Security Informatics

RESEARCH

Open Access

Organised crime and social media: a system for detecting, corroborating and visualising weak signals of organised crime online

Simon Andrews[1,2*], Ben Brewster[2] and Tony Day[2]

Abstract

This paper describes an approach for detecting the presence or emergence of organised crime (OC) signals on social media. It shows how words and phrases, used by members of the public in social media posts, can be treated as weak signals of OC, enabling information to be classified according to a taxonomy. Formal concept analysis is used to group information sources, according to crime-type and location, thus providing a means of corroboration and creating OC concepts that can be used to alert police analysts to the possible presence of OC. The analyst is able to 'drill down' into an OC concept of interest, discovering additional information that may be pertinent to the crime. The paper describes the implementation of this approach into a fully-functional prototype software system, incorporating a social media scanning system and a map-based user interface. The approach and system are illustrated using human trafficking and modern slavery as an example. Real data is used to obtain results that show that weak signals of OC have been detected and corroborated, thus alerting to the possible presence of OC.

Keywords: Organised crime, Social media, Entity extraction, Formal concept analysis

Introduction

The vociferous proliferation of the Internet, and more recently social media, into society and the everyday lives of its citizens has, over the last fifteen or so years, resulted in a sea-change in the behaviours and perceptions we have in relation to the information that is shared freely online [1]. Such behaviour has resulted in the creation of a vast repository of information that holds potential value as an intelligence resource, and the emergence of the open-source researcher as a valuable skill-set within the analytical repertoire of the police and other security agencies. Resources such as social media, RSS news feeds, interactive street-maps and online directory services all provide valuable stores of information that can be used to support existing investigative and analytical practices in response to serious and organised crime. This paper's novelty concerns the application of formal concept analysis (FCA) in combination with automated information retrieval and natural language processing (NLP) tools to identify, extract, categorise and corroborate information from open web sources which may be used to identify the early onset of organised crime. Specifically, the paper looks to identify what we will refer to as 'weak signals', and looks to transform these signals into corroborated alerts linked to the presence or emergence of organised crime activity, including gang activity, the trade and use of illegal narcotics, gun crime, human trafficking, and modern slavery. This research described in this paper forms part of a larger project; ePOOLICE (early Pursuit against Organised crime using environmental scanning, the Law and IntelligenCE systems). The project, which concluded in 2016, aimed to develop a prototype environmental scanning system, integrating a number of promising and mature technical components

*Correspondence: s.andrews@shu.ac.uk
[1] Conceptual Structures Research Group, Department of Computing, The Communication and Computing Research Centre, Sheffield Hallam University, Sheffield, UK
Full list of author information is available at the end of the article

to semantically filter information from open-sources, such as the web and social media, to identify information that may constitute weak-signals of organised crime. The project sought to identify the extent to which the organised crime threats could be detected at an early stage, prior to their development into larger more resilient criminal systems, through the automated collection and analysis of data from open sources [2]. These sources primarily contain information from news outlets, journalists and other outlets, but also take into account posts made by 'normal' citizens. The project incorporated the input of domain experts and practitioners throughout, including law enforcement agencies from the UK, Spain and Germany, and international organisations such as EUROPOL and the United Nations Interregional Organised Crime Research Institute (UNICRI). The input of these organisations in particular, alongside other practitioners from the projects advisory board, were leveraged in the extraction of requirements, using a combination of questionnaires, interviews and workshop session throughout the software development lifecycle, right through to their eventual testing and evaluation.

The concept of weak signals has been abstracted from the Canadian Criminal Intelligence Service's (CISC) definitions of primary and secondary indicators [3], and the perception that in reality there is little tangible value to be extracted from isolated indicators as there is potential for them to be symptomatic of a variety of phenomena, many of which are not necessarily in any way indicative of the presence or threat of organised crime. However, when these indicators are grouped under certain conditions, such as temporal or geographic proximity to a specific location and type of activity, they can begin to provide insight into the presence or emergence of crime. It is with this definition, and the notion of 'weak signals' that we use as the basis of this paper and the approach presented within it. In UK practice, the college of policing guidance on the use of open-source intelligence (OSINT) [4] is fairly limited, and no special provisions are made for social media services as a potential intelligence source. The main uses of open-sources in this respect are to develop an understanding of the locations relevant to a piece of analysis, to identify social and demographic changes, to identify external factors that may impact on crime, disorder and community concerns, to support and develop investigations by indicating lines of enquiry or the corroboration of other information, and to support the development of subject and problem profiles through the development of intelligence products.

Perhaps the greatest shift in the use of the internet over the last 10–15 years or so is the relative phenomenon that is the usage of Social Media among normal citizens. In the aftermath of the riots across London and a number of other English cities which followed the killing of Mark Duggan by the Metropolitan Police in 2011, a Her Majesty's Inspectorate of Constabulary (HMIC) commissioned review highlighted significant inefficiencies in the the way that authorities were equipped to deal with social media as an intelligence source [5]. Social Media Intelligence, or 'SOCMINT', provides opportunities for providing insights into events and groups, enhancing situational awareness, and enabling the identification of criminal intent, providing it is done so in a manner that is appropriately grounded in respect of human privacy rights [6]. Reports and case studies such as this have demonstrated the desire and scope for research in the field to explore its novel application in enhancing the decision making ability of key stakeholders, such as the police, with more recent events, and social media's role in the response, such as the Boston, Massachusetts bombings in 2013 [7], only strengthening the case for its use.

In order to demonstrate the potential utility of SOCMINT, in respect of identifying the presence and/or emergence of organised crime, the problem domain of Human Trafficking will be used as the exemplar use-case throughout this paper. Human Trafficking operates on a vast scale with a truly global impact, with almost every country in the world acting as an origin, transit or destination location for the movement and exploitation of human beings. Trafficking is so defined by article 3 of the Palermo protocol as the "recruitment, transportation, transfer, harbouring or receipt of persons, by means of the threat or use of force or other forms of coercion, of abduction, of fraud, of deception, of the abuse of power or of a position of vulnerability or of the giving or receiving of payments or benefits to achieve the consent of a person having control over another person, for the purpose of exploitation" [8]. In the UK, Human Trafficking and Modern Slavery has risen to prominence over recent years into a priority serious and organised crime threat, alongside issues like cybercrime, drug crime and child sexual exploitation (CSE) [9]. Although it is worth noting that due to recent legislative changes in the UK, it now falls under the banner of modern slavery which consolidates existing legislation related to slavery and to human trafficking. However, due to the trans-European scope of this paper, we will refer mainly back to the principles of the Palermo protocol, using the term 'Human Trafficking' interchangeably with 'Trafficking in Human Beings'.

Organisations such as Europol [10] are increasingly acknowledging growing criminal dependence on the internet and the increasingly trans-European perspective of serious and organised crime. These changes in the way that information is created and shared, combined with the diversity in the way that existing forms of criminality are being conducted provides the opportunity, and desire, for the development of new means to assist law enforcement in combating it. To provide one such approach to enable this, the tools described here facilitate the identification, extraction, processing, analysis and presentation of data from open sources, such as social media, that can reveal insight into the emergence and presence of crime both in an operational sense; by identifying specific phenomena that are linked to discrete types of crime, and from a strategic perspective in the identification and visualisation of strategic trends through the corroboration of different crime indicators. While at one end of the scale international intelligence agencies such as the National Security Agency's (NSA) PRISM programme are facilitating the acquisition, fusion and analysis of vast amounts of data from disparate sources [11], the (known) resources and capability of law enforcement agencies (both locally, regionally and even internationally) are recognised to be much more modest—with the use of data from open sources and social media often a manual task, and the remit of just a few specialist analysts and officers within each force [12].

Taxonomy of organised crime

In beginning to model its constituent elements it is necessary to ascertain a thorough understanding of the actual problem domain—human trafficking, by drawing upon established definitions used to describe and diagnose the problem by the practitioner base. The UNODC [8] have defined, using the UN Palermo protocol as the basis, what they refer to as, the three constituent 'elements' of trafficking, these being the 'act', 'means' and 'purpose', see Fig. 1. Firstly, the 'act' refers to what is being done, this can include context such as whether and how the victim has been recruited, transported, transferred or harboured. The question of how this is being achieved is answered by the 'means', which seeks to establish whether force is being used as the basis of manipulation, such as through kidnapping, abduction or the exploitation of vulnerabilities, or more subtle methods such as through fraud, imposing financial dependencies or coercion. The final element, the 'purpose' establishes the reason why the act and means are taking place, or to put it simply—the form of exploitation behind the act and means, be it for forced labour, sexual exploitation and prostitution, organ harvesting or domestic servitude.

This definition and categorisation provides an ideal underpinning for the formation of a taxonomy of Human Trafficking that can be used to form the basis of an approach to automatically identify and extract valuable data from open sources (Fig. 2). This taxonomy in actuality forms part of a larger model, consisting of elements of a broader range of organised crime threats, including the cultivation and distribution of illegal narcotics. The elements of the taxonomy are defined across four groups, visualised here using vertical columns. The first of these columns, starting from the left, is used as a high level categorisation used to separate between different crime types. The second deals with different elements within a specific type of crime—in this case one of the three component parts of trafficking, while subsequent columns, type and element, are used to show further more specialised aspects of these elements.

Each of the nodes contained within the taxonomy represents a ruleset designed to determine the relevance of a piece of content, in this case a particular Twitter posting's relevance to the subject matter—human trafficking. The level of specialisation of the rules themselves follows the structure of the taxonomy, moving from more generic words or phrases that may indicate Human Trafficking used at the higher levels of the taxonomy, whereas the more specialised end of the taxonomy, more nuanced rules that may allude to the presence of criminality are used. Running parallel to the process of content categorisation is the requirement for content (or named-entity) extraction. The specific content and nature of the rules is informed by the input of experts and practitioners in the field, elicited through workshops and additional desk based research around organised crime indicators. A more detailed discussion of the process of defining rules and the nature of these indicators is included in [13].

The process of extracting named-entities and facts varies in complexity, from the use of simple lookup tables containing keyword lists, to more sophisticated approaches making use of part-of-speech tagging and grammatical and predicate rules to detect the use of terms within specific contexts. Data such as this is most commonly available through social media services like Twitter. However, in reality, research suggests that less than 1% of all Twitter posts actually contain geo-location information, thus making it necessary to pursue other approaches in order to increase the amount of potentially useful data available. In this regard, a number of approaches have been used, from simple keyword

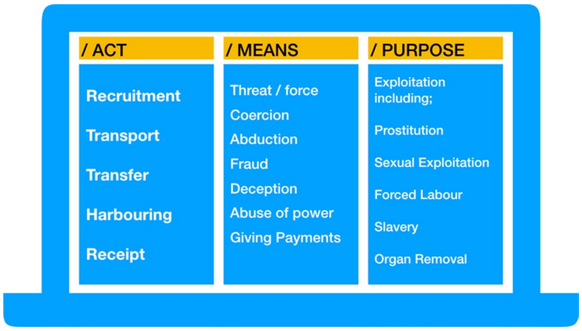


Fig. 1 The elements of human trafficking

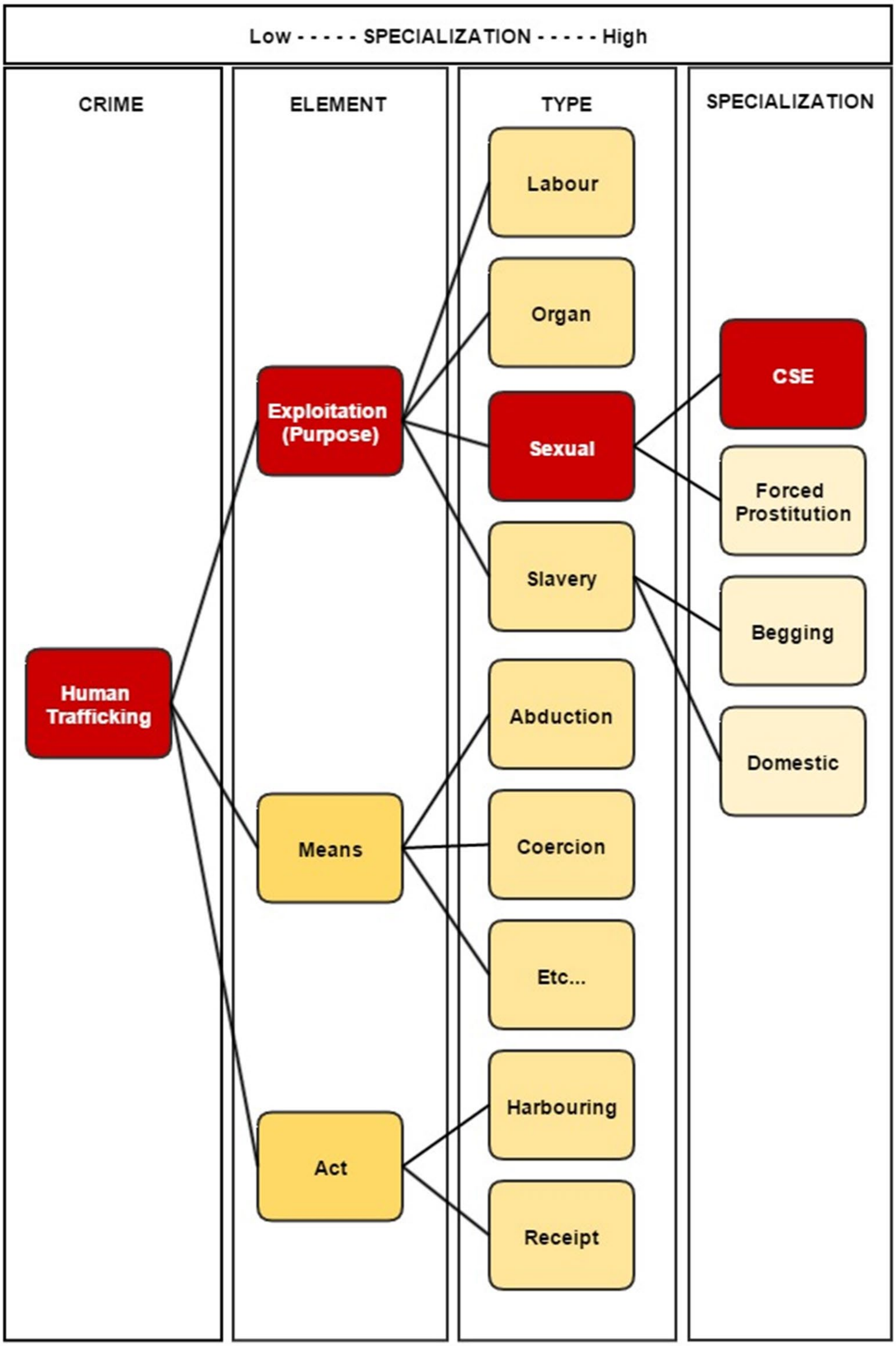


Fig. 2 Organised crime taxonomy (excerpt)

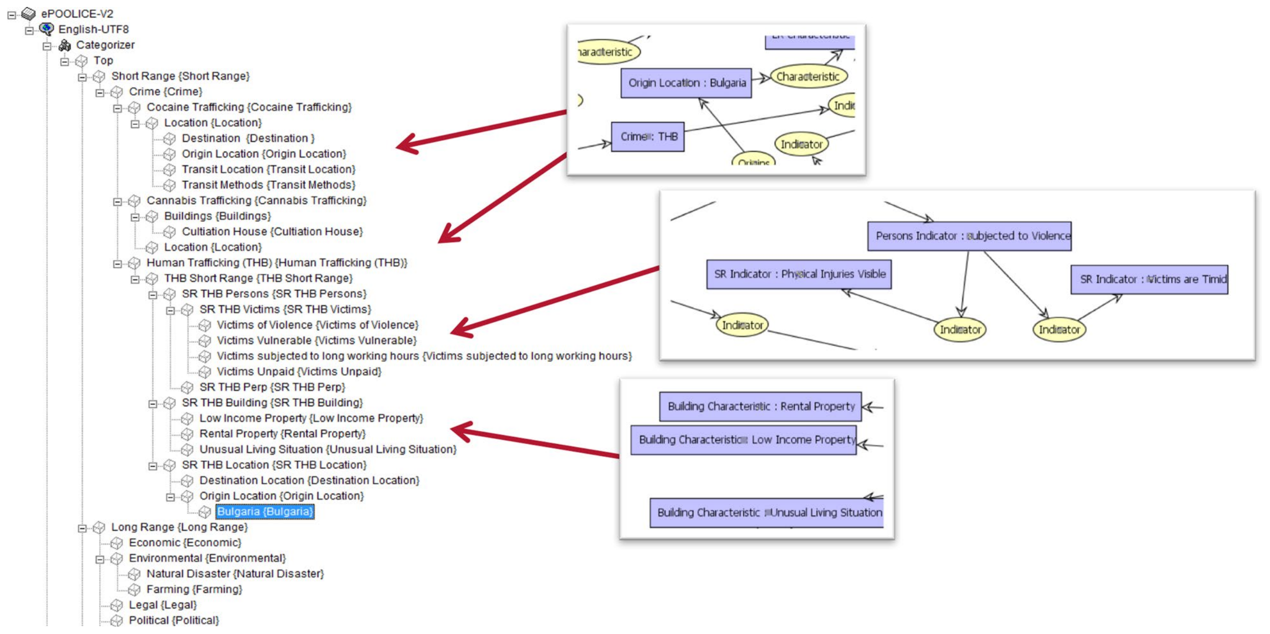
searches, to more complex algorithmic approaches. In one such example, researchers from IBM have developed an approach that uses the content of tweets that contain geotags, and then searches for similar content that does not contain them in order to assess areas where content may have originated from, as part of a larger piece of research aimed at identifying the home location of Twitter users. The tests conducted as part of the study reported accuracy of around 58% [14]. Here however, the focus is somewhat different, whereby the approach seeks to extract context in relation to named entities to enable conclusions to be inferred about events or phenomena in specific, named locations.

Weak signals of organised crime

To enable the development of a taxonomy that enables us to model and structure the information deemed useful to extract we can refer to a wealth of literature from both academic and practitioner perspectives that provide insights into the factors that contribute to and indicate organised crime. These indicators vary from high level, secondary information such as Political, Economic, Socio-cultural, Technological, Legal and Environmental (PESTLE) factors, right down to operationally oriented information that offers guidance on how to identify potential victims of trafficking. Existing models to anticipate changes and developments in organised criminality across geographic areas have focused on this kind of data alongside existing crime statistics [15].

In the past, and to some extent a problem that still exists, a lack of information and common understanding about what Human Trafficking is has hindered the impact and effectiveness of efforts to combat it [16]. Despite varied and wide-ranging counter-trafficking initiatives from NGOs, Law Enforcement and Governments, reliable information regarding the magnitude and nature of trafficking across regional and national borders is still hard to come by due to a number of issues around the sharing, fusion and understanding of data that is already being collected [17]. The purpose of the approach developed and described in this paper is not to provide a statistically accurate representation of the presence and emergence of trafficking but rather to increase access to, and usability of, data from previously untapped open-sources. In previous work, we have discussed indicators across a three-level model [13] moving from credible and accepted indicators of trafficking at level 1 of the model, through to the observations and content created online, including on social media, by citizens regarding these 'weak signals.'

In this paper we discuss the latter and, more specifically, the modelling and use of this information as 'weak signals' that allude to the presence and/or emergence of criminality in citizen generated content, whilst using the formal definitions and doctrine that exists to underpin the framework and organisation of the model itself. Perhaps the most comprehensive list of indicators comes from the UNODC [18] who provide an extensive typology categorised by different types of exploitation such as; domestic servitude, child sexual exploitation, labour exploitation and begging/petty crime—the labels used in the taxonomy structure are outlined in Fig. 2. Using sexual exploitation as an example, indicators include things such as the appearance that persons are under the specific control of another, that the person(s) appear to own little clothing, or rely on their employer for basic amenities, transport and accommodation and more. The taxonomy excerpt included in Fig. 3 shows at a high level how a taxonomy node focused on attributes which may



- ePOOLICE-V2
  - English-UTF8
    - Categorizer
      - Top
        - Short Range {Short Range}
          - Crime {Crime}
            - Cocaine Trafficking {Cocaine Trafficking}
              - Location {Location}
                - Destination {Destination }
                - Origin Location {Origin Location}
                - Transit Location {Transit Location}
                - Transit Methods {Transit Methods}
            - Cannabis Trafficking {Cannabis Trafficking}
              - Buildings {Buildings}
                - Cultiation House {Cultiation House}
              - Location {Location}
            - Human Trafficking (THB) {Human Trafficking (THB)}
              - THB Short Range {THB Short Range}
                - SR THB Persons {SR THB Persons}
                  - SR THB Victims {SR THB Victims}
                    - Victims of Violence {Victims of Violence}
                    - Victims Vulnerable {Victims Vulnerable}
                    - Victims subjected to long working hours {Victims subjected to long working hours}
                    - Victims Unpaid {Victims Unpaid}
                  - SR THB Perp {SR THB Perp}
                - SR THB Building {SR THB Building}
                  - Low Income Property {Low Income Property}
                  - Rental Property {Rental Property}
                  - Unusual Living Situation {Unusual Living Situation}
                - SR THB Location {SR THB Location}
                  - Destination Location {Destination Location}
                  - Origin Location {Origin Location}
                    - Bulgaria {Bulgaria}
        - Long Range {Long Range}
          - Economic {Economic}
          - Environmental {Environmental}
            - Natural Disaster {Natural Disaster}
            - Farming {Farming}
          - Legal {Legal}
          - Political {Political}

haracteristic
Origin Location : Bulgaria
Characteristic
Indic
Crime: THB
Indicator
Origins
Persons Indicator : subjected to Violence
SR Indicator : Physical Injuries Visible
SR Indicator : Victims are Timid
Indicator
Indicator
Indicator
Building Characteristic : Rental Property
Building Characteristic: Low Income Property
Building Characteristic : Unusual Living Situation

Fig. 3 Organised crime taxonomy

indicate an individual is vulnerable, or a potential victim of trafficking or exploitation, may contain specific rules designed to identify text which suggests they may have been subjected to violence, as one example of a weak-signal. Although in this form, these indicators are quite abstract and it can be difficult to see how they may manifest in real, open-source, data—it is possible to develop rules looking for keywords and phrases that can provide 'weak-signals' of their existence online.

In order to facilitate the identification and extraction of these weak signals in social media and other open-sources, what we will refer to as 'contextual extraction' methods [19] are used in order to identify, and subsequently extract, key entities and facts (i.e. previously unknown relationships between different entities) from the data. This approach to information extraction using natural-language processing builds upon the existing principles of template based information extraction [20], also sometimes referred to as 'Atomic Fact Extraction' [19]. These 'facts' enable the extraction of entities within a specific context, i.e. locations in relation to an arrest or type of exploitation on a per sentence basis.

A number of example entities are included in Table 1.

While in isolation, the extraction of these entities on their own does not necessarily provide much actionable information, it is possible using rules that attempt to infer relations between them to begin to make some assumptions about the data and its content. For example a single tweet may contain multiple locations and other entities, but without some means to establish a relationship between the two there is no way to automatically infer, with any confidence at least, that they are linked. Fortunately, through the use of contextual extraction, and the aforementioned 'facts' we can infer these relationships in a number of ways, using prepositions, parts-of-speech tagging and Boolean operators that specify distance between words and other parameters. As the examples discussed in this paper refer to data from Twitter only, these relationships are done on a 'per sentence' basis. Figure 4 shows how this works in practice.

Table 1 Extracted entity examples

Entity	Example
Location	Atlanta
	Boston
	San Diego
Event	Trafficking
	Arrest
	Transaction
Exploitation	CSE
	Forced labour
	Forced prostitution

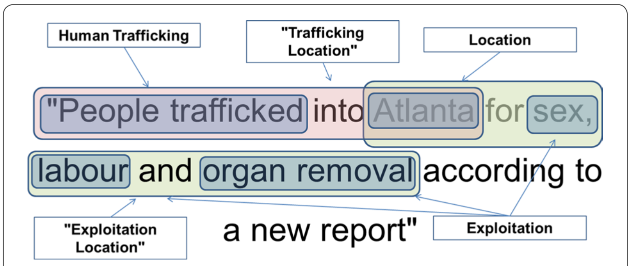


Fig. 4 Entity and fact extraction in social media

From these rules we can begin to make some assumptions about the entities being extracted. For example, it is now possible to ascertain with a degree of confidence that specific locations are in reference to a specific event. At this point, it is important to acknowledge the challenges posed by the use of SMS-language (textese) as communication via services such as Twitter do not necessarily adhere to strict grammar or syntax conventions. Although a number of novel approaches to handle this type of language are in development (see, for example [21]) due to the use of examples that use accepted, formal terminology, we do not address this issue here.

By using context–sensitive concept matching using Language Interpretation/Text Interpretation (LITI) rules [19], concepts can be also matched to the specific context for which they are being used. Using the example of a location, the use of a specific place in reference to an event or action, such as an arrest can be extracted by matching locations within proximity of text indicating an event followed by a preposition and the location. Such an approach brings about a number of potentially advantageous features. The use of contextual concept definitions enables increased levels of accuracy, and thus confidence, in the extracted data relative to the context in which it is being extracted. For example, instances of location that occur with a specified number of words, separated by a preposition, of terminology indicating an arrest, increases the level of confidence that the information is in fact in reference to the context as opposed to just being any occurrence of a location name in the text. Social media sites such as Twitter often include features that enable location information to be captured directly from the browser or mobile application.

Geolocation

To provide location-based corroboration and visualisation, automatic extraction of named locations from the various data sources is used. Working from an extensive collection of known locations, the detection and word-level extraction of countries, regions and cities is performed. However, regions and cities are extracted as-is and further contextualisation is required to disambiguate the named location from others that share the same name. For example, the term New York could easily refer to either New York City or New York State.

Named locations are not extracted as locations alone; the rules within the OC ontology are used to extract locations against a given OC context where possible. One example is 'trafficking location' which uses rules to detect a context regarding an entity (individual, group, asset etc.) being trafficked or moved. Another example is 'exploitation location' which detects a named location against OC exploitation contexts (such as forced labour).

Reverse geocoding, the process of resolving one or more geographical coordinates from a named location, is applied to acquire the additional contextual information required (country, region, city, etc.) to perform effective corroboration and map-based visualisation. Disambiguation of the location is also a resulting benefit; however this is based entirely on the popularity, size or importance of a location. For example, Washington State or Washington, DC are much larger and more widely referenced than Washington County in Alabama. If Washington, DC has the highest ranking then the location Washington will identified as Washington, DC.

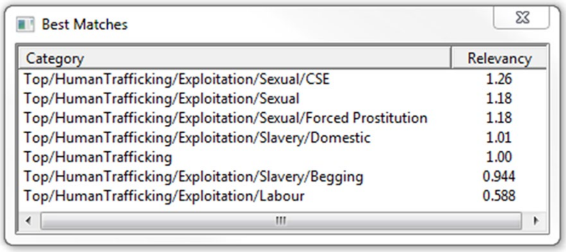
The reverse geocoding process utilises the Google Maps API. Using this API, the named location is supplied, returning a list of matching results, by order of relevance based on the most commonly searched places (i.e. popularity), the size (i.e. population), or the importance (e.g. capital). Each result provides a geographical location in decimal latitude/longitude format, along with additional hierarchical information, for example, New York City is within NY or New York State, and within the US or United States of America. Since the Scanning System is primarily investigating a US-based dataset, searches are restricted to the boundary of the US. It is important to note that a single tweet may contain multiple named locations, in which case, all of these are resolved and the tweet is processed according to each of those locations.

Categorisation and filtering

In addition to extracting facts and entities from the input data, similar techniques are also utilised as the basis of a rule-based approach to classify content against a number of pre-defined categories. Membership to one or more of these categories is then used as the basis for content filtering. If the content analysed by the crawler does not meet the criteria of at least one of the rules within the parameters defined in the content categorisation process, it is then disregarded.

Table 2 Weak signals—keywords and phrases

Weak signal	Keywords and phrases
Physical injury	Subjected to violence
	Timid
	Forced to have sex
	Women beaten
Physical appearance	Provocative dress
	Live with a group of women
	Unhappy
Unable to leave place of residence	Afraid to leave
	Under control of others
	Financially dependant
Irregular movement of individuals	Men come and go at all hours
	Women do not appear to leave
	Lots of activity at night



Category	Relevancy
Top/HumanTrafficking/Exploitation/Sexual/CSE	1.26
Top/HumanTrafficking/Exploitation/Sexual	1.18
Top/HumanTrafficking/Exploitation/Sexual/Forced Prostitution	1.18
Top/HumanTrafficking/Exploitation/Slavery/Domestic	1.01
Top/HumanTrafficking	1.00
Top/HumanTrafficking/Exploitation/Slavery/Begging	0.944
Top/HumanTrafficking/Exploitation/Labour	0.588

Fig. 5 Categorization output

The categorisation model is defined using a similar approach to the entity and fact extraction model, utilising a number of 'hand-crafted' rules organised in a hierarchical structure, with the only key difference being that rather than being designed to identify and extract specific pieces of data and/or information they aim to discern the relevance of the content against the defined topics using the same taxonomy structure defined in Fig. 2. The rules themselves use a range of techniques, again focusing on the identification of keywords and phrases. A number of examples of the phrases and keywords used as part of the categorisation taxonomy are shown in Table 2. Matched content is then assigned a 'relevancy' rating depending on the number of keywords and/or phrases that are met within the match criteria—as shown in Fig. 5.

A social media scanning system

To implement the content extraction and categorisation models, an integrated pipeline that facilitates the crawling of social media is put in place. This process manages and enables the seamless collection, restructuring, processing, filtering and output of the data in preparation for further analysis. The stages of the data preparation and processing pipeline is shown in Fig. 6.

Different data sources lend themselves to the identification and extraction of different kinds of information and indicators. For example, the data from the Web and RSS feeds used in the process allow for the extraction of data related to explicit instances of past and ongoing criminality, as reported by the media, law enforcement and other outlets. Such data can be used to enhance situational awareness, identifying key locations and trends in criminality by geographic region. Whereas as social media, although potentially useful in identifying similar information, can also be used as a source of identifying 'weak signals', due to the presence of commentary and observations provided by everyday citizens. Though it should be acknowledged that these posts are particularly uncommon when compared with the data collected from news outlets, journalists and other interest groups.

To extract and expose data to the processing pipeline and FCA components a number of approaches are used depending on the source type. Primarily, three different data types are used by the system; Social Media, Web pages and RSS feeds although this is flexible and can be adapted to include other disparate data types, both structured and unstructured. The process for extracting data from each varies fairly significantly from source to source. In the case of crawling pages from the web, a number of seed URLs from a hand-curated database are used as a starting point for the crawling process, with subsequent links within those target URLs also transversed. These seed URLs consist of local, regional and national news reporting mediums and law enforcement agency (LEA) web-pages. To filter out noise and irrelevant data an initial stage of categorisation is used to determine a base level of relevance to the domain, utilising keywords and phrases that indicate a surface level relation to criminality, disregarding any source data that does not meet the criteria defined by the rule-set. Continuing with the example of human trafficking and modern slavery that is used throughout, the initial filtering criteria contain rules that both explicitly refer to trafficking and forms of human exploitation and more subtle, nuanced, rules look for weak signals such as references to injury or signs of physical abuse, or the presence of paraphernalia that may be linked phenomena, such as drug cultivation. Similarly, a curated set of RSS feeds use a similar method to web transversal, however these feeds are monitored for changes and updates, as opposed to being crawled periodically in search of new content.

For social media however, as the main source of focus in this article, the approach is somewhat different. Utilising the 'Search API' offered by Twitter [22], queries can be made against the service's index of recent and/or popular posts from the previous seven days, with only the most relevant tweets returned from during the time period. At the time of writing, the amount of data returned is limited by the API's rate limit, currently set at 180 queries per 15 minutes. This amount is subject to change. In terms of the queries themselves, a number of pre-defined operators exist that allow for the matching of keywords, exact phrases and other operations. When dealing with larger sources of unstructured data, such as web-pages or other documents, it is often necessary to heuristically parse the content to remove additional page elements such advertisements before processing. However, as the data extracted from Twitter is already sufficiently structured, this step is not needed here.

Using the same example as in previous figures, the XML enables this content to be extracted alongside other meta-data, such as the geolocation and other details about the author, the data and time that the post was made, the date/time it was extracted by the crawler, keywords used to query the API, and a URL link back to the original content. Other metadata is also provided by default, but has been filtered as it is not essential to the process being described. An example of the XML output is shown in the code snippet below.

```
<xml version=”1.0” encoding=”utf−8”>
<article>
<authorgeolocation>Atlanta, GA</authorgeolocation>
<query>trafficking</query>
<body>People trafficked into Atlanta for sex, labour and
    organ removal according to a new report</body>
<pdate>20150129</pdate>
<categories>top\HumanTrafficking\Exploitation\sexual</
    categories>
<TraffickingLocation>Atlanta</ TraffickingLocation>
<ExploitationLocation>Atlanta</ExploitationLocation>
<ExploitationType>Sexual; Labour; Organ Harvesting</
    ExploitationType>
</article>
```

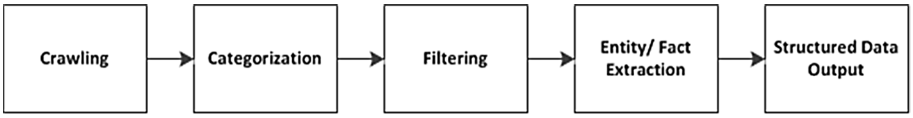


Fig. 6 Social media scanning process

Applying formal concept analysis

Background to formal concept analysis

A formal description of formal concepts [23] begins with a set of objects X and a set of attributes Y. A binary relation $I \subseteq X \times Y$ is called the *formal context*. If $x \in X$ and $y \in Y$ then xIy says that object x has attribute y. For a set of objects $A \subseteq X$, a derivation operator $^{\uparrow}$ is defined to obtain the set of attributes common to the objects in A as follows:

$$A^{\uparrow} := \left\{ y \in Y \mid \forall x \in A : xIy \right\}.$$

Similarly, for a set of attributes $B \subseteq Y$, the $^{\downarrow}$ operator is defined to obtain the set of objects common to the attributes in B as follows:

$$B^{\downarrow} := \left\{ x \in X \mid \forall y \in B : xIy \right\}.$$

(A, B) is a formal concept iff $A^{\uparrow} = B$ and $B^{\downarrow} = A$. The relations $A \times B$ are then a closed set of pairs in I. In other words, a formal concept is a set of attributes and a set of objects such that all of the objects have all of the attributes, there are no other objects that have all of the attributes and there are no other attributes that all the objects have. A is called the *extent* of the formal concept and B is called the *intent* of the formal concept.

A formal context is typically represented as a cross table, with crosses indicating binary relations between objects (rows) and attributes (columns). Table 3 is a small example of a formal context where the objects are information sources and the attributes are named entities, signals of crime categories and values present in the information sources.

Formal concepts in a cross table can be visualised as closed rectangles of crosses, where the rows and columns in the rectangle are not necessarily contiguous. $FC1$, below, is a example of a formal concept from the Organised Crime formal context above:

$$\begin{array}{l} FC1: \\ (\{Crime\ THB, Location\ A\}, \\ \{Source\ 1, Source\ 2, Source\ 5, Source\ 7, Source\ 9\}) \end{array}$$

Typically, as the number of attributes in a formal concept increase, the corresponding number of objects that share those attributes reduces—the objects become more specialised. This behaviour can be seen in $FC2$, below, with the addition of the attribute *Category Exploitation*:

$$\begin{array}{l} FC2: \\ (\{Crime\ THB, Location\ A, Category\ Exploitation\}, \\ \{Source\ 5, Source\ 7, Source\ 9\}) \end{array}$$

$FC2$ is said to be a *sub-concept* of $FC1$. These connections between formal concepts are fundamental in formal concept analysis (FCA) and can be visualised as a *formal concept tree*. Figure 7 shows a formal concept tree derived from the Oganised Crime formal context. Each node is a formal concept. The tree can be read by understanding that objects are inherited by formal concepts *from the right* and attributes are inherited by formal concepts *from the left*. Thus, for example, formal concept 1 in the tree is $FC1$, above, and formal concept 6 in the tree is $FC2$, above.

The power of FCA to cluster similar objects and to capture increasing specialism make it an ideal technique for corroborating information sources and providing 'drill down' from highly corroborated information at a general level to further, more detailed, information at a more specialised level. Furthermore, the automated tools now available for FCA make it applicable in areas where large volumes of data are being analysed.

Corroborating information using formal concept analysis

A single tweet containing a weak signal of OC is not a sensible basis for Law Enforcement Agencies (LEAs) to take action. However, if a number of sources contain weak signals of the same element of OC from the same location, then this may form a credible basis to warrant further investigation. Such corroboration can be automated by the application of FCA [23] to the structured data extracted from the information sources. FCA can be used to cluster the sources into what we will call *OC Threat Concepts* (or simply *OC Concepts*) where one shared attribute is a location and another shared attribute is an element of OC. When mining the data for formal concepts, if a minimum support is set, say of ten information sources, only OC Threat Concepts with a least ten sources will be obtained.

To carry out FCA, the structured data extracted from the information sources must be converted into a formal context using a process known as scaling [24, 25]. For example, each location in the data becomes a formal attribute in the context. If there are 100 locations in the data, there will be a corresponding 100 columns in the formal context, one for each location. The same approach applies to other named entities, such as drugs. To use ordinal data, such as dates or currency, these can be scaled using appropriate ranges or 'bins' of values.

The weak signals of elements of OC are scaled using the Taxonomy. So, for example, a weak signal may indicate Human Trafficking, and thus the source in which the signal was contained will be labeled with the formal attribute *Crime-HumanTrafficking*. Several different weak signals may all point to Human Trafficking, and thus sources containing any of them would all be labeled with the attribute *Crime-HumanTrafficking*. Other weak signals may point to more specific elements of Human

Table 3 An 'Organised Crime' formal context

	Crime THB	Crime drug trafficking	Crime money laundering	Drug cocaine	Location A	Location B	Category exploitation	Specialism sexual exploitation	Trafficking location X	Trafficking location Y
Source 1	×				×				×	
Source 2	×				×					
Source 3		×		×		×				
Source 4		×				×				
Source 5	×				×		×			
Source 6		×				×				×
Source 7	×				×		×		×	
Source 8		×	×			×				
Source 9	×				×		×	×		
Source 10		×				×				

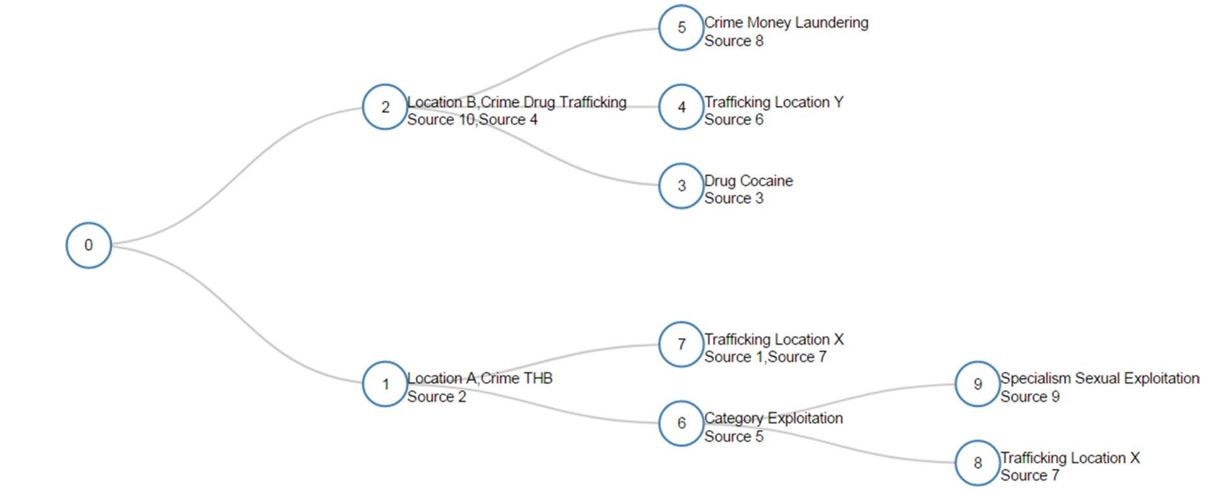


Fig. 7 A formal concept tree derived from the organised crime formal context

Table 4 A formal context scaling part of the OC taxonomy (CSE is child sexual exploitation)

OC taxonomy	Human trafficking	Exploitation	Sexual	CSE
Source containing weak signal of THB	×			
Source containing weak signal of exploitation	×	×		
Source containing weak signal of sexual exploitation	×	×	×	
Source containing weak signal of CSE	×	×	×	×

Trafficking, such as Exploitation which is a component of Human Trafficking (from the taxonomy). A source with such a weak signal will be labeled with both *Crime-HumanTrafficking* and *Element-Exploitation*. Thus the general ‘is a part of’ rule in a taxonomy becomes naturally captured by FCA (see Table 4). This type of scaling greatly adds to the process of corroboration: if there are five sources of information containing weak signals of THB and five sources containing weak signals of Sexual Exploitation, for example, this implies there are *ten* sources containing weak signals of THB.

A similar approach is used making use of information available from the Google Maps API. The deep geolocation context, particularly the hierarchical structure, allows the FCA process to cluster locations at various levels. For a given location, each element of the resolved location hierarchy is used in the formal context as shown in Table 5.

Table 5 A formal context demonstrating hierarchical location scaling

	US	New York State	New York City
Country-level	×		
State-level	×	×	
City-level	×	×	×

Using this approach, FCA detects and corroborates entities within locations at various levels in the hierarchy. The detected concepts are then viewable at the country, state and city levels and can be navigated using the formal concept tree discussed above.

Experimental implementation

Using a data set created from 29,096 tweets as information sources, obtained by scanning for tweets containing weak signals of OC, a formal context was created by scaling the extracted structured data as described above. Using a minimum support of 80 tweets, the context was mined for OC Threat Concepts using a modification of the open-source In-Close concept miner [26]. The result is visualised as a formal concept tree in Fig. 8.

In the tree, the head node is the concept containing all the tweets from concepts that satisfy the minimum support (5512 tweets) and each of the branches is to an OC concept—a concept where one attribute is a location and another is an OC. In this example, every OC is Human Trafficking as this was the type of OC being searched for by the Scanning System. The number inside each node is simply a concept ID number assigned by the concept miner. The number outside the node, below the list of attributes, is the object count (the number of tweets contained in the concept) and in each case this is above the minimum support threshold of 80. Thus concept 53, for example, has the attributes *authorlocation-Atlanta* and *Crime-HumanTrafficking*, and has 185 objects (tweets). In other words, within the data set there are 185 tweets that have the author location Atlanta and contain a weak signal of Human Trafficking. With this high level of corroboration, a police analyst will be alerted to investigate this further, and a possible next step in the investigation is automated by FCA in the form of a 'drill-down' to the OC concept's sub-concepts.

OC concept drill-down

The OC concepts in Fig. 8 contain limited information—they only have a location and the OC Human Trafficking. However, individual tweets in the OC concept may contain further information pertinent to the OC. But physically inspecting 185 tweets, although far less work than examining 29,096 tweets, is nonetheless quite time consuming. However, several tweets in the OC concept may all share the same additional information and this can be divulged by examining the sub-concepts of an OC concept. Each of the sub-concepts will have the same location and crime as the original OC concept but with one or more additional attributes from the structured data extracted from the tweets. Such a result can easily be obtained by mining the data for concepts that contain the attributes of the OC concept and at the same time reducing the minimum support required.

Figure 9 shows a concept tree with the 'Atlanta' OC concept from Fig. 8 and its sub-concepts produced when the minimum support is set to 5.

In the tree, concept 3 shows that 10 of the 185 'Atlanta' tweets also contain a reference to the drug amphetamine. They may not all contain the actual word *amphetamine*, but they will all contain a word or phrase that is commonly used to mean or refer to amphetamine. But using lists of such words and phrases, the entity extraction process carried to produce the structured data will thus label each of these tweets with the attribute *drug-amphetamine*, which in turn enables the FCA to group them together.

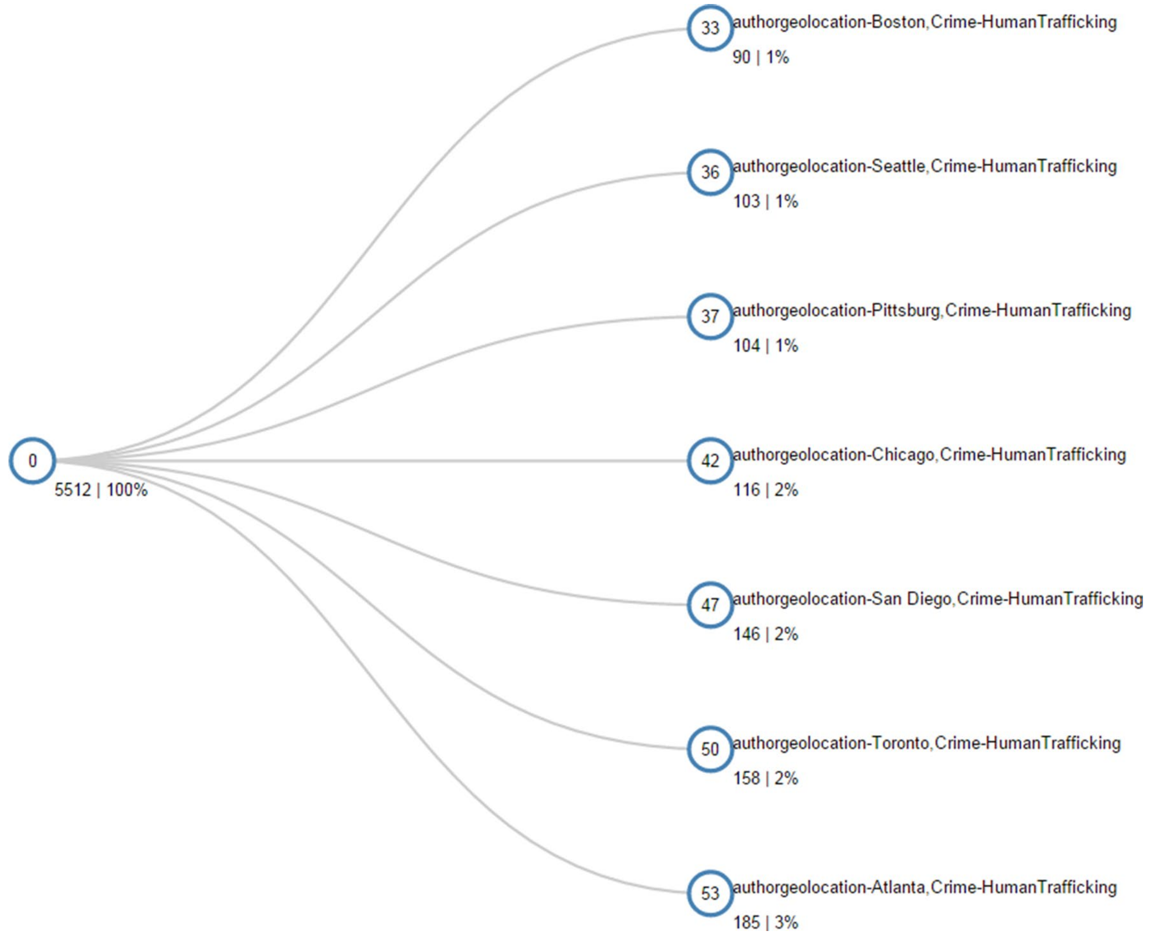

0
5512 | 100%
33
authorgeolocation-Boston,Crime-HumanTrafficking
90 | 1%
36
authorgeolocation-Seattle,Crime-HumanTrafficking
103 | 1%
37
authorgeolocation-Pittsburg,Crime-HumanTrafficking
104 | 1%
42
authorgeolocation-Chicago,Crime-HumanTrafficking
116 | 2%
47
authorgeolocation-San Diego,Crime-HumanTrafficking
146 | 2%
50
authorgeolocation-Toronto,Crime-HumanTrafficking
158 | 2%
53
authorgeolocation-Atlanta,Crime-HumanTrafficking
185 | 3%

Fig. 8 Tree of OC concepts

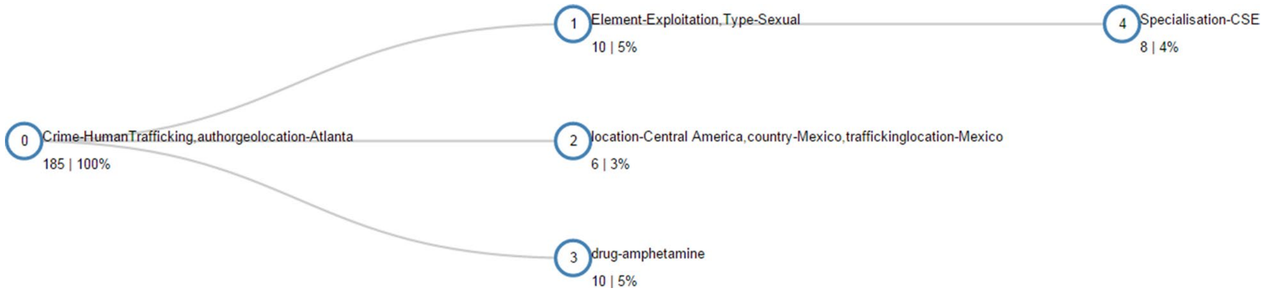

0 Crime-HumanTrafficking,authorgeolocation-Atlanta
185 | 100%
1 Element-Exploitation,Type-Sexual
10 | 5%
2 location-Central America,country-Mexico,traffickinglocation-Mexico
6 | 3%
3 drug-amphetamine
10 | 5%
4 Specialisation-CSE
8 | 4%

Fig. 9 Drill-down for the 'Atlanta' OC concept

Concept 2 shows that 6 of the 185 Atlanta tweets also contain the location Central America and the county Mexico. Furthermore, a semantic rule in the entity extraction process has determined that Mexico is being referred to in the tweet as a trafficking location and thus these tweets are labeled with the attribute *traffickinglocation-Mexico*.

Concept 1 shows that 10 of the 185 Atlanta tweets contain weak signals of the OC Human Trafficking *element* Exploitation and the *exploitation type* Sexual. Furthermore, in 8 of those 10 tweets there are weak signals of CSE, further specialising the OC.

Thus, through this simple automated process, the police analyst has potentially more information that may be pertinent to an OC and more specific information regarding the nature of the OC. Because the original OC concept involved corroboration by a large number of sources, the analyst can gain some confidence that further information contained in sub-sets of the tweets has credibility. Indeed, the analyst may now want to trace back to the original tweets (or to the text of these tweets) and, because they have been grouped together by FCA, it is simple task to provide this facility.

Implementation for end-users

The processes and components described above were implemented as a part of the European ePOOLICE Project [2]. The OC Taxonomy and entity extraction components developed by the authors were implemented in the system to provide data to be consumed by various analytic components, one of which was the FCA OC Threat corroboration component described above. The user interface was developed in close collaboration with Law Enforcement Agency end-users, using a map-based approach. The system allows a police analyst to select a region and type of OC to scan for and then acquire sources on the Internet (such as tweets) that match those search criteria. Structured data is extracted automatically from the sources, as described above, allowing the user to carry out a variety of analytic tasks and display the results in an appropriate visualisation. Figure 10 shows the user interface with the FCA 'Corroborated Threats' option selected. There are a number of other options listed on the left of the screen and these components were developed by other members of the ePOOLICE consortium [2]. However, it is outside the scope of this article to describe these other components here.

The map in Fig. 10 is displaying the OC concepts as described in the example above. Various icons are used by the system to indicate types of OC and the one here is for Human Trafficking. The analyst is able to click on an OC concept icon to display its information (essentially this is the attributes and objects of the concept) and drill down to its sub-concepts. Figure 11 shows the 'Atlanta' OC concept with its associated information, including its attributes (*crime: humantrafficking* and *location: atlanta*) and its objects (tweets), listed as URLs allowing the analyst to trace back to the original sources. The sub-concepts are displayed as icons below the original concept and Fig. 12 shows the additional information displayed when one of these sub-concepts is clicked on—in this case the attribute *drug: amphetamine*. The 10 sources that contain a reference to amphetamine are listed and at this point the analyst may wish to look at some of these. Clicking on a URL will take the user to the original source. Above the lists of attributes and sources is a list of categories. These are categories that are referred to in one or more of the sources in the OC concept, but not in all of them. Thus they give the analyst further information of interest but without an indication of their level of corroboration.

Evaluation

Although difficult to evaluate in an operational sense (we cannot, for example, act as the police in investigating organised crime) it is possible to say something about the quality of the results in terms of the accuracy of the weak signals identified. A sample of 20 inferred OC categories were selected by police experts as having the greatest potential for providing information useful to their

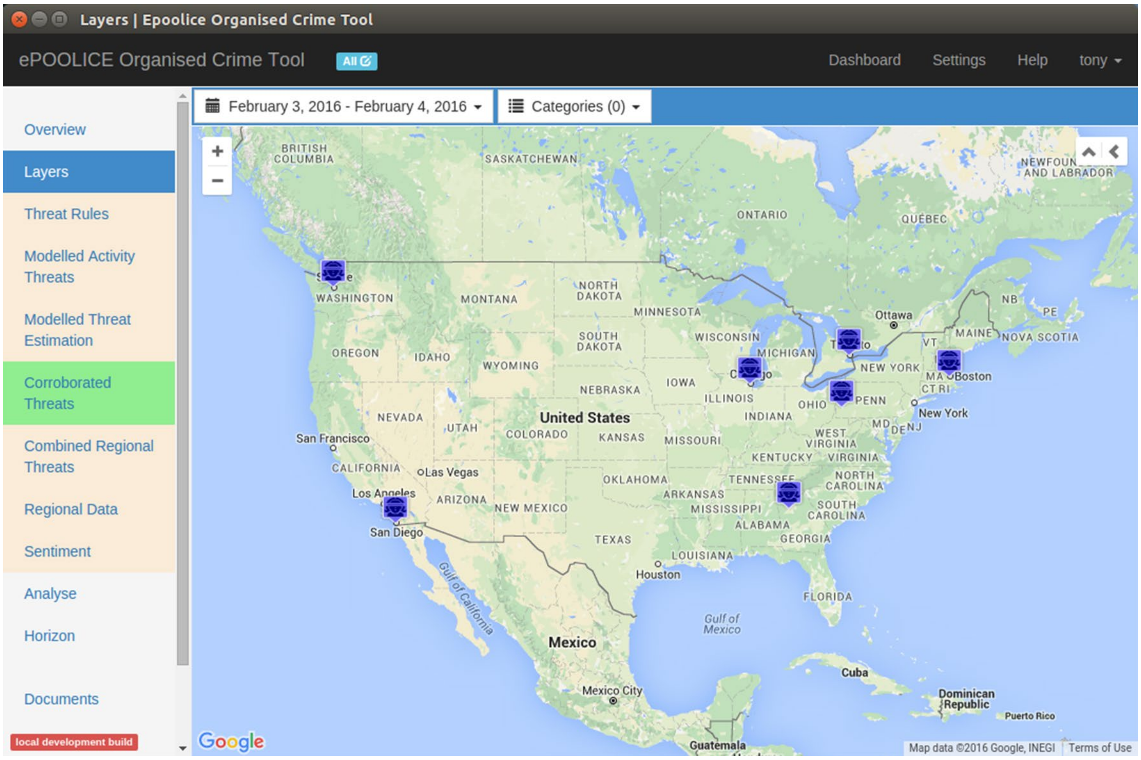
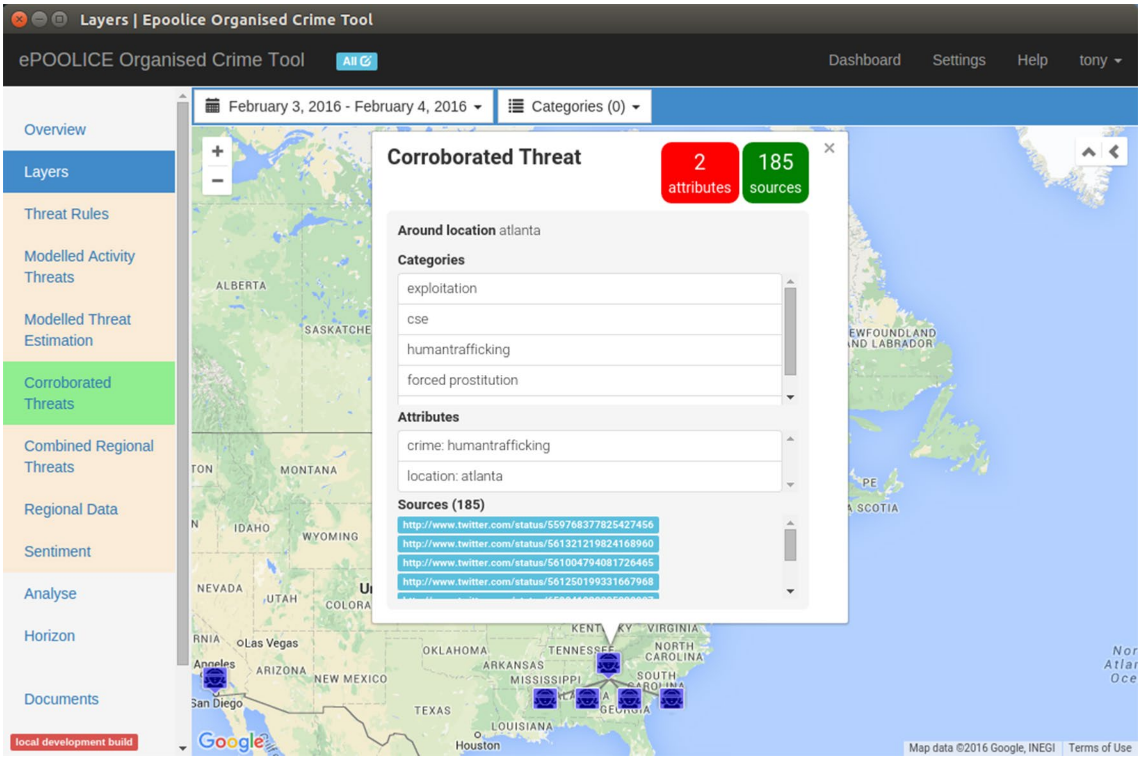


Fig. 10 ePOOLICE system showing OC concepts


Layers | Epoolice Organised Crime Tool
ePOOLICE Organised Crime Tool
All
Dashboard
Settings
Help
tony
February 3, 2016 - February 4, 2016
Categories (0)
Overview
Layers
Threat Rules
Modelled Activity Threats
Modelled Threat Estimation
Corroborated Threats
Combined Regional Threats
Regional Data
Sentiment
Analyse
Horizon
Documents
local development build
San Diego
Las Vegas
Houston
Google
Map data ©2016 Google, INEGI
Terms of Use

Corroborated Threat

2 attributes

185 sources

Around location atlanta

Categories

exploitation

cse

humantrafficking

forced prostitution

Attributes

crime: humantrafficking

location: atlanta

Sources (185)

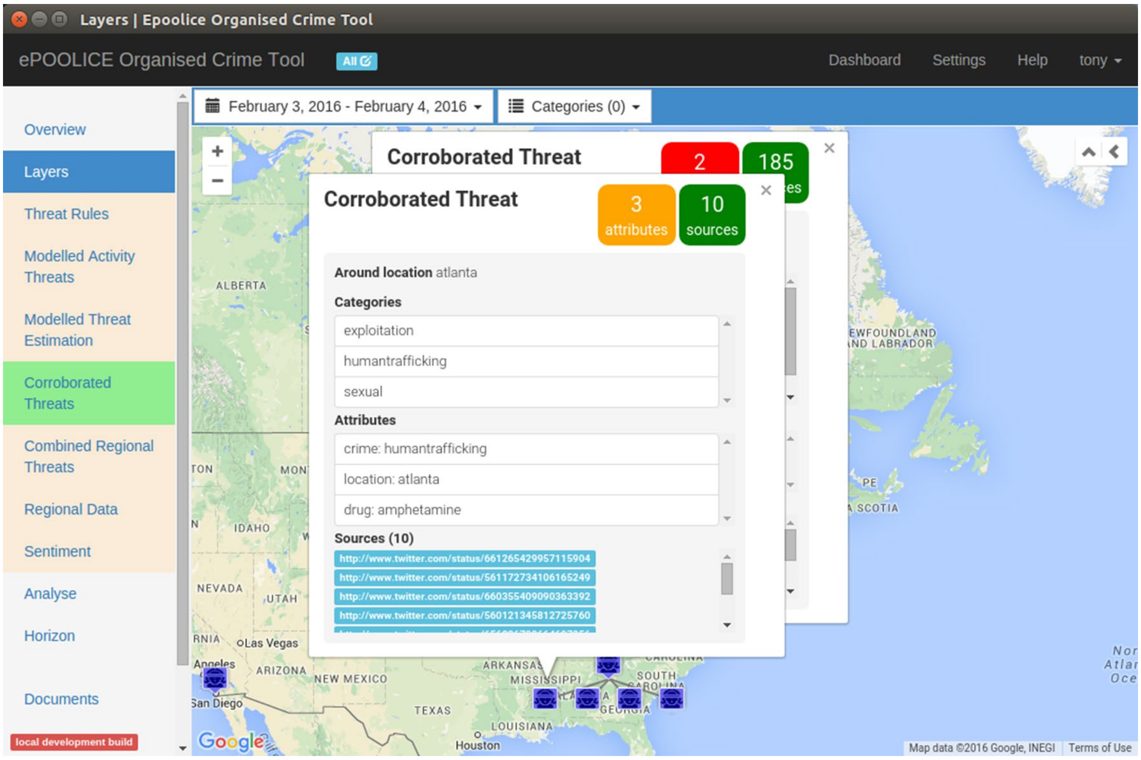
http://www.twitter.com/status/559768377825427456

http://www.twitter.com/status/561321219824168960

http://www.twitter.com/status/561004794081726465

http://www.twitter.com/status/561250199331667968

Fig. 11 ePOOLICE system showing San Diego and Atlanta OC concepts and details of the Atlanta OC concept



Corroborated Threat 2 185

Corroborated Threat 3 attributes 10 sources

Around location atlanta

Categories

exploitation
humantrafficking
sexual

Attributes

crime: humantrafficking
location: atlanta
drug: amphetamine

Sources (10)

http://www.twitter.com/status/661265429957115904
http://www.twitter.com/status/561172734106165249
http://www.twitter.com/status/660355409090363392
http://www.twitter.com/status/560121345812725760

Fig. 12 ePOOLICE system showing drill-down information

investigations and these were inspected against the original text sources, with 16 out of 20 correctly identified from weak signals as being crime related. In the other four cases, the context within which the identified words or phrases were used clearly indicated that the source was not referring to OC. Although only a small sample, this was an encouraging level of false positives. However, further evaluation is required on larger samples to produce a statistically significant result.

In order to test the accuracy of the developed content categorisation and named-entity extraction models, an evaluation set of 164 Twitter posts was selected from the existing larger corpus that is used as the exemplar throughout this paper. The particular tweets selected were done so randomly, however provision was made to remove duplicate 'retweets' or posts that were simply identical to others already contained within the evaluation set, so as to ensure a broader range of entities and rules within the model were tested. The reason for the selection of a relatively small corpus of test data was to enable the manual categorisation of the named entities contained within the data, utilising the domain knowledge embedded within the research team in order to establish, based on the rules and categories known within the model, the areas of best fit for the content under analysis. The manual assessment allowed for the use of human intuition to denote when proper nouns referring to places etc. were in fact referring to a place, regardless of capitalisation, a process much more difficult when using automated tools. Similarly the human assessment can account for errors in grammar, spelling and other issues which may have prevented a successful match using the automated tools, in addition to accounting for words or phrases that may be indicative of a particular form of exploitation or crime that may not be contained within the taxonomy and ruleset used by the automated tools. To reduce the risk of error in the manual process, the evaluators were initially briefed by a police specialist in the use of social media as an intelligence source.

The same data-set was then processed through the model using the developed NLP model and tools, and the resulting .csv outputs compared to determine the accuracy of the automated approach. The evaluation was done so on a binary basis, with a 1 value awarded in instances where the results of the analysis matched exactly, and a null value awarded if the results did not match. The results of the evaluation by output field are presented as percentages in Fig. 13 indicating the success ratio of the automated approach. A subsequent analysis was also conducted to determine the number of false-positives returned using the automated model, that is the the number of non-null values which do not match the value returned during the manual processing of the same dataset. The results of this analysis are included as a separate series in Fig. 13.

In a number of areas, the model returned results accurately 100% of the time. For entities such as ‘pdate’ (date of posting), author and organisation this was due to the rules determining their extraction simply identifying the presence of raw values from the data. In the case of pdate and author these are pre-defined fields in the source content provided through the Twitter API. For Organisation this is done through existing supervised learning techniques embedded within the SAS Content Categorisation software which are built through reading in a large dataset of organisation names against which rules are generated using well established techniques such as Hidden Markov Models and Decision Trees [27]. In instances where the automated model was 100% accurate, it is expected that the results indicate 0% false positives.

The decline in accuracy of location extraction entities continent, country and city as they become more granular can be attributed to the fact that the model uses simple lookup-tables of known locations as the basis of the extraction. While the number of continents and countries and the various names to which they are often referred is relatively finite, the number of towns and cities in the United States alone means that this simple method of extraction can be prohibitive for entities where potential range of values to be returned is vast. This difficulty is also reflected in the increasing number of false positives recorded for these entities due to the potential for place names to be the same as common nouns.

For other entities such as exploitationlocation, traffickinglocation and arrestlocation the results are again

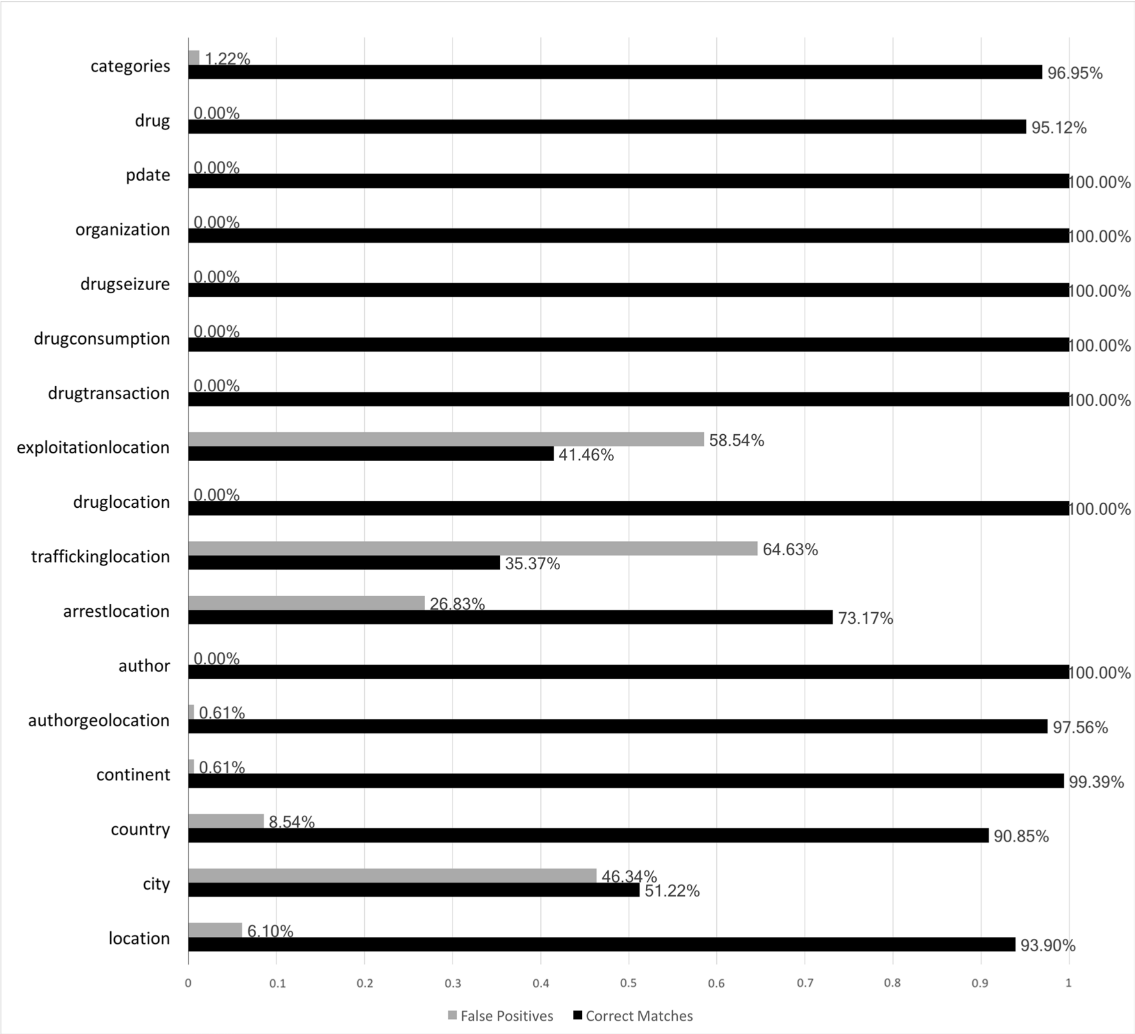


Fig. 13 Content categorisation and named-entity extraction evaluation results

varied. In this instance, predicate rules which look for separate entities such as Location alongside syntax which may indicate the presence of exploitation, arrest or trafficking within the same sentence were used. While useful to an extent, the rules by their nature are limited. The cognitive abilities of a human analyst with knowledge of the domain may be able to infer that, for example, an individual being recovered from a particular location is likely to have been exploited in that same location, whereas the rules used rigidly rely on the explicit presence of language indicating exploitation has taken place. A large number of false-positives for these entities are also returned. This again reflects the value of human cognition in the sense-making process. While the example provided in Fig. 4 demonstrates how these rules can work well, small changes to the syntax could have resulted in the return of a false positive. For instance if the text had said "Atlanta news reports Human Trafficking issues across the country" the ruleset would still return Atlanta as a trafficking location, where as a human analyst could infer that a more appropriate location would be the United States, or that it's not really possible to determine a location at all based on the information provided.

A further qualitative evaluation was provided by 24 end-users during a hands-on feedback session held in December 2015. Various law enforcement agencies from within the EU were represented, including those from regional, national and international organisations, from countries including the UK, France, Spain, Belgium and Estonia. For the evaluation, users were first given an introduction to, and demonstration of the system, by a facilitation team from the ePOOLICE project, highlighting both an overview of what it aims to do at a conceptual level and a practical guide to its features, functions and user interface. Participants were then invited to test the system, and asked to provide qualitative feedback on usability and utility. This feedback was collated using physical questionnaires at the end of the test and through an interactive debrief session where they were able to pass comments and ask further questions about the system to the facilitation team. The feedback gleaned highlighted factors such as the need to refer potentially personal and sensitive information about the origins of the identified indicators so they could be followed up, and, the complexity of the way in which the systems outputs were presented via the map-based interface. However, feedback regarding the utility of the system in providing a means to show trends in current, and alluding to emerging, forms of criminality in different geographic areas was overwhelmingly positive, especially through the utilisation of previously untapped data sources such as social media.

Related work

Until relatively recently, commercial web-crawling systems have tended to focus on providing market research for clients' products and services. A prominent example of this is the sentiment analysis work carried out by SAS using keyword search combined with taxonomy-based text-mining [28]. However, companies are adapting these systems to provide intelligence to law enforcement agencies [29]. BlueJay software, made by BrightPlanet, was able to monitor social media during high profile events and illicit activities, to aid LEAs in the collection of incriminating evidence, although the service has now been discontinued due to the increasing cost of accessing Twitter data [30]. OpenMIND advertises the ability to crawl the "deep web"—inaccessible to search engines but accessible to their software—to provide agencies with "actionable intelligence" [31]. The majority of these tools are designed to be used by the companies that made them, to provide data and intelligence to LEAs as a service. By contrast, the ePOOLICE system has been designed with the police as end-user in mind, providing constant, 'in house', monitoring and a user interface designed for the police analyst. However, LexisNexis, have also developed a "next-generation" policing platform designed for use by LEAs, for crime analysis and investigation. The platform is called Accurint Virtual Crime Center, and links different data types on people, places, vehicles, phones and other information into one visual dashboard. As such, Accurint Virtual Crime Center perhaps comes closest to what the ePOOLICE system is attempting. However, the LexisNexis approach is based on accessing and fusing data from existing databases (such as national law enforcement databases and public records databases) rather than obtaining intelligence from scanning the Internet and social media. Thus, Accurint Virtual Crime Center provides investigators with a central access point and analytics for large volumes of dispersed and disparate data, but without the immediate and real-time monitoring provided by ePOOLICE.

In terms of current research, probably the most prominent example that bears similarities with ePOOLICE is the work being carried out at Cardiff University, UK, developing a system to scan social media to predict outbreaks of hate crime [32]. Researchers are building a machine-learning based classifier to distinguish tweets displaying hateful or antagonistic views with a focus on race, ethnicity, or religion [33]. By identifying textual features of such tweets, the idea is for the classier to be used to predict off-line hate crime based on the level of on-line 'cyber-hate'. It is interesting to note that the researchers used NLP to derive syntactic grammatical relationships in a tweet that can be used as features for classification to enrich their lexicon of cyber-hate beyond simple unigram and bigram terms. This type of text analysis is central to the ePOOLICE entity and relation extraction system (as described above). However, whereas the Cardiff system is intended as a predictive tool, ePOOLICE is focused on information extraction and visualisation. It was this focus, and the proven track-record of tools such as SAS's text mining software, that provided the motivation in ePOOLICE to exclude machine learning in its processes. Nevertheless, if a predictive element was to be added to the ePOOLICE system, clearly a machine learning approach would be appropriate. Indeed, in a similar, parallel, European project (project ATHENA [34], in which the authors' organisation was a partner), machine learning was added as a means of assessing the credibility of information in social media [35].

Increasingly, organised forms of crime, particularly those which involve human exploitation; modern slavery, human trafficking and CSE take place in some capacity online. With the proliferation of these 'cyber-enabled' crimes [36], the vectors through which individuals are recruited, deceived and coerced into exploitation has expanded exponentially [37]. The use of social media, classified ads and other internet mediums is now widespread in human trafficking cases, including those associated both labour and sexual exploitation, and domestic servitude [37, 38]. The internet provides a venue for both the recruitment, deception and coercion of potential victims, and for the sale of illicit goods and services that use trafficked labour and sex-workers [39]. Meanwhile, the dark-web and private message-boards facilitate the dissemination of CSE material [40]. Though this era of online crime has opened up new commercial opportunities for criminals to expand their illicit enterprises, it has also created a resource that can be leveraged by law enforcement in response, and work to make use of this information has already begun.

Indeed, the use of software tools to extract, analyse and visualise detected indicators of human trafficking is not a process unique to the research we have presented. Existing studies, such as that by Ibanez [41], have sought to utilise open-source data from message boards and classified ads. Such information, when combined with social network and phone number analyses has been used to successfully detect the movement patterns of traffickers and their victims. While Ibanez's study, at the time of its publication, was reliant on the manual analysis of open source data—it does acknowledge the potential future utility of web crawling and natural language processing (NLP). As we have demonstrated through our research, automated data extraction and analysis opens the doors to a more expansive and efficient approach to indicator identification.

Other work, such as that by Poelmans [42] has shown the potential value of FCA in combating organised crime, and in the case of this work; the specific challenge of human trafficking. In this particular case, FCA was leveraged as part of a 'Temporal Concept Analysis' on archival policing reports. The techniques were used with some success to build a profile and eventually identify potential suspects who may be involved in human trafficking based on historical data.

Approaches which look to leverage the value of OSINT through automated crawling and analysis have also been employed by researchers working in other areas of crime and security [43]. Popular applications include those associated with Child Sexual Exploitation (CSE) material [44], through to the detection of extremist propaganda and terrorist communities [44] and the identification of civil unrest [45, 46]. While many of these approaches deal with text and metadata, other tools are being developed which seek to analyse multimedia data, such as video and images, for illicit content [47, 48].

More generally, social media data is being used in a variety of ways as researchers experiment with new and novel applications to detect and potentially prevent crime using data derived from the medium. In one such example, social media has been used as a way to try and better understand and explain the location of criminal incidents. For example, Bendler et al. [49] theorised and proved that theft and robbery incidents were more likely in locations with increased social activities, whereas the opposite was true for vehicle theft.

We can conclude from this review of related work that social media is actively being studied and used as a source of intelligence and that steps have been made to automate the process. However, the focus has been predominantly on keyword searching with the notable exception of the machine learning approach being taken at Cardiff University [33]. Thus there is a clear gap in this area for the application of NLP to obtain potentially useful intelligence from sources that would otherwise be overlooked. The development of the notion of 'weak signals', tied to a formal taxonomy of OC and a set of semantic rule, has the potential to give the analyst a situational awareness that is fuller and richer than currently possible. There is also a gap in the analytical capability of exiting approaches—the majority tend to be focused on information gathering, leaving the analyst to decide how to process and utilise the information and potentially lead to a situation of information overload, particularly given the huge amount of online content available. Here, this gap if filled with the application of FCA and the provision of a map-based interface. The FCA automatically clusters sources at geographic locations, filters out sources that are not corroborated and gives the analyst the facility to drill down into clusters of interest to reveal more detail.

Conclusion and further work

In this paper we have demonstrated how FCA can be used in combination with map based visualisation, data extraction and NLP techniques to extract and classify data to detect, through social media, the presence of corroborated organised crime threats. Establishing the idea of 'weak signals' as the presence of key words and phrases that point to criminality, the approach shows one way in which social media can be used as the basis for the development of intelligence from open-sources (OSINT). This intelligence can be used by police and other law enforcement agencies as they seek to adopt new technologies to build situational awareness to aid in the fight against transnational organised crime threats like Human Trafficking.

This paper has shown how using Named Entity Recognition (NRE) allows for the detection and extraction of named locations within a given piece of text. The process used in this paper focused on sentence-level extraction which appears to provide reasonable results on Twitter data. This is due to the limited nature of the content provided in each source, or tweet, when only the source itself is being used—for example, by not extracting further information from the user's profile or surrounding content. Using sentence-level extraction on other sources, such as news articles, often misses important contextual information given in neighbouring sentences, such as the country or state that the given named location is within. For NRE to be effective at extracting locations with greater accuracy, then full-content-level extraction would be required, which in itself introduces further challenges. Performing NRE based on a list of known locations also has the challenge of identifying words that whilst real locations, are not likely to be locations in the given context. Such words that have been observed in this project include "buy", "sell", and "god" which are all names of real locations but were being used in their normal sense. More sophisticated rule-based extraction is required to ensure the author of the content is actually referring to a location by detecting language such as "in Madrid" or even common article layouts such as "MADRID—A man was detained yesterday, suspected of being the ringleader of a major forced labour syndicate".

Disambiguating locations once extracted focuses on the challenge of handling locations with common or similar names. Typically, this is done statistically by using only most relevant or largest place which presents a risk of false positives. Probability states that when an source is talking about Washington, it is referring to Washington, DC rather than Washington County, Alabama, or any number of other places called Washington. However, this will not always be true and may not be suitable in the context of OC. Therefore, wider contextual extraction is imperative in order to provide more accurate results. The presence in the text of other location-based information could also be exploited for disambiguation, for example a mention of a building or place in Washington, DC in addition to the location Washington.

Location granularity proved to be another issue. Whilst the identification of towns and cities may be useful at a strategic level, at the operational level is it often specific houses or buildings that are important. However, there may be many houses with the same number and street name, even within the same geographical area (London Road, for example, is a very popular road name in the UK). Again, the presence in the text of other location-based information could help identify the actual house. The use of location-base coding systems, such as Postcodes in the UK or ZIP Codes in the US, may provide additional means of extracting specific locations.

Issues of context also apply to other named entities, such as drugs. Whilst we have successfully implemented comprehensive lists of drug names, including the various slang terms used, false positives will continue to be a problem unless we can detect more contextual information to confirm them. 'Weed' and 'grass' are well-known terms for the leaf form of cannabis, for example, but including information sources about gardening would clearly not useful in the context of combating OC. Further development of rule-based text mining is required to help identity OC as being the context.

Further work is also required in the detection of false information in malicious, joke or hoax sources. This area has been explored by a number of researchers, such as [50–52] using a variety of approaches including machine learning and NLP. Adding such a level of credibility assessment to the filtering process will further enhance the effectiveness of the system presented here.

Nevertheless, given the limitations above and the further work required, the current system has shown that it is quite capable of providing useful information in the detection of organised crime threats, as evidenced by the evaluation given above. The system has also been successfully adapted as part of the European ATHENA Project [34] where social media is used to provide situational awareness for crisis management [53]. The system has also be adapted as part of an Open Source Intelligence (OSINT) Hub in the Centre of Excellence in Terrorism, Resilience, Intelligence and Organised Crime Research (CENTRIC) at Sheffield Hallam University. In particular the approaches used to crawl, filter, analyse and extract data from the web and social media have been adapted for these purposes. In ATHENA, an evolved version of the same processing pipeline described here is used alongside a taxonomy and ruleset designed to identify and extract information about crises, such as natural disasters, terrorist attacks and other events. Within the OSINT hub, the same techniques are also adapted, in a more exploratory manner, as part of a wider initiative to explore the broader value and utility of information from open-sources in providing situational awareness for LEAs, and to supplement existing forms of evidence in live investigations.

In summary, this work makes a number of contribution to the field: A novel, semantic, approach to identify OC intelligence in social media via the development of a set of 'weak signals' that has the benefit of finding information that would otherwise be overlooked; a new rule-based approach to categorise and organise information into structured data; a novel FCA-based approach to deal with information overload, provide corroboration of information and provide a framework for analysis based on geographic location and information drill-drown. It is hoped that these contributions will be of interest and useful to police practitioners and researchers in this field.

Authors' contributions

SA: overall concept and system design; formal concept analysis. BB: organised crime taxonomy; social media scanning; categorisation and entity extraction. TD: software development and system integration; geolocation; user interfaces. All authors read and approved the final manuscript.

Author details

[1] Conceptual Structures Research Group, Department of Computing, The Communication and Computing Research Centre, Sheffield Hallam University, Sheffield, UK. [2] Centre of Excellence in Terrorism, Resilience, Intelligence and Organised Crime Research, Sheffield Hallam University, Sheffield, UK.

Acknowledgements

This Project has received funding from European Union Seventh Framework Programme FP7/2007-2013 under Grant Agreement No. FP7-SEC-2012-312651. This article is an expanded and updated version of a paper presented at the 22nd International Conference on Conceptual Structures, Annecy, France, 2016 [54].

Competing interests

The authors declare that they have no competing interests.

Legal and ethical disclaimer

No data that can or may be considered sensitive or personal has been handled as a result of the research undertaken. The authors do however acknowledge, despite being outside of the scope of the research present, that in practice the operational utility of such a system would be dependent on the use of data that may be considered personal and/or sensitive.

Publisher's Note

Received: 25 October 2017 Accepted: 27 November 2018
Published online: 13 December 2018

References

1. Perrin A (2015) Social media usage: 2005–2015. Pew Research Center. http://www.pewinternet.org/files/2015/10/PI_2015-10-08_Social-Networking-Usage-2005-2015_FINAL.pdf
2. Pastor RP, Larsen HL (2017) Scanning of open data for detection of emerging organized crime threats—the ePOOLICE project. In: Larsen H, Blanco J, Pastor Pastor R, Yager R (eds) Using open data to detect organized crime threats. Springer, Berlin, pp 47–71
3. CISC Strategic Criminal Analytical Services (2007) Strategic early warning for criminal intelligence. Criminal Intelligence Service Canada (CISC). http://publications.gc.ca/collections/collection_2013/sp-ps/PS64-107-2007-eng.pdf
4. College of Policing (2014) Authorised professional practice investigation guidelines. https://www.app.college.police.uk/app-content/investigations/introduction/
5. Her Majesty's Inspectorate of Constabulary (2011) The rules of engagement: a review of the august 2011 disorders. https://www.justiceinspectorates.gov.uk/hmic/media/a-review-of-the-august-2011-disorders-20111220.pdf
6. Omand D, Bartlett J, Miller C (2012) Introducing social media intelligence (SOCMINT). Intell Natl Sec 27(6):801–823
7. Gilgoff D Lee JJ (2013) Social media shapes Boston bombings response. Natl Geogr 7
8. United Nations (2000) United Nations convention against transnational organized crime and the protocols thereto. http://www.osce.org/odihr/19223?download=true
9. The Home Office (2015) The strategic policing requirement. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/417116/The_Strategic_Policing_Requirement.pdf
10. Europol (2011) EU organised crime threat assessment: OCTA 2011. File no. 2530-274. Europol, O2—Analysis & Knowledge, The Hague
11. Lyon D (2014) Surveillance, snowden, and big data: capacities, consequences, critique. Big Data Soc 1(2):2053951714541861
12. Babuta A (2017) Big data and policing: an assessment of law enforcement requirements, expectations and priorities. https://rusi.org/sites/default/files/201709_rusi_big_data_and_policing_babuta_web.pdf
13. Brewster B, Ingle T, Rankin G (2014) Crawling open-source data for indicators of human trafficking. In: Proceedings of the 7th IEEE/ACM international conference on utility and cloud computing. IEEE Computer Society, pp 714–719
14. Mahmud J, Nichols J, Drews C (2012) Where is this tweet from? inferring home locations of Twitter users. ICWSM 12:511–514
15. Williams P, Godson R (2002) Anticipating organized and transnational crime. Crime Law Soc Change 37(4):311–355
16. UNODC (2009) Global report on trafficking in persons. https://www.unodc.org/documents/Global_Report_on_TIP.pdf
17. Laczko F, Gramegna MA (2003) Developing better indicators of human trafficking. Brown J World Aff 10(1):179–194
18. UNODC (2009) Anti-human trafficking manual for criminal justice practitioners. https://www.unodc.org/documents/human-trafficking/TIP_module2_Ebook.pdf
19. Chakraborty G, Pagolu M, Garla S (2014) Text mining and analysis: practical methods, examples, and case studies using SAS. SAS Institute
20. Ritter A, Etzioni O, Clark S et al (2012) Open domain event extraction from twitter. In: Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 1104–1112
21. Owoputi O, O'Connor B, Dyer C, Gimpel K, Schneider N, Smith NA (2013) Improved part-of-speech tagging for online conversational text with word clusters. In: Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: human language technologies. Association for Computational Linguistics, pp 380–390
22. Twitter Inc. (2015) Twitter search API. https://dev.twitter.com/rest/public/search
23. Ganter B, Wille R (1998) Formal concept analysis: mathematical foundations. Springer, Berlin
24. Wolff KE (1993) A first course in formal concept analysis: how to understand line diagrams. Adv Stat Softw 4:429–438

25. Andrews S (2009) Data conversion and interoperability for FCA. In: Croitoru M, Fortin J, Jäschke R (eds) Proceedings of the 4th conceptual structures tools interoperability workshop at the 17th international conference on conceptual structures. University Higher School of Economics, Moscow, pp 42–49
26. Andrews S (2011) In-Close2, a high performance formal concept miner. In: Andrews S, Polovina S, Hill R, Akhgar B (eds) Conceptual structures for discovering knowledge—proceedings of the 19th international conference on conceptual structures (ICCS). Springer, Berlin, pp 50–62
27. Nadeau D, Sekine S (2007) A survey of named entity recognition and classification. Lingvist Investig 30(1):3–26
28. SAS (2013) SAS® sentiment analysis: fact sheet. https://www.sas.com/content/dam/SAS/en_us/doc/factsheet/sas-sentiment-analysis-104357.pdf
29. Mateescu A, Brunton D, Rosenblat A, Patton D, Gold Z, Boyd D (2015) Social media surveillance and law enforcement. Data Civ Rights 27:2015–2027
30. BlueJay (2015) What you can do with BlueJay. http://brightplanet.com/bluejay/. Accessed 7 Oct
31. OpenMIND (2018) OpenMIND. http://www.3i-mind.com/solutions/openmind/. Accessed 3rd Mar
32. Murgia M (2018) US police to scan social media for violence alerts. https://www.ft.com/content/e2c850be-80c8-11e6-bc52-0c7211ef3198. Accessed 3rd Mar
33. Burnap P, Williams ML (2015) Cyber hate speech on twitter: an application of machine classification and statistical modeling for policy and decision making. Policy Internet 7(2):223–242
34. Andrews S, Yates S, Akhgar B, Fortune D (2013) The ATHENA project: using formal concept analysis to facilitate the actions of responders in a crisis situation. In: Akhgar B, Yates S (eds) Strategic intelligence management. Elsevier, Amsterdam, pp 167–180
35. Andrews S, Day T, Domdouzis K, Hirsch L, Lefticaru R, Orphanides C (2017) Analyzing crowd-sourced information and social media for crisis management. In: Akhgar B, Staniforth A, Waddington D (eds) Application of social media in crisis management. Springer, Berlin, pp 77–96
36. Wall DS (2005) The Internet as a conduit for criminal activity (October 21, 2015). In: Pattavina A (ed) Information technology and the criminal justice system. Sage, Beverley Hills, pp 77–98. https://ssrn.com/abstract=740626. Revised 2010–2015
37. Latonero M (2011) Human trafficking online: the role of social networking sites and online classifieds. SSRN. http://dx.doi.org/10.2139/ssrn.2045851
38. Wang H, Cai C, Philpot A, Latonero M, Hovy EH, Metzler D (2012) Data integration from open internet sources to combat sex trafficking of minors. In: Proceedings of the 13th annual international conference on digital government research. ACM, New York, pp 246–252
39. Kunze EI (2009) Sex trafficking via the internet: how international agreements address the problem and fail to go far enough. J High Tech Law 10:241
40. Carback J (2018) Cybersex trafficking: toward a more effective prosecutorial response (updated) (May 19, 2018). Criminal Law Bulletin, 54. https://ssrn.com/abstract=3181311
41. Ibanez M, Suthers DD (2014) Detection of domestic human trafficking indicators and movement trends using content available on open internet sources. In: 2014 47th Hawaii international conference on system sciences (HICSS). IEEE, pp 1556–1565
42. Poelmans J, Elzinga P, Viaene S, Dedene G (2010) A method based on temporal concept analysis for detecting and profiling human trafficking suspects. In: Artificial intelligence and applications. Acta Press, Calgary, pp 1–9
43. Agarwal S, Sureka A, Goyal V (2015) Open source social media analytics for intelligence and security informatics applications. In: Proceedings of the 4th international conference on big data analytics. Springer, Berlin, pp 21–37
44. Charalambous E, Kavallieros D, Brewster B, Leventakis G, Koutras N, Papalexandratos G (2016) Combatting cybercrime and sexual exploitation of children: an open source toolkit. In: Akhgar B, Bayerl P, Sampson F (eds) Open source intelligence investigation. Springer, Berlin, pp 233–249
45. Hua T, Lu CT, Ramakrishnan N, Chen F, Arredondo J, Mares D et al (2013) Analyzing civil unrest through social media. Computer 12:80–84
46. Compton R, Lee C, Lu TC, De Silva L, Macy M (2013) Detecting future social unrest in unprocessed twitter data: "emerging phenomena and big data". In: Proceedings of the 2013 IEEE international conference on intelligence and security informatics. IEEE, pp 56–60
47. Fu T, Huang CN, Chen H (2009) Identification of extremist videos in online video sharing sites. In: Proceedings of the 2009 IEEE international conference on intelligence and security informatics. IEEE, pp 179–181
48. Agarwal S, Sureka A (2014) A focused crawler for mining hate and extremism promoting videos on YouTube. In: Proceedings of the 25th ACM conference on hypertext and social media, ACM, New York, pp 294–296
49. Bendler J, Ratku A, Neumann D (2014) Crime mapping through geospatial social media activity. In: Proceedings of the 2014 international conference on information systems. https://aisel.aisnet.org/icis2014/proceedings/ConferenceTheme/12/
50. Gupta A, Kumaraguru P, Castillo C, Meier P (2014) TweetCred: a real-time web-based system for assessing credibility of content on Twitter. In: Proceedings of the 6th international conference on social informatics (SocInfo)
51. O'Donovan J, Kang B, Meyer G, Höllerer T, Adali S (2012) Credibility in context: an analysis of feature distributions in Twitter. In: Proceedings of the 2012 international conference on privacy, security, risk and trust, at the 2012 international conference on social computing, Amsterdam, September 3–5, 2012, pp 293–301
52. Castillo C, Mendoza M, Poblete B (2011) Information credibility on Twitter. In: Proceedings of the 20th international conference on the world wide web, WWW 2011, Hyderabad, March 28–April 1, 2011, pp 675–684
53. Andrews S, Gibson H, Domdouzis K, Akhgar B (2016) Creating corroborated crisis reports from social media data through formal concept analysis. J Intell Inf Syst 47(2):287–312
54. Andrews S, Brewster B, Day T (2016) Organised crime and social media; identifying and corroborating weak signals of human trafficking online. In: Proceedings of the 22nd international conference on conceptual structures. vol 9717 of Lecture notes in computer science (Lecture notes in artificial intelligence). Springer, Berlin, pp 137–150