iii) data storage, and (iv) UI. These components, as well as the flow of interaction, are visualised in Figure 1.

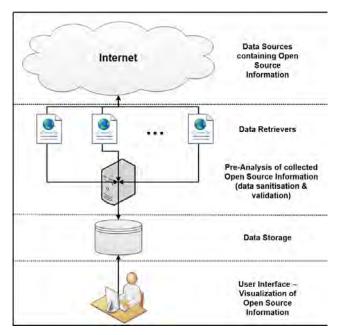


Figure 1: Architectural components of the specialised media monitoring tool

Each component presented in Figure 1 is further discussed and elaborated on in the following sections.

3.1 Selected Data Sources

A significant feature of the specialised media monitoring tool is the monitoring of pre-selected digital media platforms, which must offer current, relevant and up to date open-source information. Based on these requirements, Twitter and online (digital) newspapers were selected as the primary data sources for the specialised media monitoring tool. Twitter is a popular, prominent, and frequently used microblogging and social networking service. The platform offers insight into "what's happening in the world and what people are talking about right now" (About Twitter, 2020) by sharing messages called tweets. Online or digital newspapers are described as electronic documents that present breaking news promptly. Although various online newspapers do exist, media bias associated with the newspaper must also be taken into consideration. Media bias is defined as "political bias in journalistic reporting, in programming selection, or other in mass communication media" (Definitions for media bias, 2020). Therefore, it is important to ensure that a wide variety of online newspapers and identified open sources are covered in the collection process to compensate for bias and/or validate the information.

Collecting from both Twitter and online newspapers will ensure the specialised media monitoring tool has access to current and relevant open-source information. Although the specialised media monitoring tool currently only collects open-source information from two data sources, the extensibility and modifiability of the architectural design of the tool will permit additional data sources to be included in the future. The following section discusses the collection methodology followed to collect open-source information from the described data sources.

3.2 Collection Methodology

The collection methodology of the specialised media monitoring tool relies on the notion of data retrievers. A data retriever is an automated but dedicated script that enables the collection of open-source information from a specific data source. Data retrievers allow for the continuous collection of open-source information by following a scheduled routine to collect and extract relevant information as per a predefined schema. The specialised media monitoring tool currently consists of four unique data retrievers:

- Newspaper retriever: Python-developed script that utilizes the feedparser and newspaper libraries to
 collect the latest news articles from various newspapers via Really Simple Syndication (RSS) feeds. The
 script is scheduled to execute twice daily to ensure complete coverage of the latest news articles.
- Twitter retriever location-based: Python-developed script that uses the Tweepy library to access the Twitter Application Programming Interface (API). The script is scheduled to execute hourly, and a predefined list of locations is used to filter the collection of tweets.
- Twitter retriever keyword-based: Python-developed script that uses the Tweepy library to access the
 Twitter API. The script is scheduled to execute hourly, and a predefined list of keywords is used to filter
 the collection of tweets.
- **Twitter retriever handle-based**: Python-developed script that uses the *Tweepy* library to access the Twitter API. The script is scheduled to execute hourly, and a predefined list of Twitter handles (individuals of interest) is used to filter the collection of tweets.

The developed data retrievers permit the collection of open-source information from Twitter, as well as local and international newspaper agencies. To enable intelligence extraction from the collected open-source information, the above-mentioned data retrievers must collect more than just the tweet content or the text from a newspaper article. Data elements extracted by the newspaper retriever include the following: newspaper article title, newspaper article text, author of the newspaper article, published date, access date, and Uniform Resource Locator (URL). Data elements extracted by both Twitter retrievers include the following: Twitter handle, tweet, retweet count, "favourited" count, location, hashtags, tweet ID (the unique identifier of the tweet), user ID (the unique identifier of the Twitter user), publish date and access date.

Applying appropriate analysis and data mining techniques to the vast variety of collected data elements will enable the extraction of collated datasets for further analysis towards producing intelligence. Specific and ontology-defined extraction rules can also be set up to create dashboards to enable situational awareness of certain aspects of the security landscape. However, to ensure the accuracy and reliability of the produced intelligence, it is important to ensure the validity of the data used for analysis. To achieve this, data cleansing and validation techniques must be applied to the collected open-source information. Data cleansing is the process of detecting and correcting incomplete, incorrect, inaccurate or irrelevant parts of the collected open-source information (Rahm & Hong, 2000). Deployed data cleansing techniques forming part of the collection methodology include the following:

- Language evaluation and exclusion (currently only analysis of English news articles and tweets supported).
- Removal of incomplete or malformed data elements.
- Consistent formatting and uniformity of date elements to enable periodical searches.

Once the appropriate data cleansing techniques have been applied to the collected open-source information, the cleaned data is transferred to data storage.

3.3 Data Storage Structure

Data storage refers to the collective methods and technologies used to capture and retain digital data. Central to data storage is databases, which allows for the organised collection of data generally stored and accessed electronically (Derclaye, 2005). Although relational databases, which model data as a series of tables consisting of rows and columns (Connolly & Begg, 2003), are still widely used and recommended, a new generation of non-relational databases appeared early in the 21st century. NoSQL (or Not Only SQL) databases attempt to overcome the limitations of a relational database by utilising a distributed, non-rigid and flexible structure to store large amounts of data (Sharma & Meenu, 2012). A well-known example of a NoSQL database is MongoDB.

MongoDB is described as a document-based NoSQL database using a schemaless structure to store large amounts of unstructured data. With MongoDB, the database is a physical container for collections and each collection is a group of MongoDB documents. A document is a set of key-value pairs holding the actual data encoded in a JavaScript Object Notation (JSON) format. The document-oriented model of MongoDB provides key qualities, such as high scalability, heterogeneity and improved performance (Krishnan, Sudheep Elayidom, & Santhanakrishnan, 2016).

The obtained and pre-processed (cleaned) open-source information to be retained in data storage is not organised according to a pre-defined model. The unstructured nature of the open-source information is due to the different data sources used for the collection of the open-source information, as well as the data cleansing techniques deployed to remove unusable data elements. Therefore, the specialised media monitoring tool requires a database structure capable of handling data with a dynamic schema. The key qualities of MongoDB, as well as the ability to support advanced queries and perform complicated operations, confirm the appropriateness of MongoDB as the database for the specialised media monitoring tool.

The visualisation of the open-source information captured and stored in the database is further discussed in the following section.

3.4 User Interface Design

The intended purpose and use of the specialised media monitoring tool are to assist analysts to effectively observe situational awareness. Therefore, the UI of the specialised media monitoring tool is wisely crafted to enable usability and permit users to extract the necessary intelligence from the collected open-source information. Figure 2 presents a visual representation of the UI design for the specialised media monitoring tool.

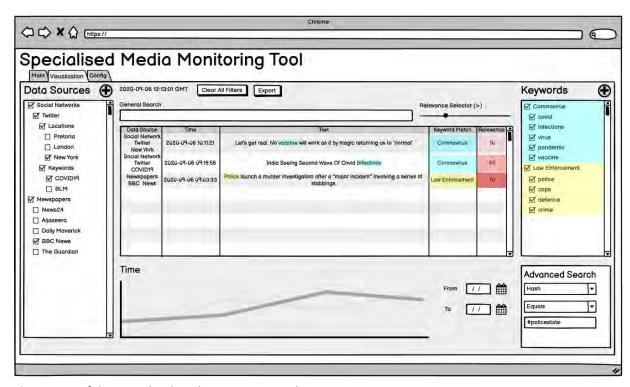


Figure 2: UI of the specialised media monitoring tool

The UI design of the specialised media monitoring tool includes the following key UI elements to allow the user to select and control the display of open-source information via the UI:

- **Data source selection**: choosing from the available data sources, either primary (Twitter or online newspapers) or pre-analysed subsets extracted from the primary data sources. The selection of the data sources will limit the amount of open-source information displayed and processed during searches.
- General searches: conduct a non-specialised search to find results using the selected data sources.
- **Pre-defined keyword searches**: conduct specialised searches by selecting pre-defined keywords from set categories (e.g., Coronavirus or Law Enforcement) to find results using the selected data sources.
- **Advanced searches**: conduct specialised searches by constructing unique queries to search for specific results from the selected data sources.
- Date-based search: conduct specialised searches using date filters to limit the amount of open-source information displayed and processed during searches.
- **Relevance selection**: limit the amount of open-source information displayed based on the quality of the information.
- Clear all filters: reset to default selections and clear all results.

- Exporting results: produced by searches and selected filters to allow for offline analysis.
- Visualisation of results: produced by searches and selected filters to simplify observation of situational awareness (illustrated in Figure 3). Visualisation techniques include
 - Word clouds a graphical representation of all words contained within the selected data sources, in which the size of each word indicates its frequency.
 - Sentiment analysis categorising the emotion, either positive or negative, of the selected data sources.

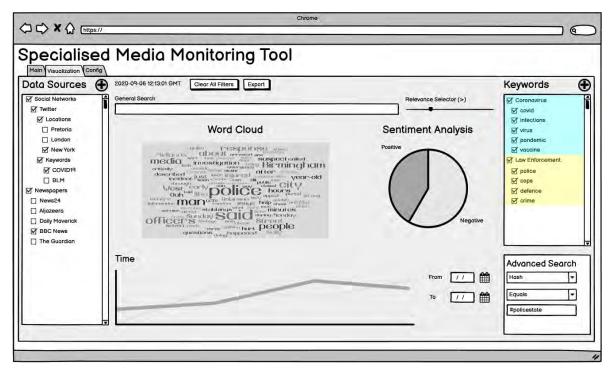


Figure 3: Visualisation of techniques of specialised media monitoring tool

The purpose of the above UI elements is to enable the usability of the specialised media monitoring tool, allowing users to monitor situational awareness in near real-time.

4. Benefits of Specialised Media Monitoring Tool

The specialised media monitoring tool is developed specifically to assist analysts to more effectively collect and process open-source information, which enables the observation of situational awareness. The primary benefit of the specialised media monitoring tool is that it assists the military or business intelligence functionary in the collection and pre-analysis of data towards producing good quality intelligence products. The Intelligence Cycle describes a set of functions of the processes used to create intelligence to enable improved decision-making. The cycle starts with a problem and consists of several processes including planning & direction, collection, processing, analysis & production, dissemination & utilisation/integration (Johnson, 1986). The specialised media monitoring tool, therefore, supports the first three processes of the Intelligence cycle.

- It enables data collection from online platforms.
- It enables directed data collection and thus data storage of only relevant data.
- Enables data mining using specialised searches.
- Enables pre-analysis by tagging keywords.
- User-friendly and client customised user interface.

The obtained intelligence products from digital platforms, such as social media and online newspapers, can enable security agencies (e.g., law enforcement or the military) to closely observe the situational awareness of a security environment. Continuous situational awareness allows for the detection of security incidents or triggers that may result in dangerous events.

A key example of the value the specialised media monitoring tool can add to the security agencies is within the capacity of early warning or profiling. Analysts could detect troublesome messages on social media platforms

and act upon them before they cause damage. An example of using social media to monitor events for situational awareness is the 2019 Christchurch shooting where two consecutive mass shootings occurred at mosques in a terrorist attack in Christchurch, New Zealand, during prayers. Parts of the mass shooting were live streamed on social media. Furthermore, a hate-filled document was also published some 10 to 20 minutes before the attack. Had law enforcement agencies been monitoring social media, such an event could have perhaps been avoided due to early warning.

Another benefit of the specialised media monitoring tool is the ability to simultaneously monitor a range of threat scenarios, such as civil unrest, rival factions (e.g., gangster groups or crime syndicates), terrorist activities, and political developments. For example, protests are usually coordinated and shared on social media platforms. Law enforcement agencies that monitor these could warn people about safety in various locations. Furthermore, analysing these social media posts could provide insights into causes for such events and key role players who require further investigation.

Finally, the features incorporated into the specialised media monitoring tool enables embassies to utilise the tool to observe country risks, political developments within a nation, as well as civil society's reaction to government initiatives.

5. Conclusion

This paper aimed to introduce the new specialised social media monitoring tool, which was designed to assist analysts to observe situational awareness. The specialised media monitoring tool guides the collection of open-source information to enable the creation of intelligence products by incorporating four distinct features: (i) monitoring pre-selected digital media platforms, (ii) geographically guide searches for information, (iii) direct the extraction of information, and (iv) structuring of the collected information. By making use of the specialised media monitoring tool, analysts can monitor open-source information, which is collected from online newspapers and social media feeds, to attain intelligence. The obtained intelligence, thus, enables the observation of situational awareness of chosen aspects within a security environment, providing analysts with insights into the causes and details of security events. Although valuable information for situational awareness is contained within the open-source data sources, it is still a complex process to extract relevant information and to detect useful patterns in the data to identify threats and to predict possible attacks. Automation is required for intelligent decision-making. For future work, the authors will investigate the use of semantic technologies to enable the software system to further process the data and annotate it with meaning.

Acknowledgements

The authors would like to thank Roelof Temmingh for sharing his technical insight, guidance and support during the design and development of the specialised social media monitoring tool.

References

About Twitter. (2020). Retrieved from Twitter: https://about.twitter.com/

Albin, S. T. (2003). The Art of Software Architecture. Indianapolis: Wiley Publishing Inc.

Best, C. (2011). Challenges in Open Source Intelligence. European Intelligence and Security Informatics Conference.

Connolly, T., & Begg, C. (2003). Database Systems. Pearson.

Definitions for media bias. (2020). Retrieved from Definitions: https://www.definitions.net/definition/media+bias

Derclaye, E. (2005). What is a Database. The Journal of World Intellectual Property, 5(6).

Forsyth, A. (1999). The Information Age: Economy, Society and Culture. Journal of Planning Education and Research, 19(2).

Gannon, J. (2014). The Strategic Use of Open-Source Information. Intelligence Community Perspective, 67-71.

Gheisari, M., Irizarry, J., & Horn, D. B. (2010). Situation awareness approach to construction safety management

improvement. *Proceedings of 26th Annual ARCOM Conference* (pp. 311-318). Leeds, UK: Association of Researchers in Construction Management.

Haris, A. R. (2016). Information Issues in Digital Era. Faculty of Information Management, Universiti Teknologi.

Hootsuite. (2020). Retrieved from Hootsuite: https://hootsuite.com/

Johnson, L. K. (1986). Making the intelligence "Cycle" work. *International Journal of Intelligence and Counter Intelligence*, 1(4), 1-23.

Krishnan, H., Sudheep Elayidom, M., & Santhanakrishnan, T. (2016). MongoDB - a comparison with NoSQL databases. *International Journal of Scientific & Engineering Research*, 7(5), 1035-1037.

Maltego. (2020). Retrieved from Maltego: https://www.maltego.com/

Maqbool, A. (2020, July 10). Black Lives Matter: From social media post to global movement. Retrieved from BBC News: https://www.bbc.com/news/world-us-canada-53273381

Heloise Pieterse et al

- Noble, D. F. (2004). Assessing the Reliability of open Source Information. Stockholm, Sweden: Proceedings of the 7th International Conference on Information Fusion.
- OSINT and Intelligence. (2013). Reuser's Information services, 1-8.
- Rahm, E., & Hong, H. D. (2000). Data Cleaning: Problems and Current Approaches. *IEEE Bulletin of the Technical Committee on Data Engineering*, 23, 3-13.
- Sharma, V., & Meenu, D. (2012). SQL and NoSQL Databases. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(8), 20-27.
- Trending now: Cops to keep a keen eye on social media. (2013, March 17). Retrieved from The Indian Express: http://archive.indianexpress.com/news/trending-now-cops-to-keep-a-keen-eye-on-social-media/1089333
- WHO Press Conference on COVID-19. (2020, March 11). Retrieved from WHO: https://www.who.int/docs/default-source/coronaviruse/transcripts/who-audio-emergencies-coronavirus-press-conference-full-and-final-11mar2020.pdf?sfvrsn=cb432bb3_2
- Williams, H. J., & Blum, I. (2018). *Defining Second Generation Open Source Intelligence (OSINT) for the Defense Enterprise*. Santa Monica: RAND Corporation.