Automated Text Analysis for Intelligence Purposes: A Psychological Operations Case Study

Stefan Varga (iD)**, Joel Brynielsson** (iD)**, Andreas Horndahl** (iD)**, and Magnus Rosell** (iD)

Abstract With the availability of an abundance of data through the Internet, the premises to solve some intelligence analysis tasks have changed for the better. The study presented herein sets out to examine whether and how a data-driven approach can contribute to solve intelligence tasks. During a full day observational study, an ordinary military intelligence unit was divided into two uniform teams. Each team was independently asked to solve the same realistic intelligence analysis task. Both teams were allowed to use their ordinary set of tools, but in addition one team was also given access to a novel text analysis prototype tool specifically designed to support data-driven intelligence analysis of social media data. The results, obtained from the case study with a high ecological validity, suggest that the prototype tool provided valuable insights by bringing forth information from a more diverse set of sources, specifically from private citizens that would not have been easily discovered otherwise. Also, regardless of its objective contribution, the capabilities and the usage of the tool were embraced and subjectively perceived as useful by all involved analysts.

S. Varga
KTH Royal Institute of Technology, Stockholm, Sweden

Swedish Armed Forces Headquarters, Stockholm, Sweden
e-mail: svarga@kth.se

J. Brynielsson (✉)
KTH Royal Institute of Technology, Stockholm, Sweden

FOI Swedish Defence Research Agency, Stockholm, Sweden
e-mail: joel@kth.se

A. Horndahl · M. Rosell
FOI Swedish Defence Research Agency, Stockholm, Sweden
e-mail: andreas.horndahl@foi.se; magnus.rosell@foi.se

1 Introduction

Vast amounts of easily accessible data on the Internet is generated every day by news outlets, individuals, and other information sources. The traffic volumes that produce this sea of data, too big for any human or group of humans to process without help, are predicted to increase even more in the future [13]. In other words, *information overload*, which simply put is about receiving too much information (to handle) [20], sometimes impairs the human ability to make use of available information. As the right pieces of information have the potential to be of value to some individuals or groups, it is desirable to possess a capability that takes advantage of the possibilities of online data to the largest extent possible.

A set of tools that enable users to sift through huge amounts of data in search of sought after information that is relevant to them, is software in the form of data mining and other analytical tools. It is generally thought that the end outcomes for various stakeholders will improve when such tools and techniques are employed in a systematic fashion. Even if that assumption seems highly plausible, there appears to be limited research that either confirms or proves such a hypothesis wrong. On the contrary, the (positive) value of text mining and similar techniques are often taken as a given fact in the literature.

This work seeks to examine the actual contribution of the use of a text mining tool for a real life intelligence task. In the literature much attention has been paid to the investigation of matters related to the *usability* of software in terms of design and function [44]. When it comes to the actual *usefulness* of software in relation to some task, considerably less scholarly articles are available. Furthermore, it is not always obvious that computer software contribute to the productivity of organizations at all [35, 45]. It is obviously of value to know if a piece of software actually contributes to overall organizational goals or not. In this chapter the usefulness of a piece of text analysis software for the purpose of intelligence analysis is examined.

1.1 The Intelligence Field

Because the context of this case study is set in an intelligence analysis setting, some words about the intelligence field is given to enlighten the reader.

It seems that intelligence is not consistently defined in the literature, as many articles within the field begin with quite extensive discussions about basic definitions [6]. In the national level context, however, Bimfort already in 1958 [7] suggested that intelligence is about collecting and processing information about foreign countries and their agents, that is needed by a government for its foreign policy and national security goals. This definition will stand for the purpose of this chapter as well.

According to U.S. doctrine Joint Publication 2-01 [51], which has similar definitions as other countries, the objective of joint intelligence operations in a

military context is to provide accurate and timely intelligence to commanders that gives them an understanding of the operational environment with, in particular, regard to adversary forces, capabilities, and intentions. The goal of intelligence analysis is therefore to produce accurate assessments of the current state of affairs, as well as sufficiently good estimates about the future that can be of use to various decision-makers [14].

Although this mission is simple enough in theory, (good) analytical work is not trivial. A number of uncertainties affect the quality of the final analysis, and the process itself is not straightforward as it typically involves a mixture of imaginative and critical reasoning [8]. Sometimes analysts are in possession of a piece of information, and in search for missing pieces to fit into a hypothesis. At the same time the analysts may also search for supportive evidence for multiple hypotheses [8]. In addition, most intelligence processes strive to collect information from several mutually independent sources in order to, hopefully, reduce uncertainty. Open source information that is readily available and easily accessible is certainly often used in this respect.

In a military intelligence context the scope of the analytical interest for different "situations" as discussed above, varies with the hierarchical levels of war, e.g., the tactical, the operational, and the strategic level, which highlights diverse aspects of awareness. Although the exact meaning and classification of the different levels vary between nations and organizations, the existence of multiple hierarchical levels is uncontroversial. In the following, Swedish doctrine is used to exemplify. The differences and boundaries between the levels have diminished over time due to the dynamics and complexities of modern conflicts. The same pieces of information can sometimes answer to the intelligence requirements of multiple levels. The greatest distinguishing factor between the levels is the time perspective [48]. The goal of the tactical level intelligence function is to support the individual (military) unit in its planning and mission execution. The tactical level deals with a limited geographical region on the battlefield and the character of the questions that need answers are concrete. The relevant timeframe for these activities is the immediate; it ranges from a day to a week, and sometimes from an hour to days [48]. The goal of the operational level intelligence function, on the other hand, is to support ongoing or planned (military) operations. Here the area of operation (theatre) is the region of interest. The questions are diverse, both concrete and abstract [24], and the timeframe is the intermediate; it typically ranges from a week to months [48]. At the strategic level the goal is to answer questions that may be abstract and cover a diversity of different topics with unclear relations, and to produce intelligence in the form of estimates that can be used to, e.g., create policies and military plans, and to inform about security measures on a national level. The timeframe for strategic intelligence ranges from months to several years [10]. The strategic level, thus, concerns and anticipates events of far-reaching political, diplomatic, social, economic, and military significance, that often revolve around the questions of war, peace, and stability [37].

With the wide variety of requirements as indicated above, a national level intelligence agency needs to have different types of intelligence sources at its

disposal. A commonly accepted taxonomy divide some of the source types into open source intelligence (OSINT), human intelligence (HUMINT), measurements and signatures intelligence (MASINT), signals intelligence (SIGINT), and imagery intelligence (IMINT) [14, p. 104]. With the rapid growth of the Internet and the availability of data, OSINT has grown to become an important collection discipline for intelligence purposes; not only for government intelligence functions, but also for civilian use [46]. OSINT, more specifically, is intelligence that fulfills specific intelligence requirements and is produced from publicly available information from both traditional media and web-based sources, that is, information that anyone can lawfully obtain by request, purchase, or observation [51]. Glassman and Kang [22] characterize the baseline work methods for OSINT work to be (1) the search for relevant information, (2) the organization of the information, and (3) the differentiation of it, with the goal of transforming (converting, translating, and formatting) text, graphics, sound, and motion video in response to users' intelligence requirements [51].

Heuer [27] suggests that there are in general two fundamental but different methods to address intelligence analysis problems: the conceptually driven approach, and the data-driven approach. The conceptually driven approach requires the presence of an analytical schema, e.g., some model, and the results of the analysis are directly correlated to the availability and the quality of data that is fed into the model. For data-driven analysis, on the other hand, there may not be a well-developed analytical model available, but rather an abundance of available data. In this latter case the challenges are not mainly about acquiring the data, but rather to find and select the relevant pieces of information that can be used to form sensible hypotheses for the intelligence problem at hand. A potential source of error when using the conceptual approach, is that research has shown that long-standing general beliefs and preconceptions by analysts affect the results of assessments even if there is available data that support other outcomes [33].

To this respect the availability of vast amounts of data that the Internet provides, has changed the work for intelligence analysts [18], and is a good match for the requirements of data-driven analysis. In particular, it has been found that social media posts provide relevant data that may be exploited for intelligence analysis. Several types of analyses can be made based on social media data, e.g., text analysis, social network analysis, and trend analysis [47].

1.2 Research Questions and Outline

The overall purpose of the research presented in this chapter is to investigate how automated text analysis can contribute to solve ordinary intelligence analysis tasks. To do this, a case study that sought to answer the following two research questions was performed:

– Is PhraseBrowser, a specific instantiation of a text analysis tool, perceived as useful for solving typical analytical tasks?
– Does the use of the text analysis tool improve the quality of a typical analytical deliverable?

To address the research questions, a research design that asked analysts to make an intelligence assessment of a psychological operations case was created. The case of the poisoning of the former Russian intelligence officer Sergei Skripal and his daughter Yulia in Great Britain was chosen. The rationale for choosing this incident was that it was well covered by multiple media outlets of different types as well as other data producing sources. The topic was reasonably well within the ordinary field of interest for the studied subjects, the scope of the intelligence analysis requirement was fairly realistic, and some amount of different narratives and counter-narratives, e.g., obfuscation, could be expected to be launched and spread to flourish on the Internet.

The remainder of this chapter is structured as follows. Section 2 provides an overview of the area of psychological operations, and related work. Then, in Sect. 3, a description of the text analysis prototype tool PhraseBrowser follows. Next, Sect. 4 describes the undertaken methodology covering the research design, the observational study setup, and the execution of the study. In Sect. 5 the results are presented. Then, the findings as well as issues of validity are discussed in Sect. 6. Finally, the conclusions are presented in Sect. 7.

2   Background

After having introduced the intelligence *field* above, this section describes the framework for intelligence *analysis*, psychological and influence operations, and lists some related work.

2.1   Intelligence Analysis

The basic process for producing intelligence is about collecting information, making analyses, and then creating an intelligence product that can form the basis for decision-making. It is primarily a cognitive process, an activity that takes place in the analyst's own head. It is not always clear exactly how the assessment goes and how it ought to be done in different situations, but on a methodological level there are strong links between the intelligence profession and that of scientific work: intelligence work is largely about formulating hypotheses, and examining and falsifying these hypotheses if possible. But there are also large differences relative to scientific work in that intelligence work has a connection to operational work and the need to deliver forecasts within a given timeframe based on currently available information, regardless of whether it is judged to be of sufficient quality or not.

Rather than having the character of a "secret science" *in itself*, the characteristic features of intelligence work can thus be said to be about the (indeed often secret) *application* of scientific methods and approaches on information and intelligence questions that are operational and strategic, rather than scientific, in nature. In an analysis of the intelligence subject, and in an attempt to go from established practice to a more structural approach of it, Agrell and Treverton [1, p. 279] state this in the following terms:

> Intelligence analysis has the potential to become an applied science. Its purpose would be managing the uncertainty in assessments of threats and possibilities based on incomplete, unreliable, or uncertain data in a context in which demand requires those assessments irrespective of the limitations. Defined in these terms, intelligence analysis stands out as a genuine cross-disciplinary science in-being, with a theoretical basis and a set of methods not limited to any single subject matter or field of analysis but rather adapted to every specific application.

As noted, management of information and its related uncertainty play a central role, and the means to measure precision, quality, and utility—so-called information awareness [5]—is crucial.

2.2 *Psychological Operations*

The *root causes* for armed conflicts, or indeed any controversies between humans, may be of different kinds, e.g., ideological differences, competition for scarce resources, etc., but the *end goal* in conflicts is always to impose the will of one party on the other(s) in one way or the other, as expressed by, e.g., the ancient Chinese war theorist Sun Tzu [49], as well as the highly influential Preussian war theorist Carl von Clausewitz [54]. But there are other methods to affect the will of an opponent than by using physical force: psychological and influence operations are ways to achieve this.

Psychological operations, Psyops, within the context of military operations is a part of offensive information operations that aim to influence perceptions, attitudes and ultimately change the behavior of foreign approved target audiences [50]. Another name for Psyops that is sometimes used, is military information support operations, MISO [52]. Such activities are sometimes seen as a key enabler in a military commander's campaign plan [4, 50], and they can be conducted with both short-term and long-term goals, according to doctrine. The overarching information operations field also involves other activities such as civil affairs, computer network attack, deception, destruction, electronic warfare, operations security, and public affairs [4]. Psyops and related activities, however, are not only part of U.S. and Nato doctrine, but integral parts of Russian, Chinese, and other countries' national security or military doctrines as well [4]. A related term is influence operations, which consists of similar types of activities, which are not necessarily conducted in conjunction with military operations. Influence operations primarily consist of non-kinetic, communications-related, and informational activities that aim to

affect cognitive, psychological, motivational, ideational, ideological, and moral characteristics of a target audience [36]. Influence operations can be carried out by government organizations other than the military, as well as by civilian information outlets. Furthermore, operations can be conducted openly (with a named source), covertly, and clandestine.

The boundaries between roles, mandates, and who does what, as well as between military and civilian organizations, with regard to influence operations is not always clear-cut [36]. For the sake of simplicity, only the term Psyops is used in this text. The approved targets for operations can be individual persons, groups and networks, adversary leadership coalitions, and the (mass) public [36]. On the defensive side of Psyops, a possible task would be to determine if oneself, or some other target audience, is on the receiving end of adversarial operations. Such operations could aim to change the perceptions/views, clog the perceptions, or delegitimize credible news outlets [39]. To accomplish such an analysis, Nato doctrine [39], which is used to highlight Psyops principles here, prescribes that a detailed examination of source, content, audience, media, and effect (SCAME) is carried out. The SCAME template can also be used in the planning process for offensive operations.

2.3   Related Work

There seems to be a general consensus about the perceived usefulness of data mining tools, in that they can help find useful information in many instances. Examples of precisely how such tools contribute to the information gathering efforts, however, and more to the point, the quality improvement of the end products, are harder to find. A study commissioned by the British Government [38], however, concluded that text mining contributes with several benefits in the context of academic research. The study found one improvement to be overall increased researcher efficiency and research quality. Mining was further attributed to bring about the ability to "unlock hidden information", i.e., insights about underlying non-obvious connections between texts, as well as the capacity to develop new knowledge and the ability to explore "new horizons" [38, p. 19]. The conclusions were based on results from several case studies.

Pal [41] listed application areas where data mining tools were of great use both within the public and the commercial sector. For the public sector, fields such as, for example, scientific enquiry and research analysis, criminal investigation and homeland security, via health insurance and healthcare applications, were found to benefit from text mining. Within the purely commercial sector, areas such as customer segmentation and targeted marketing, finance, etc., were mentioned.

Several scholars have noted that publicly available data on the Internet can be systematically used for intelligence purposes [18, 40]. It has been observed that some of the automated data processing techniques used for systematic processing of corporate data, i.e., business intelligence, have migrated into security intelligence [10], and that the further development of these techniques has the potential to

provide a competitive edge [12, 31, 42]. There are, however, challenges as of how to find and extract relevant and meaningful pieces of information relative to the task at hand.

Automated text analysis using methods and tools from the field of natural language processing has been proposed as a way to off-load some of the selective and interpretative work from human intelligence analysts. With regard to more closely related government and military intelligence tasks, Guo et al. [23] did work on entity extraction from human-generated tactical reports to support intelligence analysis. They extracted entities such as organizations, locations, persons, etc., with promising results. Razavi et al. [43] sought to extract information about risks in maritime operations.

Other examples with applications from the commercial sector include work by He et al. [25] that explored how the use of text mining was useful for companies in the pizza industry. They concluded that text mining of social media adds useful pieces of information, for example, in companies' quests to understand the pizza market. Alex et al. [2] were able to show an increase in efficiency, e.g., by the reduction of work time, in a biomedical data system scenario where natural language processing technology was used for curation.

3   PhraseBrowser

This section serves to describe the scope and function of the prototype tool PhraseBrowser. PhraseBrowser is continuously developed at the Swedish Defence Research Agency. It is one of several prototypes in a framework, aiming to highlight the possibilities given by web and text analysis tools to analysts. The development process of PhraseBrowser, to this respect, provides opportunities for researchers and practitioners to engage in mutually beneficial discussions.

3.1   Overview

PhraseBrowser is a text analysis prototype tool designed to support analytical work by processing Twitter data. The idea is to provide the user/analyst with several perspectives of the collected data through predefined themes. The different perspectives provide an overview of the data that may guide an analyst to select interesting subtopics and drill down further to find specific contents of interest. Hence, the prototype tool may be useful both for monitoring a subject or area of interest, and for conducting research to answer specific questions. Although the prototype tool comes with predefined themes, it is a relatively easy and straightforward process to quickly add simple first versions of new themes. This versatility makes the tool relevant for analytical work that concerns a wide variety of topical issues as well as for time-sensitive tasks.

3.2 Phrases

PhraseBrowser presents phrases to the user sorted by so-called phrase types. A phrase is defined as a sequence of one or more words, and each phrase has a type. A few example phrases (and their phrase types) from the set of tweets used in the study are: "lab says" (phrase type: "General Phrase"), "Boris Johnson lied about Skripal" ("Explicit Untruth"), "UK" ("Location"), "Yulia" ("Person"), and "30 questions" ("Counted Thing"). See Table 1 for more examples of phrase types and phrases. At the time of the study there were more than 50 different phrase types of varying quality available to the analysts.

In the interface the user can choose a type, e.g., "Counted Things", leading to that a list of phrases of that type are presented along with statistics on how many tweets the phrases were used in. To read the texts containing any particular phrase, e.g., "30 questions", the user simply clicks on that phrase. Each piece of text works as a link to the original tweet on Twitter, where more context may be found by studying the content of corresponding accounts, etc.

The phrases may be identified using any automatic method. In the version of the tool used in the observational study presented herein, a third party library was used to identify entities (see Sect. 3.5), and the rule language described in Sect. 3.6 was used for all other phrase types.

3.3 Predefined Phrase Types and Filtering

In the current study the analysts worked with a set of predefined phrase types, and were not able to alter them or add new ones. Table 1 displays some of these phrase types, and in the following they are described a bit further:

"General Phrases" tries to capture any kind of content based on part of speech tags. This is an example of a phrase type that results in many phrases—perhaps too many. It would probably be useful to exchange or complement this phrase type with a machine learning method. For now this is the most general phrase type, primarily used to explore content without looking for any of the specific content that most of the other phrase types try to capture.

"Counted Things/Persons" is defined using other phrase types capturing counts, and at the same time "things" and/or "persons". One possible use of this phrase type is to look for differing numbers being given in some context. Sources may exaggerate the number of protesters at an event, for instance.

"Entities" such as "Person", "Location", and "Organization" are found by an entity detector (see Sect. 3.5). These entities are reused by several of the other phrase types, e.g., the "Counted Persons" phrase type mentioned above.

"Explicit Untruths" captures phrases that use any word in a long list of words explicitly related to deception, propaganda, misinformation, fake news, etc. The idea is that it is potentially interesting whenever someone writes that something is

Table 1 Examples of themes/types available in the PhraseBrowser tool along with phrases extracted from the data set used in the study

Theme/Type	Phrase	#
General Phrases	*All*	2,806,084
	lab says	9297
	used in Skripal poisoning	8842
	produced in Russia	5284
	disinformation campaign in Britain	4326
	was in US	4205
	Porton Down research laboratory	3401
	research laboratory has told Sky News	3373
	to Sergei Skripal	3122
	that Christopher Steele	2803
	of Yulia Skripal	2563
Counted Things/Persons	*All*	105,931
	60 Russian diplomats	1406
	30 questions	927
	two weeks	861
	14 simple questions	823
	hundred narratives	383
	two BBC colleagues	294
	20 European countries	280
	2800 Russian bots	232
	23 British diplomats	212
	two people poisoned	203
Entities	*All*	5,601,461
	Skripal	373,735
	Russia	113,772
	UK	107,982
	Yulia	97,560
	Salisbury	32,903
	Porton	25,029
	Novichok	22,236
	Putin	22,017
	OPCW	19,292
	Theresa May	14,136
Explicit untruths	*All*	62,579
	propaganda	4532
	Boris Johnson lied about Skripal	1641
	Moscow's lies	427
	Kremlin propaganda	417
	UK lies	204
	Theresa May is lying	155
	"Russia bot" narrative	153
	the Skripal narrative	142
	lying about the source of Novichok	111
	Downing Street spin-master	82

a lie, fake, misinformation, etc.: either the statement is true, or the person writing it has an agenda... Hence, these phrases often contain suspicious/interesting statements that can potentially be used as a starting point for further analysis. In Sect. 3.7 some simplified rule examples for "Explicit Untruths" are presented.

To drill down into the data, each phrase and phrase type may be used as a filter that narrows the search to only include tweets containing the chosen phrases. The resulting smaller data set can then be studied using the other available types and phrases. For instance, filtering the data using the phrase "Theresa May" and then looking at the phrase type "Explicit Untruths", one would only obtain the "Explicit Untruths" that co-occur with "Theresa May".

3.4 PhraseBrowser System Overview

Figure 1 shows an overview of the PhraseBrowser system. Data is downloaded in real time using the Twitter Streaming API and/or RSS feeds. The data is continuously processed by several analysis components, and stored in an Elasticsearch[1] database search engine. The system architecture and construction is scalable and adapted to parallelization, using Docker[2] for packaging of subsystems and Kafka[3] for distributing data between the subsystems.

The analysis phase contains several subsystems that process the data and adds information to it. The original data (including metadata, if any) is stored by the storage component along with the result of the analysis as (additional) metadata.
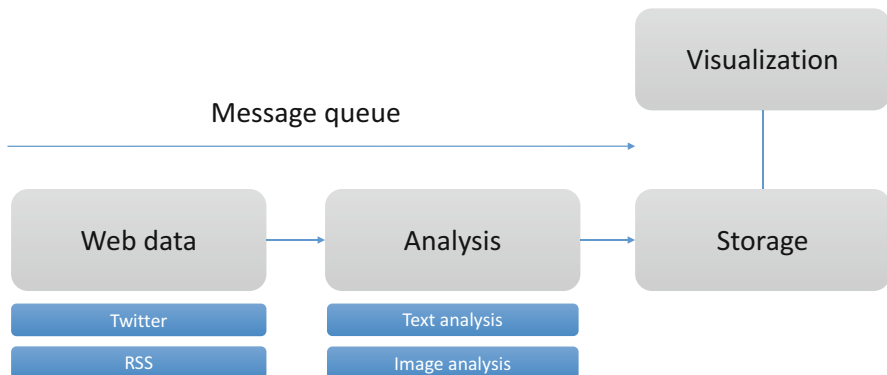

Fig. 1 PhraseBrowser system sketch. The prototype tool can handle Twitter and RSS data, and analyzes both text and images

---

By filtering on the metadata, different parts of the stored data can be retrieved and visualized in the user interface.

PhraseBrowser can be run in real time on a data stream. However, depending on the hardware used, the analysis may not keep up with the stream. In such cases PhraseBrowser continuously processes the latest tweet or RSS update, meaning that certain pieces of data may never be processed. In the present observational study a single ordinary PC was used to run the search query "Skripal,skripal" using the Twitter Streaming API, leading to that 5% of the tweets were discarded.

3.5   Text Processing

Figure 2 magnifies the part of the (text) analysis steps relevant to this study. For detecting most phrases, a rule-based approach [30] is applied. Each piece of text is first run through a natural language processing (NLP) library to divide it into sentences and tokens, lemmatize tokens and determine part of speech, and detect entities. The output from the NLP library is transformed into a simple specific text format representing each sentence in the input data. The formatted text is then sent to the rule language engine, resulting in a set of phrases for each sentence. The phrases are added to the metadata for the tweet/text as described in Sect. 3.4.

Through this construction, the rule language engine is separated from the NLP library and the only thing that has to be done to try a new library is to specify the transformation from the NLP library output format to the specific text format needed by the rule language engine. Transformations have been specified for a few different libraries. For the current work TwitIE [9], a GATE [15] pipeline specialized for microblog texts, was used.

The transformation into the previously mentioned specific text format includes turning the entities detected by the NLP library into phrases, with phrase types corresponding to the entity type. Hence, these phrases are originally detected by whatever method the library uses. For entities this is usually a combined method based on both dictionaries and machine learning.

The rule language, described further in Sect. 3.6, is flexible and constructed to make it possible to detect anything from single words using simple word lists to more complex phrases. It is language dependent in the sense that for any new language it is necessary to create a parallel text processing pipeline according to
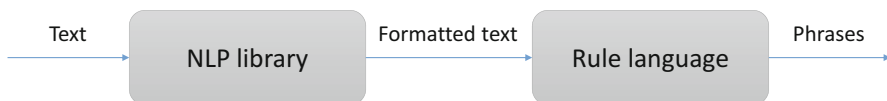
Text → **NLP library** → Formatted text → **Rule language** → Phrases

Fig. 2 PhraseBrowser text processing. Each piece of text is processed by the NLP library, followed by a transformation of the NLP library output into a specific text format. Based on the formatted text, the rule language engine produces phrases for each sentence

Fig. 2. However, after an initial learning period, it is easy to quickly create basic phrase types for a new language. So far processing pipelines and rules for English and Swedish have been implemented.

3.6 PhraseBrowser Rule Language

The rule language can be compared to regular expressions, but on token (word) level rather than character level. The concept is inspired by many previous such pattern detection methods, like for instance Hearst's patterns for finding hyponymy relations [26, 32].

Each phrase type is defined using several rules. The rule language allows for references to other phrase types, making it possible to reuse solutions and create more sophisticated rules. The rule engine processes each sentence by applying the rules in order of complexity, making sure that a rule referring to another rule is applied after the rule it is referencing has already been applied.[4] Every result is temporarily kept in a data structure containing information about position in the sentence, so that results do not need to be found twice. This makes the rule engine efficient, and less likely to become the bottleneck of the larger system. However, with a huge set of rules and/or rules that are too generic and always find phrases, the rule engine could still become a problem.

During the text analysis the rule language is applied to a single sentence at a time. If larger text blocks were to be considered, i.e., several sentences are used, too many phrases would be selected. Hence, to capture information that is distributed over several sentences, other methods need to be used. Such methods were, however, not used in the study presented herein, and will therefore not be discussed further.

3.7 PhraseBrowser Rule Language Examples

Table 2 shows a few simple examples of rules written with the rule language. The left column contains rules for two different example phrase types called "violence" and "violence_in_location". The first phrase type is a word list that simply detects whenever the listed words appear, as in the example sentences in the right column. The rule language has several features that allow for creating more useful word lists, such as using the lemma instead of the actual token and using part of speech tags.

The phrase type "violence_in_location" contains two rules that reference the "violence" phrase type and the entity type "LOCATION". The two rules also allow

---

[4]Circular references are not allowed.

Table 2 Simple examples of the PhraseBrowser rule language

#	Phrase type and rules	Example text
	name:violence	
1	violence	London has seen a lot of **violence**.
2	fist fight	There is a **fist fight** in London.
	name:violence_in_location	
3	!a(LOCATION) !any[*] !a(violence)	**London has seen a lot of violence**.
4	!a(violence) !any[*] !a(LOCATION)	There is a **fist fight in London**.

The left column shows simple rules, with rule numbers for convenience. The phrase type "violence_in_location" reuses the phrase type "violence". The right column provides examples of applying the rules. The boldface part of the example sentences is what would be found by the rule in the same row

for any number (zero or more) of any kind of token in between.[5] The first rule in "violence_in_location" (row 3 in Table 2) therefore could be read as:

> a location entity, followed by zero or more appearances of any token, followed by either "violence" or "fist fight".

The second rule (row 4 in Table 2) can be interpreted analogously. In practice the "violence_in_location" rules would detect too many uninteresting phrases, as sentences may be long and the mentioning of a location does not necessarily relate to where the "violence" is taking place. To overcome this problem, the rule language has features for stopping phrases that contain certain tokens (or phrase types) between parts of the rules, and it also allows for requiring the presence or absence of certain tokens (or phrase types) within the current sentence.

In Table 3 some simplified examples of the rules for the "Explicit Untruths" phrase type are presented. The rules show the expressive power of the rule language: if the building blocks are well thought-out, the rules can become capable of detecting many different kinds of relevant phrases. It is often easier to split more complex phenomena into parts. The "Explicit Untruth" rules used in the study, for instance, are split into a few subtypes. The precise rule structure can be accomplished using linguistic insights, but more importantly it should be based on the data at hand and be useful for the analyst.

It is imperative that either the analyst is actively involved in the creation of the rules or works in a team with an expert, since otherwise the analyst may interpret the results erroneously. An analyst at the least needs to be made aware of the limitations of the rule language. Trial and error brings the analyst a long way in creating rules that find many relevant examples, though. The precision can become high enough, while the recall obviously cannot be guaranteed.

---

[5]It is possible to specify any number of repetitions, like for instance two to three tokens, as well as specifying that only tokens from other phrase types are allowed in the repetition.

Table 3 Simplified examples of the "Explicit Untruth" rules

#	Phrase type and rules	Example text
	name:u_basic	
1	lie	Boris Johnson **lied** about Skripal
2	propaganda	It's all Kremlin **propaganda**
3	spread rumor	John **spreads rumors** about Paul
4	misinformation	The fight against **misinformation**
	name:u_obj	
5	!a(PERSON)	**Boris Johnson** lied about **Skripal**
6	!a(LOCATION)	It's all **Kremlin** propaganda
	name:untruth	
7	!a(u_basic)	That's just **lies**
8	!a(u_obj) !a(u_basic)	It's all **Kremlin propaganda**
		Many **UK lies** today
9	!a(u_obj) 's !a(u_basic)	Have you heard **Moscow's lies**?
		We are used to **John's misinformation**
10	!a(u_obj) !a(u_basic) about !a(u_obj)	**Boris Johnson lied about Skripal**
		John spreads rumors about Paul

The left column shows simple rules along with rule numbers, and the right column shows examples of applying these rules. The boldface part of the text examples is what would be found by the rule in the same row. Only the last phrase type, "untruth", is presented to the user. The others provide partial solutions. Note that rules 8–10 have been applied to two texts each, exemplifying the expressive power of the rule language

3.8 New and Improved Rules

An advantage of the PhraseBrowser rule language as well as other similar rule languages, is that they allow a user to quickly add rules that capture new themes of interest. The ability to modify the rules is an important feature of the tool in situations when the predefined themes do not match a specific topic and there are time constraints involved. Quickly adding a first version of a new theme is as easy as adding a word list containing some keywords. Although the obvious aim is to capture precisely everything that is relevant to a specific investigation, it is better to be able to retrieve at least some amount of desirable information through the inclusion of new rules, than risk ending up with insufficient data or no data at all. At the same time it is important to minimize the retrieval of noise, e.g., irrelevant data, which risks to burden the analysts tasked with interpreting the results.

In the present observational study the analysts worked with a set of predefined phrase types, and could not alter them or add new ones. If they had had that possibility, after studying the data they would perhaps have added phrase types for "poisons" and "lab results" to capture more of what was written about those topics. They could then, for example, easily have gone on to study what was written about persons or untruths mentioned in connection to those topics using the predefined phrase types.

While studying the results of using a first version of a rule, it is quite common to realize ways to adapt and extend it. Interesting phrases that appear in the texts that were captured can be added to the rule. Rules that capture irrelevant phrases can be altered using different features of the rule language. To this end, a user interface for interactive development of rules has been implemented for PhraseBrowser. This interface allows the analyst to iteratively refine the phrase type to capture more relevant data.

If a specific information requirement occurs in several investigations or over time, more effort could be put into studying the resulting data to improve the information gathering. Any number of methods could be used, including creating more sophisticated rules or training a machine learning model suited for the task. The goal is always to assist the analysts, and the rule language helps to put a first attempt in place quickly. A more sophisticated method (based on rules, machine learning, or any other method) can always be combined with the other phrase types to allow for varied ways to study the data. Also, whenever a new phrase type is added, many new combinations become possible. These combinations as well as the separate phrase types can be invaluable when a new subject needs to be studied.

3.9 PhraseBrowser Interface

Figure 3 shows the PhraseBrowser interface. Here 1,140,608 tweets have been downloaded using the Twitter search query "Skripal,skripal", as can be seen in the top gray area. In the left part (the blue area entitled "Phrases") the phrase type "Person" has been chosen and in the list of identified persons "Boris Johnson" is selected (the line is shaded). The tweets in the right part consequently all contain "Boris Johnson". Should any of the tweets seem interesting, it is possible to read them in context on Twitter by clicking on them.

Both the list of phrases and the list of tweets are usually much longer than what can be seen in the interface without scrolling through the lists. The count displayed next to the phrases denotes the number of tweets the phrase appears in, and the number presented before each tweet denotes the number of retweets.

To the left of each phrase in the left part of the user interface there are two buttons. Using the right button the user can plot the number of appearances of the phrases over time (not shown in this book chapter). When pressing the left button for a phrase a filter temporarily removing all tweets not containing that particular phrase is applied, which also removes all phrases not co-occurring with that particular phrase. In the example in Fig. 3 all phrases of type "Explicit Untruth" have been applied as filter, meaning that all displayed tweets contain both "Boris Johnson" and a phrase of the type "Explicit Untruth". This filter is shown in the top gray area as "All Untruth".

Several filters can be used at the same time. For instance one could go on and filter on "Boris Johnson" in addition to "All Untruth" and then look at the phrase type "Location" to see which locations are mentioned in tweets containing both of
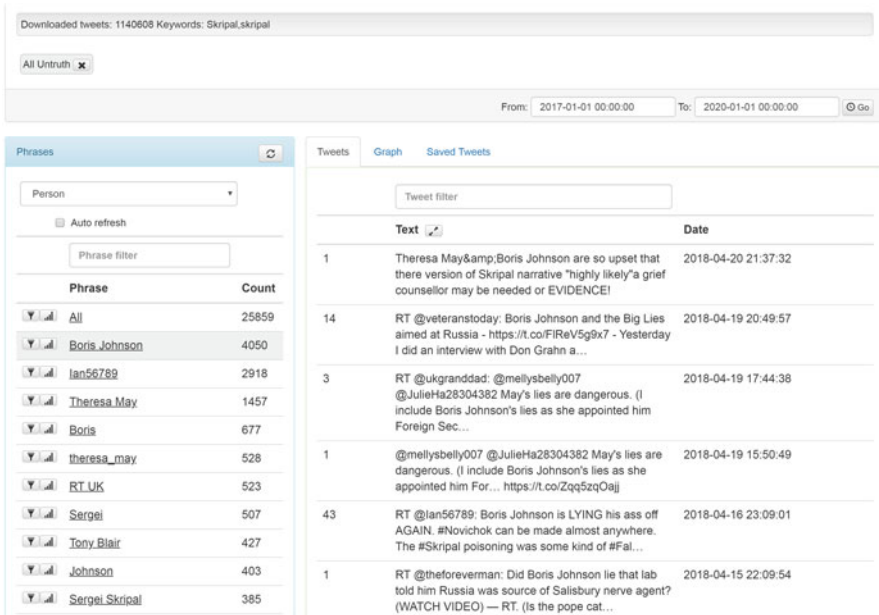
Fig. 3 The PhraseBrowser user interface. The gray area at the top is the filter area, showing the Twitter search query, the number of downloaded tweets, as well as any filters that are active. In the blue "Phrases" area to the left the user can choose a phrase type ("Person" in the figure), and the corresponding phrases are displayed at the bottom. The "Tweets" area to the right displays the tweets that contain the chosen phrases in the blue area (filtered by the filters in the gray area). Both the "Phrases" and the "Tweets" areas have filter search boxes that allow the user to search among the phrases and tweets

these filters. Each filter can also be switched to its opposite, so that, for instance, only tweets not containing "Boris Johnson" are shown.

The example in Fig. 3, as described in the previous paragraphs, shows one snapshot of the work an analyst could use the tool for. By filtering on "All Untruth" he/she solely gets tweets containing explicit untruths. These tweets may be interesting since they are likely to contain accusations regarding the truthfulness of statements centered around the Twitter search topic. To get more perspectives on the accusations, the analyst could use several of the other phrase types. For instance he/she would likely use the phrase type "General Phrases" to obtain an overview of the content. In Fig. 3 the phrase type "Person" has been used, so that the list of phrases to the left shows person names that have been seen in tweets containing the accusations. For some reason the analyst takes particular interest in "Boris Johnson" here, and reads the tweets containing both this name and explicit untruths. This may provide an insight that leads the analyst to follow up by other means or use other phrase types to browse the data.

4 Method

This section describes research design, observational study setup and execution, as well as contextual factors that were present at the time of the study.

4.1 Research Design

The observational study participants ($N = 8$) were evenly split into two teams. One team was equipped with their usual set of tools, and the other team also had their ordinary tools, but was in addition provided with the PhraseBrowser text analysis tool. Hence, the general methodological approach was to collect data and compare the results of the two teams. The prototype tool had previously been used and evaluated at the unit by their personnel for some time.

The goal was to find a group of study participants that could be assumed to benefit a great deal from the use of text mining tools in their ordinary line of work. To this respect a military unit from the Swedish Armed Forces agreed to participate in the study. One of the normal tasks of the unit is to make intelligence assessments of the information environment. The participants consisted of active duty and reserve officers, as well as permanent and part-time soldiers and civilians. The aim was to put together uniform teams with respect to their educational level, experience, and gender. The team compositions were suggested by a seasoned analyst from the military unit, who consequently had some knowledge about the research design, and therefore was not allowed to personally participate actively in the observational study. The teams were constituted as indicated in Tables 4 and 5.

Table 4 Team without the PhraseBrowser tool

Age	Gender	Military category	Work role/title	Months on job	Educational level
34	Male	Civilian	Section chief	30	B.Sc.
35	Male	Civilian	Analyst	24	B.Sc.
34	Female	Civilian	Analyst	12	M.Sc.
27	Female	Civilian	Analyst	15	M.Sc.

The designated team leader in the observational study is enlisted as the topmost person in the table

Table 5 Team with the PhraseBrowser tool

Age	Gender	Military category	Work role/title	Months on job	Educational level
26	Female	Civilian	Analyst	18	B.Sc.
29	Male	Civilian	Analyst	3	M.Sc.
37	Female	Reserve	Analyst	24	University
23	Male	Soldier	Analyst	48	High school

The designated team leader in the observational study is enlisted as the topmost person in the table

4.2 Intelligence Assessment Task

The analytical questions for the teams concerned psychological operations. The case of the alleged poisoning of the former Russian intelligence officer Sergei Skripal and his daughter Yulia was used to provide a realistic backdrop for the intelligence assessment questions. This event occurred on March 4, 2018 in Salisbury, England.

As the observational study task evolved around a Psyops scenario it was assumed that text analysis would be helpful mainly for *source* and *content* analysis (see Sect. 2.2). For *source* analysis it was hypothesized that PhraseBrowser may be of use in the search for actors and authors of information used in a Psyops campaign. For *content* analysis it was hypothesized that PhraseBrowser can help extract factual information related to a campaign. The following questions/tasks were phrased:

1. Identify and document alternative explanations that contradict the British explanation.
2. Identify possible explanations and messages that might be part of a Russian influence campaign.
3. Identify and document possible sources and routes for dissemination of pro-Russian messages according to question/task 2 above.

The two teams were asked to produce an intelligence assessment consisting of a maximum of three sheets of A4 paper. They were instructed to collect data from Swedish or English language information resources only. Every information element used in the final assessment, was to be adequately referenced and, if possible, easily retrievable.

4.3 Study Setup

The observational study was conducted in situ at the military unit. The two teams resided in rooms that are normally used for similar (intelligence) work. The observational study was carried out during a single day, with additional data collection being done the day after.

The Twitter data used was selected according to the principle of relevance sampling [34]. The search keywords "Skripal,skripal" were used, which were judged to be discriminatory enough to capture all relevant tweets concerning the poisoning event. The search query resulted in a data set of 1,140,608 tweets that was downloaded through the Twitter Streaming API between March 19 and April 23, 2018, meaning that the data collection began some days after the event.

To answer the research questions, two main types of data collection were carried out: (1) the subjective views of the usefulness of the software [16] were collected from the personnel who participated in the observational study, and (2) the objective views of their performance were judged by external observers, i.e., the rating of the quality of the intelligence assessment deliverables. Observations of the work

conducted by the teams were also made. Hence, the data was collected in four different ways:

1. The team members answered questions about the perceived usefulness of PhraseBrowser.
2. Four (4) subject matter experts, SMEs, were asked to judge the overall quality of the deliverables. They were to highlight what they saw as interesting pieces of information. The experts came from two different departments within the Swedish Armed Forces Headquarters and an independent think tank. Their experience from working as Russia analysts were 8, 10, 10, and 23 years.
3. All the tweets that were used by the teams were put in random order, and all participants were asked to judge the value of each individual tweet. They judged each tweet with regard to its perceived value (1–4), and its perceived area of use, e.g., (1) whether it added new hitherto unknown information, and/or, (2) if it served as a pointer to other related information sources, and/or, (3) if it (the account) was judged to be a new source itself.
4. One senior researcher per team was deployed to observe the work processes.

4.4   Observational Study Execution

The teams received the intelligence task simultaneously, and at the same time they were also given time for preparation. During this 1 h preparation, they had to designate a team leader, make an overall working plan, and make a time schedule. At least one person per team was required to read through the background material on the Skripal case from Wikipedia.[6]

Both teams were allowed to use all available information sources. Both looked into other sources than Twitter. Both teams used web browsers checking sites like google.com, hashtags.org, tweetdeck.twitter.com, etc. They also read the official government sites of various countries and traditional news outlets.

5   Results

In this section the data that was collected and the observations that were made during and after the observational study are presented.

---

[6]https://en.wikipedia.org/wiki/Sergei_Skripal [May 2, 2018].

5.1 Perceived Usefulness

The analysts who used PhraseBrowser expressed that they liked to use the tool in general. Among the perceived benefits they noted that it can be useful as events unfold in an area of interest. The tool was judged to speed up the whole data collection phase of an intelligence task. Another recurring theme in the observations about its usefulness was the breadth of the collection that enables the analyst to discover a multitude of perspectives, and allows the analyst to get an overview of an issue quickly. One respondent specifically pointed out that one of the strengths of the tool was that it steers analysts to look at data without the restraints of their preconceptions. Another useful feature of the text analysis tool that was voiced, was that it added valuable perspectives from apparently private citizens (and their accounts) in addition to the more readily available and accessible information from mainstream news and information outlets.

5.2 Subject Matter Expert Evaluation

The four SMEs, who normally read material in other formats that are not put together under time constraints, had some general remarks about the format of the deliverables, namely that they were short and not very well laid out. The SMEs also pointed out that they knew most of the information already, even though there were pieces of information that they had no prior knowledge about. Three of the four experts stated that the deliverable produced by the PhraseBrowser-team consisted of more alternate explanations and was "more detailed". Otherwise, the experts did not find any significant differences between the deliverables.

5.3 Information Fragment Value

The two teams drew information from a total of 25 unique tweets that contributed to their assessments. The team with PhraseBrowser used 15 tweets, and the team without the tool found ten relevant tweets that they used. The sample size of the data that was collected was too small to draw any general conclusions. However, some inferences can be made based on the collected data:

1. The ratio of very valuable tweets (rated with a score of 4), was almost equal between the team with PhraseBrowser (50%) and the team without it (47%).
2. At least one member of the team without PhraseBrowser stated that 90% of the tweets that they used brought hitherto unknown information. The corresponding number for the team with PhraseBrowser was 67%.
3. The members of the team with PhraseBrowser reported that the tweets that they used pointed to other potential sources to a larger extent than the other team.

5.4   General Observations of Work

It could be observed that the participants were well trained and had a common understanding of the work process, and worked according to a well-established staff methodology, which meant that they could quickly start working together and act as a team. Concerning working procedures, many similarities between the two groups could be seen with regard to how they approached the task and planned their work.

Both groups planned the work according to roughly the same schedule with two major work shifts with an interim discussion in between, and then a final synthesis discussion before the preparation of the final intelligence report. At the planned interim synthesis discussions, the analysts told what they had found as a basis for being able to draw common conclusions and align the further work. After the final work shift, longer discussions were held focusing on the possible explanations for the poisoning of Skripal, as a basis for writing the end intelligence report.

A whiteboard was used to note the conditions for the task with regard to available time and preconditions regarding limitations, intelligence/information, success factors, and immediate actions to be taken. For sharing joint work and working together, a large screen that everyone could easily see was used. Before lunch, plans were made regarding who should monitor which sources, the cutoff time for further information collection, and how the work was going to be logged.

During the afternoon the team leaders led the work and moderated the discussions that needed to be conducted. Discussions included, for example, which sources and objects that ought to be prioritized and monitored, and information sharing among the analysts concerning, for example, different spellings and synonyms to be used for the information collection ("Skripal", "Julia", and "Sergei" are equally interesting words to look for). During the work alternative explanations of the cause of the poisoning of Skripal were listed on the whiteboard along with preliminary intelligence confidence levels. These confidence levels were successively updated during the exercise based on which and how many sources that spoke for or against the respective explanations.

Both teams were judged to work systematically and be led by competent team leaders. They followed their work plans very well. Due to the severe time constraints there was some amount of stress in both teams, but the atmosphere was calm and professional. Both teams planned for and had a short 15-min break during the observational study, meaning that they invested an equal amount of time in solving the task.

5.5   Observations of Work Related to the Use of PhraseBrowser

The team with PhraseBrowser chose to designate one of the most proficient users of PhraseBrowser as team leader, which resulted in that two individuals with limited experience to operate the tool were tasked to do so. It was noted that they did

not take advantage of all of the useable features of the software, and that they could not operate the software in an optimal manner. Otherwise, no significant differences in the working conditions between the teams were observed. Members of both teams explicitly strived to only take information that was collected during the observational study into account, effectively suppressing their prior knowledge of the case.

The team with PhraseBrowser used the tool to find different explanations of the Skripal incident that were proposed in the Twitter data, and to some extent to also understand whether these explanations were used in some information operation. The latter is obviously very difficult, and proved too hard to achieve during the short time allotted. The team used the tool similarly to the description in Sect. 3.9, trying several different phrase types and the filter functionality. Some of the phrase types proved useful, such as "General Phrases", "Entities", and "Explicit Untruths", as described in Sect. 3.3. Among other phrase types that were experienced as useful was one that tries to capture citations and statements, one trying to find accusations, and a few more specific phrase types for finding expressions of aggressions and tensions between different actors.

For each phrase type the team looked at phrases in order of how many times they had been used. When finding an interesting phrase, they sometimes looked at a graph of how often it had been used over time. For each interesting phrase the team always read one or more tweets these phrases appeared in. They collected the tweets that were interesting enough, taking the number of retweets into account. If these tweets had a link in them they followed that link, sometimes resulting in useful longer media articles.

6 Discussion

In this section the theory for measuring perceived usefulness as well as some validity aspects of the study are discussed. A few notes about OSINT as a data source for text mining are also provided, and the section concludes with some proposals and ideas for improvement of the PhraseBrowser tool.

6.1 Theory

This case study has aimed to examine *usefulness*, a subset of the overarching problem of user acceptance—why people embrace or reject computers and computer software. It has been shown that there are numerous variables that affect this acceptance. The widely cited technology acceptance model, TAM, conceived by Davis et al. [17], is used to model user acceptance. Davis [16] divides the variables that affect acceptance into two main categories, the *perceived ease of use* and the *perceived usefulness*. The perceived ease of use is defined by Davis [16, p. 320]

as the "degree to which a person believes that using a particular system would be free of effort", or in other words: how easy it is to use a particular system. By contrast, perceived usefulness is defined as "the degree to which a person believes that using a particular system would enhance his or her job performance" [16, p. 320]. Later Venkatesh and Davis [53] presented an extended model, that they called TAM2, that among other developments, e.g., the introduction of social influence processes, divided the concept of perceived usefulness into four factors: job relevance (to what extent the proposed system is able to support one's job), output quality (how well the system performs), result demonstrability (to what extent performance can be attributed to the system), and the perceived ease of use, which in TAM2 is a direct determinant for perceived usefulness as well as for the intention to use, and ultimately—user behavior. With the extended model the performance dimension was emphasized. In this study the aim has primarily been to investigate the job relevance and the output quality aspects in terms of this model. The result demonstrability aspect was not emphasized, and ease of use questions were not considered at all.

6.2 *Validity of the Study*

The usefulness, or functional validity of some solution, is hard to measure because it is highly context dependent. A viable option is to compare the solution with another competing solution [34], which is what was intended here. A case study is by design a "small-*N*" study [21] that cannot be expected to provide conclusions with statistically significant results. On the other hand, case studies have other strengths, such as that they can provide valuable insights that may be used as a basis for further research.

The participants were operational personnel who ranked the observational study task to be on average realistic (scale: completely unrealistic/somewhat realistic/realistic/very realistic) compared to their normal tasks, and its relevance to be high to very high (scale: limited/some/high/very high). In this respect, i.e., operational personnel solving realistic tasks in a realistic setting, the results ought to be judged to have a high ecological validity [11].

A threat to the validity of the results, however, is the prior knowledge of the case by the observational study participants which could have affected the results. It was inevitable that the teams used their background knowledge of the case, even though they actively tried to avoid doing so. Therefore an effort was made to establish the level of participant background knowledge of the Skripal case. Some familiarity of the case was expected, but the main intention was to make sure that the two teams overall had a reasonably equal level of experience. Five of the eight participants stated that they had followed the events of the case briefly, while three stated that they had followed them closely (scale: not at all/briefly/closely/very thoroughly). Three of the participants answered that they had followed the case in other languages than English (e.g., in Russian and German).

There are numerous factors that affect team performance. To reach reliable conclusions in the observational study it was important to account for such contextual factors, and isolate the variables of interest, i.e., to the extent possible the variables related to the contribution of the text analysis tool only. Among several models that seek to analyze performance shaping factors to minimize human errors and optimize performance [3], the cognitive reliability and error analysis method (CREAM) of Hollnagel [28] lists such factors. The CREAM model lists "common performance conditions", in three categories: human, technological and organizational. Here we extract the identified factors of importance from the CREAM framework. CREAM suggests that eleven factors are of importance: (1) availability of resources, (2) training and experience, (3) quality of communication, (4) human-machine interface and operational support, (5) access to procedures and methods, (6) conditions of work, (7) number of goals and conflict resolution, (8) available time (time pressure), (9) circadian rhythm, (10) crew collaboration quality, and (11) quality and support of organization.

Equalization of the CREAM factors for the teams to the widest possible extent was made. The teams were given the same conditions except for factors 1 (available resources), and 4 (human-machine interface and operational support), where one team was given PhraseBrowser, and the other one was not. Here it can be noted that it was both unexpected and unfortunate that the team equipped with PhraseBrowser did not chose to assign the most proficient PhraseBrowser user(s) to operate the software. If the software would have been used more to the full extent of its capabilities, perhaps a bigger difference between the teams would have been the result. However, the research team decided not to interfere with the internal work processes, i.e., the division of labor, of either team. It was not possible to make factor 2 (training and experience) identical for the two teams, but a justification for the team composition has been presented in this chapter. Factor 8 (available time) was the same for both teams, but the participants stated that (on average) at least four times as much time would be needed to solve the given task sufficiently well. Thus, it is reasonable to assume that the limited time affected the quality of the outputs for both teams negatively but equally. Another aspect with regard to time that may have affected the results, is that the data collection of tweets started only some time, i.e., around 2 weeks, after the incident took place.

6.3 Open Source Intelligence as a Data Source for Text Mining

Since the main information source for the intelligence assessments was OSINT, a few words on OSINT as a source is in place. One of the drawbacks that was found some time ago and that has already been mentioned, is that the sheer amount of data can be overwhelming for individuals or even organizations [29]. Moreover, even if a huge amount of data is easily accessible, much data still remain out of reach for practitioners due to various constraints, e.g., lack of access to closed forums, pay-for services, and password-protected sites that are prohibited, and there are

also other legal constraints such as copy restrictions that may hinder access [29]. Another drawback of publicly available information is that the quality of it can be questionable—after all, anyone can post just about anything on the Internet without discrimination.

A significant part of the OSINT problem is how to handle unstructured, soft data, and Dragos [19] noted that there are multiple other uncertainties beside the credibility of sources in such data, for example, intrinsic properties such as the ambiguity of natural language and the presence of possible inconsistencies in the provided information. On a smaller scale, practical obstacles such as the handling of several obscure languages and the prevalence of "slang" language pose problems in the information organization work phase when it comes to unifying and cleaning the data for further processing [29].

6.4 Potential Improvements of PhraseBrowser

During and after the observational study, the participating analysts provided several suggestions on how PhraseBrowser could be improved. Some of these thoughts are related to and discussed here, along with a discussion on the improvements that have been made since the observational study, as well as some plans for future work.

The analysts found the phrase types and phrases useful. However, as there are already many phrase types to choose from, it is not always easy to understand what all of them were created to support. The analysts would have liked to have short explanations of the phrase types, and also some indication about how well-developed they are. The latter could partially be answered by displaying the number of rule lines in the phrase types.

The participants also realized the potential benefits of being able to create their own phrase types, and asked for this functionality. Since then a first version of a user interface for rule development has been implemented. To be expressive, the rule language is still, however, somewhat complex, so only dedicated analysts, perhaps with some computer science knowledge, can be expected to use it to its full potential.

PhraseBrowser is focused on providing an overview of the textual content in tweets. The analysts would have liked to have access to more sources within the interface, such as other social media platforms, news media, etc. Since then partial support for RSS feeds has been implemented.

During the study the analysts found interesting tweets that led them to interesting Twitter accounts. They would have liked to get more information about these accounts within the prototype, and be able to search for specific accounts in the data. This functionality will possibly be implemented in a separate, complementary prototype.

There are many ways to filter data that are implemented in the PhraseBrowser prototype, some of which were also mentioned by the analysts, such as URLs and location information derived from accounts. It is also possible to look at data during

smaller or larger time intervals and for chosen phrases, but the analysts would have liked this functionality to be more developed.

As described in Sect. 3.4, PhraseBrowser does not necessarily analyze every tweet when the data stream is too large compared to the computing power. Also, using phrases as filters, as described in Sect. 3.3, reduces the set of tweets under investigation. These facts are currently reflected to some extent in the user interface, but the analysts would have liked them to be more prominent.

PhraseBrowser, and the set of prototype tools it is part of, allows for a great deal of flexibility. After using the possibilities for a while, certain usage patterns might emerge. For instance, it could under certain circumstances prove beneficial to filter data by "Explicit Untruths" and a particular list of persons, before trying to explore what is written. If further methods for filtering data, such as by image content or user account meta data, are available, even more complex usage patterns may prove useful. Such usage patterns could potentially be executed in advance, with the results presented using a digital dashboard.

7 Conclusions

This chapter presents the results of a case study that seeks to investigate the perceived usefulness of a text mining tool and how it affects the quality of the end result (output) of a realistic intelligence assessment task. Conclusions from a case study like this, however, should be interpreted in light of the limited scope of the study, e.g., the number of participants and the study design at large, and be regarded as preliminary. As outlined in Sect. 1, two research questions have governed the present study:

– Is PhraseBrowser, a specific instantiation of a text analysis tool, perceived as useful for solving typical analytical tasks?

It was found that all analysts that used or had used the PhraseBrowser tool previously, *liked* the tool and subjectively *perceived* it as useful. The main benefit was that it was thought to provide analysts with an opportunity to get an overview of an issue quickly. The results indicate that its main contribution is to highlight pointers to other sources making it possible to conduct further searches, and to a lesser extent also to find unique information and new Twitter sources.

– Does the use of the text analysis tool improve the quality of a typical analytical deliverable?

Based on the inputs from the study participants and the SMEs, it was not possible to discern any major quality differences in the intelligence assessment deliverables. The assessment of the deliverable from the team that operated the tool reported that it contained more diverse and "more detailed" pieces of information than the deliverable from the team that did not have the tool.

In the future the research methodology ought to be developed further, and more observational studies with other groups of analysts should be undertaken. A specific observational study could be to examine the usefulness of the PhraseBrowser tool for intelligence analysis of ongoing events. It should also be noted that the value of PhraseBrowser was examined relative to the use of a range of other pieces of software, e.g., the ones that the team that did not use the tool had at its disposal. To further strengthen the results of this study, future research should strive to also establish the function of the additional software used, in some detail.

References

1. W. Agrell, G.F. Treverton, The science of intelligence: reflections on a field that never was, in *National Intelligence Systems: Current Research and Future Prospects*, ed. by G.F. Treverton, W. Agrell, Chap. 11 (Cambridge University Press, New York, 2009), pp. 265–280

2. B. Alex, C. Grover, B. Haddow, M. Kabadjov, E. Klein, M. Matthews, S. Roebuck, R. Tobin, X. Wang, Assisted curation: does text mining really help?, in *Proceedings of the Pacific Symposium on Biocomputing (PSB 2008)* (World Scientific, Singapore, 2008), pp. 556–567. https://doi.org/10.1142/9789812776136_0054

3. M.A.B. Alvarenga, P.F.F. Frutuoso e Melo, R.A. da Fonseca, A review of the models for evaluating organizational factors in human reliability analysis, in *Proceedings of the 2009 International Nuclear Atlantic Conference (INAC 2009)* (Brazilian Nuclear Energy Association (ABEN), Rio de Janeiro, 2009)

4. L. Armistead (ed.), *Information Operations: Warfare and the Hard Reality of Soft Power*, Issues in Twenty-First Century Warfare (Brassey's, Washington, 2004)

5. S. Arnborg, J. Brynielsson, H. Artman, K. Wallenius, Information awareness in command and control: precision, quality, utility, in *Proceedings of the Third International Conference on Information Fusion (FUSION 2000)*, vol. 2 (IEEE, Piscataway, 2000), pp. ThB1/25–32. https://doi.org/10.1109/IFIC.2000.859871

6. M. Bang, Military intelligence analysis: institutional influence. Ph.D. thesis, National Defence University, Helsinki, October 2017. http://urn.fi/URN:ISBN:978-951-25-2930-80

7. M.T. Bimfort, A definition of intelligence. Stud. Intell. **2**(4), 75–78 (1958)

8. M. Boicu, G. Tecuci, D.A. Schum, Intelligence analysis ontology for cognitive assistants, in *Proceedings of the Third International Ontology for the Intelligence Community Conference (OIC-2008)* (George Mason University, Fairfax, 2008), pp. 31–35

9. K. Bontcheva, L. Derczynski, A. Funk, M.A. Greenwood, D. Maynard, N. Aswani, TwitIE: an open-source information extraction pipeline for microblog text, in *Proceedings of the Ninth International Conference on Recent Advances in Natural Language Processing (RANLP 2013)* (Association for Computational Linguistics, Stroudsburg, 2013), pp. 83–90

10. A. Breakspear, A new definition of intelligence. Intell. Natl. Secur. **28**(5), 678–693 (2013). https://doi.org/10.1080/02684527.2012.699285

11. U. Bronfenbrenner, Toward an experimental ecology of human development. Am. Psychol. **32**(7), 513–531 (1977). https://doi.org/10.1037/0003-066X.32.7.513

12. J. Brynielsson, A. Horndahl, L. Kaati, C. Mårtenson, P. Svenson, Development of computerized support tools for intelligence work, in *Proceedings of the 14th International Command and Control Research and Technology Symposium (14th ICCRTS)* (U.S. Department of Defense CCRP, Washington, 2009). Paper no. 48

13. Cisco, Cisco visual networking index: forecast and trends, 2017–2022. White paper C11-741490-00, February 2019

14. R.M. Clark, *Intelligence Analysis: A Target-Centric Approach*, 5th edn. (CQ Press, Washington, 2016)

15. H. Cunningham, V. Tablan, A. Roberts, K. Bontcheva, Getting more out of biomedical documents with GATE's full lifecycle open source text analytics. PLOS Comput. Biol. **9**(2), e1002854 (2013). https://doi.org/10.1371/journal.pcbi.1002854

16. F.D. Davis, Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q. **13**(3), 319–340 (1989). https://doi.org/10.2307/249008

17. F.D. Davis, R.P. Bagozzi, P.R. Warshaw, User acceptance of computer technology: a comparison of two theoretical models. Manag. Sci. **35**(8), 982–1003 (1989). https://doi.org/10.1287/mnsc.35.8.982

18. M. Degaut, Spies and policymakers: intelligence in the information age. Intell. Natl. Secur. **31**(4), 509–531 (2016). https://doi.org/10.1080/02684527.2015.1017931

19. V. Dragos, An ontological analysis of uncertainty in soft data, in *Proceedings of the 16th International Conference on Information Fusion (FUSION 2013)* (IEEE, Piscataway, 2013), pp. 1566–1573

20. M.J. Eppler, J. Mengis, The concept of information overload: a review of literature from organization science, accounting, marketing, MIS, and related disciplines. Inf. Soc. **20**(5), 325–344 (2004). https://doi.org/10.1080/01972240490507974

21. A.L. George, A. Bennett, *Case Studies and Theory Development in the Social Sciences* (MIT Press, Cambridge, 2005)

22. M. Glassman, M.J. Kang, Intelligence in the internet age: the emergence and evolution of open source intelligence (OSINT). Comput. Hum. Behav. **28**(2), 673–682 (2012). https://doi.org/10.1016/j.chb.2011.11.014

23. J.K. Guo, D. Van Brackle, N. LoFaso, M.O. Hofmann, Extracting meaningful entities from human-generated tactical reports. Procedia Comput. Sci. **61**, 72–79 (2015). https://doi.org/10.1016/j.procs.2015.09.153

24. W.M. Hall, G. Citrenbaum, *Intelligence Analysis: How to Think in Complex Environments* (Praeger, Santa Barbara, 2009)

25. W. He, S. Zha, L. Li, Social media competitive analysis and text mining: a case study in the pizza industry. Int. J. Inf. Manag. **33**(3), 464–472 (2013). https://doi.org/10.1016/j.ijinfomgt.2013.01.001

26. M.A. Hearst, Automatic acquisition of hyponyms from large text corpora, in *Proceedings of the 15th International Conference on Computational Linguistics (COLING-92)*, vol. 2 (Association for Computational Linguistics, Stroudsburg, 1992), pp. 539–545

27. R.J. Heuer, Jr., *Psychology of Intelligence Analysis* (Center for the Study of Intelligence, Central Intelligence Agency, Washington, 1999)

28. E. Hollnagel, *Cognitive Reliability and Error Analysis Method (CREAM)* (Elsevier, Oxford, 1998). https://doi.org/10.1016/B978-0-08-042848-2.X5000-3

29. A.S. Hulnick, The downside of open source intelligence. Int. J. Intell. CounterIntell. **15**(4), 565–579 (2002). https://doi.org/10.1080/08850600290101767

30. J. Jiang, Information extraction from text, in *Mining Text Data*, ed. by C.C. Aggarwal, C. Zhai, Chap. 2 (Springer, Boston, 2012), pp. 11–41. https://doi.org/10.1007/978-1-4614-3223-4_2

31. F. Johansson, J. Brynielsson, P. Hörling, M. Malm, C. Mårtenson, S. Truvé, M. Rosell, Detecting emergent conflicts through web mining and visualization, in *Proceedings of the 2011 European Intelligence and Security Informatics Conference (EISIC 2011)* (IEEE, Piscataway, 2011), pp. 346–353. https://doi.org/10.1109/EISIC.2011.21

32. D. Jurafsky, J.H. Martin, *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*, 2nd edn. (Prentice Hall, Upper Saddle River, 2009)
33. C. Koopman, J. Snyder, R. Jervis, Theory-driven versus data-driven assessment in a crisis: a survey of international security readers. J. Confl. Resolut. **34**(4), 694–722 (1990). https://doi.org/10.1177/0022002790034004006
34. K. Krippendorff, *Content Analysis: An Introduction to Its Methodology*, 2nd edn. (SAGE, Thousand Oaks, 2004)
35. T.K. Landauer, *The Trouble with Computers: Usefulness, Usability, and Productivity*, A Bradford Book (MIT Press, Cambridge, 1995)
36. E.V. Larson, R.E. Darilek, D. Gibran, B. Nichiporuk, A. Richardson, L.H. Schwartz, C.Q. Thurston, *Foundations of Effective Influence Operations: A Framework for Enhancing Army Capabilities* (RAND Corporation, Santa Monica, 2009). https://www.rand.org/pubs/monographs/MG654.html
37. K. Lim, Big data and strategic intelligence. Intell. Natl. Secur. **31**(4), 619–635 (2016). https://doi.org/10.1080/02684527.2015.1062321
38. D. McDonald, U. Kelly, The value and benefit of text mining to UK further and higher education. Digital Infrastructure Directions Report, Doc# 811, Version 1.1, JISC, March 2012. https://www.jisc.ac.uk/reports/value-and-benefits-of-text-mining
39. NATO, Allied joint doctrine for psychological operations. Publication AJP-3.10.1(A), NATO Standardization Agency, Brussels, October 2007
40. D. Omand, J. Bartlett, C. Miller, Introducing social media intelligence (SOCMINT). Intell. Natl. Secur. **27**(6), 801–823 (2012). https://doi.org/10.1080/02684527.2012.716965
41. J.K. Pal, Usefulness and applications of data mining in extracting information from different perspectives. Ann. Libr. Inf. Stud. **58**(1), 7–16 (2011)
42. N.A. Pollard, On counterterrorism and intelligence, in *National Intelligence Systems: Current Research and Future Prospects*, ed. by G.F. Treverton, W. Agrell, Chap. 6 (Cambridge University Press, New York, 2009), pp. 117–146
43. A.H. Razavi, D. Inkpen, R. Falcon, R. Abielmona, Textual risk mining for maritime situational awareness, in *Proceedings of the 2014 IEEE International Inter-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)* (IEEE, Piscataway, 2014), pp. 167–173. https://doi.org/10.1109/CogSIMA.2014.6816558
44. F.E. Ritter, G.D. Baxter, E.F. Churchill, *Foundations for Designing User-Centered Systems: What System Designers Need to Know about People* (Springer, London, 2014). https://doi.org/10.1007/978-1-4471-5134-0
45. D.E. Sichel, *The Computer Revolution: An Economic Perspective* (Brookings Institution Press, Washington, 1997)
46. R.D. Steele, *On Intelligence: Spies and Secrecy in an Open World* (AFCEA International Press, Fairfax, 2000)
47. S. Stieglitz, L. Dang-Xuan, A. Bruns, C. Neuberger, Social media analytics: an interdisciplinary approach and its implications for information systems. Bus. Inf. Syst. Eng. **6**(2), 89–96 (2014). https://doi.org/10.1007/s12599-014-0315-7
48. Swedish Armed Forces, Försvarsmaktens underrättelsereglemente [The armed forces' intelligence manual]. Publication M7739-353025, Headquarters, Stockholm, December 2010
49. S. Tzu, *The Art of War*, trans. by S.B. Griffith (Clarendon Press, Oxford, 1963)
50. U.S. Department of Defense, Joint doctrine for information operations. Joint Publication 3.13, Joint Chiefs of Staff, Washington, October 1998
51. U.S. Department of Defense, Joint and national intelligence support to military operations. Joint Publication 2-01, Joint Chiefs of Staff, Washington, July 2007
52. U.S. Department of Defense, Information operations. Joint Publication 3.13, Joint Chiefs of Staff, Washington, November 2012

53. V. Venkatesh, F.D. Davis, A theoretical extension of the technology acceptance model: four longitudinal field studies. Manag. Sci. **46**(2), 186–204 (2000). https://doi.org/10.1287/mnsc.46.2.186.11926
54. C. von Clausewitz, *On War*, trans. by M. Howard, P. Paret (Princeton University Press, Princeton, 1976). Originally published 1832