Developing Methods to Support Social Media Intelligence Analysis

Daniela Miele, Lauren Glenister, Angela Woods
Soar Technology, Inc.
Orlando, FL
daniela.miele@soartech.com,
lauren.glenister@soartech.com,
angela.woods@soartech.com

ABSTRACT

With the recent expansion of social media and the massive outreach these platforms have, the need for social media monitoring has become more important. The specific intent being analysis and sensemaking of actions taken by entities, organizations, and networks that may have the ability to influence narratives or audiences. To achieve this, the Department of Defense (DoD) has made progress in the field concerning analysis of social media intelligence (SOCINT) where intelligence analysts engage in an iterative sensemaking process, encompassing procedures such as locating, gathering, and organizing data. As a result of these actions, the main goal is to develop schemas and hypotheses that support identification of future courses of actions aiding in mitigation strategies.

Current SOCINT tooling and technologies are geared towards the locating and gathering phases of analysis. These tools work towards gathering data related to identified topics from social media platforms such as Twitter, Facebook, and Reddit. However, when it comes to the organization of the data, intelligence analysts often resort to tooling not designed for this phase of analysis, such as Microsoft Word or PowerPoint. This paper will focus on potential ways that sensemaking and schema development can be integrated into SOCINT training and tooling. Specifically, through the use cases of intelligence analysis tooling that aims to eliminate this gap by targeted support mechanisms. For example, the implementation of an ontology for structuring and filtering data, as well as the integration of visualizations to identify patterns among the collected data. The objective of investigating training procedure modifications and enhancements is to evaluate mechanisms that are hypothesized to aid analysts in better formulating schemas and interpreting datasets. As a result of this investigation, tooling tailored to schema development can be provided to encourage proper selection for courses of action and analysis directions.

ABOUT THE AUTHORS

Daniela Miele is a Research Associate at Soar Technology, Inc and is focused on the of human-computer interaction. She is currently pursuing a Master of Science in Human Factors at Embry-Riddle Aeronautical University, Worldwide. She has earned a Bachelor of Science in Psychology, Human Factors Track and a minor in Cognitive Sciences at University of Central Florida.

Lauren Glenister is a Research Scientist at Soar Technology Inc. and is focused on social media, cognitive psychology, and training. She is currently pursuing a Doctorate in Modeling and Simulation at University of Central Florida. She has earned both a Master of Science in Software Engineering and Bachelor of Science in Human Factors Psychology from Embry-Riddle Aeronautical University, Daytona Beach.

Angela Woods is a Software Architect at Soar Technology with over 24 years of software engineering experience with emphasis on designing system architectures for real-time uses including intelligent training, agent-based simulation, data analytics and game development. She is an expert in dynamic adaptation techniques and her most recent work has included detecting semantically meaningful events in timeseries data streams and using machine learning techniques to build causal models that can augment adaptive intelligent training systems. She excels at coordinating technical solution development on interdisciplinary teams that include both technical and non-technical contributors and has served as architect on over $50M of government funded research.

Developing Methods to Support Social Media Intelligence Analysis

Daniela Miele, Lauren Glenister, Angela Woods
Soar Technology, Inc.
Orlando, FL
daniela.miele@soartech.com,
lauren.glenister@soartech.com,
angela.woods@soartech.com

INTRODUCTION

Analysis of Social Media Intelligence (SOCINT) is a growing priority within the Department of Defense (DoD) in support of information operations (IO), providing important information on an adversary group's or audience's demographics, size, organizational structure, areas of activity, and network reach (Marcellino et al., 2017). In an effort to reduce workload and increase the efficiency of intelligence analysts' tasks, there has been a focus on the integration of how new technology and automation advancements, such as natural language processing (NLP), collaborative environments, and machine learning algorithms, could be integrated into daily analyst workflows. This paper will focus on use cases of intelligence analysis tooling that works to bridge the gap between the collection and distilling of social media content to the sensemaking, schematizing, and hypothesizing phases of intelligence analysis. Specifically, these tools use automation that aids intelligence analyst needs by enabling analysts to 1) easily and collaboratively link social media accounts to account classifications, groups, screenshots, information maneuvers, and analyst insights and 2) visualize various relationships identified within their own personal dataset to indicate next steps and support the reasoning for selected courses of action within the analyzed domain.

Understanding Analyst Workflow and Challenges

To execute a successful effort in designing social media intelligence analysis tooling for intelligence analysts, an understanding of the types of tasking, decision making process, and workflow of an analyst must be attained. This understanding allows for the breakdown of workflow into digestible categories that allow for the identification of the different types of tooling and functionalities needed to support the entire process from data gathering to a finalized decision. This decomposition identifies how to group tasks and develop tools that each are aimed at serving a specific purpose of intent to support the overall goal of the analysis. In addition to the decomposition of functionalities and tasks, pain points found within an analyst's typical workflow are identified so that future SOCINT tooling can be designed with the intent of overcoming identified problem areas.

Analyst Workflow

Research on the workflow and decision-making process of an intelligence analyst has led to the development of the Sensemaking Loop for Intelligence Analysts model, Figure 1 (Pirolli & Card, 2005). This model describes the nonlinear data flow from raw data sources to a reportable product. The model represents a 14-step iterative process at encompasses six core aspects: external data sources, shoebox, evidence files, schemas, hypotheses, and presentations. Among these core categories is a cyclical process that involves various efforts including but not limited to searching and filtering, reading, and extracting, schematizing, building a case, and storytelling. This process is non-linear, therefore the steps will likely be repeated several times or involve backtracking among the steps (i.e., moving from reading and extracting back to searching and filtering for more data) throughout the analysis process as more details are uncovered. Once relations begin to be pinpointed and important information is pulled, the analyst can begin to form a schema (i.e., frame of reference) and search for evidence that backs their claims or hypothesis among the curated information.

With this model in mind, the future of SOCINT tooling can be generated based on the interactions among the six core aspects as each are distinct processes found within the overarching analysis. For example, various tools have been developed to meet the searching and filtering effort that occurs between the creation of the shoebox via external data sources. Another tool can be developed that focuses solely on the development of schemas after reviewing the information inputted to the shoebox and evidence file. Lastly, a SOCINT tool that aids analysts in building their case and telling the overall story could be helpful to reduce workflow and better organizing information.

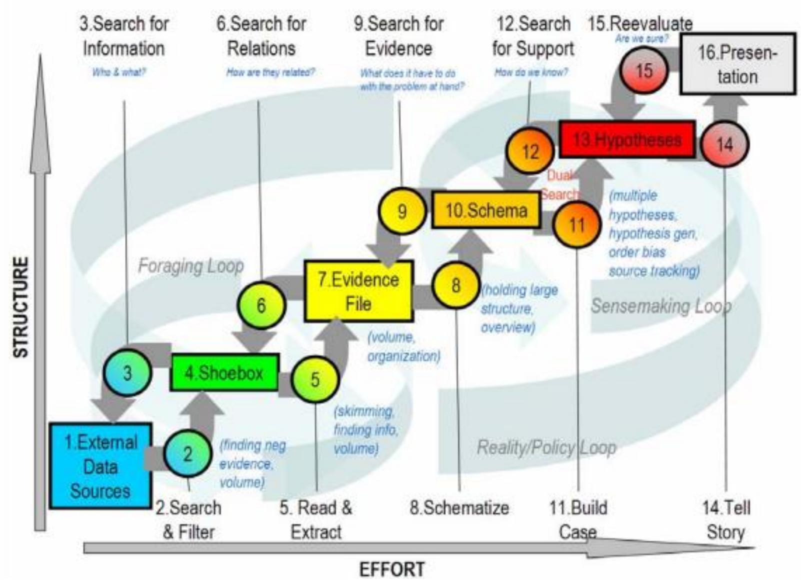

Figure 1. Sensemaking Loop for Intelligence Analysis (Pirolli & Card, 2005)

Challenges

Beyond the multi-step workflow for an Intelligence Analyst, there is a series of cognitive challenges that have been identified by the research community. Table 1 summarizes these findings.

Table 1. Challenges faced by an Intelligence Analysts

Challenges	Description
Time Pressure	Timelines have been condensed and decision makers want reports quicker. (Hutchins et al., 2006).
Lack of Expertise	Analysis are being completed by analysts that do not have expertise in a specific domain or data type (e.g., classified content, open source, interviews, reference material etc.) (Patterson et al., 2001).
Synthesizing Multiple Sources of Information	Analyst must combine several sources of data with variable levels of "validity and reliability" (Hutchins et al., 2006).
Coping with Uncertainty	There is a relationship between the context of the data and the perspective of the one interpreting the data. Depending on the task and goal of the analyst the same data may be dismissed by one and be important to another. "When high levels of uncertainty are present regarding the situation, the ability to interpret the data based on context sensitivity is likely to be diminished" (Hutchins et al, 2006).
High Cognitive Workload	Analyst must deal with the "continual onslaught of information" coupled with labor intensive data entry. This can cause analysts to be inundated with data that must be meticulously organized to derive a reportable conclusion (Hutchins et al. 2006)
Potential for Error	Due to the high workload and lack of expertise, analysts may miss important information due to focusing on other pieces of data. This can lead to the different analysts deriving different conclusions from the same dataset (Hutchins et al. 2006).

Data Overload	Advances in technology and the access to social media has led to more data channels being available for interpretation and analysis. The data abundance compounds many of the other challenges noted in this table. (Hutchins et al. 2006). The amount of data that must be assessed causes a bottleneck in the analysis process where analyst spend more time doing data entry that assessing the data entered.
Complex Human Judgements	The determination of the "validity, reliability and relevance to the particular event" is unique to each analyst. In team environments, this can cause disagreements in analyses products.

These challenges can compound each other making the analysis process more difficult. For example, according to Hutchins et al (2006), the various types of data resources require expertise to understand and derive meaningful results. However, commonly these data resources are assessed by those without expertise in that domain leading to cognitive bias because "greater weight [is given] to the types of information they understood and less weight to less understood types of information" (Hutchins et al. 2006). An analyst may anchor a conclusion based on a misplaced value in certain information and due to data overload, time pressure, and high cognitive workload does not have a chance to review their claim before presenting the results.

The Gap in SOCINT Tooling

As the SOCINT field continues to grow, there is a more imminent need for proper tooling to satisfy the requirements and needs of the intelligence analyst's daily responsibilities. While there have been many programs to develop social media analytics, e.g., algorithms to query and process raw social media data, less attention has been put towards developing post-analytic tools that incorporate SOCINT into intelligence analyst sensemaking and report generation workflows. Specifically, analysts are left engaging with Government-Off-The-Shelf (GOTS) tools that are not tailored to meet their needs such as Microsoft Word, PowerPoint, and Excel. The box in Figure 2 illustrates where these GOTS are currently used in the model. These tools are adapted to fit analyst workflows by being used to record processes, investigations, discoveries, and the provenance of the data collected. For example, analysts have utilized Microsoft PowerPoint to provide a high-level overview of specific entities (i.e., accounts, groups) that are to be further pursued, in addition to being a means for presentation and information transfer among team members.

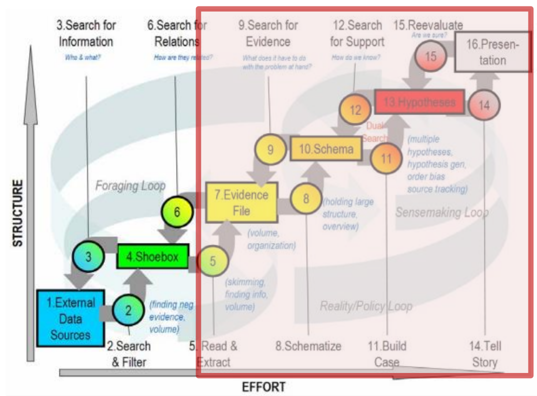

Figure 2. Current Integration of GOTS in the Analysis Process

While these tools can be adapted for analyst use, they are not ideal as they lack the features and functionality analysts need to collect in-depth information, make dynamic connections, examination emergent characteristics, and work collaboratively towards a mutual goal. Moreover, they do not address the challenges noted in Table 1 or adhere to the intelligence analyst sensemaking process discussed in Figure 1. As a result of this disconnect between SOCINT tooling and analyst needs, the currently available GOTS tools require significant labor to use, hinder team workflows, and disregard the known challenges faced by analysts. These future technologies could aid in the analyst's workflow by providing easy shortcut methods via dynamic linking, and NLP mechanisms to equip analysts with ways to navigate their dataset more efficiently.

EFFORTS TOWARD FUTURE SOCINT TOOLING USE CASES

The following use cases focus on prototype applications that are being developed to support the conceptualization of how future SOCINT tooling could aid analyst workflows and address the challenges faced. Both applications are presented as use cases to outline the specific features and functionalities future SOCINT tools could possess and the impact they would have on an intelligence analyst's workflow and progress.

Use Case #1: Analyst's Scratchpad

The analyst scratchpad, Figure 3, a cloud -based SOCINT analysis and synthetization tool that enables analysts to link social media accounts easily and collaboratively to screenshots and analyst insights. As shown in Figure 3, an analyst can use the scratchpad (left side of Figure 3) to input text and images of key insights. The right side of Figure 3 illustrates the pull-out drawer that can assist analyst in maintain situational awareness of the information already collected about an account or group of accounts. This tool consists of several features and capabilities based on the needs and challenges of the SOCINT analyst including an ontology, dynamic linking, collaborative environment, data management, input suggestions, entity cards, and tracking threads of analysis.

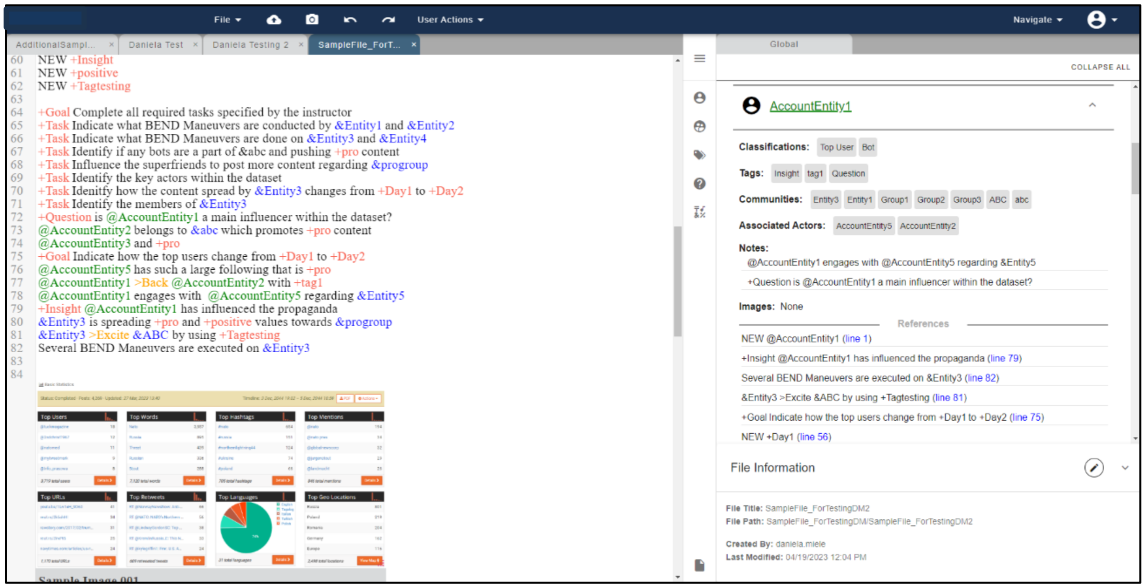

Figure 3. Analyst Scratchpad Tool with Entity Card Displayed

Structured Ontology

To equip analysts with a structured yet flexible methodology, the scratchpad leverages a custom ontology that provides users with the ability to create their own folksonomy, or a self-generated classification system for certain data types. Specifically, this approach is geared towards analysts and equips them with five high-level data types that can be added and dynamically linked within the scratchpad (see Table 2). The data types can be categorized into two categories: user driven and predetermined. The data types that are user driven allow the user to create their own folksonomy. The user driven data type provides the flexibility for analysts to make new accounts, groups, and tags as tasks evolve. The predetermined data types are rigid and cannot be changed without administrator intervention. These rigid data types are derived from a set lists of classification types or model of various influence maneuvers identified by Dr. Kathleen Carley and Lt. Col. David Beskow (2019).

Table 2. Scratchpad Data Types

Data Type	Entity Creation	Token	Definition
Account	User driven	@	Specific accounts the user is tracking
Group	User driven	&	A series of accounts the user wants to categorize together
Tag	User driven	+	A term the user wants to use to describe content found.
Account Classification	Predetermined	~	Characteristics of an account (e.g., bot, cyborg, top user, etc.)
BEND Maneuver	Predetermined	>	Influence Maneuvers that accounts and groups leverage

A unique token for each of the five data types allows for the unstructured user inputs to the scratchpad to be structured for future use. These tokens are color coordinated to provide feedback to the user that the token and entity was registered by the scratchpad. The usage of the tokens and the colorations are illustrated in Figure 3.

Dynamic Linkage

Each line in the scratchpad determines the dynamic linkages that are generated. Figure 4 shows a snippet from the above Figure 3 and the resulting linkages just for the account AccountEntity1.

Data Type	Linkages
Account	AccountEntity2, AccountEntity5
Group	Entity5
Tag	Insight, tag1
BEND Maneuver	Back

Figure 4. Dynamic Linkages for the Account AccountEntity1 from Lines 77 - 79

As the user inputs their text and image insights, the dynamic linkages are made which organizes the unstructured notes into a structured format. The structured data is saved in a knowledge graph that expands as the user inputs more data. The table on the right of Figure 4 only displays the results for the account AccountEntity1; however, these linkages occur for each entity for each data type. Since these linkages are made while the user is inputting information, they can review the structured format in the knowledge graph rather than needing to review a series of text files to assess their analysis. These capabilities afford analysts the ability to move faster from data entry to data analysis.

Collaborative Environment

The scratchpad tool is also a cloud-based application that allows for work to be shared among a team. Team members can view and (with permission) edit scratchpad files made by other members of their team. The dynamic linkages created are also shared within a team. This allows for users to have different focus points about the same account or topic, but the information to be merged and organized into one knowledge graph. This collaborative approach could support uncovering emergent details from the analysis. For instance, if one analyst is focused on content related to Topic1 and GroupA and another analyst is investigating Topic2. Analyst 2 finds evidence that Topic2 is also related to GroupA. This information would be linked together to reveal to both analysts that Topic1 and Topic2 are related to GroupA through a share knowledge graph. This integration of information illustrates the emergent qualities of the dataset that a collaborative environment can lead to.

Data Management

The user driven data types provide a method for the user to be able to create entities for accounts, groups, and tags as needed. However, this folksonomy approach can cause an issue for data management. For example, a user could have the tag *UnitedStates* in their analysis and as they continue, they make the tag *USA* and *America*. The user knows they are referring to the same topic. However, three different terms are being used which causes three different entities to be created under the Tag data type. Capabilities are being developed to allow the user to combine these three entities so that one entity can be the primary entity and the other two can be the secondary entities. This data management issue is even more evident in team environments in which different users may refer to the same topic with different terms. The necessity of the data management capability is essential with the usage of a user driven approach to provide clarity and precision to the analysis process.

Suggestion Capabilities.

The ability to manually enter text and images can lead to errors in misspellings or misremembering entity previously identified. The scratchpad tool incorporates two approaches for error handling: drop down list of known entities while the user types and using natural language processing.

Figure 5 depicts an example of the dropdown functionality implemented into the tool. As the user types, the scratchpad will attempt to keyword match the text to known entities in the knowledge graph. (Note: The user can identify a new entity to user driven data types by typing the word "NEW" prior to the new entity). If the entity is on the list, the user can

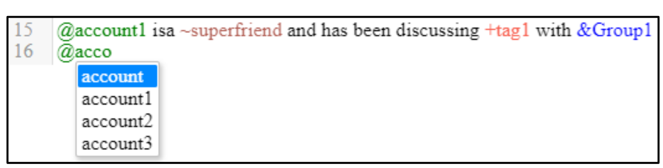

Figure 5. Dropdown to Aid Users in Remembering their Previously Identified Entities.

then simply choose it from the dropdown using their cursor, arrow keys on the keyboard or finish typing out the term manually. This approach only handles keyword exact matching, another capability is implemented to support close to matches (i.e., misspellings).

NLP is used to offer recognized entity suggestions when a user misspells a term or misses the expected token prior to the entity. The SOCINT tool (the scratchpad) will run a check on all words typed into the scratchpad against known entities for each data type. If there is a recognized term with a missing token (e.g., the known account “AccountEntity3” is typed without the “@”), shown in Table 2, or the term is misspelled (e.g., the known tag “tag1” is typed “tga1”) then a suggestion will be given to the user. Figure 6 shows these two examples of suggestion capabilities.

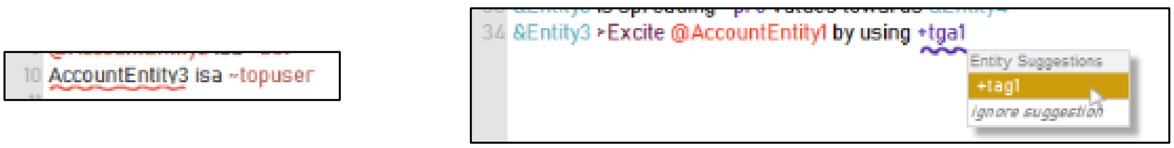

Figure 6. Examples of Entity Suggestions with NLP

Machine learning can be leveraged further by offering suggestions of terms that are used often that could become recognized entities. For instance, the user has used the term “voters rights” multiple times throughout their analysis. The scratchpad could offer a suggestion to make a tag “+voters_rights” to allow the user to find all related text easier and link data types on the same line.

Entity Cards

Based on the intended use of SOCINT tools, there is a high likelihood that multiple analysts (both within and between teams) capture the same information or related information pertaining to a specific entity (i.e., account, group, hashtag) as the purpose of the analysis revolves around the same topic. Therefore, a need to consolidate all information connected to a specific entity and present the output in one location emerges. The scratchpad’s feature built in baseball like entity cards allow this consolidation to be realized. Entity cards are automatically generated “cards” with all identified and linked information pertaining to either a specific account or group the analyst has added within the scratchpad. As analysts input text and images the tool records the information and dynamically links content to specific accounts and groups to generate summaries of profiles within a card-like manner. Account and Group Entity Cards, shown in Figure 7, are displayed within the scratchpad to allow the analyst to review content inputted from them and their teammates in real time. Within each entity card, under the reference section, is a traceback feature that can provide a link back to where the information was first typed. This could give insight on what the analyst’s perspective when the information was originally recorded.

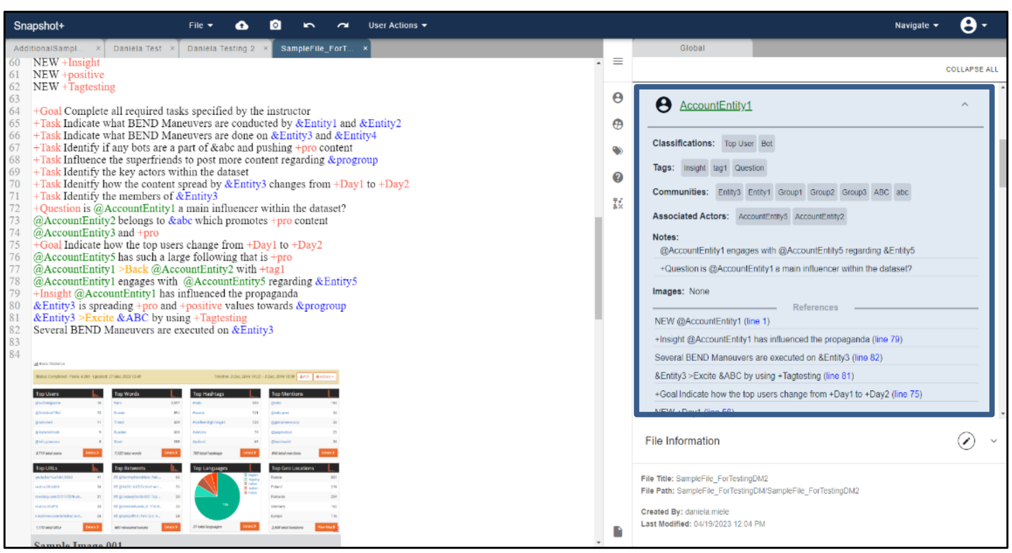

Figure 7. Entity Card for AccountEntity1 is Shown on the Right Panel of the Scratchpad .

Tracking Threads of Analysis

The scratchpad affords the ability for two types of information the be recorded: (1) the *content* the analyst is inputting and (2) the order the analyst is inputting content. The first type of information has been dynamically linked (discussed in a prior section), shown in entity cards (discussed in a prior section), and visualized in various widgets (discussed in a subsequent section). The second type of information provides insight on the approach a user took throughout the analysis process. Figure 8 illustrates a visualization based on this second type of information which includes the type of actions a user took, what type of activity, and in what time sequence. Types of Actions describe the major actions a user can take including making certain types of linkages or a new entity for a user driven data type. There are three types of activities that can be associated to an action type: add, edit, delete. For example, a user can add a new account entity, or a user can delete a line in the scratchpad that removes a link between an account and group. These actions are represented as dots and the activity type is represented as the color of the dots. Figure 8 shows these colored dots over a time sequence for a specific user. This visualization is an example of how the data collected in the scratchpad could be used; however, this is only an initial visualization.

Beyond this visualization, this collected data could be used to trace an analyst's thread of analysis. The "threads of analysis" is a term that is used to explain how an analyst approaches an investigation. It is thought that an analyst will follow a hypothesis or insight and then go on to another other (Hutchins et al., 2006). Tracking this process could provide insight to analyst perspective and could indicate how the analyst moved through their train of thought on a certain topic. The tracking of the threads of analysts could provide new areas of research including understanding how varied levels of analysts approach the same task, identification of weaknesses in an argument, and potentially identifying optimal approaches for analysis which could be used for future training.

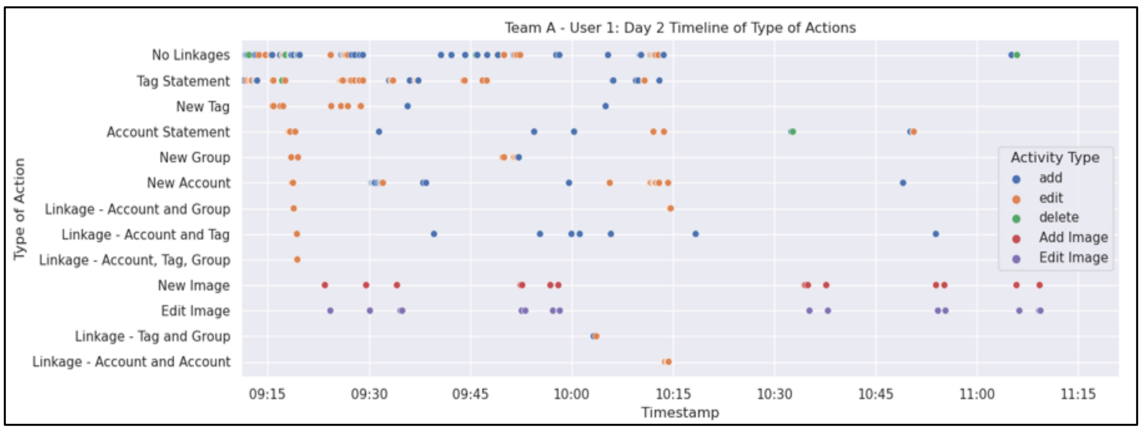

Figure 8. Output of Tracking the Analysis Process Through Performed Actions

Use Case #2: Analyst's Workspace

The analyst's workspace refers to a UI client that provides analysts the ability to view information added into the scratchpad in visual representations. The goal of the workspace is two-fold: (1) Aims to provide analysts with a testbed for identifying further connections or points for analysis based on patterns documented in visual representations. This continues the progress toward schematizing and developing hypotheses for future investigations and (2) The workspace equips analysts with an easy-to-use structure for presenting material found and documented in the scratchpad for their tasked analysis. Workspace visualizations are dependent on the data types and linkages inputted into the scratchpad which are then organized into a knowledge graph to create outputs in the form widget types including nodal graphs, lists, entity cards, and notes (see Figure 9). The note widget is a freeform text widget that allows the user to input text to explain insights found. The rest of the widgets (nodal graphs, lists, and entity cards) are based on the scratchpad files the user chooses to import for analysis, these are discussed in detail below.

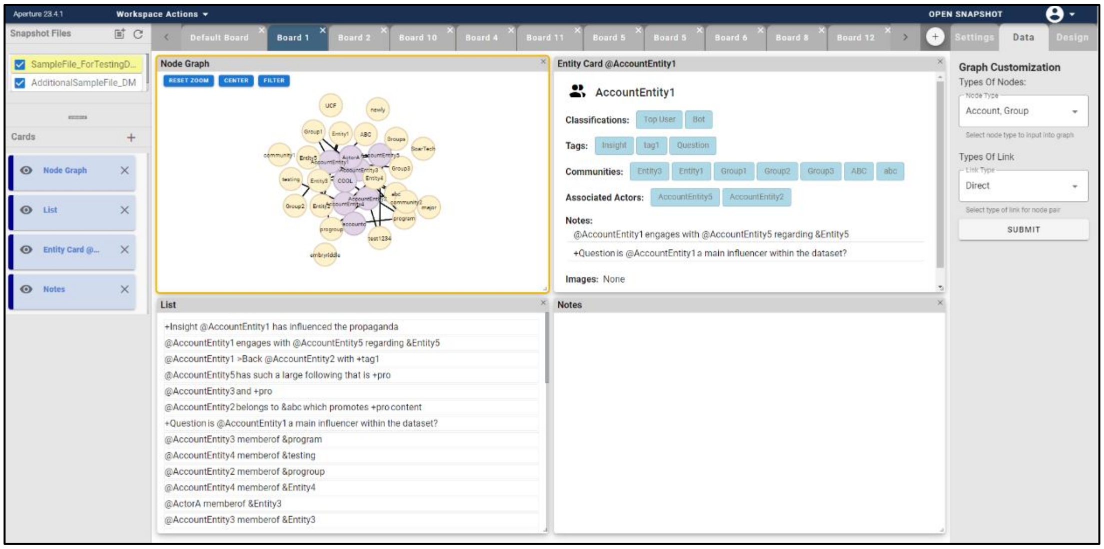

Figure 9. Analyst Workspace Tool Nodal Graph Visualization (Account, Groups, Tags)

Nodal Graph Widget

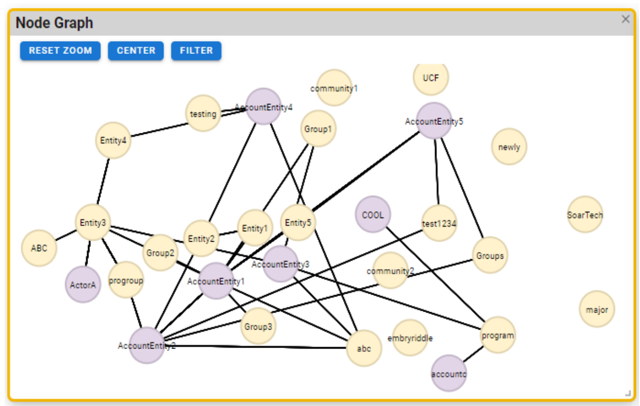

Figure 10. Nodal Graph Widget Example

Nodal graphs are complex visualizations that allow that display the interrelationships of data between different nodes (data types) connected by lines indicating a relationship or link is present. Intelligence analysis work typically occurs within teams comprised of several members, resulting in replicated data, or repeatedly mentioned entities that an analyst may not identify as influential or significant until all data is combined and compared. Figure 10 shows how nodal graphs allow all information related or linked to a specific entity to be viewed in one location. For example, three different analysts may be mentioning the same account known as “@AccountEntity4”, but the data is spread across several files so the analysts may not realize how important this entity is. By constructing a nodal graph that illustrates all connections related to all accounts, the analyst may identify that there is a significant node cluster surrounding “@AccountEntity4". This would be indicative of analysts needing to further investigate the individual and their role within the dataset and decision making that needs to occur. Nodal graphs also provides analysts with the ability to engage in design decisions that will aid in making their visual more salient or represent different patterns of information that supply different meanings to the context of the graph. This can include the ability to dynamically scale the size of nodes based on the number of connections, alter the thickness of the line representing connections based on the number of connections, hide unwanted node clusters, and designate colors for salience.

List Widget

The utilization of lists to communicate findings is very powerful as it identifies the contents of the dataset or evidence file constructed in a straightforward manner that eliminates ambiguity. The workspace list widget is depicted in Figure 11. In relation to social media intelligence analysis work, analysts may be interested in viewing only the accounts they classified as top users to pinpoint which accounts need to be closely monitored. Workspace lists provide analysts with preset controls including the specification of the data type(s) to view, Boolean logic filtering, and item selections to easily set up the desired list. For example, if the analyst adds several accounts and marks them as members of a specific group known as "&group1", the analyst may only want to identify those group members in one consolidated area. To accomplish this, the visualization tool will allow the analyst to select the data type of accounts and then filter by "&group1". The incorporation of visuals allows for greater flexibility to be added throughout future iterations of SOCINT tooling that equips analysts with visualization aids. These future functionalities could include the inclusion of reference numbers (i.e., @accountA was referenced 16 times within a single scratchpad file and 55 times globally) or linkage counts (i.e., @accountA was linked to 10 groups) to better indicate what entities are the most influential or most mentioned.

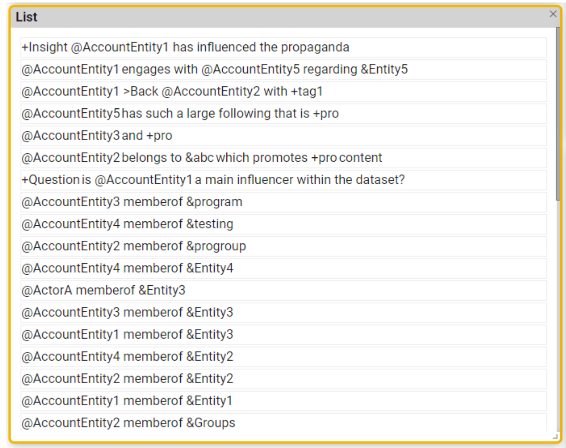

Figure 11. List Widget Example

Entity Card Widget

As mentioned, and described in relation to the scratchpad features, the workspace tool also supports the visualization aid of Entity Cards. Within the analyst's workspace this feature allows analysts to view automatically generated "cards" with all identified and linked information pertaining to either a specific account or group analysts have added to the selected or specified scratchpad files. This creates an easy reference for analyst to have all related information to an account or group in one location rather than the information being scattered throughout several scratchpads and across analysts.

REVISITING ANALYST CHALLENGES

Given the nature of intelligence analysis being a nonlinear process that is typically influenced by a strict timeline, analysts must be able to move quickly throughout the entire processes from data collection and entry to data analysis with the goal of identifying patterns and next steps. This setting yields common challenges indicated by the research (Hutchins et al., 2006; Patterson et al., 2001) which can be categorized into two overarching general associations – externally and internally induced challenges. The following sections will indicate how the features available within the prototype tools of the analyst's scratchpad and workspace aid in the moderation of these identified analyst challenges.

Externally Induced Challenges

External challenges refer to influences dictated by the given setting, scenario, or environment that the analyst themselves have no control over. Challenges influenced by external factors include time pressure, high workloads, and uncertainty.

Intelligence analysts mainly work in team settings where there are several analysts working on a specific task with a single intent. The need to delegate tasking is highly important to ensure that the goal is accomplished within the time constraints (pressure) the analysts are working under. By equipping SOCINT tools on a platform that encourages collaboration, analysts can more effectively divide tasking and then easily reunite to discuss, compare, and combine data near the end of an analysis. Without a system or tool that encourages collaboration, analysts may spend more time developing a protocol for dividing work, therefore reducing the time they have available to spend on the analysis and increasing uncertainty of knowing what their teammates are actively working on.

The structured ontology allows for the alleviation of uncertainty by specifically allowing analysts to engage in both a predetermined and user driven data framework that allows for the easy aggregation of data across files, analysts (users), and teams. This decreases the level of uncertainty that can be inputted into the dataset as they are limited to the five key data types authorized by the ontology. The feature of dynamic linkages among the available data types specifically targets the moderation of time pressure and workload challenges experienced by analysts. By incorporating the automation of dynamic linking of information through analyst's text inputs, there is a significant time reduction needed by the analyst to manually scan the created documents to identify all associations needed for a successful analysis. Once the data is linked, analysts can quickly transition to the analysis and pattern recognition actions for their assigned task.

The workspace visualization features allow for reduction of both workload and a solution to a scenario with increased time pressure. Specifically, there are various ways in which the data can be represented to elicit pattern detection allowing for a more fine-tuned approach to the analysis by being able to dissect the content quicker than other standard tooling could provide analysts. In addition, developed visuals can be used to guide analyst analyses by being recycled, modified, or recreated across tasks and teams. For example, an intelligence analysis may occur over the course of several days and to compare the data and modifications to the datasets analyzed across each day, the same visuals could be used and replaced with the various scratchpad files made for each day. This would reduce the workload of analysts by eliminating the need to spend too much time making a debrief or replicating previous actions to obtain comparison data.

Internally Induced Challenges

Internally induced challenges refer to individual or unique personal characteristics that are difficult to control or standardize across analysts or teams. These challenges include the potential for error, expertise level, and complex human judgements.

Incorporating NLP features such as suggestion capabilities in the design of future analyst tooling (i.e., the scratchpad) can aid in overcoming these types of challenges. Future NLP work could include the investigation and validation of the techniques outlined in this paper by the comparison of analyst productivity and accuracy with and without these functionalities embedded in their analysis tools. This future research could include a concentration on the ability to overcome human made errors that may influence the integrity of the information collected for analysis. Humans are not infallible and will inevitable may minor inconsistencies throughout their note taking that could influence conclusions drawn at a later point in time. For example, an analyst will likely be under temporal pressure and may be rushing to collect evidence leading to the mistyping or misspelling of entity names. If this error is not caught in the moment in which it is executed, there is a high likelihood that the author will failure to recognize this was a mistake and other analysts may take the text at face value.

Additionally, the use of a structured ontology and data management system helps drive analysts into a standardized framework and organization pattern for the collected evidence from various data sources. The implementation of an ontology allows for analysts of different expertise levels to have and enter in unstructured text or thoughts into a tool and have it then structured in a format that can be understood by individuals of all expertise levels. Error potential is then reduced once the information is developed into a standardized format that all analysts can understand regardless of who recorded the information. Once generalizability is easily incorporated into the data, all analysts can then use a visualization tool (i.e., the workspace) to then analyze the linkages and relationships found within the dataset. Specifically, this feature becomes powerful for reducing error and overcoming human judgements as it allows for analysts of different expertise levels to analyze the same data from different angles, maximizing the judgements made regarding the information.

CONCULSION

The current field of social media intelligence analysis provides valuable insights into entities that need further research and investigation to reduce tensions that may arise among communities of conflicting values or outlooks. However, with the rapid growth of social media users and the exponential expansion of social media platforms, substantive analyses of these communications can become quite difficult if intelligence analysts are not equipped with the proper tooling and mechanisms to accomplish the intended goal(s). While the current analysis execution consists of SOCINT tooling for data collection, there is a lack of methodologies that support the organization of information and the identification of patterns that may be collected. This forces intelligence analysts to exhibit non-standardized workflows and the inconsistent use of tools that inhibit analysts' ability to efficiently and effectively identify entities to monitor or pursue in relation to a given task. Specifically, the crucial processes of sensemaking accomplished via the analysis of collected data may be compromised as incorrect associations or missed connections may occur because of the unorganized high volume of data, leading to inappropriate schema development. Therefore, there is an essential need to standardize the intelligence analysis process that is congruent with their goals and provide analysts with an easy-to-use system to accomplish data organization on a large scale. The goal of this paper was to identify features and mechanisms that may be suitable to include in future SOCINT tooling applications that will primarily aid in the mitigation of complex challenges intelligence analysts have been documented to experience during data collection and synthesis. Overall, by targeting the moderation of these specific challenges research over the years has indicated in future SOCINT designs, tooling can directly cater to analyst workflow and needs promote more efficient individual and collaborative experiences that contribute towards reaching valuable conclusions or patterns that may have not been indicated otherwise.

ACKNOWLEDGEMENTS

The research reported in this paper was funded by the Office of Naval Research (ONR), Contract Numbers N68335-21-C-0840, N68335-20-C-0307, and N68335-20-C-0085. The authors thank teammates and collaborators for contributions to concepts and implementations reported here. We also thank Dr. Rebecca Goolsby for contributions to problem definition and solution context that informed our approach. The views expressed in this paper are solely those of the authors, and do not necessarily reflect the opinions of the Office of Naval Research, or any other Department of Defense agency, unless stated in official directives.

REFERENCES

Beskow, D. M., & Carley, K. M. (2019). Social cybersecurity: an emerging national security requirement. Carnegie Mellon University Pittsburgh United States.

Hutchins, S. G., Pirolli, P. L., & Card, S. K. (2007). What makes intelligence analysis difficult?: A cognitive task analysis. In *Expertise out of context* (pp. 283-318). Psychology Press.

Marcellino, W., Smith, M. L., Paul, C., & Skrabala, L. (2017). Monitoring social media: Lessons for future department of defense social media analysis in support of information operations. RAND NATIONAL DEFENSE RESEARCH INST SANTA MONICA CA.

Patterson, E. S., Roth, E. M., & Woods, D. D. (2001). Predicting vulnerabilities in computer-supported inferential analysis under data overload. *Cognition, Technology & Work*, *3*, 224-237.

Pirolli, P., & Card, S. (2005, May). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In *Proceedings of international conference on intelligence analysis* (Vol. 5, pp. 2-4).

Russell, D. M., Stefik, M. J., Pirolli, P., & Card, S. K. (1993, May). The cost structure of sensemaking. In *Proceedings of the INTERACT'93 and CHI'93 conference on Human factors in computing systems* (pp. 269-276).